IEEE SOSE 2012

IEEE SOSE 2012
7th INTERNATIONAL CONFERENCE ON SYSTEM OF SYSTEMS ENGINEERING

JULY 16-19, 2012
GENOA, ITALY

Proceedings

Conference Chair welcome message

On behalf of the University of Genoa, I would like to welcome you to IEEE SoSE2012, the Seventh International Conference on System of Systems Engineering in Genoa, Italy, technically sponsored by IEEE System, Man, and Cybernetics Society and by IEEE Reliability Society.

IEEE SoSE2012 addresses branches of numerous engineering fields such as control, computing, communication, information technology with different applications such as manufacturing, defense, national security, aerospace, aeronautics, energy, environment, healthcare, and transportation. The conference theme is "SoSE in cooperative and competitive distributed decision making in complex dynamic systems", focusing on SoSE control and simulation methodologies to support decisions in different application fields, among others transport, energy, industrial and environmental risk management.

We hope you enjoy the content of the Conference and have ample opportunity to network with your international colleagues.

Thanks to all authors, keynote speakers and presenters for their effort and considerable papers in the fields on System of Systems Engineering.

A special thank to Mo Jamshidi who has given Italy the fantastic opportunity to organize IEEE SoSE2012 and to Michael Henshaw, general IEEE SoSE 2012 co-chair, who has provided the unique opportunity to have a direct link to the state of the art of SoSE projects in Europe. A special thank to Paola Girdinio, dean of the Faculty of Engineering of the University of Genoa, to have supported IEEE SoSE2012 with important industrial links, and a final thank to Chiara Bersani and Hanane Dagdougui to have shared with me the extraordinary experience of organizing the IEEE SoSE2012 program.

Roberto Sacile
Department of Communication Computer and System Sciences
University of Genoa, Italy

General Conference Chair
2012 IEEE International Conference on System of Systems Engineering

Founding Chair
Mo Jamshidi, University of Texas San Antonio, USA

Conference General Co-Chairs
Roberto Sacile, University of Genoa, Italy
Michael Henshaw, Loughborough University, UK

Chair for the liaisons with Italian SOSE industries
Paola Girdinio, University of Genoa, Italy

Program Co-Chairs
Ahmed Ouammi, CNRST, Rabat, Morocco
Riccardo Minciardi, University of Genoa, Italy

Finance co-Chairs:
Roberto Sacile, University of Genoa, Italy
Mo Jamshidi, University of Texas San Antonio, USA

Tutorial and Organized Sessions Chairs
Hanane Dagdougui, University of Genoa, Italy
Franck Guarnieri, Mines ParisTech, France

Review Process Chairs
Hanane Dagdougui, University of Genoa, Italy
Chiara Bersani, University of Genoa, Italy
Huseyin Dogan, Loughborough University, UK

Local Arrangements
Chiara Bersani, University of Genoa, Italy
Hanane Dagdougui, University of Genoa, Italy
Anita Trotta, University of Genoa

Publication Chairs
Hanane Dagdougui, University of Genoa, Italy
Chiara Bersani, University of Genoa, Italy

Industrial Liaisons
Garry Roedler, Lockheed Martin
Domenico Pizzorni, Eni, Italy

USA Liaison

Ferat Sahin, Rochest Institute of Technology, USA

Asia & Pacific Liaison

Saeid Nahvandi, Deakin University, Australia

Syoji Kobashi, University of Hyogo, Japan

Technical Program Committee

Chiara Bersani, University of Genoa, Italy
Hanane Dagdougui, University of Genoa
Huseyin Dogan, Loughborough University, United Kingdom
Michael Henshaw, Loughborough University, United Kingdom
Mo Jamshidi, University of Texas at San Antonio, USA
Ahmed Ouammi, CNRST, Morocco
Roberto Sacile, University of Genoa, Italy
Ferat Sahin, Rochester Institute of Technology, USA

Reviewers

Datu Agusdinata, Purdue University, USA
Marco Aiello, University of Groningen, Netherlands
Ali Benmeddour, National Research Council, Canada
Chiara Bersani, University of Genoa, Italy
Kul Bhasin, NASA Glenn Research Center, USA
Marcus Bjelkemyr, Royal Institute of Technology, Sweden
Ruxandra Botez, Ecole de Technologie supérieure, Canada
Azedine Boulmakoul, Faculté des Sciences Technique de Mohammedia, Morocco
Raymond R. Buettner, Jr., Naval Postgraduate School, USA
Reggie Cole, Lockheed Martin Corporation, USA
Hanane Dagdougui, University of Genoa, Italy
Charles E. Dickerson, Loughborough University, United Kingdom
Huseyin Dogan, Loughborough University, United Kingdom
Alex Duffy, University of Strathclyde, United Kingdom
Nil Ergin, Penn State University, USA
Timothy J Eveleigh, The George Washington University, USA
Marco Fossa, University of Genoa, Italy
Emmanuel Garbolino, MINES ParisTech, France
Mauro Giacomini, University of Genoa, Italy
Jeffrey Hayden, PresciPoint Solutions, L.L.C., USA
Michael Henshaw, Loughborough University, United Kingdom
Paul C. Hershey Raytheon, Inc. , USA
Philip John, Cranfield University, United Kingdom
Roy Kalawsky, Louborough University, United Kingdom
Gary Langford, Naval Postgraduate School, USA
Stefano Massucco, University of Genoa, Italy
Riccardo Minciardi, University of Genoa, Italy
Georgios Moschoglou, The George Washington University, USA
Gerrit Muller, Buskerud University College, Norway
John Osmundson, NPS, USA
Ahmed Ouammi, CNRST, Morocco
Richard Payne, Newcastle University, United Kingdom
Shrisha Rao, International Institute of Information Technology, India
George Rebovich Jr., The MITRE Corporation, USA
Michela Robba, University of Genoa, Italy
Claudio Roncoli, University of Genoa, Italy
Roberto Sacile, University of Genoa Italy
Simona Sacone, University of Genoa, Italy
Ferat Sahin, Rochester Institute of Technology, USA
Yashar Sahraei Manjili, The University of Texas at San Antonio, USA
Silvia Siri, University of Genoa, Italy
Eva Trasforini, CIMA, Italy
Jie Xu, Leeds University, United Kingdom

Keynote Speakers

Michael G.H. Bell, Imperial College, United Kingdom
Francesco Benvenuto, AnsaldoNucleare, Italy
Fabrizio Bianco, AnsaldoNucleare, Italy
Antonino Carmelo Arecchi, SELEX Sistemi Integrati SpA
Maria Pia Fanti, Polytechnic of Bari, Italy
Antonella Di Fazio, Telespazio, Italy
Luigi Fortuna, Università degli Studi di Catania, Italy
Mattia Frasca, Università degli Studi di Catania, Italy
Michael Henshaw, Loughborough University, United Kingdom
Mo Jamshidi, University of Texas at San Antonio, USA
Nadia Mazzino, Ansaldo STS, Italy
Fabrizio Morciano, SELEX Sistemi Integrati Spa, Italy
Rush D. Robinett III, Sandia National Laboratories, USA
Werner Steinhögl, ICT Embedded Systems and Control, European Commission, directorate general Information Society and Media (INFSO)
Massimiliano Zazza, Ministry of Infrastructure and Transport, Italy
Driss Zejli, Centre National de Recherche Scientifique et Technique (CNRST), Morocco

Venue: NH Hotel Marina, Porto Antico, Genoa

Monday, July 16

8:00 AM - 6:00 PM

Registration at Lounge NH Marina Hotel

9:00 AM- 9:30 AM

Welcome to IEEE SoSE2012
Room: Mediterraneo

Chairs: Mohammad Jamshidi (University of Texas at San Antonio, USA), Roberto Sacile (University of Genoa, Italy), Michael Henshaw (Loughborough University, United Kingdom)

9:30 AM - 10:30 AM

Keynote Speaker: "Networked Microgrids: Implementing the Smart Grid with Renewable Energy? " by Rush D. Robinett III
Room: Mediterraneo

How will the integration of large percentages of renewable energy sources affect the electrical power grid? More specifically, how will replacing fossil-fuel burning generators with intermittent renewable energy sources impact the "Smart Grid", notably in the form of additional energy storage, feedback control systems, and networked microgrids? Dr. Rush D. Robinett III will explore these timely issues, explaining how the brute force, dispatchable fossil-fuel-based electricity generation of today will give way to a "Smart Grid" based on two-way power and information flow in the form of networked microgrids.

Rush D. Robinett III is Senior Manager of the Grid Modernization and Military Energy Systems Group at Sandia National Laboratories, focusing on the research and development of microgrids and networked microgrids. Rush began his career at Sandia in 1988 as a Member of the Technical Staff, working on the Star Wars Program. In 1995, he was promoted to Distinguished Member of Technical Staff and shortly thereafter to Technical Manager of the Intelligent Systems Sensors and Controls Department within the Robotics Center. In 2002, Rush was promoted to Deputy Director and Senior Manager of the Energy and Infrastructure Future Group, where he developed new opportunities in distributed, decentralized energy and transportation infrastructures with a focus on entropy and information metrics. Rush has authored more than 100 technical articles, including three books and holds seven patents. He has three degrees in Aerospace Engineering: a BS and Ph.D. from Texas A&M University and an MS from the University of Texas at Austin.

10:30 AM - 11:15 AM

Keynote Speaker: "The shift of Morocco to renewable energies: towards a system-of-systems perspective" by Driss Zejli
Room: Mediterraneo

Today, Morocco imports almost 96% of its commercial energy needs. However, it has a very large potential for harnessing renewable energy sources. Wind and solar energy could play an important role in the decarbonization of the Moroccan electricity generation reducing hence the energy bill of the country. The shift to a significantly larger solar and wind energies will require the creation of a robust renewable energies system of systems that should operate jointly with a central electricity generation.

Driss Zejli is the Head of the Renewable Energy Economy and Technologies Unit of the National Centre of Scientific and Technical Research of Morocco since 2005. He is author and co-author of many papers, conferences and books on both renewable energies and renewable energy powered desalination. He participated in many national and international research projects.

11:15 AM - 11:45 AM

Coffee Break
Room: Terrazza Acquamarina

11:45 AM - 12:20 PM

Keynote Speaker: "Distributed Renewable Energy Systems: System of Systems Based Intelligent Management of Micro-Grids" by Mo Jamshidi
Room: Mediterraneo

Sustainable energy has slowly but surely becoming relevant source of energy in many parts of the world. From US to Europe to Asia and South Pacific and on to Middle East, photovoltaic and wind energies are the desired sources. Such new and expanded sources of energy is fast creating numerous sources of distributed generations and creating a great number of micro-grids of energy. However, intermittency nature of both of these sources would make it very difficult to rely on for instantaneous need for electricity in urban or rural areas. In this presentation, fundamental aspects of cyber-physical systems will first be discussed and then a management scheme for microgrids will be presented based on a Fuzzy Logic-based framework is proposed for control of Battery Storage Unit in Micro-Grid Systems to achieve Efficient Energy Management.

Mo Jamshidi (Fellow IEEE, Fellow ASME, Fellow AAAS, Fellow TWAS, Fellow NYAS, Assoc. Fellow-AIAA) is the Lutcher Brown Endowed Chaired Professor of the University of Texas Systems and working at the University of Texas, San Antonio, TX, USA. He has over 560 technical publications including 58 books (10 text books), research volumes, and edited volumes. He is the Founding Editor or co-founding editor or Editor-in-Chief of 5 journals including IEEE Control Systems Magazine. He is the founding editor-in-chief of the IEEE Systems Journal (inaugurated in September 2007). Dr. Jamshidi is a Fellow or member of 8 societies and academies. He is the recipient of the IEEE Centennial Medal and IEEE Control Systems Society Distinguished Member Award and the IEEE CSS Millennium Award. He is currently on the Board of Governors of the IEEE Society on Systems, Man and Cybernetics and the IEEE Systems Council. He is an Honorary Professor at three Chinese and one Australian Universities. In October 2005 he was awarded the IEEE's Norbert Weiner Research Achievement Award.

12:20 PM - 1:00 PM

Keynote Speaker: "How to behave nicely in a Systems of Systems society" by Michael Henshaw
Room: Mediterraneo

We don't have engineering problems; we have enough good engineers to solve any that come along. Our problem is managing the people!" is a statement that we hear over and over again from those involved in managing the development and operation of Systems of Systems (SoS). It is not wholly true, of course, but in many important SoS in military and civilian endeavours, the role of humans both in terms of causing difficulties and in recovering from them are dominant features of SoS operation. George Rebovich's Jr. insightful comment that from the single-system community's perspective, its part of the SoS capability represents additional obligations, constraints and complexities is the key consideration in terms of avoiding problems and achieving more effective operations; but overcoming this problem is far from easy. We shall begin with a consideration of some of the major challenges to coping in a SoS environment. These concern management of security and safety, particularly in the situation of increasingly networked Information and Operational Technologies (IT/OT integration). Loss of situational awareness of participants in a SoS appears to be a significant factor in catastrophic failures; this will also be discussed using examples. The additional challenges due to autonomous decision making by non-human agents will be reviewed. Perhaps an under-appreciated aspect of SoS is their ability to introduce previously unconsidered ethical conundrums that arise for the three aspects noted above. Having set the scene with the introduction of problems, we shall consider the ways in which these may be mitigated by identification of desired behaviours and how they can be encouraged. We shall comment on the use of open architectures as a means for enabling partnerships among system owners and operators and the opportunities for incentivising better SoS behaviours. Ultimately, the presentation will leave the question unresolved, but will identify the key research themes that should be addressed to enable systems engineers and others to mitigate bad behaviours and encourage more effective SoS operation.

Michael Henshaw received a BSc. (first class) and a PhD in Applied Physics from University of Hull, UK. He also holds an MBA. He pursued research in Plasma Physics in both Hull and York Universities before moving to British Aerospace as an Aeronautical Engineer. Following seventeen years working for BAE Systems in Aeronautical and, later Systems, engineering he joined Loughborough University as Professor of Systems Engineering in Aug. 2006. He is a co-chair of IEEE Technical Committee on Systems of Systems Engineering and an Associate Editor of both the IEEE Systems Journal and the Royal Aeronautical Society Journal. He is also Academic Director of INCOSE UK. He is a member at large of the NATO RTO Systems Concepts and Integration Panel and chairman of Working session A on Capability Planning and Management Allocation. His research has focused on Systems of Systems with particular concentration on the themes of Network Enabled Capability, Through Life Systems Management, Autonomous Systems and latterly Cyber Security. In all of these areas complexity and interoperability between systems are the major challenges being addressed.

1:00 PM - 2:00 PM

Lunch Buffet
Room: Terrazza Acquamarina

2:00 PM - 4:00 PM

Smart Grids including Renewable Energies

Room: Mediterraneo

Chair: Michael Henshaw (Loughborough University, United Kingdom)

2:00 Exploring Human Factors Effects in the Smart Grid System of Systems Demand Response
Michael Miller (Georgia Institute of Technology, USA); Kelly Griendling (Georgia Institute of Technology, USA); Dimitri Mavris (Georgia Institute of Technology, USA)

2:30 A System of Systems Model for the Control of the University of Genoa Smart Polygeneration Microgrid
Stefano Bracco (University of Genoa, Italy); Federico Delfino (University of Genoa, Italy); Fabio Pampararo (University of Genoa, Italy); Michela Robba (University of Genoa, Italy); Mansueto Rossi (University of Genoa, Italy)

3:00 Intelligent Decision Making for Energy Management in Microgrids with Air Pollution Reduction Policy
Yashar Sahraei Manjili (The University of Texas at San Antonio, USA); Amir Rajaee (The University of Texas at San Antonio, USA); Mohammad Jamshidi (University of Texas at San Antonio, USA); Brian T Kelley (University of Texas at San Antonio, USA)

3:30 Optimal Control Strategy for the Wind Power Exchanges in a Network of Microgrids
Hanane Dagdougui (University of Genoa, Italy); Ahmed Ouammi (University of Genoa, Italy); Roberto Sacile (University of Genoa, Italy)

Smart Grids including Renewable Energies

Room: Jonio

Chair: Huseyin Dogan (Loughborough University & BAE Systems, United Kingdom)

2:00 Development of Innovative Systems for Operation and Control of Electric Power Distribution Networks
Roberto Caldon (University of Padova, Italy); Stefano Massucco (University of Genoa & IEES Lab., Italy); Carlo Alberto Nucci (University of Bologna, Italy); Fabrizio Pilo (Dipartimento Di Ingegneria Elettrica Ed Elettronica, Italy); Paola Verde (University of Cassino, Italy)

2:30 Towards a Cooperative Life Cycle Documentation for Distributed Renewable Energy Power Plants
Johannes Schmidt (Institute for Applied Informatics at the University of Leipzig, Germany); Antonius van Hoof (Baden-Wuerttemberg Cooperative State University (DHBW) Stuttgart, Germany)

3:00 Analysis of Moroccan Wind and Solar Potential Using Artificial Neural Network Approach
Hanane Dagdougui (University of Genoa, Italy); Ahmed Ouammi (University of Genoa, Italy); Driss Zejli (CNRST, Morocco); Roberto Sacile (University of Genoa, Italy)

3:30 Supervision, Analysis and Control System of Photovoltaic Power Plants
Marco Raggio (University of Genoa, Italy); Paolo Lambruschini (University of Genoa, Italy); Rajiv Bajpai (University of Genoa, Italy); Abhishek Sharma (University of Genoa, Italy)

Sustainability including Renewable Energies

Room: Tirreno

Chair: Roberto Sacile (University of Genoa, Italy)

2:00 Developing an ICT-Enabled, Anti-Prophetic Approach to Sustainable Cities
Eleanor Cosgrave (University of Bristol, United Kingdom); Theo Tryfonas (University of Bristol, United Kingdom); Kirsten Cater (University of Bristol, United Kingdom)

2:30 Towards an Intelligent System for Cognitive Behavioral Treatment
Panagiotis Karampelas (Hellenic American University, USA); Ioanna Laniti (Hellenic American University, USA); Despina Konstas (Hellenic American University, USA); Panagiotis Kalagiakos, Dr (Hellenic American University, USA)

3:00 Socio-technical Considerations for Enterprise System Interfaces in Systems of Systems
Murray Sinclair (University of Loughborough, United Kingdom); Carys Siemieniuch (Loughborough University, United Kingdom)

3:30 Astronomic Sun Tracker Performance and Solar Energy Collection Comparison for Different Italian Sites
Marco Fossa (University of Genoa, Italy); Claudio De Domenico (University of. Genoa, Italy)

4:00 PM - 4:30 PM

Coffee Break
Room: Terrazza Acquamarina

4:30 PM - 6:30 PM

Complex Dynamic Systems
Room: Adriatico
Chair: Bingfeng Ge (National University of Defense Technology & University of Waterloo, Canada)

4:30 The Viewband Concept: Introducing Life-cycle Modeling in Enterprise Architectural Frameworks
Daniele Gianni (European Space Agency, The Netherlands); Andrea D'Ambrogio (University of Rome TorVergata, Italy)

5:00 Human Performance Modeling in System of Systems Analytics
Craig Lawton (Sandia National Laboratories, USA); John Gauthier (Sandia National Laboratories, USA)

5:30 Efficient Systems Analysis by Combining SysML and Coevolution
David Morgan (BAE Systems, United Kingdom); Antony Waldock (BAE Systems, United Kingdom); David Corne (Heriot-Watt University, United Kingdom)

6:00 Basic Guidelines for Simulating SysML Models: An Experience Report
Mara Nikolaidou (Harokopio University of Athens, Greece); George Dimitrios Kapos (Harokopio University of Athens, Greece); Vassilis Dalakas (Harokopio University of Athens, Greece); Dimosthenis Anagnostopoulos (Harokopio University of Athens, Greece)

Smart Grids including Renewable Energies
Room: Mediterraneo
Chair: Stefano Massucco (University of Genoa & IEES Lab., Italy)

4:30 System of Systems Information Interoperability Using a Linked Dataspace
Edward Curry (National University of Ireland, Ireland)

5:00 Business Interactions Modeling for Systems of Systems Engineering: Smart Grid Example

Edin Arnautovic (Masdar Institute of Science and Technology, UAE); Davor Svetinovic (Masdar Institute of Science and Technology, UAE); Ali Diabat (Masdar Institute, UAE)

5:30 On Dynamic Models for Wind Farms as Systems of Systems
Kalev Rannat (Tallinn University of Technology, Estonia); Merik Meriste (Tallinn University of Technology, University of Tartu, Estonia); Leo Motus (Estonian Academy of Science & Tallinn University of Technology, Estonia); Jürgo S Preden (Tallinn University of Technology, Estonia)

6:00 Ground Properties Evaluation for the Design of Geothermal Heat Pump Systems and Uncertainty Measurement During the Thermal Response Test
Marco Fossa (University of Genoa, Italy); Davide Rolando (University of Genoa, Italy)

7:00 PM - 8:00 PM

Terrazza Acquamarina Welcome Cocktail
Room: Terrazza Acquamarina

Tuesday, July 17

8:00 AM - 8:30 AM

Registration at Lounge NH Marina Hotel

8:30 AM - 10:30 AM

Risks and Transportation
Room: Jonio
Chair: Carys Siemieniuch (Loughborough University, United Kingdom)

8:30 Towards Automatic Identification System of Maritime Risk Accidents by Rule-Based Reasoning Knowledge
Bilal Idiri (Mines ParisTech, France); Aldo Napoli (MINES PARISTECH, France)
9:00 Real-time Risk Definition in the Transport of Dangerous Goods by Road
Chiara Bersani (University of Genoa, Italy); Claudio Roncoli (University of Genoa, Italy)
9:30 Integration of a Bayesian Network for Response Planning in a Maritime Piracy Risk Management System
Xavier Chaze (Mines ParisTech & Centre de Recherche sur les Risques et les Crises (CRC), France); Amal Bouejla (Mines-ParisTech, France); Aldo Napoli (MINES PARISTECH, France); Franck Guarnieri (MINES PARISTECH, France)
10:00 Access and Monitor Vulnerability of Urban Metro Network System in China
Zhiru Wang (Southeast University Nanjing, P.R. China); Qiming Li (Southeast University Nanjing, P.R. China); Jingfeng Yuan (Southeast University Nanjing, P.R. China); Zhipeng Zhou (Southeast University Nanjing, P.R. China); Ruoyu Jia (Southeast University Nanjing, P.R. China); Ying Lu (Southeast University Nanjing, P.R. China)

Risk Management
Room: Adriatico
Chair: Murray Sinclair (University of Loughborough, United Kingdom)

8:30 DRIHM: Distributed Research Infrastructure for Hydro-Meteorology
Antonio Parodi (CIMA Research Foundation, Italy); Nicola Rebora (CIMA Research Foundation, Italy); Dieter A Kranzlmüller (Ludwig-Maximilians-Universitaet (LMU) Muenchen, Germany); Andrea Clematis (CNR, Italy); Michael Schiffers (Ludwig-Maximilians-Universität München, Germany); Antonella Galizia (CNR, Italy); Daniele D'Agostino (Italian National Research Council, Italy); Alfonso Quarati (CNR, Italy); Pierre-Henri Cros (CERFACS, France); Quillon Harpham (HR-Wallingford, United Kingdom); Bert Jagers (DELTARES, The Netherlands); Emanuele Danovaro (Italian National Research Council, Italy); Tatiana Bedrina (CIMA Research Foundation, Italy)
9:00 Modeling and Simulation for System Reliability Analysis: The RAMSAS Method
Alfredo Garro (University of Calabria, Italy); Andrea Tundis (University of Calabria, Italy)
9:30 SoS in Disasters: Why Following the Manual Can Be a Mistake
Antonella Cavallo (The University of Adelaide, Australia); Vernon Ireland (The University of Adelaide, Australia)
10:00 Distributed Control System (DCS) for Cernavoda NPP (Romania)

Francesco Benvenuto (Ansaldo Nucleare, Italy); Fabrizio Bianco (Ansaldo Nucleare, Italy)

Security and Safety for Complex System of Systems
Room: Tirreno
Chair: Jon Holt (Atego, United Kingdom)

8:30 The Holistic Military Capability Life Cycle Model
Jukka Anteroinen (National Defence University of Finland, Finland)

9:00 Aligning Analysis and Engineering Decision-Making Within a Military Distributed System of Systems
Jon Salwen (The MITRE Corporation, USA); Murray Daniels (The MITRE Corp., USA); Jeffrey Higginson (The MITRE Corp, USA); Tim W Rudolph (US Air Force & Electronic Systems Center, USA)

9:30 Systems Approach to the Safety of Complex Technical Facilities
Mikhail Belov (IBS, Russia)

10:00 Research on Capability Requirements Generation of Weapon System-of-systems Based on CRTAM Model
Yajie Dou (National University of Defense Technology, P.R. China); Long Li (National University of Defense Technology, P.R. China); Qingsong Zhao (National University of Defense Technology, P.R. China); Yingwu Chen (National University of Defence Technology, P.R. China)

Transportation Systems
Room: Mediterraneo
Chair: Mohammad Jamshidi (University of Texas at San Antonio, USA)

8:30 The Port as a System of Systems: a System Dynamics Simulation Approach
Claudia Caballini (University of Genoa, Italy); Simona Sacone (University of Genoa, Italy); Silvia Siri (University of Genoa, Italy)

9:00 Freeway Networks as Systems of Systems: An Event-Triggered Distributed Control Scheme
Antonella Ferrara (University of Pavia, Italy); Alberto Nai Oleari (University of Pavia, Italy); Simona Sacone (University of Genoa, Italy); Silvia Siri (University of Genoa, Italy)

9:30 A Selex-SI Solution to Enable Distributed Decision Making in Vessels Traffic Management
Aniello Napolitano (SESM Scarl, Italy); Stefano Gelli (Selex Sistemi Integrati, Italy); Dario Di Crescenzo (SESM s. c. a. r. l., Italy)

10:00 Operational and Real-Time Team Decision Problems in the Risk-Averse Transportation of Dangerous Goods by Road
Claudio Roncoli (University of Genoa, Italy); Michael Bell (Centre for Transport Studies, United Kingdom); Roberto Sacile (University of Genoa, Italy)

10:30 AM - 11:00 AM

Coffee Break
Room: Terrazza Acquamarina

11:00 AM - 12:00 PM

Keynote Speaker: "Cooperative and distributed systems for integrated logistics services" by Massimiliano Zazza & Antonella Di Fazio

Room: Mediterraneo

The freight transport benefits from advanced and innovative services based on ICT (Information and Communication Technologies. In fact, ICT equipped freight objects (such as vehicles, containers and shipments) which know their present location and are autonomous, self-aware and capable to communicate with their environment increase safety and security (at parking areas and on the move) that are critical for example in the transportation of precious and dangerous goods, allow efficiency of processes to be improved and enable new value adding features. Moreover, cooperative and distributed systems integrating mobile communications, advanced satellite navigation technologies (using EGNOS - European Geostationary Overlay Service - and Galileo), onboard processing and sensors, geographic information and enhanced digital maps, support all transport users, and improve traffic efficiency, safety and security. In the last decade, various projects have been implemented in Europe aimed at the development of interoperable platforms providing a large variety of efficiency, mobility, comfort and safety related added value services for the freight transport market. In this paper, two cases involving Italian stakeholders are presented, in the light of the European vision concerning the European-Wide Service Platform (EWSP): ¬- European projects devoted to the use of EGNOS for the road freight transport ¬- The on-going initiatives conducted in Italy in relation to the Italian ITS (Intelligent Transport Systems) Directive and the National Logistics Plan (NLP).

Antonella Di Fazio has a Degree in Physics. She works in the GNSS Infomobility Business Unit in Telespazio (a Finmeccanica/ Thales company), in charge of innovative applications & services. Since 2000 she has been the program and technical coordinator of European R&D projects, dealing with Satellite Navigation technologies applications and services in the transport domain (primarily road, freight and logistics sectors). In the last eight years she has been involved in activities devoted to the use of the European GNSS (EGNOS/Galileo) for regulated applications and services, in particular in road and land transport domains (such as dangerous goods transport, road charging, city logistics, regulated fleet management, freight security and intermodal transport). Antonella Di Fazio is also:- Member of the Board and the Technical & Scientific Committee of TTS Italia (the Italian ITS Association), in charge of satellite navigation technologies in ITS- Expert in TEN-T Expert Group on ITS & New technologies, set up as part of the TEN-T policy review.

Massimiliano Zazza Born on 18 July 1960 in Rome (Italy). He hold a Master’s Degree in Mathematics in 1992. From 1992 to 1995 he has developed algorithms in railway traffic optimization models and in 2009 he attended the European Senior Civil Servant course of excellence - SSPA. He is with Ministry of Infrastructure and Transport since 2006 where he is Head of ICT & ITS Project Unit, Responsible for MIT in ITS national/european projects and Responsible for MIT in the development of the Logistics National Platform. From 2002 to 2006 he worked for Ministry of Infrastructure and Transport as Head of ICT Project Unit, Project Manager in experimental project for transport of dangerous freight, Project manager for the development of the national WAN for the Ministry and Project manager for implementation of ERP system for HR and management control (SAP). From 1999 to 2001 he worked for Ministry of Transport and Navigation as Member of Y2K National Unit, Head of ICT Project Unit. He worked also on Policy of development IT systems and Planning and management of security assessment activities. From 1992 to 1998 we worked for South East Railways – ROME (ITALY) as Head of IT Unit developing Planning of BPR activities and IT systems.

12:00 PM - 1:00 PM

Keynote Speaker: "Making Decisions in Hazardous Transport Network" by Michael G.H. Bell
Room: Mediterraneo

This presentation reviews strategies for decision-making in transport networks subject to various forms of hazard. The range of hazards covered range from high probability / low consequence incidents to low probability / high consequence incidents. The key concepts of hyperarc, hyperpath and hypertour are introduced and the way they can be used to represent decision-making under uncertainty is discussed. Routeing and scheduling algorithms for hypernetworks are presented and discussed with respect to their structure and efficiency. Two applications will be used for illustration; the first is risk averse navigation in congested road networks, where the focus is on high probability / low consequence incidents, while the second is hazmat transportation, where the focus is on low probability / high consequence incidents. There is a discussion about how developments in information and communication technologies are transforming decision-making in uncertain networks and how these changes may be represented in hypernetworks. The presentation concludes by highlighting areas where future progress in the field can be expected. Michael Bell is Professor of Transport Operations and Director of the Port Operations Research and Technology Centre (PORTeC) at Imperial College London. His research and teaching interests span ports and maritime transport, logistics, transport network modelling, traffic engineering and control, intelligent transport systems, transit assignment and journey planning (in no particular sequence).

Michael Bell is Professor of Transport Operations and Director of the Port Operations Research and Technology Centre (PORTeC) at Imperial College London. His research and teaching interests span ports and maritime transport, logistics, transport network modelling, traffic engineering and control, intelligent transport systems, transit assignment and journey planning (in no particular sequence).

1:00 PM - 2:00 PM

Lunch Buffet
Room: Terrazza Acquamarina

2:00 PM - 2:30 PM

Keynote Speaker: "Wayside Monitoring Trains and Infrastructures: Information Management in a Railway Control Centre" by Nadia Mazzino, Ansaldo STS, Italy
Room: Mediterraneo

Control Centre collects information from different Wayside Monitoring Systems acquiring multimedia data and measurements related transit trains. Control Centre of Wayside Monitoring Systems allows operators to get real time status of the transit train and infrastructure, detecting defects and giving information/images to help the operators to take the most suitable decisions in order to avoid accidents. Moreover, managing the collected data, the railway maintenance process can be improved.

Nadia Mazzino hold a Laurea degree in Math in 1990 and she works in AnsaldoSTS since 1993. From 1993 to 1998, she worked on the requirement definition to the delivery of a new traffic command and control system. She was involved in the project of Roma Termini, studying and developing innovative graphic representation for the users to automatically command train movements. From 1999 to 2006 she was involved in a big project for Italian Railway: eight centers for command and control of traffic, PIS, D&M, Security. Each center manages about 80 stations. AT the beginning she worked on the detail project in the development area. Since 2002 she started the delivery activities, giving also assistance to the client, projects for new requirements, definition of scheduling. Since 2003 she has the responsibility of coordinating engineering department activities on all these systems. She was in a “continuous empowering project” to find out new solutions to make the product much more complete. From 2007 to nowadays, she works in a new area of the company (Innovation&Competitivness) she has the responsibility of the development of innovative projects and of the business for railway customers. The main projects are in the security area (development of a new Security Management System), risk reduction area (diagnostic wayside systems, level crossing protection).

2:30 PM - 3:00 PM

Keynote Speaker: "The Monitoring and Control of Processes and Systems in the Nuclear Power Plants and Facilities" by Francesco Benvenuto & Fabrizio Bianco, Ansaldo Nucleare, Italy
Room: Mediterraneo

As a result of the Chernobyl accident, a Shelter was constructed over the destroyed buildings in order to create a barrier between the damaged Chernobyl Nuclear Power Plant (ChNPP) Unit 4 and the external environment. Because of the risks related to the large quantities of nuclear material (especially long-lived radioactive materials) located in the Shelter and to its structural conditions, the EU and others international contributors funded an international plan, called Shelter Implementation Plan (SIP), to solve the problems of ChNPP Unit 4 "Shelter". In this plan framework, a specific task for monitoring the Shelter Object conditions was planned. The Integrated Automated Monitoring System (IAMS), developed by a Consortium of companies, led by Ansaldo Nucleare, is the system capable of monitoring the parameters of nuclear, radiation and industrial safety of the Shelter, providing information on the status of Shelter Object, local area and Chernobyl NPP site.

3:00 PM - 4:00 PM

Keynote Speaker: "Requirement Development Management for Large Systems, Modelling the Software Architecture of Large Systems, MDA Approach for the Realization of Large Systems, A Model-driven Approach for Configuring and Deploying Systems of Systems" by Antonino Arecchi and Fabrizio Morciano, SELEX SI, Italy
Room: Mediterraneo

SELEX Sistemi Integrati (SELEX SI) is involved in the design, development and integration of Systems of Systems for Naval Defence, Homeland Defence, Homeland Security, Air Traffic Management and Maritime Domain Awareness. SELEX SI is focusing attention on requirement management, architecture modelling and model driven approach for the realization of systems of systems. A Company Unified Requirements Platform, has been set up to manage the requirements of all the business Products at the different levels, ensuring the bidirectional traceability of requirements, keeping track of their evolutions and changes, enabling the requirements reuse across the projects and their customizations for the specific Customer needs. Guidelines have been developed and applied to ensure a homogeneous UML modelling approach to the definition of reference architectures for different domains, which are used as an inspiration in the design of concrete architectures for specific projects. Requirements and design models are the main inputs for system implementation, including automatic generation of interfaces through Model Driven Architecture approach, which is used also in the integration, deployment and maintenance of Large Systems to produce configuration data and feed testing and integration tools.

Antonino Arecchi is currently the head of System Architecture and Integration department in the Defence System Business Unit of Selex Sistemi Integrati. He works in the company since 1988, mostly in the naval domain, participating to the main naval missile defence, coastal surveillance and naval combat system projects developed by Selex Sistemi Integrati in the past 20 years. In his career he has covered the roles of FREMM program manager, head of Combat System Engineering department, head of Combat Management Systems department, project manager of Horizon Combat Management System. Before the current job, he has worked in the Software Architecture department where he has contributed to the set-up of the Company Unified Requirements Platform and the definition of architecture modelling guidelines.

Fabrizio Morciano is head of of Software for Flight Data Plans and Tools department. He works in Selex Sistemi Integrati since 2002 following activities related to middleware for Air Traffic Management. He has been working on Cardamom middleware (a CORBA, CCM middleware) for 8 years in the role of Technical Architect and he represented Selex Sistemi Integrati at Object Management Group for 5 years following the standardization process of Enhanced View Of Time specification, Load Balancing specification and Application Management and System Monitoring specification

4:00 PM - 4:30 PM

Coffee Break
Room: Terrazza Acquamarina

4:30 PM - 6:30 PM

Risks and Transportation
Room: Mediterraneo
Chair: Craig Wrigley (Lockheed Martin UK, United Kingdom)

4:30 A System of Systems Approach to Near Miss Accidents in the Transport of Dangerous Goods by Road
Silvia De Nadai (University of Genoa, Italy); Francesco Parodi (University of Genoa, Italy); Domenico Pizzorni (ENI, Italy)

5:00 Intelligent Transport Systems (ITS) Applications on Dangerous Good Transport on Road in Italy
Mauro Benza (University of Genoa, Italy); Chiara Bersani (University of Genoa, Italy); Massimo D'Incà (University of Genoa, Italy); Claudio Roncoli (University of Genoa, Italy); Anita Trotta (University of Genoa, Italy); Domenico Pizzorni (ENI, Italy); Stefania Briata (ENI, Italy); Riccardo Ridolfi (ENI, Italy)

5:30 Predictive Transportation Control of a Complex Dynamical System for High TP and Short TAT
Yoshiyuki Tajima (Hitachi, Ltd., Japan); Takashi Noguchi (Yokohama Research Laboratory, Hitachi, Ltd., Japan); Takashi Fukumoto (Hitachi, Ltd., Japan)

6:00 Modeling and Optimization of Aircraft Trajectories: a Review
Maria Pia Fanti (Politecnico di Bari, Italy); Giovanni Pedroncelli (University of Trieste, Italy); Gabriella Stecco (University of Trieste, Italy); Walter Ukovich (Univbersity of Trieste, Italy)

4:30 PM - 6:00 PM

Risk Management
Room: Adriatico
Chair: Antony Waldock (BAE Systems, United Kingdom)

4:30 Causal Factors Behind the Failed FiReControl Project: a Large-Scale System-of-Systems
Cornelius Ncube (Bournemouth University, United Kingdom)

4:52 Enhancement of Ontology with Spatial Reasoning Capabilities to Support Maritime Anomaly Detection
Arnaud Vandecasteele (Mines ParisTech & Centre de Recherche sur les Risques et les Crises (CRC), France); Aldo Napoli (MINES PARISTECH, France)

5:15 A New Embedded E-Nose System to Identify Smell of Smoke
Salaheddin Sadeghifard (South Pars Gas Complex, Iran); Leili Esmaeilani (South Pars Gas Complex, Iran)

5:37 A System of Systems for Air Quality Decision Making
Claudio Carnevale (University of Brescia, Italy); Giovanna Finzi (University of Brescia, Italy); Enrico Pisoni (University of Brescia, Italy); Marialuisa Volta (University of Brescia, Italy)

Security and Safety for Complex System of Systems
Room: Tirreno
Chair: Ella-Mae Hubbard (Loughborough University, United Kingdom)

4:30 Technology Contribution Rate: Concepts, Framework and Case Study
Hanlin You (National University of National Defense, P.R. China); Qingsong Zhao (National University of Defense Technology, P.R. China); Jiang Jiang (Department of Management, P.R. China); Yanjing Lu (National University of Defence Technology, P.R. China)

5:00 Technology System of Systems: Concepts and Hierarchical Structure
Leilei Chang (National University of Defense Technology, P.R. China); Yanjing Lu (National University of Defence Technology, P.R. China); Qingsong Zhao (National University of Defense Technology, P.R. China); Jiang Jiang (Department of Management, P.R. China)

5:30 Research on Evolving Capability Requirements Oriented Weapon System of Systems Portfolio Planning
Zhou Yu (College of Information System and Management， National University of Defense Technology, P.R. China); Kewei Yang (National University of Defence Technology, P.R. China)

4:30 PM - 6:30 PM

Security and Safety for Complex System of Systems

Room: Jonio

Chair: Cornelius Ncube (Bournemouth University, United Kingdom)

4:30 Using Indicators for System Complex Safety
Tullio Tanzi (Mines ParisTech, France); Raoul Textoris (Mines ParisTech, France)

5:00 Advancing the Defense in Depth Model
Stephen L Groat (Virginia Tech, USA); Randy Marchany (Virginia Tech, USA); Joseph G Tront (Virginia Tech, USA)

5:30 Threshold Design for Low Cost Wave Sensors Through Statistical Analysis of Data
Maricris C. Marimon (Nara Institute of Science and Technology, Japan); Kenji Sugimoto (Nara Institute of Science and Technology, Japan)

6:00 Shutdown Reduction Methods for Compressors in Condensate Stabilization Units
Elaheh Esfandiari Jahromi (Gas Refinery Staff & South Pars Gas Complex, Iran); Leili Esmaeilani (South Pars Gas Complex, Iran); Salaheddin Sadeghifard (South Pars Gas Complex, Iran); Alireza Niknam (South Pars Gas Complex, Iran)

Wednesday, July 18

8:00 AM - 8:30 AM

Registration at Lounge NH Marina Hotel

8:30 AM - 10:30 AM

Decision Making in System of Systems
Room: Mediterraneo
Chair: Alvaro Miyazawa (University of York, United Kingdom)

8:30 A Framework for Enabling an Integrated and Proactive Decision Making in Airport Systems
Dario Di Crescenzo (SESM s. c. a. r. l., Italy); Aniello Napolitano (SESM Scarl, Italy); Massimo Loffreda (SESM s. c. a. r. l., Italy)

9:00 The Impact of Multi-Institutional Semi-Structured Learning Environments (MISSLE)
Raymond R. Buettner, Jr. (Naval Postgraduate School, USA)

9:30 System of Systems to Provide QoS Monitoring, Management and Response in Cloud Computing Environments
Paul C. Hershey (Raytheon, Inc., USA); Shrisha Rao (International Institute of Information Technology, Bangalore, India); Charles B. Silio Jr. (University of Maryland at College Park, USA); Akshay Narayan (International Institute of Information Technology - Bangalore, India)

10:00 Design and Realization of the Simulation Component-based Parallel Framework for HLA Federate
Jing Zhang (Science and Technology on Complex Systems Simulation Laboratory, P.R. China); Zhang Yingchao (Science and Technology on Complex Systems Simulation Laboratory, P.R. China); Yu Qin zhang (Yard 10, AnXiangBeiLi, Chao Yang District, Beijing, P.R. China); Guan Chuan fang (Yard 10, AnXiangBeiLi, Chao Yang District, Beijing, P.R. China); Li Wei (Yard 10, AnXiangBeiLi, Chao Yang District, Beijing, P.R. China); Peng Yong (NUDT, Changsha, Hunan, P.R. China)

Robotics
Room: Jonio
Chair: Hanane Dagdougui (University of Genoa, Italy)

8:30 Reconstitution of Electromyographic Signals From Pen-Tip Velocity
Inès Chihi, I. (Ecole Nationale d'Ingènieur de Tunis & Tunis, Tunisia); Afef Abdelkrim (Ecole Supérieure de Technologie et d'Informatique & Ecole Nationale d'Ingénieurs de Tunis, France); Mohamed Benrejeb (ENIT, Universiy of Tunis, Tunisia)

9:00 Validation of Swarms of Robots: Theory and Experimental Results
Luis Mendez (Carleton University, Canada); Sidney Givigi (Royal Millitary College of Canada, Canada); Howard Schwartz (Carleton University, Canada); Alain JG Beaulieu (Royal Military College of Canada, Canada); Gerard Pieris (Defence R &D Canada, Canada); Giovanni Fusina (Defence R&D Canada - Ottawa, Canada)

9:30 An Autonomous Image-guided Robotic System Simulating Industrial Applications

Raza Ul Islam (COMSATS Institute of Information Technology, Pakistan); Jamshed Iqbal (COMSATS Institute of Information Technology, Pakistan); Sarah Manzoor (COMSATS Institute of Information Technology, Pakistan); Aayman Khalid (COMSATS Institute of Information Technology, Pakistan); Sana Khan (COMSATS Institute of Information Technology, Pakistan)

10:00 A System Architecture for Heterogeneous Moving Objects Trajectory Models Using Different Sensors
Azedine Boulmakoul (FST Mohammedia MOROCCO, Morocco); Lamia Karim (FST Mohammedia MOROCCO, Morocco); Ahmed Lbath (University of Grenoble 1 - Joseph Fourier - LIG Lab, France); Adil Elbouziri (FST Mohammedia MOROCCO, Morocco)

Software Architecture

Room: Tirreno

Chair: Huseyin Dogan (Loughborough University & BAE Systems, United Kingdom)

8:30 Exploiting Cloud Computing for Enabling Distributed Testing of Complex Systems:The SELEX-SI Roadmap
Gabriella Carrozza (SESM Scarl, Italy); Massimo Loffreda (SESM s. c. a. r. l., Italy); Vittorio Manetti (SESM scarl, Italy)

9:00 Automated Context Aware Composition for Convergent Services
Armando Ordonez (University of Cauca, Colombia); Juan Carlos Corrales (University of Cauca, Colombia); Paolo Falcarin (University of East London, United Kingdom)

9:30 Executable System-of-Systems Architecting Based on DoDAF Meta-model
Long Li (National University of Defense Technology, P.R. China); Yajie Dou (National University of Defense Technology, P.R. China); Bingfeng Ge (National University of Defense Technology & University of Waterloo, Canada); Kewei Yang (National University of Defense Technology, P.R. China); Yingwu Chen (National University of Defence Technology, P.R. China)

10:00 A Data-Centric Executable Modeling Approach for System-of-Systems Architecture
Bingfeng Ge (National University of Defense Technology & University of Waterloo, Canada); Hipel Keith (University of Waterloo, Canada); Long Li (National University of Defense Technology, P.R. China); Yingwu Chen (National University of Defence Technology, P.R. China)

SoSE Strategies

Room: Adriatico

Chair: Simon Ford (University of Cambridge, United Kingdom)

8:30 Governance Mechanisms in System of Systems
Hamid R. Darabi (Stevens Institute of Technology, USA); Alex Gorod (The University of Adelaide, Australia); Mo Mansouri (Stevens Institute of Technology, USA)

9:00 Approaches in Addressing System of Systems
Vernon Ireland (The University of Adelaide, Australia); Antonella Cavallo (The University of Adelaide, Australia); Amina Omarova (The University of Adelaide, Australia); Yasmin Ooi-Sanches (Thales, Australia); Barbara Rapaport (The University of Adelaide, Australia)

9:30 Understanding the Dynamics of System-of-Systems in Complex International Negotiations
Barbara Rapaport (The University of Adelaide, Australia); Vernon Ireland (The University of Adelaide, Australia); Alex Gorod (The University of Adelaide, Australia)

10:00 Kony 2012 Movement Through a System of Systems Engineering Lens
Ryley Smithson (University of Adelaide, Australia)

10:30 AM - 11:00 AM

Coffee Break
Room: Terrazza Acquamarina

11:00 AM - 12:30 PM

Decision Making in System of Systems
Room: Jonio
Chair: Cornelius Ncube (Bournemouth University, United Kingdom)

11:00 Dynamic Modularity: A Distributed Decision Mechanism in System of Systems
Babak Heydari (Stevens Institute of Technology, USA); Kia Dalili (Stevens Institute of Technology, USA)
11:22 ARCNET: A Systems-of-Systems Architecture Resource Collaborative Network Evaluation Tool
Jean C Domercant (Georgia Institute of Technology & Aerospace Systems Design Laboratory, USA); Kelly Griendling (Georgia Institute of Technology, USA)
11:45 A Mathematical Model for Formulating Interdependence of Autonomy and Belonging in System of Systems
Hamid R. Darabi (Stevens Institute of Technology, USA); Mo Mansouri (Stevens Institute of Technology, USA); Alex Gorod (The University of Adelaide, Australia)
12:07 Integrated Approach and Decision Support Algorithms for Complex Systems Effectiveness Evaluation
Simeone Solazzi (SELEX Sistemi Integrati, Italy); Francesco Ciambra (SELEX Sistemi Integrati, Italy); Michele Sinisi (SELEX Sistemi Integrati, Italy)

11:00 AM - 1:00 PM

Software Architecture
Room: Tirreno
Chair: Vincenzo Arrichiello (SELEX Sistemi Integrati, Italy)

11:00 Self-aware Architecture to Support Partial Control of Emergent Behavior
Leo Motus (Estonian Academy of Science & Tallinn University of Technology, Estonia); Merik Meriste (Tallinn University of Technology, University of Tartu, Estonia); Jürgo S Preden (Tallinn University of Technology, Estonia); Raido Pahtma (Tallinn University of Technology, Estonia)
11:30 Application of Component Engineering to the Design of Holistic Spell Checking Algorithm
Leena Alhussaini (University of Edinburgh, United Kingdom)
12:00 The Method of Analyzing Mapping Between Capability and Performance Index Based on DSM/DMM Models
Kewei Yang (National University of Defence Technology, P.R. China); Yanjing Lu (National University of Defence Technology, P.R. China); Jie Mao (National University of Defense Technology, P.R. China); Zhiwei Yang (National University of Defense Technology, P.R.

China); Long Li (National University of Defense Technology, P.R. China); Qingsong Zhao (National University of Defense Technology, P.R. China)

12:30 Embedded Concurrent Computing Architecture Using FPGA
Muataz Hameed Salih (UniMap, Malaysia); Badlishah R Ahmad (Universiti Malaysia Perlis, Malaysia); Abid Yahya (University Malaysia Perlis, Malaysia); Mohd. Arshad (USM, Malaysia)

11:00 AM - 12:35 PM

Contract-based Modelling and Analysis Tehnologies for Systems-of-Systems

Room: Mediterraneo
Chair: Steve Riddle (Newcastle University, United Kingdom)

11:00 Features of CML: a Formal Modelling Language for Systems of Systems
James Woodcock, A. Cavalcanti (University of York, United Kingdom); J. Fitzgerald (Newcastle University, United Kingdom); P. Larsen (Aarhus University, Denmark), Alvaro Miyazawa (University of York, United Kingdom), S. Perry (Atego, United Kingdom)

11:31 COMPASS Tool Vision for a System of Systems Collaborative Development Environment
Joey Coleman, Anders Malmos, Peter Gorm Larsen (Aarhus University, Denmark); Jan Peleska (University of Bremen, Germany); Ralph Hains (Atego, United Kingdom); Zoe Andrews, Richard Payne (Newcastle University, United Kingdom); Simon Foster, Alvaro Miyazawa (University of York, United Kingdom); Cristiano Bertolini, André LR Didier (Federal University of Pernambuco, Brazil)

12:03 Extending VDM-RT to Enable the Formal Modelling of System of Systems
Claus Nielsen (Aarhus University, Denmark); Peter Gorm Larsen (Aarhus University, Denmark)

11:00 AM - 1:00 PM

SoSE Strategies

Room: Adriatico
Chair: Ahmed Ouammi (University of Genoa, Italy)

11:00 System of Systems: "Defining the System of Interest"
Vishal Barot (Loughborough University, United Kingdom); Andrew Kinder (Loughborough University, United Kingdom); Michael Henshaw (Loughborough University, United Kingdom); Carys Siemieniuch (Loughborough University, United Kingdom)

11:30 Influence Strategies for Systems of Systems
Nirav Shah (MIT & Systems Engineering Advancement Research Initiative, USA); Joseph Sussman (MIT, USA); Donna H. Rhodes (Massachusetts Institute of Technology, USA); Daniel E Hastings (MIT, USA)

12:00 Modelling Systems-of-Systems: Issues and Possible Solutions
Craig Wrigley (Lockheed Martin UK, United Kingdom)

12:30 A Semantic Mediation Framework for Architecting Federated Ubiquitous Systems
Georgios Moschoglou (The George Washington University, USA); Timothy J Eveleigh (The George Washington University & CACI International, USA); Thomas Holzer (The George Washington University, USA); Shahryar Sarkani (The George Washington University, USA)

1:00 PM - 2:00 PM

Lunch Buffet
Room: Terrazza Acquamarina

2:00 PM - 4:00 PM

Complex Dynamic Systems
Room: Adriatico
Chair: Roberto Sacile (University of Genoa, Italy)

2:00 An Application to Two-Hop Forwarding of a Model of Buffer Occupancy in ICNs
Marco Cello (University of Genoa, Italy); Giorgio Gnecco (University of Genoa, Italy); Mario Marchese (DIST- University of Genoa, Italy); Marcello Sanguineti (University of Genoa, Italy)

2:30 Natural Language Processing Based Services Composition for Environmental Management
Armando Ordonez (University of Cauca, Colombia); Juan Carlos Corrales (University of Cauca, Colombia); Paolo Falcarin (University of East London, United Kingdom)

3:00 Proposal of SymbiosisADS Concept and Negotiation Support Methods for Cooperative Resource Allocation
Koichiro Iijima (Hitachi, Ltd., Japan); Takashi Fukumoto (Hitachi, Ltd., Japan); Michiki Nakano (Hitachi, Ltd., Japan)

3:30 System of Systems Complexity and Decision Making
Zhang Yingchao (Science and Technology on Complex Systems Simulation Laboratory, P.R. China)

Decision Making in System of Systems
Room: Jonio
Chair: Raymond R. Buettner, Jr. (Naval Postgraduate School, USA)

2:00 Modeling System of Systems Acquisition
Nil Ergin (Penn State University, USA); Paulette Acheson (Missouri University of Science and Technology, USA); John Colombi (Air Force Institute of Technology, USA); Cihan H Dagli (Missouri University of Science and Technology, USA)

2:30 The Originating Concept: a Foundation for System of Systems Architecting Decision Making
Vincenzo Arrichiello (SELEX Sistemi Integrati, Italy)

3:00 An Interoperable Reconstruction and Recovery Decision Support Tool for Complex Crises Situations
Francesca Matarese (SESM Scarl a Finmeccanica Company, Italy); Dario Di Crescenzo (SESM s. c. a. r. l., Italy); Antonio Strano (SESM Scarl a Finmeccanica Company, Italy); Florence Aligne (Thales Research and Technology, France); Juliette Mattioli (Thales Research & Technology France, France)

3:30 A Systems-Of-Systems Approach to the Development of Flexible Cost-Effective Training Environments"
Luminita Ciocoiu (Loughborough University, United Kingdom); Michael Henshaw (Loughborough University, United Kingdom); Ella-Mae Hubbard (Loughborough University, United Kingdom)

Manufacturing in SoSE
Room: Tirreno
Chair: Kelly Griendling (Georgia Institute of Technology, USA)

2:00 A System of Systems Architecture Framework (SoSAF) for Production Industries
Asif Mahmood (Politecnico Di Torino, Italy); Francesca Montagna (Politecnico di Torino, Italy)

2:30 System-of-system Approaches and Challenges for Multi-Site Manufacturing
Simon Ford (University of Cambridge, United Kingdom); Ursula Rauschecker (Fraunhofer IPA, Germany); Nikoletta Athanassopoulou (IfM-ECS, United Kingdom)

3:00 System of Systems Thinking in Product Development Processes: A System Dynamic Approach
Alemu Moges Belay (University of Vaasa, Finland); Petri Helo (University of Vaasa, Finland); Torgeir Welo (Department of Engineering Design and Materials, Norway)

3:30 A Mathematical Framework for the Planning and Control of Complex Systems
Davide Giglio (University of Genoa, Italy); Simona Sacone (University of Genoa, Italy); Silvia Siri (University of Genoa, Italy)

Contract-based Modelling and Analysis Tehnologies for Systems-of-Systems
Room: Mediterraneo
Chair: Steve Riddle (Newcastle University, United Kingdom)

2:00 Model-based Requirements Engineering for System of Systems
Jon Holt (Atego, United Kingdom)

2:40 Technical Challenges of SoS Requirements Engineering
Stefan Hallerstede (Aarhus University, Denmark); Finn Hansen (Aarhus University, Denmark); Jon Holt (Atego, United Kingdom); Rasmus Lauritsen (Aarhus University, Denmark); Lasse Lorenzen (Bang and Olufsen, Denmark); Jan Peleska (University of Bremen, Germany)

3:20 Interface Specification for System-of-Systems Architectures
Richard Payne (Newcastle University, United Kingdom); Jeremy Bryans (Newcastle University, United Kingdom); John Fitzgerald (Newcastle University, United Kingdom); Steve Riddle (Newcastle University, United Kingdom)

4:00 PM - 4:30 PM

Coffee Break
Room: Terrazza Acquamarina

4:30 PM - 5:30 PM

Keynote Speaker: "New Consensus Protocols for Agent Networks with Discrete Time Dynamics and Distributed Task Assignent" by Maria Pia Fanti
Room: Mediterraneo

In recent years the study of the consensus problem in multi-agent systems has received a great effort by the scientific community involving several fields and many applications. More precisely, in networks of autonomous agents, “consensus” means to reach an agreement regarding a certain quantity of interest that depends on the state of all the agents. The solution of the consensus problem is obtained in presence of severely restricted communication capability. Moreover, a consensus algorithm (or protocol) is an interaction rule that specifies the information exchange between an agent and all of its neighbors on the network. This lecture intends to present the theoretical frameworks for posing and solving consensus problems in two different contexts. Firstly, we consider networked systems with discrete time dynamics that in the related literature are described by directed or undirected graphs and the associated graph Laplacian matrix. Since the standard protocols exhibit low speed of reaching a consensus for particular topologies of the graphs, new classes of convergent consensus algorithms are proposed. The talk shows the convergence properties of such iterative schemes and specifies the protocols that each agent has to use. Secondly, the lecture deals with the consensus problem over network of agents with quantized state variables. In particular, a novel distributed algorithm is presented for a multi-agent assignment problem in which a group of agents has to reach a consensus on an optimal distribution of tasks among themselves. The problem is formalized as a distributed consensus strategy that is applied to networks where each agent can perform a maximum number of tasks, each task can be assigned to a subset of agents and a cost is associated to the pair agent-task. The task assignment problem is solved when all the agents agree on the assignment that guarantees an optimal (or suboptimal) total load.

Maria Pia Fanti received the Laurea degree in electronic engineering from the University of Pisa, Italy in 1983 and was a visiting researcher at the Rensselaer Polytechnic Institute of Troy, New York, in 1999. Since 1983 she has been with the Department of Electrical and Electronic Engineering of Politecnico di Bari (Italy), where she was assistant professor from 1990 till 1998 and an associate professor from 1998 till April 2012. Now she is full professor of System and Control Engineering in the same department. Her research interests include discrete event systems, Petri net, sensor networks, management and modeling of automated manufacturing systems, automatic guided vehicle systems, railway and traffic networks, supply chains, and healthcare systems. She has published around 180+ papers and one textbook on these topics. She was General Chair of the 2nd IFAC Workshop on Dependable Control of Discrete Systems, the 2010 IEEE Workshop on Health Care Management, and the 2011 IEEE Conference on Automation Science and Engineering.. Prof. Fanti is Associate Editor of IEEE TRANS. ON Systems, Man, and Cybernetics Part A and Enterprise Information Systems, guest editor of IEEE TRANS. ON Automation Science and Engineering, Enterprise Information Systems. She is vice-chair of the IEEE Italy Section, Co-Chair of the IEEE SMC Technical committee on Discrete Event Systems, Chair of the Central & Southern Italy SMC Chapter, member of the IFAC Technical Committee on Discrete Event and Hybrid Systems.

5:30 PM - 6:30 PM

Keynote Speaker: "Cooperative and competitive distributed circuits and systems: from Art to Time" by Luigi Fortuna & Mattia Frasca

Room: Mediterraneo

The research discussed in the proposed lecture deals with a bridge between electronic nonlinear circuits and some aspects of human thinking. Several experiments obtained in our laboratory, where cooperative and competitive distributed circuits and systems are used, are discussed. The case studies we present explain in detail as aesthetics can be viewed as an emergent behaviour of dynamical systems. In particular, the following experiments will be dealt with: 1) The generation of three dimensional shapes by 3D large scale arrays of dynamical systems with chaotic behaviour . 2) The generation of patterns equal to that appearing in modern art by using Cellular Nonlinear Networks (CNNs) . 3) The 3D dynamical artistic trajectories generated by using synchronized circuits in controlled robots that can be related to kinematic art. The three case studies can be considered examples of the fact that art arises as an emergent behaviour. Moreover, patterns generated by a one-dimensional array of cooperative cells arranged in the so-called Wolfram Machine will be also introduced. Finally, by using cooperative three-cells based electronic circuits with chaotic behavior the universal concept of "Time" will be discussed.

Luigi Fortuna was born in Siracuse, Italy, in 1953. He received the degree of electrical engineering (cum laude) from the University of Catania, Catania, Italy, in 1977.He is a Full Professor of system theory with the Università degli Studi di Catania, Catania, Italy. He was the Coordinator of the courses in electronic engineering and the Head of the Dipartimento di Ingegneria Elettrica Elettronica e dei Sistemi (DIEES). Since 2005, he has been the Dean of the Engineering Faculty, Catania. He currently teaches complex adaptive systems and robust control. He has published more than 450 technical papers and is the coauthor of ten scientific books, among which: Chua's Circuit implementations (World Scientific, 2009), Bio-Inspired Emergent Control of Locomotion Systems (World Scientific, 2004), Soft-Computing (Springer 2001), Nonlinear Non Integer Order Circuits and Systems (World Scientific 2001), Cellular Neural Networks (Springer 1999), Neural Networks in Multidimensional Domains (Springer 1998), Model Order Reduction in Electrical Engineering (Springer 1994), and Robust Control-An Introduction (Springer 1993). His scientific interests include robust control, nonlinear science and complexity, chaos, cellular neural networks, softcomputing strategies for control. robotics, micronanosensor and smart devices for control, and nanocellular neural networks modeling. Dr. Fortuna was the IEEE Circuits and Systems (CAS) Chairman of the CNN Technical Committee, IEEE CAS Distinguished Lecturer from 2001 to 2002, and IEEE Chairman of the IEEE CAS Chapter Central-South Italy.

Mattia Frasca received his Ph.D. in Electronics and Automation Engineering in 2003, at the University of Catania, Italy, where he is currently research associate. His scientific interests include nonlinear systems and chaos, Cellular Neural Networks, complex systems and bio-inspired robotics. He is involved in many research projects and collaborations with industries and academic centres. He is referee for many international journals and conferences. He was in the organizing committee of the 10th "Experimental Chaos Conference" and co-chair of the "4th International Conference on Physics and Control". He is coauthor of three research monographs (with World Scientific): one on locomotion control of bio-inspired robots, one on self-organizing systems and one on the Chua's Circuit. He published more than 150 papers on refereed international journals and international conference proceedings and is co-author of two international patents.

8:45 PM - 11:59 PM

Gala Dinner at Genoa Aquarium "Eating with the Sharks"

Thursday, July 19

9:00 AM - 9:45 AM

Keynote Speaker: "Programme of the European Embedded Systems Engineering objective, European Commission INFSO G3" by Werner Steinhögl
Room: Mediterraneo

10:00 AM - 10:15 AM

Coffee Break
Room: Terrazza Acquamarina

10:15 AM - 12:15 PM

State of the art of some SoSE EU projects
Room: Mediterraneo
Chair: Michael Henshaw (Loughborough University, United Kingdom)

10:15 Development of Strategic Research and Engineering Roadmaps in Systems of Systems Engineering and Related Case Studies: The ROAD2SOS Project
Simon Ford (University of Cambridge, United Kingdom)

10:45 Comprehensive Modelling for Advanced Systems of Systems: The Compass Project
John Fitzgerald (Newcastle University, United Kingdom)

11:15 Designing for Adaptability and Evolution in Systems of Systems Engineering: The DANSE Project
Michael Winokur (Israel Aerospace Industry, Israel)

11:45 Trans-Atlantic Research and Education Agenda in Systems of Systems: The T-AREA-SoS Project
Carys Siemieniuch (Loughborough University, United Kingdom)

12:15 PM - 12:45 PM

Debate: the prioritisation of SoS research for future benefit to Commerce and Society
Room: Mediterraneo

Panel members are: Prof. Roberto Sacile (Genoa University), Dr. Dan DeLaurentis (Purdue University), Dr. Cornelius Ncube (Bournemouth University), Dr. Werner Steinhögl (European Commission).

12:45 PM - 1:00 PM

Concluding remarks and arrivederci to IEEE SOSE2013 - Hawaii
Room: Mediterraneo

A B C D E F G H I J K L M N O P Q R S T U V W Y Z

2012 7th International Conference on System of Systems Engineering

Smart Grids including Renewable Energies

Smart Grids including Renewable Energies

Smart Grids including Renewable Energies

Risks and Transportation

Risk Management

Security and Safety for Complex System of Systems

Transportation Systems

Security and Safety for Complex System of Systems

Security and Safety for Complex System of Systems

Decision Making in System of Systems

Robotics

Software Architecture

SoSE Strategies

Decision Making in System of Systems

Software Architecture

Contract-based Modelling and Analysis Technologies for Systems-of-Systems

SoSE Strategies

SoSE Strategies

Complex Dynamic Systems

Decision Making in System of Systems

Manufacturing in SoSE

Contract-based Modelling and Analysis Technologies for Systems-of-Systems

Exploring Human Factors Effects in the Smart Grid System of Systems Demand Response

Michael Z. Miller
Georgia Insititute of Technology
Aerospace Systems Design Laboratory
Atlanta, GA, U.S.A.
mzmiller@asdl.gatech.edu

Kelly Griendling
Georgia Insititute of Technology
Aerospace Systems Design Laboratory
Atlanta, GA, U.S.A.
kgriendling@asdl.gatech.edu

Dimitri N. Mavris
Georgia Insititute of Technology
Aerospace Systems Design Laboratory
Atlanta, GA, U.S.A.
dimitri.mavris@aerospace.gatech.edu

Abstract – *Human factors can play a significant role in the behavior of a system of systems. This is particularly true in collaborative system of systems, an example of which is the Smart Grid Demand Response. The Smart Grid Demand Response example is used in this paper to develop and prototype an approach to examine the impact of human factors (e.g., news stories and human interactions) on system of systems performance (e.g., peak load). The approach presented makes use of an agent-based model combined with probabilistic analysis to show the sensitivity of the performance of the Demand Response to uncertainty in human's decision to participate in a direct load control program. The human's decision could be influenced by several factors, including interactions with other humans and the news. Since most of the information regarding human factors is not known a priori, a sensitivity analysis was conducted in order to examine emergent behavior. The results of this study show that human factors can have a significant impact on a collaborative system of systems. In this study, news stories and human interactions have a significant influence on the effectiveness of the notional Demand Response program modeled.*

Keywords: System of Systems, Smart Grid, Demand Response, Agent-Based Modeling, Human Interactions.

1 Introduction

The field of system of systems (SoS) has developed over the last decade as a result of the emergence of information technologies allowing for systems to have virtual (rather than physical) interfaces. SoS are comprised of individual, independently useful systems that are brought together to achieve a purpose greater than can be achieved by the component systems alone [1, 2, 3]. SoS have been loosely categorized into four types: virtual, collaborative, acknowledged, and directed based on the level of decision-making and acquisition power held by individuals or organizations within the SoS, with virtual having the most distributed decision making and directed having the most centralized decision making [3]. The Smart Grid, as envisioned in this paper, is what is known as a collaborative SoS, where independent users choose to conform to a standard technology (the Smart Grid Demand Response (DR)) to reap the benefits of the SoS [3]. One key challenge with collaborative SoS is that if the adoption rates are too low, the all-around benefits are lost. As such, the decision of users to participate, and the decision of power generators/distributors to uphold the Smart Grid DR is dependant on the participation of some critical mass of users that allows for a large enough reduction in the demand curve at peak times. This reduction in the peak demand is advantageous for multiple reasons, including that backup generators do not have to fire, thus reducing the cost of providing electricity, as well as reducing the environmental impacts. These cost savings are the main incentive for the users to enroll. However, there are some negative effects to users as well. Consumers who opt-in would choose to give up some level of control over the appliances in their home, such as allowing for leeway in the temperature settings on the thermostat. Furthermore, there may be an initial investment required on the part of the users to obtain the technology, although an alternate model may be for power companies to subsidize the upfront cost as a way to entice users to participate. It can be seen that the consumer has many considerations before deciding whether or not to enroll.

There are many engineering challenges unique to SoS, and especially collaborative SoS, but the most relevant challenge to this work is the autonomy of the human elements. In this case, users of the electricity grid can choose to participate in a Smart Grid program, and their decisions will drive the success or failure of the system. Too few agents will not achieve a significant enough reduction in the peak load. This paper describes the development of an Agent-Based Modeling (ABM) approach to explore human factors elements influencing the success of the Smart Grid. There are two primary elements being explored: first, the impact of users interacting with one another and sharing opinions about the Smart Grid, thereby influencing neighbors and friends to participate or not; and second, the impact of positive or negative press regarding the Smart Grid, and the impact of this press on users' decisions to participate. An ABM approach was selected because this type of modeling approach allows for a study of individual behaviors and interactions between individuals leading to emergent behaviors within the population, and thus was deemed most appropriate.

1.1 Smart Grid Demand Response

The current United States energy grid was conceived almost a century ago and is currently reaching its physical and technological limits [4]. The next generation of energy grid is often referred to as the Smart Grid, which utilizes two-way communication between the residential energy-consumer and the energy producer/distributor. This enables many other Smart Grid technologies, such as: Fault Isolation, Load Balancing, Distributed Generation/Storage, DR, etc. [4]. Of particular interest in this paper are those technologies associated with DR, which work to reduce peak loads by allowing the power generation/distribution companies to have some level of influence over consumer's energy consumption [5]. Reducing the peak load can prevent power companies from having to use expensive, inefficient peaking plants in order to meet the energy demand. However, the success of these technologies is dependent on the willingness of a critical threshold of users to participate, and thus these technologies require an understanding of the Smart Grid as a SoS. The SoS is comprised of the combination of the systems, technologies, supply and distribution companies, and organizational and individual users of electricity. One of the most interesting and challenging aspects of studying this SoS is the interaction between the human elements and the performance of the technologies within the SoS. Many of the proposed technologies for the Smart Grid depend on gaining sufficient participation from residential energy consumers, who can independently and autonomously decide to invest in and use these technologies.

Understanding and designing for this human element is critical to the success of the potential future Smart Grid. Peak load reduction can only be obtained if a critical mass of users chooses to opt-in to use the technologies. However, the decision of users to opt-in or opt-out can be influenced by a variety of factors, some of which power generator/distributors can influence and some of which they cannot. For example, companies can provide incentives to opt-in, by providing lower rates to customers who choose to opt-in or by increasing electricity prices during peak hours. On the other hand, public opinion and the media influence consumers' opinions independently of the company's control, and it is these factors that will be the focus of this study.

A previous paper by Cooksey and Mavris [6] examined the trade between the amount of control required and the number of participants required to achieve sufficient peak load reduction, and took first steps at exploring how the human element affected the success of Smart Grid. This paper largely focused on elements within the power generator/distributor's control; these elements were primarily the level of control taken by the power company and how this influenced user decisions. This paper attempts to extend that work by exploring the factors that may cause users to choose to opt-in or opt-out of a Smart Grid DR program, such as the impact of a news story that casts these technologies in a negative or positive light. Ongoing work is currently underway to combine these two approaches to provide a complete analysis framework for exploring the human factors elements of Smart Grid performance.

Implementing DR requires unique evaluation because it seeks to change the current behavior of residential energy-consumers, while maintaining the current profitability of the energy producer/distributor. The reduction in peak load can occur in one of two ways: Direct-Load Control (DLC) or Variable-Rate Pricing [7]. This paper focuses on the use of DLC, which allows the energy producer/distributor to control specific appliances owned by the residential energy-consumer. This paper does not explore the trade between the level of control given to the energy producer and the decision of consumers to opt-in, but rather on how factors external to the energy producer's control can influence the decision of consumers to participate and thus influence the success of the DLC program.

2 Methodology

In order to investigate the effects of consumer interactions and news on the enrollment of consumers participating in a DR program, a model that is able to capture both the physics of the DR, the individual behaviors of consumers, and the interactions of consumers with each other is required. Due to the high level of uncertainty in human behaviors, even after assuming all consumers behave rationally, it was necessary to use a probabilistic approach to modeling consumers' decision making. Building on the work of Cooksey and Mavris [6], consumer preference to enrolling in a DR program is dictated by top-level factor, β; β can vary between 0 and 100% and represents the amount of control that a consumer is willing to give a power generator/distributor. The β factor is not only unique to individual consumers, but the power generator/distributor also have a minimum β value (set-β) that represents the amount of control needed for a DR program to theoretically be effective (not accounting for the impact of consumer enrollment).

The comparison between a consumer's β value and the set-β is used to determine the initial enrollment of consumers in a DR program. In this study, β was assumed to be a static value, which means that it will not affect a consumers' willingness to enroll or drop out of a DR program. However, future studies will include the variation of this parameter when determining enrollment levels in the DR program. β was set in this case to give an initial enrollment of approximately fifty (50) percent of the population. In addition, there is a probability that a DLC event will occur on any given day within the simulation.

Willingness to change enrollment status based on consumer interaction or a news story was modeled probabilistically, allowing for the investigation of the effects of social influences (e.g., agent interaction and news) on enrollment in a DR program. A probability of interaction was assigned to agents, which represented the probability that agents interact with each other and discuss enrollment in DLC. If two agents interacted and had differing enrollment statuses, there was some probability that each would change their enrollment status based on the interaction.

In order to model the impact of enrollment based on the news, a beta distribution was created to model the probability that a DLC-related news story was run. If a news story was run, it could be negative with probability p or positive with probability 1-p. If an agent hears this news story, and their enrollment status is opposite what is suggested by the news story, the agent will change their enrollment status with probability q. Since these probabilities are not known, an approach to investigate the sensitivities was taken instead. After creating the M&S environment, it was necessary to explore the design space in order to identify any interesting behavior. The details of the model and the exploration of the design space are further explained in the following section.

3 Simulation Approach

As previously stated, the main intent of the M&S environment is to investigate the effects of consumer interaction and news on consumers' decisions to enroll in a DR program. Because consumers' reactions are not known a priori, the modeling technique will need to not only capture individual consumer actions, but also allow for this information to be aggregated in order to see the effects of enrollment on the effectiveness of a DR program. These considerations led to the choice of ABM to build the M&S environment. In order to study the dynamic behavior of the enrollment status and peak load reduction over time, the simulation was executed over thirty days of consumer activity. One of the main drawbacks of ABM is that model complexities can increase simulation run time beyond acceptable limits. In order to account for this trade-off between model complexity and simulation run time, many agent actions/interactions were modeled using probabilistic Discrete Event Simulation (DES). The use of DES allowed for a time step of thirty minutes to be used, which also decreased simulation run time.

The agents are modeled in a way to minimize simulation run time by only accounting for behavior that directly uses electricity, interacting with other agents, or being influenced by a news story. Most of these behaviors are modeled by using uniform distributions, but the frequency of news stories is dictated by a beta distribution; this was done in order to investigate the effects of clustered news stories against news stories that are more evenly spread out throughout the simulation. As previously stated, most of these behaviors were modeled using probabilistic DES. Since the main focus of this paper is investigating the effects of social interactions on DR enrollment, electricity consumption was normalized using the peak electricity consumption, averaged over ninety-days with no DR events. This allows for comparisons between the percentages of electricity consumed between different days of the simulation.

In order to fully investigate the design space, a 40-case space-filling Design of Experiments (DoE) was run, varying the variables across given ranges, which is shown in Table 1. Due to the stochastic nature of the model, each DoE case was repeated 30 times with a different random seed. Thus, each DoE case results in a distribution of possible behaviors across the stochastic effects. The results were then used to gain insight regarding the interaction between social factors and enrollment levels and the overall impact on peak reduction.

Table 1. Variable ranges for DoE to investigate sensitivities and design space exploration.

Variable	Min.	Max.
Prob. of Interaction	0%	25%
Prob. of Changing Mind from Interaction	0%	100%
Alpha (Beta Distr. Shape Parameter)	1	10
Beta (Beta Distr. Shape Parameter)	1	10
Prob. of Bad News Story	0%	100%
Prob. of Changing Mind from News Story	0%	100%
Prob. of a DR Event	0%	50%
Random Seed	1	30

4 Results

4.1 Single Model Execution

The effects of agent interactions and news stories impact the effectiveness of a DR program (i.e., reduction of the peak load) because they affect the number of agents enrolled in this program. The results of this can be seen in Figure 1, where the peak load occurs at 19:30 hours for all three days. This figure shows the results from a 3-day simulation run in which there are both a negative news story on Day 1 and a positive news story on Day 2. For this simulation run, it was assumed: a 40% probability that an agent would change their mind based on a news story; a 10% probability that agents would interact; a 10% probability that an agent would change their mind based on an agent interaction, which have a 10% probability of occurring; and a minimum β value of 0.5, which is set by the power generator/distributor.

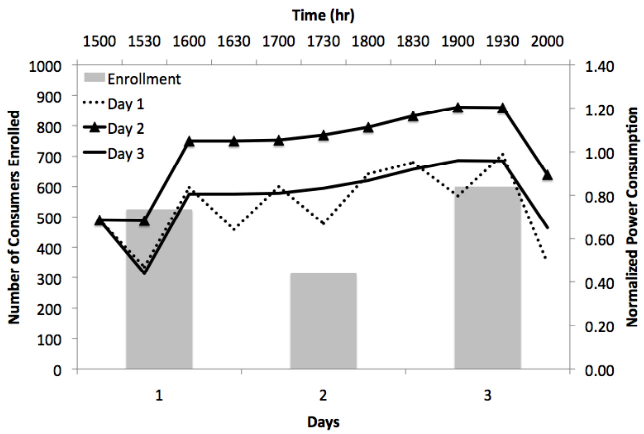


Figure 1. Normalized power curves for three days with DR events. The peak power varies with the number of consumers enrolled in a DR program.

As previously stated, the probabilities about consumers' willingness to change their opinion is not known, which means that Figure 1 is not able to capture all aspects of the problem. In order to fully understand this problem, it will be necessary to see the effects of varying these probabilities; this will be done using a DoE in order to explore the entire design space.

4.2 Design Space Exploration and Sensitivity Analysis

In order to better understand the impact of agent interactions and press on the performance of a Smart Grid DR program, a stochastic analysis was performed according to the DoE described above. The results of the analysis show that the impact of press is much more significant than the impact of agent interactions. Figure 2 shows the relationship between the mix of news stories run during a 30 period (represented by the probability that a news story is negative where a probability of 1 signifies that all news stories are negative, a probability of 0 signifies that all news stories are positive, and a probability in between represents a mix of positive and negative news stories where the percent of stories which are negative is equal to the probability shown) against the enrollment levels at day 1, day 15, and day 30. In addition, the chart is divided according to the frequency with which news stories occur, with the probability of a news story being the likelihood that a news story is presented each day. As expected, more frequent news stories cause a more significant relationship between the type of news and the enrollment, with negative news resulting in lower enrollment and positive news resulting in higher enrollment. Day 1 represents the start of simulation, in which approximately half of the population is enrolled in the DLC. The points on the graph represent the mean over the 30 repetitions and the error bars represent one standard error from the mean. The impact of agent interactions on the enrollment level in the DLC after 30 days was insignificant, regardless of how often agents interacted with each other when there are no news stories. However, if positive or negative news story causes a significant shift in the enrollment one way or the other, the agent interactions can act to exaggerate the impact of the news story, causing an overall higher or lower overall enrollment level than would be seen by the impact of the news alone.

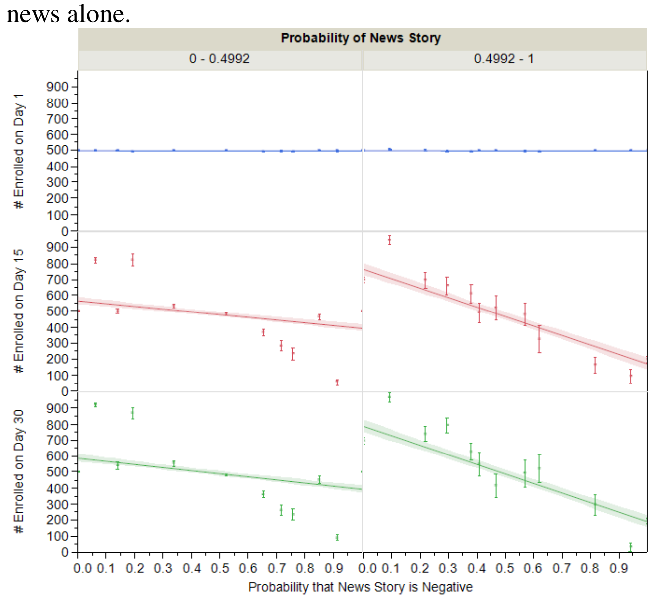


Figure 2. Impact of news stories on enrollment levels. Error bars are constructed using one standard error from the mean.

Given that news stories (which are largely beyond the control of an electricity supplier) can have a significant impact on the consumer's decision to enroll or not enroll in a DLC system, the next question is how significant that effect is on the peak load. Figure 3 shows the relationship between the enrollment level and the peak load, where the points represent the mean over the repetitions and the error bars represent one standard error from the mean. Since the loads have been normalized, a load of one (1) represents the load, which is seen with the baseline DLC enrollment. Thus, enrollment below the baseline will cause an increase in the peak load and enrollment above the baseline causes a decrease in the peak load. As can be seen from the figure, there is a linear trend between the enrollment level and the peak load, such that for every 100 people who choose to enroll in the program there is approximately a 4% decrease in the peak load. Further examination of this figure also shows a trend of diminishing returns on peak reduction when the enrollment is greater than approximately 90%. This is significant because it suggests that there is a finite limit to the amount of peak reduction that can be achieved, regardless of the enrollment.

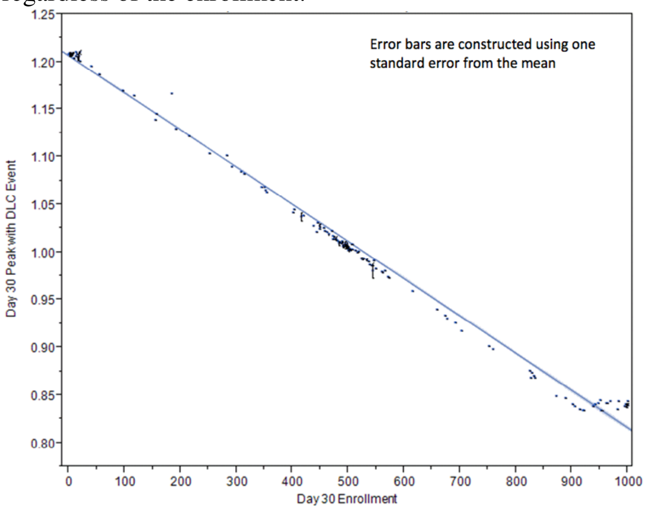


Figure 3. Relationship between enrollment and peak load on 30th day of simulation, for days in which a DR event occurs.

It is also interesting to examine the overall trends in enrollment. Figure 4 shows a frequency diagram of the enrollment on the first day of the simulation and the enrollment on the last day of the simulation across all cases. Although all cases start out with approximately half of the population enrolled, there is a large spread of enrollment rates by the last day of the simulation. Of particular interest is the tri-modal nature of this frequency diagram. Across all of the variations performed, there were three occurrences that occurred most frequently: no enrollment after thirty days, full enrollment after thirty days, or approximately 50% enrollment after thirty days. This suggests that there may be three modes. Further examination of the data reveals that the case of no enrollment corresponds (as expected) to those cases in which there is simultaneously a high probability of negative news and a high probability that agents will change their mind based on that news. Likewise, full enrollment corresponds to the cases in which there is simultaneously a high probability of positive news and a high probability of agents changing their mind based on that news. The cases in the middle group correspond to cases in which there is no news, there is no probability of agents changing their enrollment status, or the news is roughly evenly mixed between positive and negative news stories. While this result is not unexpected, it suggests that the news plays an extremely significant role in the success or failure of the DLC concept. The cases which have similar enrollment to the original baseline are those in which news does not play a role in the decision process, either because of a lack of news or evenly mixed news. The impact of news strongly dominates the impact of agent interactions, as well as the impact of the occurrence of DLC events.

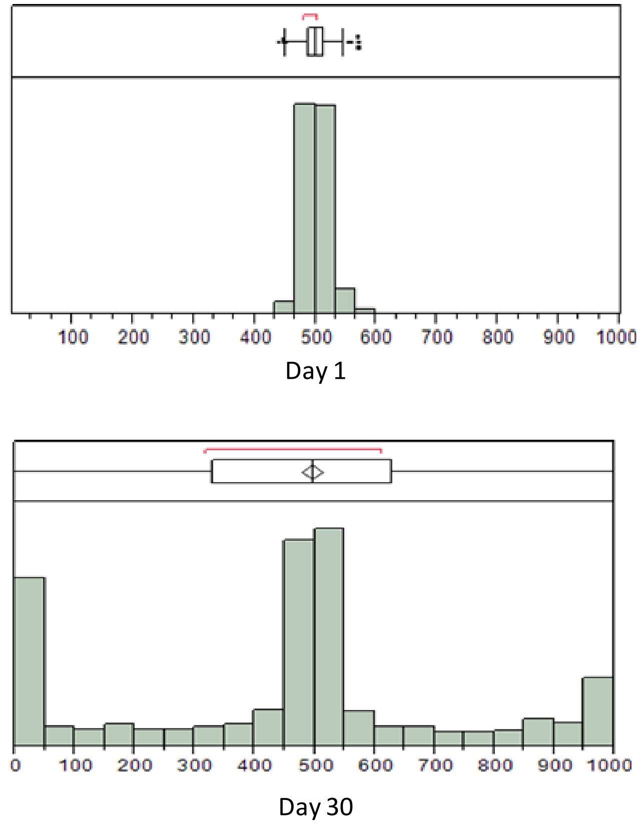


Figure 4: Comparison of enrollment on day 1 to enrollment on day 30 across all cases

Based on the information presented in this section, one can see the impact news stories have on enrollment and thus peak reduction; consider the following notional example: a power generator/distributor wants to achieve a peak reduction of 5%. Based on Figure 3, this requirement will only be met with an enrollment of approximately 68%. By examining Figure 2, one can see that this could only happen when there is a high probability of a news story, and a low probability of that news story being negative (i.e., the power generator/distributor will only achieve this reduction if there are a lot of positive news stories). Based on this notional example, there is a need for a power generator/distributor to engage in some form of a public relations ad campaign.

5 Conclusion

The approach demonstrated here provides a first cut at a methodology to analyze the impact of human factors on the performance of the Smart Grid DR SoS. It has often been hypothesized the effect of human elements can have a significant impact on SoS performance and emergent behaviors, and the methodology presented here provides a way to account for these factors when performing engineering analysis on a SoS. The initial results of this particular study confirm the hypothesis that human factors are a key driver for SoS performance. These results also show that the effect of external factors such as news on consumer opinion can have a significant impact on the success or failure of a DR program, although the model needs to undergo additional verification and validation to ensure the accuracy of the results. However, the study demonstrates the feasibility of using a probabilistic approach to include the effects of human behavior into SoS analysis, and provides a springboard for further research in this area. The next step in this work will be to compare the sensitivity of the performance to the human elements to the sensitivity of the performance to the technical elements (e.g., variations to the Smart Grid architecture and the performance of technologies) to understand which elements are the key drivers of performance. This type of analysis will enable trades between effort spent on technological improvements and effort spent on marketing and recruitment of participants to a DR program.

6 Future Work

One of the biggest concerns with implementing improvements on any agent-based model is that increasing the fidelity of the model can drastically increase simulation run time. Keeping this in mind, one of the first improvements that can be made is to link utility functions of the agents to their decision about whether or not to enroll in a DR program. In order to do this, modifications need to be incorporated into the utility function proposed by Cooksey and Mavris [6] in order to account for external influences (e.g., news stories and agent interactions). Another area of improvement would be incorporating other types of DR schemes, including Variable-Rate Pricing. A last suggestion for improvements to this model would be differentiating between possible public relations campaigns by the power producer/distributor and news stories; this would be especially beneficial because these companies are not able to control the news or human interactions, but they can perform their own advertising. As for future work on with respect to SoS, the most obvious area of improvement would be applying this methodology to other collaborative SoS. A further improvement would be modifying this methodology in order to account for human factors in other types of SoS.

7 References

[1] Haskins, Cecilia., ed. 2011. *INCOSE systems engineering handbook: A guide for system life cycle processes and activities*. Version 3.2.2. San Diego: International Council on Systems Engineering

[2] Sage, Andrew P., and Cuppan, Christopher D. 2001. “On the systems engineering and management of systems of systems and federations of systems.” *Information Knowledge Systems Management* 2 (4): 325-345.

[3] U.S. Department of Defense. 2008. *Systems engineering guide for systems of systems.* Washington DC: Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics.

[4] Smartgrid.gov. “Smartgrid.gov: What is the Smart Grid?” Accessed November 2, 2011. http://www.smartgrid.gov/the_smart_grid#smart_grid

[5] US Department of Energy. 2006. Benefits of Demand Response in Electricity Markets and Recommendations for Achieving Them: A report to the United States Congress Pursuant to Section 1252 of the Energy Policy Act of 2005.

[6] Cooksey, K.D., Mavris, D. "Game theory as a means of modeling system of systems viability and feasibility," *Aerospace Conference, 2011 IEEE* , vol., no., pp.1-11, 5-12 March 2011

[7] Borenstein, S., Jaske, M. and Rosenfeld, A. 2002. Dynamic Pricing, Advanced Metering, and Demand Response in Electricity Markets. UC Berkeley: Center for the Study of Energy Markets. Retrieved from: http://escholarship.org/uc/item/11w8d6m4.

[8] World Economic Forum. 2010. Accelerating Successful Smart Grid Pilots. Retrieved from: http://www3.weforum.org/docs/WEF_EN_SmartGrids_Pilots_Report_2010.pdf

A System of Systems Model for the Control of the University of Genoa Smart Polygeneration Microgrid

Stefano Bracco*, Federico Delfino**, Fabio Pampararo**, Michela Robba***, Mansueto Rossi**

University of Genoa – Italy
* DIME - Savona Campus, University of Genoa, Via Magliotto 2, 17100 Savona
** DITEN - Savona Campus, University of Genoa, Via Magliotto 2, 17100 Savona
*** DIST - Savona Campus, University of Genoa, Via Magliotto 2, 17100 Savona

Abstract - *Experimental tests and demonstration projects are very useful to derive new methods and tools for the optimal control of smart grids. In this work, the University of Genoa Smart Polygeneration Microgrid (SPM) is firstly presented, in connection with the different sub-systems that compose the overall system. Then, a simplified mathematical dynamic model, that can be used for optimal control purposes, is described. Finally, a dynamic optimization problem is formalized and solved.*

Keywords: *Smart microgrid, polygeneration, renewable energy, optimal control.*

1 Introduction

Many national and international research programs are aiming at developing innovative technologies and new energy management strategies in order to reach the targets set out by EU in the 20-20-20 Directive. The development of the renewable energy sector, the concept of sustainable energy, and the use of technologies for distributed generation have focused attention on smart grids. Microgrid research fits very well with ongoing smart grid activities throughout the world and several challenges need to be investigated [1]. Microgrids integrate different distributed energy sources and energy storage devices, and need intelligent management methods and efficient design in order to meet the needs of the area they are located in [2]. To this end, complex decision problems can be formulated for a portion or the overall microgrid system; in this context, experimental research activities and demonstration projects are very useful to develop new methods and tools for the optimal control of smart grids. Lidula and Rajapakse [3] present a review of existing microgrid test networks around the world (North America, Europe and Asia) and some innovative simulation models present in literature.

The University of Genoa Smart Polygeneration Microgrid (SPM) was born from the “2020 Energy” Project (Italian Ministry of Education, University and Research funding) at the Savona University Campus [4]. The SPM, like other microgrids around the world, is characterized by distributed generation units, electrical storage devices, and a variety of thermal and electrical loads. The SPM’s peculiarity is due to: the set of generation units and storage systems for both electrical and thermal energy production that make it a complete test-case; the possibility of defining and updating a software for the SPM control; a fast telecommunication network; and the integration with the research activities of the Engineering Faculty. The SPM can be used for two main purposes: demonstration and teaching activities, and to test models, methods and tools related to the research challenges of smart grids.

In this paper, first of all, the different characteristics of the SPM are highlighted through the description of the sub-systems that compose the overall system (i.e., the electrical sub-system, the thermal sub-system, the control sub-system). Then, a simple decision model (that can be included in the control sub-system) is described and solved as an example of the coordination strategy that can be implemented for the different available technologies. Specifically, the decision model is a dynamic one and can be used for optimal control purposes. State and control variables, objectives, and constraints are formalized in order to minimize the SPM's overall operating costs. Furthermore, CO_2 emissions of the SPM are evaluated. Finally, future developments are reported in connection to the formalized decision model.

2 The SPM system model

A microgrid can be seen as a “system of systems” in which different subsystems have to interact: for example, in [5], a system of systems approach for a microgrid made up of a power system and a management, operation, and control system is described. The Savona University Campus and the scheme of the SPM are shown in Figure 1. It is possible to note the main following sub-systems:

- the *SPM electrical sub-system* characterized by: a cogeneration Capstone C65 microturbine (65 kW_e in cogeneration mode), two parabolic solar concentrators (CSPs) coupled with Stirling engines (1 kW_e each), two wind turbines (3 kW_e each), a photovoltaic field (49.9

kW_e), electrical vehicle charging stations, smart meters, inverters, storage batteries (capacity of about 100 kWh_e), a dedicated grid connected to the existing Campus grid and to the public distribution network;

- the *SPM thermal sub-system* characterized by: the C65 microturbine (112 kW_{th} in cogeneration mode), the two CSPs (3 kW_{th} each), two adsorption chillers (32.5 kW_{th} of cooling capacity), the thermal storage systems, and a dedicated district heating network connected to the Campus existing one;
- the *SPM control sub-system*, which includes local controllers, a communication network, a central controller, a data storage system, and a software for the SPM supervisory control.

Figure 1. The map of the University of Genoa SPM

Moreover, the SPM is connected to the subsystems of the Savona University Campus, which can be basically subdivided into:

- the *Campus electrical sub-system* (also including power units) which is characterized by a cogeneration micro gas turbine (Capstone C30 model, 28 kW_e) and the local distribution network;
- the *Campus thermal sub-system* characterized by the C30 gas turbine (74 kW_{th} in cogeneration mode), a district heating network, two 440 kW_{th} traditional boilers fed by natural gas and heat pumps for building conditioning.

Of course, the interaction between electrical and thermal sub-systems of the SPM is mainly due to the presence of the cogeneration units (the gas turbines and the CSP system); furthermore, the SPM and the Campus subsystems are linked together because both the SPM and the Campus power generation units are used to fulfill the whole energy demand (both thermal and electrical). The control subsystem manages the entire system with the aim of optimizing and monitoring energy production, storages, and power flows by the minimization of the annual operating costs and the environmental impact.

In the following, a simplified mathematical formalization of the SPM sub-systems is reported as an example of an optimization model, that can be integrated with the SPM software for the supervisory control. In fact, one of the main advantages of the SPM is its flexibility, with specific reference to the possibility of integrating the management software with innovative optimization algorithms based on both centralized and decentralized optimal control. In this work, a simple centralized decision problem for optimal control purposes is described.

2.1 The electrical sub-system

As previously described, the SPM electrical subsystem is characterized by several generation units (fed by fossil fuel or renewable), loads and storage systems acting on a dedicated net connected to the Campus electrical grid and to the external power network. The net is essentially characterized by a number of bus-bars (nodes) connected by a ring distribution network. In principle, at each node, a power balance for incoming and outcoming power flows could be considered. Furthermore different levels of detail could be used to model this sub-system: for instance, if the longitudinal parameters (resistances, with associated losses, and reactances) of the cables connecting the bus-bars and those of the transformer connecting the grid to the external power network were considered, a power flow would be required at each simulation step, involving also the reactive power fluxes. Moreover, if a detailed representation of the units' dynamic behaviour were necessary, a quite complex electromechanical model, taking into account gas turbines and inverters dynamic models, together with their governors, controllers, etc., should be adopted. In the proposed application, given the relatively short connections between bus-bars and, most important, the simulation step of 15 minutes, far larger than the time required by the electrical sub-system to reach the steady state, this subsystem can be modeled as a single bus-bar, considering the active power balance only.

The formalized decision model includes only a sub-set of the SPM system (the C65 gas turbine, the photovoltaic plant, the battery and the external grid) and considers an electrical system characterized by a single bus-bar to which loads, generation units, the electrical Campus subsystem and the external power network are connected.

In CHP (Combined Heat and Power) applications the C65 gas turbine produces both electrical and thermal energy and, thus, it can be considered a part of both the electrical and the thermal subsystem of the SPM. As known, both the power output and the efficiency of microturbines strongly depend on environmental conditions, mainly on ambient temperature: as reported in [6], [7], [8] and [9], at elevated inlet air temperatures they decrease. Basing on the approach suggested in [6] and [7], the following relations have been used in the model to estimate the electrical power at full load $P_{el\_full,k,t}$ and the corresponding electrical efficiency $\eta_{el\_full,k,t}$ (at each time interval (t,t+1)) of the k-th gas turbine, as a function of ambient temperature θ_t over the discretized time horizon, that corresponds to the number T of time intervals of length Δt :

$$P_{el\_full,k,t} = \begin{cases} P_{el\_\max,k} \text{ if } \theta_t \le \theta^P_{\min,k} \\ f_k(\theta_t) \text{ if } \theta_t > \theta^P_{\min,k} \end{cases} \quad k=1,\ldots,K \quad t=0,\ldots,T-1 \tag{1}$$

$$\eta_{el\_full,k,t} = \begin{cases} \eta_{el\_\max,k} \text{ if } \theta_t \le \theta^\eta_{\min,k} \\ g_k(\theta_t) \text{ if } \theta_t > \theta^\eta_{\min,k} \end{cases} \quad k=1,\ldots,K \quad t=0,\ldots,T-1 \tag{2}$$

where $P_{el\_max,k}$ and $\eta_{el\_max,k}$ are the maximum values attained by the gas turbine electrical power and efficiency, whereas f_k and g_k indicate decreasing polynomial functions reported by gas turbine manufacturers ([6] and [7]). The model considers also the possibility to operate gas turbines at partial loads: in these conditions, it is necessary to take into account that the electrical efficiency diminishes when power output decreases. In particular, the following relation has been applied:

$$\eta^*_{el,k,t} = h_k(p^*_{el,k,t}) \quad k=1,\ldots,K \quad t=0,\ldots,T-1 \tag{3}$$

where h_k are functions derived from gas turbine performance curves, while $\eta^*_{el,k,t}$ and $p^*_{el,k,t}$ are:

$$\eta^*_{el,k,t} = \frac{\eta_{el,k,t}}{\eta_{el\_full,k,t}} \quad , \quad p^*_{el,k,t} = \frac{P_{el,k,t}}{P_{el\_full,k,t}} \tag{4}$$

$\eta_{el,k,t}$ and $P_{el,k,t}$ being respectively the actual values in time interval (t,t+1) electrical efficiency and power output.
Finally, the primary energy flow $P_{PE,k,t}$ (natural gas flow rate multiplied by its lower heating value) that is used in the gas turbine is given by:

$$P_{PE,k,t} = \frac{P_{el,k,t}}{\eta_{el,k,t}} \quad k=1,\ldots,K \quad t=0,\ldots,T-1 \tag{5}$$

The electrical sub-system also includes a battery, described by the following equation [10]:

$$SOC_{t+1} = SOC_t - \overline{\eta}_t \frac{P_{S,t}}{CAP}\Delta t \quad t=0,\ldots,T-1 \tag{6}$$

where the storage state of charge at time $t+1$ (SOC_{t+1}) depends on: the state of charge at time t (SOC_t), the active power $P_{S,t}$ (positive if delivered by the storage, negative if injected), the battery rated capacity CAP, the time interval Δt, and the performance coefficient $\overline{\eta}_t$ which is equal to 0.85 when $P_{S,t}$ is negative and to 1.15 otherwise.

As regards the photovoltaic system, its power output $P_{PV,t}$ is an input of the model and it has been evaluated taking into account the plant characteristics (azimuth and tilt angles, type of modules) and the location (solar irradiance, ambient temperature, shadowing effects). Finally, there is the variable $P_{NET,t}$ which indicates the power exchanged with the external grid (withdrawn if positive, injected if negative).

2.2 The thermal sub-system

The thermal subsystem is here considered neglecting heat losses in the district heating network and the dynamics due to the heat storage system. The following generation units have been considered: the microturbines and the Campus boiler.
For the microturbines, as suggested in [6], the thermal power output $P_{th,k,t}$ in time interval (t,t+1) has been calculated, conservatively, by using the same thermal to electric output ratio ε_k for all power levels:

$$P_{th,k,t} = \varepsilon_k P_{el,k,t} \quad k=1,\ldots,K \quad t=0,\ldots,T-1 \tag{7}$$

where:

$$\varepsilon_k = \frac{P_{th\_nom,k}}{P_{el\_nom,k}} \quad k=1,\ldots,K \tag{8}$$

$P_{th\_nom,k}$ and $P_{el\_nom,k}$ being respectively the gas turbine nominal thermal and electrical power output (evaluated at ISO conditions).

The thermal power output $P_{th,B,t}$ of the boiler has been calculated by the following equation:

$$P_{th,B,t} = \eta_B \cdot P_{PE,B,t} \quad t=0,\ldots,T-1 \tag{9}$$

where $P_{PE,B,t}$ indicates the boiler power input and η_B is the boiler efficiency, assumed equal to 0.9.

2.3 The control sub-system

The SPM control subsystem includes local controllers, a communication network, and a central controller. The control network uses both communication protocols Modbus and RS485, and the protocol IEC 61850 that, in future years, should become a reference standard for communications and control architectures in the low-

voltage smart grid sector. The control sub-system is composed by the following different subsystems:
- the local controllers sub-system, which includes the interfaces with the field, composed by those devices that directly interact with the electrical network (RTUs-Remote Terminal Units) with actions on measurement of relevant parameters (i.e., current, voltage, temperature, etc.) and on the different actuators (e.g., switches);
- the communication sub-system that includes the communication network over which there is the bidirectional data exchange between the control room and the devices on the field;
- the central controller sub-system which is the "system intelligence" that allows the supervision and control necessary to guarantee an effective management and coordination of the overall system.

The software for the overall system management guarantees the SPM operations, monitoring and alarms management, through the connections with the RTUs and local control panels of devices in field. Moreover, it allows adding new components, models and tools useful for the optimization and control. A review about recent contributions, challenges and research needs in control of microgrids with storage is reported in [1]. One of the main challenges is that of optimizing the system operation of different plants both in centralized and decentralized architectures. Moreover, in literature, despite many contributions related to planning decision problems, there are few articles in the field of real time optimal control of a mix of renewable power plants integrated in an electrical network. In this work, a dynamic decision model is presented, with specific reference to a portion of the SPM.

3 The optimization problem

The main goal of the proposed optimization algorithm is that of determining the optimal values over time of the microturbines and boiler electrical and thermal power output, of the storage injection/withdrawal, and of the electrical power exchange with the external grid, according to the time-varying thermal and electrical loads, fuel and electricity prices, available energy forecasts. Different performance indexes can be formalized as a function of the chosen parameters, state and control variables, to optimize the overall system management, namely: costs related to purchasing of electricity and natural gas; benefits due to electricity sale and incentives for local consume of produced energy; the carbon footprint of the overall system. In this work, operating costs and benefits are considered as objectives, while the carbon footprint is evaluated for the optimal solution.

3.1 State and control variables

State variables are represented by the storage state of charge SOC_t, while control variables can be divided in primary ($P_{PE,k,t}$, $P_{PE,B,t}$, $P_{NET,t}$ and $P_{S,t}$) and secondary ($P_{el,k,t}$, $P_{th,k,t}$, $P_{th,B,t}$) control variables. Moreover, a binary control variable, $\delta_{k,t}$, has to be introduced: it is set to 1 if the *k-th* microturbine works in time interval (t, $t+1$), and 0 otherwise.

3.2 The objective function

The overall objective function, to be minimized, is given by the SPM operating costs (C_{TOT}) over the time optimization horizon, that are given by the sum of boiler costs (C_B), microturbine costs (C_K), and costs (C_{NET}) and benefits (B_{NET}) related to the electricity exchange with the net. That is:

$$\min C_{TOT} = \min\{C_B + C_K + C_{NET} - B_{NET}\} \tag{10}$$

where:

$$C_B = \sum_{t=0}^{T-1} P_{PE,B,t} \cdot TES_{pp} \cdot \Delta t \tag{11}$$

$$C_K = \sum_{k=1}^{K} C_k = \sum_{k=1}^{K} \left\{ \sum_{t=0}^{T-1} \frac{P_{el,k,t} \cdot \Delta t}{\eta_{el,k,t} \cdot LHV} \cdot CNG \right\} \tag{12}$$

$$C_{NET} = \sum_{t=0}^{T-1} C_{NET,t} \cdot \max(P_{NET,t}, 0) \cdot \Delta t \tag{13}$$

$$B_{NET} = \sum_{t=0}^{T-1} C_{NET} \cdot \min(P_{NET,t}, 0) \cdot \Delta t \tag{14}$$

In the Eqs. (11-14): TES_{pp} is the thermal energy service purchasing price of the boiler (0.0853 €/kWh$_{PE}$); LHV is the natural gas lower heating value (9.7 kWh$_{PE}$/m^3); $C_{NET,t}$ is the electricity purchasing price (which depends on time because two different prices have been considered: one for peak hours, from 8 a.m. to 8 p.m., and one for off-peak hours, the remaining time) expressed in €/kWh$_e$; C_{NET} is a medium price of the electricity sold to the external grid; CNG is the gas unitary cost (€/m^3) for cogeneration gas turbines that, in accordance with the Italian legislation on cogeneration, can be calculated as:

$$CNG = (0.25 \cdot \eta_{el,k,t} \cdot LHV - 0.12) \cdot (NG_{ppwf} - NG_{pp}) + NG_{pp} \tag{15}$$

NG_{pp} and NG_{ppwf} being the natural gas purchasing price with and without fee (0.7 and 0.427 €/m^3).

3.3 The constraints

Constraints are formalized to fulfill technical and environmental requirements and energy needs. As regards power balance constraints, the following equations should be satisfied:

$$D_{el,t} = P_{NET,t} + \sum_{k=1}^{K} P_{el,k,t} + P_{PV,t} \qquad t=0,\dots,T-1 \tag{16}$$

$$D_{th,t} \geq \sum_{k=1}^{K} P_{th,k,t} + P_{th,B,t} \qquad t=0,\dots,T-1 \tag{17}$$

where $D_{el,t}$ and $D_{th,t}$ are, respectively, the electrical and the thermal power demand. Furthermore, each microturbine is characterized by a "technical minimum power" ($P_{min,el,k}$), below which the machine is shut down in order to avoid high CO emissions; consequently, the following constraints have to be satisfied:

$$P_{el,k,t} \geq P_{\min,el,k}\delta_{k,t} \quad k=1,\dots,K \quad t=0,\dots,T-1 \tag{18}$$

$$P_{el,k,t} - M\delta_{k,t} \leq 0 \qquad k=1,\dots,K \qquad t=0,\dots,T-1 \tag{19}$$

where M is a big number.
Then, there is the constraint over the battery state of charge:

$$0 \leq SOC_t \leq 1 \qquad t=0,\dots,T-1 \tag{20}$$

Finally, Eqs. (1-9) have to be taken into account together with constraints over the plants' size, that result in lower and upper bounds of the variables $P_{el,k,t}$, $P_{th,k,t}$, $P_{th,B,t}$, $P_{S,t}$.

3.4 The carbon footprint assessment

The SPM CO_2 emissions have been calculated through the following equations:

$$E_{CO_2} = E_{CO_2,B} + E_{CO_2,K} + E_{CO_2,NET} \tag{21}$$

where:

$$E_{CO_2,B} = \sum_{t=0}^{T-1} P_{PE,B,t}\, f_e\, f_o\, \mu_1\, \Delta t \tag{22}$$

$$E_{CO_2,K} = \sum_{k=1}^{K}\left\{\sum_{t=0}^{T-1} P_{PE,k,t}\, f_e\, f_o\, \mu_1\, \Delta t\right\} \tag{23}$$

$$E_{CO_2,NET} = \sum_{t=0}^{T-1} \max\left(0, P_{NET,t}\right) f_{e,NET}\, \mu_2\, \Delta t \tag{24}$$

where: f_e and f_o are the emission (56 t_{CO2}/TJ_{PE}) and the oxidation factor (0.995) of the natural gas, μ_1 e μ_2 are conversion factors, and $f_{e,NET}$ is the emission factor of the national electrical mix (0.465kg_{CO2}/kWh_e).

4 Results

The developed decision model has been tested with data of a specific day in December for the Savona University Campus. Specifically, Figure 2 reports the available loads for electrical and thermal power, ambient temperature, and available data for electrical power coming from the PV.

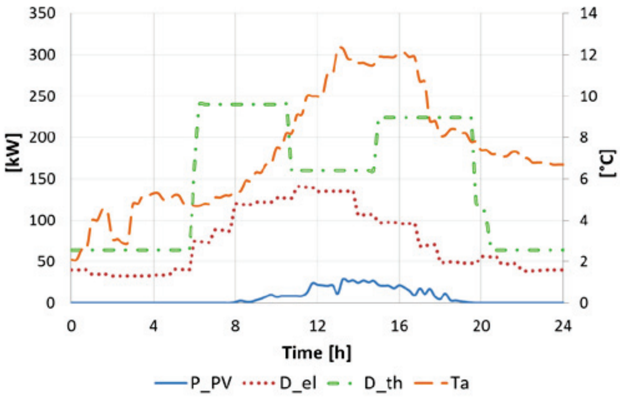


Figure 2. Power from PV, thermal and electrical loads and ambient temperature

The decision model has been applied and optimal results have been found for a time horizon of 24 hours within a time discretization of 15 minutes. However, the decision model could be inserted in a MPC (Model Predictive Control) central controller that iteratively finds the optimal solution, in which suitable optimization and simulation horizons have to be set. In scenario *a*, only one gas turbine (Capstone C65), the boiler and the storage system have been considered in the optimization problem. The optimal value of the daily operating cost is equal to 475 €, and CO_2 emissions are equal to 1.38 t. Optimal results for the control variables for generation units are detailed in Figure 3, while in Figure 4 the results for the storage state of charge are reported.

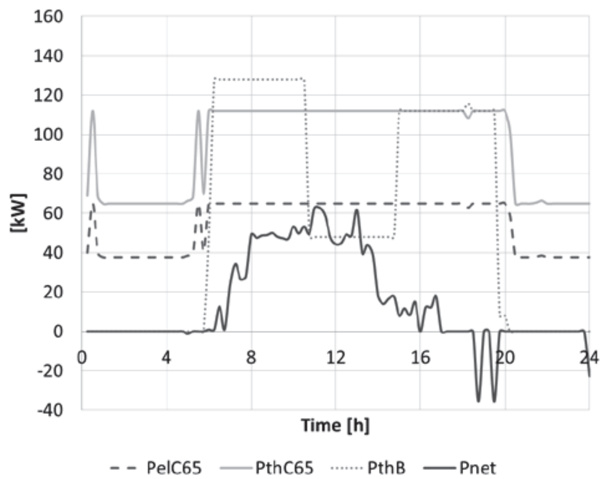


Figure 3. Optimal results (scenario *a*)

It is possible to calculate the C65 utilization factor:

$$U_{C65} = \frac{\sum_{t=0}^{T-1} P_{el,C65,t} \cdot \Delta t}{P_{el\_nom,C65} \cdot 24} \quad (25)$$

that results equal to 0.84. Only 1% of the thermal energy produced by C65 and the boiler is rejected.

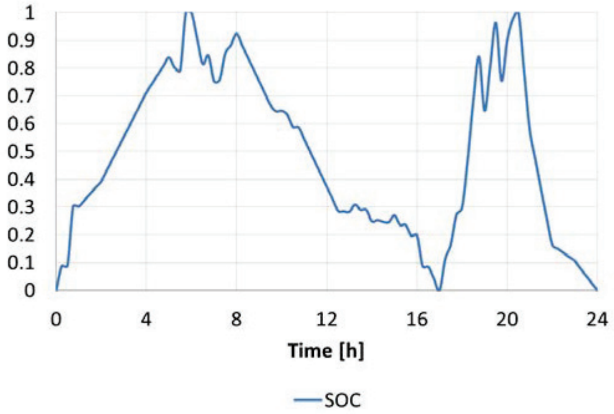


Figure 4. Optimal results: the battery state of charge

In scenario *b*, the control variable $P_{S,t}$ for the storage has been set equal to zero. The daily operating cost is of 479 € and CO_2 emissions are equal to 1.43 t. Optimal results are reported in Figure 5.

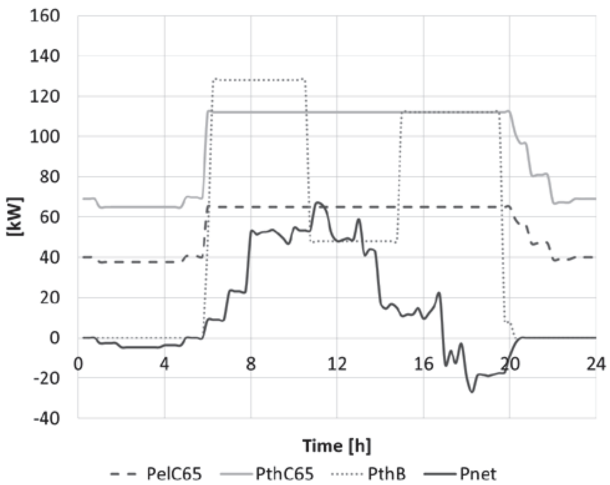


Figure 5.Optimal results (scenario *b*)

5 Conclusions

The University of Genoa SPM has been described in connection with the different sub-systems that compose the overall system. A dynamic optimization problem has been formalized, as an example of the possible algorithms that can be inserted in the whole software for the SPM control sub-system. Future developments will regard the definition of a multi-objective optimization model for the overall system, and of solution methods that can be used for real time optimal control. Moreover, models to represent the stochastic behavior of available resources, demand and prices will be defined. Finally, a comparison between centralized and decentralized decision architectures will be developed.

References

[1] R. Zamora, A.K. Srivastava, “Controls for microgrids with storage: Review, challenges, and research needs.” *Renewable and Sustainable Energy Reviews*, Vol. 14, pp. 2009-2018, 2010.

[2] R. Minciardi, M. Robba, “Dynamic optimization for distributed energy production”, Proc. of IFAC 2011, Milan, 2011.

[3] W.A. Lidula, A.D. Rajapakse, “Microgrids research: A review of experimental microgrids and test systems.” *Renewable and Sustainable Energy Reviews*, Vol. 15 pp. 186-202, 2011.

[4] S. Bracco, F. Delfino, F. Pampararo, M. Robba, M. Rossi, “Economic and Environmental Performances Quantification of the University of Genoa Smart Polygeneration Microgrid”, IEEE International Energy Conference & Exhibition (Energycon 2012), Florence, 2012.

[5] L.R. Phillips, “The microgrid as a system of systems”, *Systems of Systems Engineering – Principles and Applications,* Taylor & Francis Publishers, London, UK, 2008.

[6] Capstone Turbine Corporation, Technical Reference: Capstone Model C65 Performance, USA, 2008.

[7] Capstone Turbine Corporation, Application Guide: Model C65 Integrated CHP (ICHP), USA, 2006.

[8] M. Casisi, P. Pinamonti, M. Reini, “Optimal Lay-Out and Operation of Combined Heat & Power (CHP) Distributed Generation Systems”, *Energy*, Vol. 34, pp. 2175-2183, 2009.

[9] E. Macchi, S. Campanari, P. Silva, *Natural Gas Microcogeneration*, Polipress, Milan, 2006.

[10] T.K. Brekken, , A. Yokochi, A. von Jouanne, Z.Yen, H.M. Hapke, D.A. Halamay, “Optimal Energy Storage Sizing and Control for Wind Power Applications”, *IEEE Transactions on Sustainable Energy,* Vol. 2, pp. 69-77, 2011.

Intelligent Decision Making for Energy Management in Microgrids with Air Pollution Reduction Policy[1]

Yashar Sahraei Manjili*†, Amir Rajaee*, Mo Jamshidi*, Brian T. Kelley*
*Department of Electrical and Computer Engineering
University of Texas at San Antonio
San Antonio, TX
†sca102[AT]my.utsa.edu

Abstract— **Fuzzy Logic-based decision-making framework is implemented for energy management in microgrid systems in order to meet targets such as providing local consumers with required energy demand and making good revenue for the microgrid owner under a time-varying electricity cost policy while helping reduce negative environmental effects due to air polluting sources of electrical energy such as coal fire plants which operate in the main grid in order to provide local microgrid loads. Typically, a microgrid system has two modes of operation. It either works synchronously with the main grid or operates independently from the utility grid in an isolated mode. Distributed renewable energy generators including solar, wind in association with batteries and main grid supply power to the consumer in the microgrid network. One day period is divided to a finite number of time slots. The Fuzzy intelligent approach implemented in this article determines the rate at which power has to be delivered to/taken from the storage unit during the next time slot depending on the electricity price per kWh of energy, local load demand, electricity generation rate through renewable resources, and air pollution factor which are sampled at predetermined rates. Cost function is defined as the sum of balance/revenue due to electricity trade between microgrid and the main grid, which includes the power provided to local load and distribution losses. Five different scenarios are considered for local load and microgrid assembly operation. Measures of balance/revenue will be extracted to represent benefits of using Fuzzy logic for energy management in microgrids with air pollution reduction policy.**

Keywords- Microgrid Network, Intelligent Fuzzy Decision-Making, Power Flow Analysis, Time-Varying Electricity cost.

I. INTRODUCTION

Microgrid is a small-scale electrical grid that is designed to provide energy and distribute it between local loads. A microgrid is an aggregation of multiple distributed generators (DGs) such as renewable energy sources, conventional generators, in association with energy storage units which work together as a power supply network in order to provide both electric power and thermal energy for small communities which may vary from one common building to a smart house or even a set of complicated loads consisting of a mixture of different structures such as buildings, factories, etc [3]. Typically, a microgrid operates synchronously in parallel with the main grid. However, there are cases in which a microgrid operates in islanded mode, or in a disconnected state from the main grid [1]. Auction-based theory for pricing strategy in solar powered microgrid is studied in [2]. In [3], Authors considered Fuzzy decision-making to control battery storage unit in microgrid considering an ideal storage unit both with no maximum limit for the amount of energy stored in the battery and with maximum limit, and investigated the overall costs and profits the Fuzzy approach could bring to the system.

In this article, when the microgrid is connected to the main grid and is working synchronously with it, we assume the flow of electrical power can be either from the main grid to the microgrid or vice-versa [3]. Whenever the flow of electric power is from microgrid towards the main grid, the microgrid, or in general the customer, is making profit by selling energy to the main grid. Without loss of generality, we have assumed that for each time instant the electricity cost rate for buying energy from the main grid is equal to that of the electrical energy sold to the main grid. Demand side management is not implemented since it is assumed that the main restriction is to always provide local load with whatever amount of energy it requires. In section II, model of the microgrid for this study will be introduced. The cost function and control policy will be determined in Section III. Intelligent decision-making both with and without policies on air pollution reduction will be discussed in section IV. Section V includes simulation results and discussions on pros and cons of using intelligent Fuzzy energy management for microgrids.

II. SYSTEM MODEL

The model used for simulation of the microgrid network can generally be assumed as a three-bus power network which summarizes all the busses on the network as three different types. One of the busses in the distributed generation model is assumed to serve the renewable generators which include either solar farm, wind farm, or any other renewable generation units either in association with battery storage unit or without any storage. Another bus is assumed to be there as the representative for connection of the main grid (utility) to the local microgrid which will provide the complement part of the power demanded by the local load that renewable

[1]This work has been supported, in part, by a grant from CPS Energy, San Antonio, TX through Texas Sustainable Energy Research Institute; University of Texas at San Antonio.

electricity generation system cannot afford or will deliver the excess power from microgrid's side to the main grid. The third bus will be the summary of all busses for local loads to which the demanded power is to be provided. This load can be anything from a common building or a smart house, to a group of factories, or a mixture of all mentioned. Figure 1 shows an overall microgrid schematic including renewable electricity generators and storage unit, utility, and local load.

Figure 1 Microgrid Network General Schematic

There are five simulation scenarios in this article, specifications of which are given in Table 1. The parameters mentioned in table 1 will be introduced and used in the next sections:

Table 1. Characteristics of Five Simulation Scenarios

	Microgrid Model Elements	Fuzzy Inputs	Fuzzy Output
Scenario 1	Main grid Local Load	---	---
Scenario 2	Main grid Local Load Renewables	---	---
Scenario 3	Main grid Local Load Renewables Battery Storage	---	---
Scenario 4	Main grid Local Load Renewables Battery Storage Fuzzy Control	Pr(t) $P_R(t)$ $P_L(t)$	$P_B(t)$
Scenario 5	Main grid Local Load Renewables Battery Storage Fuzzy + Pollution Control	Pr(t) $P_R(t)$ $P_L(t)$ C(t)	$P_B(t)$

In the following, characteristics of the three buses (see Figure 2) in network model are mentioned for each scenario:

A. Characteristics of Buses in Scenario No 1

For scenario number one we only assume our network to be consisted of two busses as follows:

- First bus is of type Slack (reference) and is used as the Utility (grid) bus.
- Second bus is of type PV used as the Local Load bus.

Hence, we assume there is not renewable generation units installed and no battery storages are available and the local load, i.e. the plant only is supplied by the main grid which was the typical case prior to introduction of renewable resources and storage units to the industry. This is also the current case where no microgrids are available.

B. Characteristics of Buses in Scenario No 2

We assume there are renewable resources employed on the microgrid without any storage units available. The characteristics of the three buses in the microgrid Network model simulated in this article are as follows in the second scenario:

- Bus 1 is a PQ bus and is used as the bus for renewable generation unit.
- Bus 2 will be the Slack (reference) bus and is used as the Utility (grid) bus.
- Bus 3 is of type PV and is used as the Local Load bus.

C. Characteristics of Buses in Scenario No 3

The storage unit is also assumed to be at hand in addition to the renewable resources. Bus characteristics of the three buses in the microgrid Network model simulated in this scenario are as follows:

- Bus 1 is a PQ bus and is used as the bus for renewable generation unit and finite-capacity battery storage unit.
- Bus 2 will be the Slack (reference) bus and is used as the Utility (grid) bus.
- Bus 3 is of type PV and is used as the Local Load bus.

This must be considered that no intelligent control approach is employed yet.

D. Characteristics of Buses in Scenario No 4

In this scenario almost everything is the same as scenario three except for the fact that the intelligent decision-making approach, i.e. Fuzzy control, is also employed to provide the system with reduced costs and increased benefits. Therefore, buses will have following characteristics:

- Bus 1 is a PQ bus and is used as the bus for renewable generation unit and finite-capacity battery storage unit.
- Bus 2 will be the Slack (reference) bus and is used as the Utility (grid) bus.
- Bus 3 is of type PV and is used as the Local Load bus.

In this scenario, there are three input variables to the intelligent control system which including electricity price, local load demand, and renewable electricity generation rate.

E. Characteristics of Buses in Scenario No 5

This scenario is similar to the previous one with another input variable added to the Fuzzy controller called the air pollution index. The rules of the Fuzzy inference engine are also modified in such a way to take into account environmental concern besides providing the microgrid with reduced costs and increased benefits. The main objective in this scenario is still to provide local load with the required power demand. Making revenue for microgrid owner and reducing the air pollution compared to the first scenario are both of second priority and a compromise has to be made for these two secondary goals when generating the Fuzzy rule-base. Characteristics of buses in this scenario are:

- Bus 1 is a PQ bus and is used as the bus for renewable generation unit and finite-capacity battery storage unit.
- Bus 2 will be the Slack (reference) bus and is used as the Utility (grid) bus.
- Bus 3 is of type PV and is used as the Local Load bus.

This must be noted that battery storage units are assumed to be ideal batteries, i.e. no dynamic transient conditions for changes in the amount of stored energy in batteries are assumed. However, the energy capacity of the storage unit is assumed to be limited, i.e. that is a finite energy capacity battery.

III. PROBLEM STATEMENT

Time-varying pricing policy for electricity purchase means the bid price is not constant during the day time. The update duration of electricity cost is assumed to be 15 minutes. This implies that the money consumers have to pay to the utility for the same amount of energy used during different time-intervals might be different. Therefore, a function is defined to take into account the difference between amount of power given to the utility from the microgrid, and the amount of power taken from the utility by the microgrid. This function give us a cumulative sum of the amount that consumer must pay to the utility or, in some circumstances, the consumer obtains by selling the electricity to the utility due to proper purchase, storage, consumption, and sale policy. Eq. 1 represents this cost function:

$$Cost = \sum_{k=1}^{T} (\Pr(k) \times \Delta t \times S_U(k)) \qquad \text{Eq. 1}$$

where the electricity price rate Pr(k) is the sell and bid price per kilowatt-hour of electrical energy. $S_U(k)$ is the amount of power transferred from/to microgrid to/from the main grid during k^{th} 15-minute period. If power is received from the Grid $S_U(k)$ will be positive, and if power is delivered to the grid $S_U(k)$ will appear with a negative sign. Δt is the duration of the time interval which here is assumed to be 15 minutes. Therefore, there will be $T = \frac{24(h)}{\Delta T} = \frac{24(h)}{15(min)} = \frac{24(h)}{0.25(h)} = 96$ time intervals for each of which the electricity cost will be determined during the 24 hour day period.

Air pollution is taken into account in this article as the factor of environmental health threat. Eq. 2 represents the standard measure used for computing air pollution called Air Quality Index (AQI) introduced by Environmental Protection Agency (EPA):

$$AQI_{pollution} = \frac{Pollutions\ Data\ Reading}{Standards} \times 100 \qquad \text{Eq. 2}$$

EPA calculates AQI for five major air pollutants, including ground-level ozone, particle pollution (also known as particulate matter), carbon monoxide, sulfur dioxide, and nitrogen dioxide, using the pollutant data reading updated every hour during the day. EPA has released a new rule to regulate CO2 emissions from power plants. The new rule requires power plants to meet an output-based standard of 1,000 pounds of CO2 per megawatt-hour (MWh) of electricity produced. Constraining relationship between the energy taken from the main grid and the CO2 added to the air is assumed is shown in Eq. 3:

$$p = \psi E \qquad \text{Eq. 3}$$

Where p represents how much CO2 in unit pound is added to the air, $E=\int P(t)dt$ is the energy generated by the power plant during a specific time when P(t) stands for the function representing output electrical power of the plant in Megawatts, and ψ is the restricting coefficient which is assumed to be equal to $1000\left(\frac{lb}{MW}\right)$.

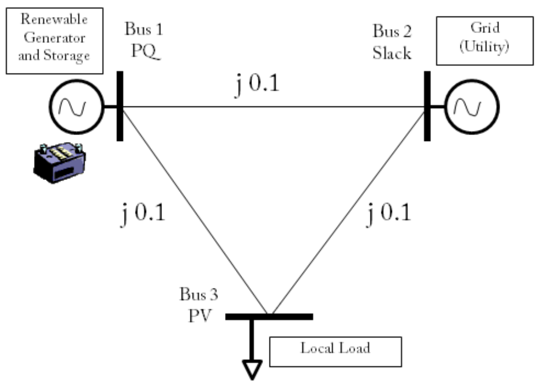


Figure 2 Three Bus Model of microgrid

Figure 2 represents the three-bus model used for simulation of the microgrid in different scenarios along with the branch impedances and types of the buses.

Mathematical representation of the active power sold to or purchased from the grid can be expressed as shown in Eq. 4:

$$P_U(t) = P_L(t) + P_{loss}(t) - P_B(t) - P_R(t) \qquad \text{Eq. 4}$$

$$s.t. \begin{cases} if\ P_B(t) < 0\ energy\ is\ being\ stored\ in\ battery \\ if\ P_B(t) > 0\ energy\ is\ being\ drawn\ from\ battery \end{cases}$$

Where $P_U(t)$ stands for the amount of power given to the microgrid, $P_U(t) > 0$, or taken out from it, $P_U(t)<0$, at time instant t. $P_L(t)$ is the demanded power of the local load at time instant t. $P_{loss}(t)$ shows the distribution loss due to branch impedances at time instant t. $P_B(t)$ is the rate at which energy is given to storage unit, i.e. $P_B(t) < 0$, or is taken from it, i.e. $P_B(t) > 0$, at time instant t. Accordingly, $P_R(t)$ represents the electrical power at the output of renewable generators at time instant t which is either equal to or greater than zero.

The value $P_B(t)$ will be determined by the fuzzy controller for each 15 minute interval at the beginning of the interval based on samples of the three, except for the scenario five which has four, input variables to the Fuzzy system including electricity cost, renewable electricity generation rate, local load demand, and, only for scenario number five, air pollution.

Hence, based on the value of $P_B(t)$ calculated by intelligent Fuzzy controller for each 15-minute interval, and the continuously sampled values of $P_L(t)$ and $P_R(t)$, values of $P_U(t)$ and also $P_{loss}(t)$ can be determined by a power flow calculation algorithm in power networks since the impedances of the branches are known.

Power flow calculation and analysis in the microgrid is the basic tool to simulate the whole system. There are a number of well-known methods for calculation of power flow in the distributed generation network [4]. Four different types of busses are generally considered in a distributed generation network, the characteristics of which will be calculated in power flow algorithms. These four types include PQ, PV, Slack, and isolated [5, 6]. For the simulation purposes of this paper, Gauss-Seidel iterative algorithm is implemented to do the power flow calculation. [6]

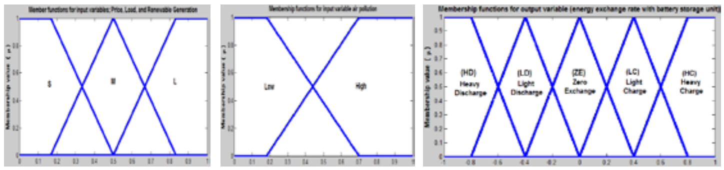
IV. Intelligent Decision-Making

Fuzzy logic [7] is used for control and energy management by determining the flow of power to/from the battery storage unit in order to improve the value of the cost function in Eq. 1. The three input variables to the Fuzzy inference engine for scenario 4 include electricity cost per kWh or Pr(t), renewable electricity generation rate or $P_R(t)$, local load demand or $P_L(t)$. The Fuzzy inference engine serves as the controller which determines a measure of power that must be sent to/taken from the battery unit during each 15 minute period, based on the samples of the three input variables at the beginning of that period.

In scenario 5, a fourth input variable will be fed to the Fuzzy inference engine called air pollution measure or C(t). For the purposes of this study, an exemplary pollution profile is generated for a typical 24 hour period in order to examine capabilities of the different scenarios. C(t) is assumed to be the average amount of CO2 on global area, not only at specific points around the polluting power plants or only around microgrid local loads. Hence, a simplified discrete-time mathematical representation for air pollution update is represented in Eq. 5 and Eq. 6:

$$C(k+1) = C(k) + \Delta C \qquad \text{Eq. 5}$$
$$\Delta C = \Delta p(k) - \Delta r(k) \qquad \text{Eq. 6}$$

where $C(k)$ represents the measure of pollutant, here CO2, concentration at the end of k^{th} 15-minute time interval. ΔC stands for the change in the CO2 measure during the k^{th} time interval. $\Delta p(k) = p(k+1) - p(k) = \psi \int_{k\Delta t}^{(k+1)\Delta t} (\sqrt{P_U^2(t) + Q_U^2(t)})dt$ represents the measure of the amount of CO2 added to the air during k^{th} time interval due to operation of the main grid's power plants when Δt represents the duration of each time interval, i.e. 15 minutes. $\Delta r(k)$ represents the removal term of pollution associated with chemical reactions and pollution's dispersion in the atmosphere during k^{th} time interval.



(a) (b) (c)

Figure 3 Fuzzy Membership functions for input and output variables of the Fuzzy Controller; (a) price, load and generation (b) air pollution (c) output

Fuzzy membership functions for the four input variables and the only output variable of the Fuzzy inference engine are shown in figure 3.

The numerical values for the three input variables price, load, and generation are normalized to the [0 1] interval, and then are Fuzzified using three Fuzzy sets defined as Low (L), Medium (M), and High (H) as can be seen in figure 3a. Air pollution has two membership functions defined as Low (L) and High (H) represented in figure 3b. After Fuzzification, the input variables will be fed to Fuzzy inference engine where the rule-base is applied to them and the Fuzzy output will be determined based on human reasoning. There is only one output variable for the Fuzzy controller which determines the amount of power to be exchanged with the battery during the next 15-minute interval. As represented in figure 3c, output variable Fuzzy set has five membership functions called Heavy Discharge (HD), Light Discharge (LD), Zero Exchange (ZE), Light Charge (LC), and Heavy Charge (HC). The power drawn from the batteries can be used to help the renewable electricity generation unit provide the local load with required demand, can be sold to the main grid, or can be partially used for both reasons [8]. The role of Fuzzy inference engine is critically important for obtaining satisfactory results. For example, some rules can be as follows in different scenarios:

IF the *Price* is ***Medium***, **AND** the *Renewable Generation Rate* is ***Low***, **AND** the *Load* is ***Medium***, **THEN** the *Battery* should be ***Lightly Discharged***.

IF the *Price* is ***Medium***, **AND** the *Renewable Generation Rate* is ***Low***, **AND** the *Load* is ***Medium***, **AND** the *Air Pollution* is ***High***, **THEN** the *Battery* should be ***Heavily Discharged***.

The primary objective in these simulations is to provide the local load with all the power it demands at any circumstances. Under low-price electricity conditions, the action decided by the rules might even sometimes require the microgrid network to purchase energy from grid and store it in the battery storage unit since the main point here is that the electricity price is low. This consequently results in more degree of freedom for the system to sell energy to the main grid during high-price periods, even under cases of high local load demand. Hence, having feasible rules predefined for the Fuzzy system helps improve the cost function drastically. The proposed approach may even sometimes result in so that the microgrid owner makes some revenue instead of paying to the utility, while provides the local load demand to the fullest extent.

V. SIMULATION RESULTS

The simulation is done on the three bus system shown in figure 2 for the duration of one week period. The Gauss-Seidel algorithm is implemented using Matlab for power flow calculation [9]. Some exemplary data are generated for electricity price rate, load demand profile, renewable electricity generation rate, and air pollution. Air pollution is updated after every time interval using Eq. 5 and Eq. 6 mentioned in section IV. For scenarios 2 to 5, resulting air pollution is compared to that of scenario 1 and the difference is represented in the simulation results as a measure called air pollution change. In the same fashion, peak pollution change refers to the difference between the peak values of pollution during the one week period of simulation for scenarios 2 to 5 with that of scenario 1. Final diagrams represent unit-less measures of balance/revenue, air pollution change, and peak pollution change so that we can compare outcomes of different scenarios for a single microgrid.

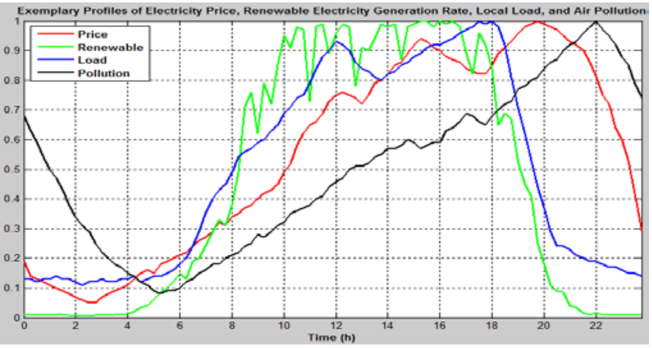


Figure 4 Profiles of Price, Renewables, Load, and Pollution

The normalized exemplary profiles for the four input variables, all or some of which are fed to the Fuzzy controller in different scenarios, are shown in figure 4 for a typical 24-hour period. These variables include electricity price, renewable electricity generation rate, local load demand, and air pollution. The data is generated arbitrarily for simulation purposes only considering similarity to the real world issues and with regard to the fact that the peak electricity consumption duration of the whole region of interest for the main grid occurs around 7:30 pm where the electricity price reaches its peak value.

Simulation results for five different scenarios are represented in figures 5 to 9. This must be noted that in this study, the renewable electricity generation plant is assumed to be able to fully provide microgrid's local load at its maximum generation rate conditions. Eq. 7 shows the relation between balance, distribution loss and the overall cost of electricity.

$$Balance = Cost - Loss \qquad \text{Eq. 7}$$

$$Loss = \sum_{t=1}^{T}\left(\Pr(t) . S_L(t)\right) \qquad \text{Eq. 8}$$

Where Cost is calculated using Eq. 1 and represents the amount that the microgrid owner has to pay to the main grid, if Cost > 0, or will get from the main grid, if Cost < 0. Loss stands for the overall sum of multiplication of the electricity price and wasted power on distribution branches, i.e. $S_L(t)$, for all 15-minute periods. Loss will always be greater than or equal to zero. Balance will then be the measure of the amount than microgrid owner had to pay to the main grid,i.e. Balance>0, or the amount of revenue that microgrid owner will get from the main grid, i.e. Balance<0, in case the power network were lossless or could be assumed to be lossless.

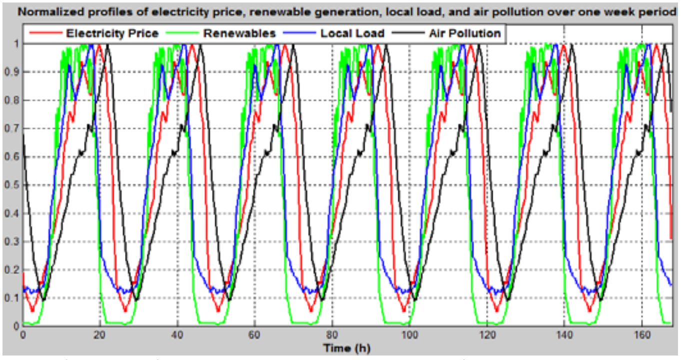


Figure 5 Price, Renewables, Load, and Pollution over one week

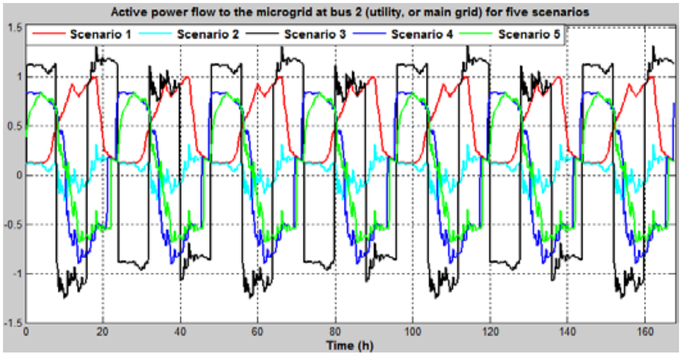


Figure 6 Power Exchange with the Utility or Main Grid; Five scenarios

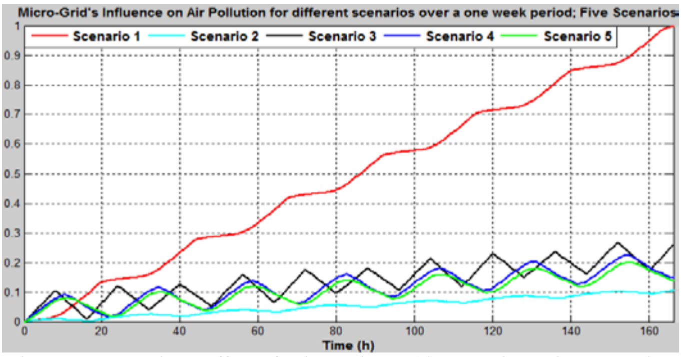


Figure 7 Normalized Effect of Microgrid on Air Pollution; Five scenarios

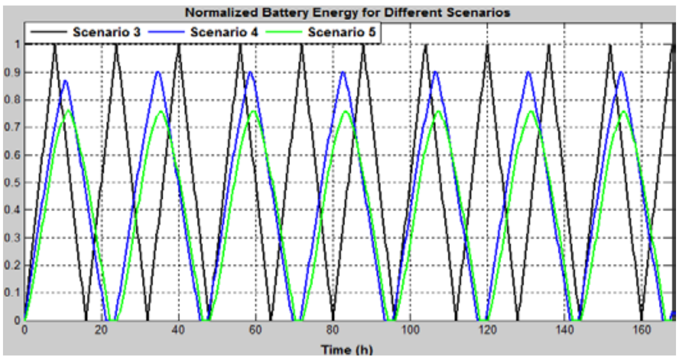


Figure 8 Normalized Battery Unit's Energy over a week; Five Scenarios

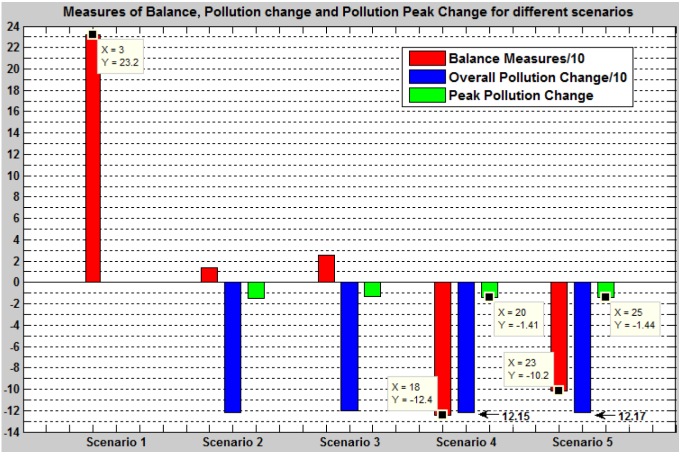


Figure 9 Final Measures of Balance, Pollution and Peak Pollution for

The center of gravity, i.e. centroid, defuzzification is used for computing crisp values of the output variable from union of the curves obtained by Fuzzy rules.

Figure 5 represents profiles of the four input variables of the Fuzzy system for the period of one week. Some factors of randomness and intermittency are associated with electricity price, local load, and renewable electricity generation rate in order to provide more realistic situations for the simulation. Air pollution is updated after each time interval using Eq. 5 and Eq. 6. In figure 6, power flow at bus number 2 which is the connection point between microgrid and the main grid is shown for five scenarios. It can be seen that in scenario 3 where the microgrid has renewable electricity generation unit and battery storage without any intelligent control system applied, there are cases when the power flow exceeds the value 1 p.u. which may be undesired for the system. This happens because of the fact that the storage unit is predefined to start from an initial condition and be charged to its full capacity before starting to discharge the stored energy to the microgrid. Hence, this can be concluded that applying an efficient control method to microgrid is of utmost priority when storage unit exists in the system. Figure 7 represents the normalized curves indicating amount of CO_2 added to the environment for different scenarios. Scenario 1 where no renewable generation system and no batteries are involved has the worst effects on environment, and scenario 2 which includes only the renewable generation unit without any storage units in the microgrid, has the best results in this regard. Also, scenario 5 which incorporates the air pollution control in Fuzzy inference engine stands right after scenario 2 on reducing CO_2 emission. The point is that in scenario 5 there are lots of profit for the microgrid owner and the consumers which is not the case for scenario 2. Normalized battery energy is depicted in figure 8 for three scenarios 3, 4, and 5. Figure 9 represents the three final measures of balance, pollution change, and peak pollution change for all five scenarios. The pollution measures represent how much other scenarios incorporate in increasing or decreasing the CO_2 emissions compared to scenario 1.

VI. CONCLUSION

The proposed Fuzzy logic-based approach is applied for energy management in a generic model of microgrid systems by controlling the flow of energy to/from storage unit. Five scenarios were considered for simulation. Measures of balance, pollution change, and peak pollution change represented that using efficient battery storage in association with renewable electricity generation units and intelligent decision-making approach can eliminate all the balances that the microgrid owner has to pay to the utility for electricity consumption and even bring revenue for the microgrid owner if the appropriate rule-base is defined for the inference engine. This must be noted that in this study, renewable electricity power plant life-cycle and costs, and also the costs associated with battery storage purchase, installation, and maintenance are not considered. Scenario 5 reveals the fact that applying the air pollution control policy to the Fuzzy inference engine rules will result in less pollution compared to all other scenarios, however, the balance measure will drop to some extent in comparison to scenario 4 which brings most financial benefits to the microgrid owner.

REFERENCES

[1] C. Cho; J. Jeon; J. Kim; S. Kwon; K. Park.; S. Kim, "Active Synchronizing Control of a Microgrid", IEEE Transactions on Power Electronics, issue 99, pp., 2011

[2] I. Maity, S. Rao, "Simulation and Pricing Mechanism Analysis of a Solar-Powered Electrical Microgrid", IEEE Systems Journal, Vol. 4.

[3] Y. S. Manjili; A. Rajaee; M. Jamshidi; B. Kelley, "Fuzzy Control of Electricity Storage Unit for Energy Management of Micro-Grids[1]", World Automation Congress (WAC), Mexico, 2012. (Accepted)

[4] N. L. Srinivasa Rao; G. Govinda Rao; B. Ragunath, "Power Flow Studies Of The Regional Grid With Inter State Tie-Line Constraints" IEEE Conference on Power Quality, pp. 165-171, 2002.

[5] R.D. Zimmerman; C.E. Murillo-Sánchez; R.J. Thomas, "MATPOWER's Extensible Optimal Power Flow Architecture", IEEE Power and Energy Society General Meeting, pp. 1-7, July 26-30 2009.

[6] R.D. Zimmerman; C.E. Murillo-Sánchez; R.J. Thomas, "MATPOWER Steady-State Operations, Planning and Analysis Tools for Power Systems Research and Education", IEEE Transactions on Power Systems, vol. 26, no. 1, pp. 12-19, Feb. 2011.

[7] S. L. Chang; L. A. Zadeh, "On Fuzzy Mapping and Control", IEEE Transactions on Systems, Man and Cybernetics, pp. 30-34, 1972.

[8] Z. Wang; R. Yang; L. Wang, "Intelligent Multi-agent Control for Integrated Building and Micro-grid Systems", IEEE PES Innovative Smart Grid Technologies (ISGT), pp.1-7, 2011.

[9] G.M. Gilbert; D.E. Bouchard; A.Y. Chikhani, "A Comparison Of Load Flow Analysis Using Distflow, Gauss-Seidel, And Optimal Load Flow Algorithms", IEEE Canadian Conf. on Elec. and Comp. Eng, 1998

Optimal control strategy for the wind power exchanges in a network of microgrids

Hanane Dagdougui
DIST Faculty of Engineering,
University of Genoa, Italy
hanane.dagdougui@unige.it

Ahmed Ouammi
CNRST, TEER
Rabat, Morocco
ouammi@cnrst.ma

Roberto Sacile
DIST Faculty of Engineering,
University of Genoa, Italy
roberto.sacile@unige.it

Abstract

Grid integration of wind power plants is becoming the main growing form of electricity generation among renewable energies. Recently, more interest is given to microgrids due to its potential benefits. The optimal operation of interconnected microgrids aims the matching of energy production with a dynamic load. In this paper, a control strategy for the wind power exchanges in a network of microgrids is formalized and tested on a Moroccan case study. The control variables are the instantaneous flows of wind power in the network of microgrids. The state is represented by the energy stored in each microgrid. This paper aims to minimize the variations of the energy stored in each storage device from a reference value, as well as to minimize the exchange of power between the microgrids.

Keywords: Wind energy; Optimal control; Smart microgrid; Morocco

1. Introduction

In recent years, global concerns about renewable energy alternatives have grow significantly, due to fossil fuel depletion, increased air pollution, climate changes and the rapid economical and societal development. In fact, renewable energy resources have been perceived serious attention in last decades due to their great potential in leading to a sustainable energy systems. The main problem of these energy resources is that their energy are intermittent, thus making the power generated fluctuant and uncertain.

Installation of renewable energy causes frequency fluctuation and distribution voltage fluctuation because output power from renewable source fluctuates due to weather condition.

Therefore it is possible to reduce electricity cost by achieving load following control using power storage facility [1]. It is necessary to smooth power flow from distribution system to achieve above technical problems and reduce electricity cost. Because of the above factors, smart grid concept is developed which cooperatively balances supply-demand between power supply side and power demand side [1].

The smart grid is a modern electric power grid infrastructure for improved efficiency, reliability and safety, with smooth integration of renewable and alternative energy sources, through automated control and modern communications technologies. Specifically, smart grid enabling new network management strategies provide their effective grid integration in distributed generation (DG) for demand side management and energy storage for DG load balancing, etc.[2].

The traditional electrical grid involves large, centralized power plants that feed power over an electro-mechanical grid to end users using one-directional power flows. While the traditional electrical grid has been successful, in recent years there have been numerous calls for the development of “smart” electrical grid by expanding the traditional electrical grid with distributed, medium-scale renewables-based energy generation systems and digital technologies [3]. DGs are expected to play an important role in future electricity supply and low carbon economy. However, high penetration of DGs into the grid environment will bring new challenges for the safe and efficient power system operation [4]. In fact, the interconnection of small and modular generation and energy storage systems to low or medium voltage distribution networks forms a new type of power system, named the microgrid (MG) which is an electrical load that can be controlled in magnitude [5]. The most important character of MGs is that the power generators are distributed and located in close proximity to the energy users. It can be interconnected to the larger electricity network, or can operate independently in a deliberate and controlled way. The interest in MG increases due to its potential benefits to provide reliable, secure, efficient and sustainable electricity from renewable energy sources [5].

In the literature, many authors [6-9] have studied the technological and methodological development of microgrids systems. This paper outlines the problem of a cooperation of a network of microgrid, where each one exploits wind sources as a source of power generation, and exchange real-time information on power demand and local energy production. The connections among microgrids might be used to exchange power and to decrease the variability of the load of the power exchanged with the connection to the main grid. The objective is to minimize the power exchanges among the grids, and to make each local storage system works around a proper optimal value. The problem is formalized as a model to support optimal decisions in a regional network of renewable power microgrids.

2. Optimal Control Strategy

The microgrids are modeled as a network of power generation, where each microgrid is composed by a wind power generation, an energy storage system and a cluster of households. The grid has the possibility either to put or to get power from one or more connections to other grids, thus under the external and internal power flows adopted.

The local power production of the grid is wind based, whose power production is supposed to be fully exploited. It is also supposed that the user demand can be fully satisfied. The main decisions are whether to store instantaneous exceeding energy production or to send it to some of the grid connections, or, alternatively, in case of lack of energy, whether it is convenient to acquire energy from some other grids or to use (if any) the energy stored in the local energy storage system.

These decisional aspects - specifically under a collaborative framework which is of interest for example in a regional network of microgrids - and their optimal solution are described hereinafter. Under this modeling vision, under a terminology viewpoint, each grid will be referred hereinafter as a “smart power microgrid” the connection to the local energy provider will be referred to as a connection to the “main grid”, and the overall set of microgrids connected among them and to the main grid will be referred to as the “network”.

Each smart power microgrid is supposed to be connected to a regional network of similar grids, and, at least for one microgrid, to one main grid. This network is modeled as a directed graph G=(V,E), where V is the set of vertex with cardinality S, representing either microgrids or the main grid, and E is the set of edges with cardinality W, representing the power connections existing among the vertexes. The microgrids are supposed to be subject to the following discrete time state equation [10, 11]:

$$z_{t+1} = Az_t + (in_t - out_t + Bu_t)\Delta t$$
$$t = 0..T-1 \qquad (1)$$
$$z_0 = z0$$

where

$u_t \in R^W$ (kW), decisional variables, is the vector of power flows sent to (or received by, when the element of u_t is a negative value) other grids in time interval $(t, t+1]$. Specifically, the generic j-th element is the directed flow of power in link j in time interval $(t, t+1]$.

$z_t \in R^{S-1}$ (kWh), state variables, is the vector of energy storage inventory of each microgrid at instant t. Specifically, the generic i-th element is the energy inventory in the storage of microgrid i at time instant t, with respect to an optimal working level $z^{*,i}$ for the storage technology present in that microgrid. In this respect, it is supposed that each element of z_t may assume both positive and negative values.

$in_t \in R^{S-1}$ (kW) is the vector of stochastic processes of power flow in input to the microgrid in time interval $(t, t+1]$ as given by the renewable energy sources exploited in each microgrid.

$out_t \in R^{S-1}$ (kW) is the vector of stochastic power demand processes of each microgrid in time interval $(t, t+1]$.

$0 \leq \alpha \leq 1$ is a parameter representing the efficiency of the device used to store energy.

A is a $(S-1)\text{x}(S-1)$ diagonal matrix describing, in each diagonal element α_{ii}, the efficiency of the energy storage technology in the i-th grid. In this respect, it holds that $0 \leq \alpha_{ii} \leq 1$.

B is the $(S-1)\text{x}W$ incidence matrix, representing the network topology, such that each element $b_{i,j} = -1$ if there is an edge (that is a power connection) leaving the i-th microgrid, 1 if it enters the i-th microgrid and 0 otherwise.

Δt is the optimization time interval.

Under the hypothesis that in_t and out_t can be forecasted on a given interval $[0, T]$, with a certain degree of uncertainty, both of them can be split into their deterministic (respectively in_t^d and out_t^d) and stochastic (respectively ω_t^{in} and ω_t^{out}) vector components as described in equation (2) and (3):

$$in_t = in_t^d + \omega_t^{in} \qquad (2)$$
$$out_t = out_t^d + \omega_t^{out} \qquad (3)$$

Let e_t be the vector of energy balance in each microgrid given by :

$$e_t = in_t - out_t \quad (4)$$

Under a simplifying hypothesis, let $e(t)$ be a vector whose elements are represented by a Gaussian white noise, where:

$$E\{e_t\} = \mu_t = in_t^d - out_t^d \quad (5)$$

So, μ_t is a completely known function on the given interval $[0, T]$, while ω_t, defined in (6), is a zero-mean normal distribution vector with variance n, not correlated with μ_t.

$$\omega_t = e_t - \mu_t \quad (6)$$

Under the assumption quoted above, and, to simplify the notation, assuming $\Delta t = 1h$, equation (1i) represents the state equation of the energy storage of the microgrid network:

$$z_{t+1} = Az_t + Bu_t + \mu_t + \omega_t \qquad t = 0..T-1 \quad (1i)$$
$$z_0 = z0$$

where here:

$\mu_t \in R^{S-1}$ is a vector whose i-th element is the deterministic energy balance of microgrid i in time interval $(t, t+1]$, as resulting from the difference of the predictions of RES power supply and user demand.

$\omega_t \in R^{S-1}$ is a vector whose i-th element is the stochastic error in the prediction of the deterministic energy balance μ_t of microgrid i in time interval $(t, t+1]$, and modeled as a zero-mean normal distribution vector with variance n, not correlated with μ_t.

Supposing that there is a perfect knowledge of the state of each local storage, and under a cooperative strategy among the grids, whose aim is to maintain an optimal level of energy in the distributed local storage systems, as well as to achieve a low flow of power among the grids, the following objective is formulated.

$$\min J(z, u) = E\{\textstyle\sum_{t=0}^{T-1} c(z_t, u_t) + z_T' M_T z_T\} \quad (7)$$

where

$$c(z_t, u_t) = z_t' M z_t + u_t' N u_t \quad (8)$$

M is a $(S-1)x(S-1)$ matrix, related to the cost of an exceeding/lacking quantity of energy stored in each energy storage device. This matrix is supposed to be $M > 0$, and constant for each instant $t \neq T$.

M_T, $M_T > 0$, has the same definition of M, but it is only defined for instant $t = T$.

N is a WxW matrix, $N > 0$, related to the cost of the power sent on each edge of the network, whose elements are constant for each time interval t.

The problem is so completely defined by the cost function (7) and by equation (8), subject to the state system (1i).

The problem is a linear quadratic Gaussian (LQG) problem that is a "non-standard" LQG due to the presence of the known input μ_t in the state equation.

Theorem

For the problem, defined by (1i), (7) and (8) the optimal control is (the proof is available in [10, 11]) :

$$u_t^* = K_t(z_t - z_t^{d2}) + K_t^g g_{t+1} \quad (9)$$

K_t is a $Wx(S-1)$ matrix given by

$$K_t = -(N + B'P_{t+1}B)^{-1}(B'P_{t+1}A) \quad (10)$$

where P_{t+1} is a $(S-1)x(S-1)$ matrix given by the discrete time algebraic Riccati equations (DARE):

$$P_t = M + A'P_{t+1}(I + BN^{-1}B'P_{t+1})^{-1}A \quad (11)$$
$$P_T = M_T$$

K_t^g is a $Wx(S-1)$ matrix given by

$$K_t^g = (N + B'P_{t+1}B)^{-1}B' \quad (12)$$

the vector z_t^{d2} is given by

$$z_{t+1}^{d2} = Az_t^{d2} + \mu_t \qquad t = 0..T-1$$
$$z_0^{d2} = z0 \quad (13)$$

and the vector g_t is given by:

$$g_t = (A' - A'P_{t+1}(I + BN^{-1}B'P_{t+1})^{-1}BN^{-1}B')$$
$$g_T = M_T z_T^{d2} \quad (14)$$

3. A case study: A Network of three Microgrids (Morocco)

Three microgrids each one composed by 500 households (Tetouan (G1), Tanger (G2) and Essaouira (G3)) can produce power for their local user community using wind energy. It should be noted that the micrograids are connected and the third one is connected to the main grid, with which it may exchange power. Fig. 1 shows the existing connections between microgrids.

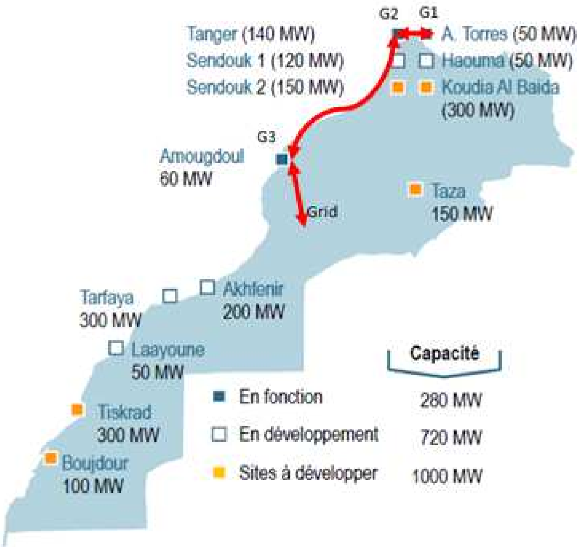


Figure 1. Directed connections planned among the microgrids.

3.1 Electrical energy demand of the households

The electric energy consumption per habitant in Morocco is given by Table 1. It shows that the consumption per habitant increases from 1.3 to 2 kWh which remain very low consumption. The real daily and monthly consumptions of a household (four persons) in Morocco are presented in Figures 2 and 3. The electrical energy consumption model for the households is based on the above statistics, assuming that each microgrid in the case study is composed by 500 households. Figure 4 shows the demand of each microgrid.

Table 1. Electrical energy consumption in Morocco [12].

	Consumption (kwh/hab)	Daily (kwh/hab)	Monthly (kwh/hab)
2002	483	1.3	39.7
2003	515	1.4	42.3
2004	545	1.5	44.8
2005	584	1.6	48.0
2006	631	1.7	51.9
2007	666	1.8	54.7
2008	694	1.9	57.0
2009	710	1.9	58.4
2010	744	2.0	61.2

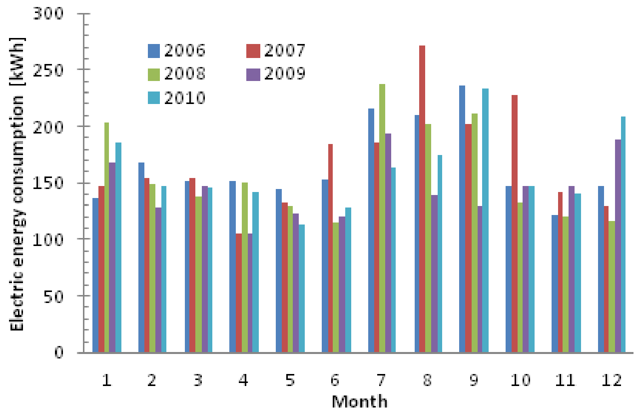


Figure 2. Monthly electrical consumption of a household (four persons) in Morocco

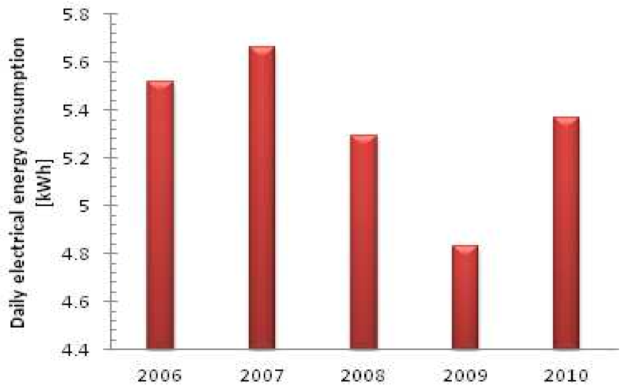


Figure 3. Daily electrical consumption of a household (four persons) in Morocco

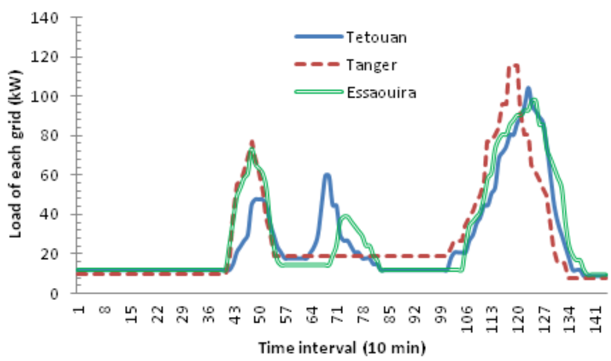


Figure 4. Load of the microgrids

3.2 Wind condition data

The annual wind speed characteristics of the three sites have been analyzed. The available mean annual wind speed of Tetouan, Essaouira and Tanger is equal respectively to 11, 8.5 and 7 m/s. It can be observed that the first site is more promising than the others. Figure 5 shows the wind speed profile of the sites used in this simulation.

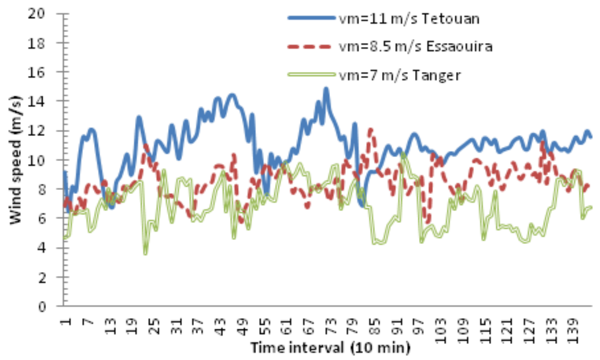
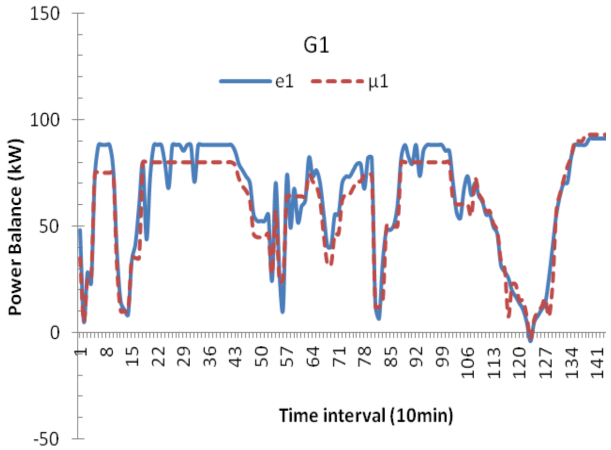


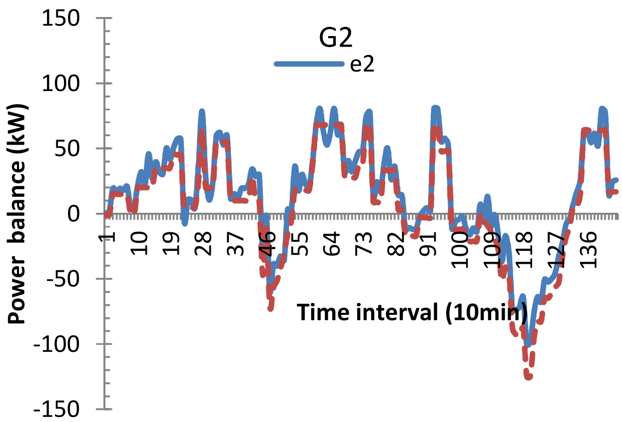
Figure 5. Wind speed profile of the sites

4. Results and discussion

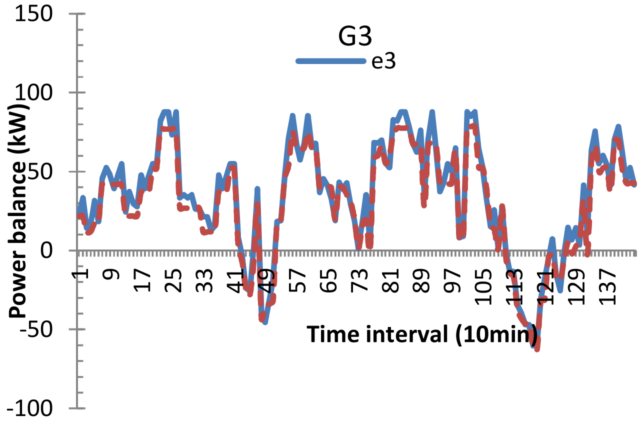
The network to be studied is composed by S=4 vertexes, where the 4-th vertex is associated to the main grid. In the network there are W=3 links, including the new connection whose performance has to be evaluated.



(a)



(b)



(c)

Figure 6 (a, b, c). The prediction of μ_t^i , and the a posteriori true power balance e_t^i for the microgrid $i = 1,..3$ (in top down order)

References

[1] Kenichi Tanaka, Akihiro Yoza, Kazuki Ogimi, Atsushi Yona, Tomonobu Senjyu, Toshihisa Funabashi, Chul-Hwan Kim. Optimal operation of DC smart house system by controllable loads based on smart grid topology. Renewable Energy 39 (2012) 132-139

[2] Vehbi C. Güngör, Dilan Sahin, Taskin Kocak, Salih Ergüt, Concettina Buccella, , Carlo Cecati, , andGerhard P. Hancke. Smart Grid Technologies: Communication Technologies and Standards. IEEE Transaction on Industrial Informatics, vol. 7 2011.

[3] Wei Qi, Jinfeng Liu, Panagiotis D. Christofides. A distributed control framework for smart grid development: Energy/water system optimal operation and electric grid integration, Journal of Process Control 21 (2011) 1504–1516.

[4] Amjad Anvari Moghaddam a, Alireza Seifi a, Taher Niknam, Mohammad Reza Alizadeh. Pahlavani Multi-objective operation management of a renewable MG (micro-grid) with back-up micro-turbine/fuel cell/battery hybrid power source. Energy 36 (2011) 6490-6507.

[5] Mohammadi M., Hosseinian S.H., Gharehpetian G.B. GA-based optimal sizing of microgrid and DG units under pool and hybrid electricity markets, Electrical Power and Energy Systems.

[6] X. Li, Yu-Jin Song, Soo-Bin Han, "Study on Power Quality Control in Multiple Renewable Energy Hybrid MicroGrid System Power Tech", 2007 IEEE Lausanne, 2007, pp 2000 - 2005 DOI: 10.1109/PCT.2007.4538625.

[7] X. Yu, A.M. Khambadkone, H. Wang, and STS Terence. Control of Parallel-Connected Power Converters for Low-Voltage Microgrid—Part I: A Hybrid Control Architecture, IEEE Transaction on Power Electronics,Vol. 25, No. 12, December 2010.

[8] Il-Yop Chung, W. Liu, DA. Cartes, and K. Schoder, Control Parameter Optimization for a Microgrid System Using Particle Swarm Optimization, IEEE International Conference on Sustainable Energy Technologies. DOI: 10.1109/ICSET.2008.4747124. 2008 , P: 837 – 842.

[9] D. Zhu, R.Yang, G. Hug-Glanzmann, “ Managing Microgrids with Intermittent Resources: A Two-Layer Multi-Step Optimal Control Approach, ”North American Power Symposium (NAPS), 2010.

[10] Minciardi R, and Sacile R. Optimal control in a cooperative network of smart power grids IEEE Syst. J 2011;doi: 10.1109/JSYST.2011.2163015

[11] Dagdougui H, Ouammi A, Sacile R. Optimal control of a regional network of power microgrids. IEEE International Systems Conference Montreal, Quebec, Canada, 2011.

[12] http://www.mem.gov.ma/

Development of innovative systems for operation and control of electric power distribution networks: management and optimal use of distributed generation and of renewable energy resources

R. Caldon (1), S. Massucco (2), C.A. Nucci (3), F. Pilo (4), P. Verde (5)

(1)University of Padova, (2)University of Genova, (3) University of Bologna,
(4) University of Cagliari, (5) University of Cassino
Italy

roberto.caldon@unipd.it; stefano.massucco@unige.it, carloalberto.nucci@unibo.it, pilo@diee.unica.it, verde@unicas.it

Abstract - *The integration of distributed energy resources for the production of electrical energy is related to the evolution of future power distribution networks in order to achieve: flexibility (meeting final user's requirements); accessibility (allowing access to local generation and particularly to renewables); reliability (ensuring the highest possible security and power quality levels); economy (allowing an adequate management of energy in an efficient and competitive way).In this framework, the paper presents some results of a joint research project involving five Italian universities. The focus is on the proposition and development of the main functionalities implemented into a centralized Distribution Management System (DMS) for the operation and control of the energy resources connected to distribution networks (generation, storage and loads) and of the network itself (automation systems, protection, etc.). Architecture, hierarchical structure and relevant functions for the control, operation and management of the active distribution network, are illustrated.*

Keywords: *Distributed Generation, active distribution networks, renewable energy sources.*

1 Introduction

The integration of distributed energy resources for the production of electrical energy is related to the evolution of future power distribution networks in order to achieve: flexibility (meeting final user's requirements); accessibility (allowing access to local generation and particularly to renewables); reliability (ensuring the highest possible security and power quality levels); economy (allowing an adequate management of energy in an efficient and competitive way) [1].

These requirements can be satisfied through the attainment of the following two main tasks:

- the development of innovative methodologies and technologies for the operation and control of the energy resources connected to distribution networks (generation, storage and loads) and of the network itself (automation systems, protection, etc.);
- the definition of standards, protocols and adequate market regulatory structures that would make it possible the evolution from the old to the new network asset.

The automation of the electrical system is integrated in a structured control hierarchy in order to satisfy various necessities relevant to the delivery of electrical energy [2]. For the management of electric distribution networks a structure similar to well known EMSs (Energy Management Systems) adopted for the transmission network, may be suitably adapted. In analogy to EMSs, electric distribution networks will use similar distributed structures suitably coordinated at different levels. In this respect, the term DMS (Distribution Management System) is generally used to identify such a specific structured control hierarchy for electric distribution networks [3].

In this framework, this paper describes the proposition and development of a DMS of which control, operation and management functions have been developed at the Authors' laboratories. Its interface with remote monitoring systems (of both network and dispersed resources) suitably developed by means of innovative technologies represents a key feature. In this respect the use of advanced Phasor Measurement Unit devices, developed at one of the laboratories involved in the research project, is discussed in the paper. Finally, as the development of the DMS functions is related to the regulatory scenarios and, at the same time, the evolution of the future distribution networks is influenced, for a fixed scenario, by the DMS itself, the paper addresses also the issue of the definition of a framework of regulatory scenarios in which the DMS and its control and operation functions are expected to be tested and assessed.

The paper is structured in the following way: in the second section, the architecture, control hierarchy, and functions of the DMS are presented along with its control and operation functions. In the third section the main functionalities of the DMS for what concerns the state estimation and the optimization of the distribution network are illustrated. In this section the use of advanced Phasor Measurement Unit devices, developed at one of the laboratories involved in the research project will be illustrated too. A further subsection is devoted to the discussion on power quality issues while the fifth section is devoted to the Conclusions.

2 DMS Architecture

It is widely recognized that local controls are no longer suited to mitigate voltage regulation issues caused by bidirectional power flows or to manage faults in distribution systems with high shares of DER (Distributed Energy Resources) [4]. The debate on decentralized or centralized control systems for Smartgrid and active distribution system is still open. There is a general consensus on the need of hierarchical control, but the techniques for the system optimization span from central control systems to autonomous agents [5]. Autonomous agents have the great benefit to reduce the need of broadband communication systems, but in many cases there is the need of a central coordination that reduces the value of agents. The centralized control systems with hierarchical structure are quite easy to implement, provided that a clear regulation framework is established in order to allow the DSO (Distribution System Operators) to make transparent and market based decisions. Anyway, in order to improve the reliability local resources have to be equipped with local intelligence capable to make decisions when the communication flow from the control center is interrupted. A centralized active management scheme generally consists of:

- a control center sited in a relevant PCC (Point of Common Coupling), e.g. in the primary substation; in this control center there are at least a DMS and a Distribution System Estimator (DSE);
- the DER local controllers (LC) that send/receive communication signals to/from the control center (i.e. place bids for the next time interval and receive the control actions for DERs), and also can take a decision if the signal from the control center is interrupted;
- a measurement system, consisting of a few measurement devices in the field; it is able to send measurement signal to the DSE;
- a communication system synchronized with a GPS system for time reference and the exchange of measurement data and control signals between control center and LCs.

Ad hoc DSE algorithms that provide the real-time status of the network, by gathering data from the distributed measurement system (insufficient at distribution level) and other available information retrieved from historical data (pseudo-measurements) have to be used in the DSE frame of the control center [6], [7]. The DMS, supervises the operation of the electric distribution network and, if necessary, modifies the set points of DERs (e.g., generators, storage devices), and responsive loads according with the results of the optimization. DERs (DG owners and Responsive Loads-RL) send day by day to the control center bids for the one day-ahead active and/or reactive power generation or load demand. Furthermore, they also offer their support to the active distribution networks (ADNs) operation for the next time interval in the intra-day market, by offering changes to production schedule and/or load demand. DSOs may adjust the day-ahead scheduling paying producers and RLs when their set points are to be changed, according to the regulatory environment. Moreover, during the day DSO provides to the DERs the control actions for the active management of the network, based on the results of an intra-day optimization. Finally, if the active management reaches an advanced level of implementation, also the network reconfiguration can be profitably exploited.

The DMS solves the optimization problem running in real-time on a dedicated DSP or on industrial computers that can be sited in the primary substation. Once the time horizon (typically one day) and the time interval (e.g. 1 hour, but even shorter) are defined, at the beginning of the time interval the DMS receives the status of the network, the technical constraints, as well as the market prices and information on energy trades. Furthermore, the DMS collects bids from DERs for the next time interval. The new set points and the current network topology are hold until the end of the time interval, when new data are gathered from the network and used for a new optimization.

In order to describe more in details the typical structure of DMS, it is worth noting that electric distribution network can be managed from a single control center and/or from distributed control centers on the territory [3], [8]. As known, control centers use a so-called SCADA (Supervisory Control and Data Acquisition) - and are based on the communication between the control center itself and the primary equipment that can be controlled (generation, breakers, on load tap changers, etc). These devices must necessarily be equipped with actuators that allow carrying out the operation demanded from the control center. The communication between the control center and the actuators is possible by use of secondary devices called Intelligent Electronic Devices (IED).

Two are the typical approaches in the development of DMS applications: (a) “GIS-centric” solution; (b) “SCADA oriented” solution.

The first solution makes reference to the term GIS - Geographical Information System and is more oriented to the off-line management of the distribution systems. The more typical functions of the distribution systems, like fault detection and network restoration have been implemented in the past thinking more to the territory and therefore with requirements of having a clear picture of the distribution network using software solutions more oriented to off-line control and management.

Figure 1 illustrates the general structure of a possible advanced DMS.

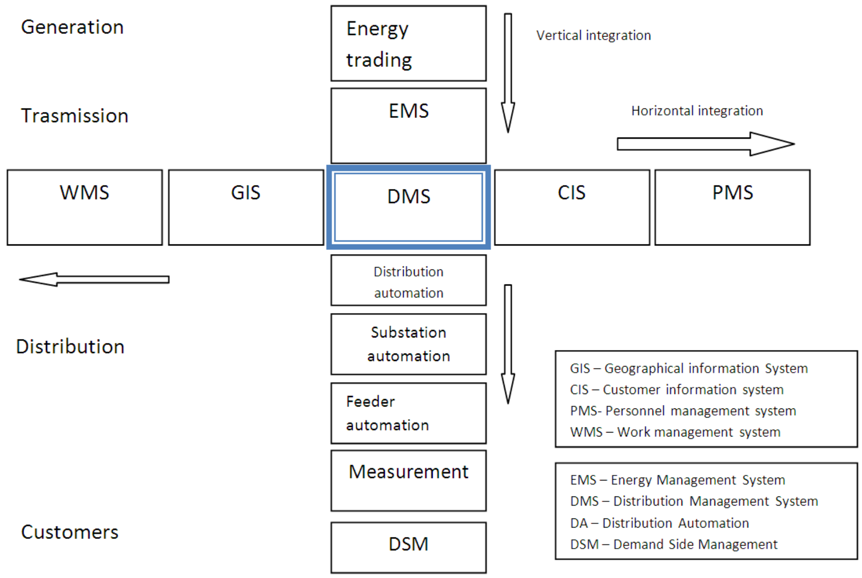


Fig. 1 - Vertical and horizontal integration in distribution networks (adapted from [8]).

3 State Estimation and Optimization in DMS

3.1 State Estimation of Active Distribution Networks

Innovative management strategies in active distribution networks require a reliable State Estimation (SE) of the system in terms of line flows and bus bar voltages. Traditional SE algorithms, adopted in transmission systems, appear hardly applicable in distribution networks because of the lack of available real-time measurements, with a consequent high level of uncertainty resulting in the estimations [10].

At present, in the distribution networks, the availability of real-time measurements is often limited to the Primary Substation (PS) (MV bus-bar voltage and P-Q flows on the outgoing feeders). In this case a large number of so-called pseudo-measurements has to be used to guarantee the network observability [11]. Pseudo-measures may represent estimates relevant to load consumptions or to DG injections and are usually afflicted by a high uncertainty, due to the impracticability of a widespread measurement system. High levels of uncertainty negatively affect the accuracy of voltage estimates, preventing an effective control of network voltages. Two different strategies may be adopted to ensure a suitable reduction of the estimated voltages uncertainties:

– reduction of pseudo-measurements uncertainty by using load modelling techniques [12].
– introduction of further remote on-line measurements. In particular, further voltage measurements in critical nodes produces the best results in reducing the degree of uncertainty [13]; to this aim a suitable algorithm has been developed to provide the optimized measurements location [14],[15].

The method evaluates the network operational state by computing a suitable parameter, denoted as "variance moment", in accordance with the time-variable probabilistic value of loads demand and generators production respectively. The uncertainties of the active and reactive injections in each node and its distance from the Primary Substation define the "variance moment" parameter.

The calculation of the variance moment is based on a matrix computation able to handle both network data (topology, composition, status of switches) and load and DG characteristics. The load uncertainty is considered equal to three times the standard deviation of the probabilistic distribution representing such load (and similarly for the generating plants). Active and reactive consumptions are separately modelled. The uncertainties on power injections are represented node by node in the diagonal matrix U_L (1), where N is the number of nodes while σ_{Pi2} and σ_{Qi2} are the variances of the overall active and reactive injected power at the ith node. The additional measurement points have been optimally allocated adopting the following procedure.

$$\mathbf{U_L} = \begin{bmatrix} \frac{\left(\sigma_{P1}^2 - j\sigma_{Q1}^2\right)}{V_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{\left(\sigma_{PN}^2 - j\sigma_{QN}^2\right)}{V_N} \end{bmatrix} \begin{matrix} 1^{st}\ node \\ \\ N^{th}\ node \end{matrix} \qquad (1)$$

The method requires that one measurement is available at least in the power system (in case of complete absence of remote voltage measurement). If other bus-bars are already equipped with on-line voltage measurements, so this nodes are introduced in the process as forcing parameters equal to one in the N_Vx1 matrix V, naming as N_V the number of on-line voltage measurements. The value of N_V determines the size of the four sub-matrixes in which the admittance matrix Y is split, as in (2).

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y_1} & \mathbf{Y_2} \\ \mathbf{Y_3} & \mathbf{Y_4} \end{bmatrix} \begin{matrix} 1:\mathrm{N_V} \\ (\mathrm{N_V}+1):\mathrm{N} \end{matrix} \quad \begin{matrix} 1:\mathrm{N_V} & (\mathrm{N_V}+1):\mathrm{N} \end{matrix} \tag{2}$$

The value of N_V defines also the matrix U_L* which is extracted from UL as in (3) (distribution of power injections uncertainty), while matrix V* is consequentially, determined as in (4), is a (N-NV)x1 vector and represents a sort of "performance drop" in the estimated voltage accuracy.

$$\mathbf{U_L} = \begin{bmatrix} \left(\sigma_{P1}^2 - j\sigma_{Q1}^2\right) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mathbf{U_L^*} \end{bmatrix} \begin{matrix} (N_V+1):N \end{matrix} \quad (N_V+1):N \tag{3}$$

$$\overline{\mathbf{V}}^* = -\left(\mathbf{U_L^*} + \mathbf{Y_4}\right)^{-1} \cdot \mathbf{Y_3} \cdot \overline{\mathbf{V}} \tag{4}$$

Finally the variance moment M is obtained as in (5) for each node where a voltage measurement is not performed (by definition, nodes 1...N_V, where a voltage measurements is performed, have variance moment equal to one). M is a (N-N_V)x1 matrix containing in the tth row the value of the performance parameter of the (N_V+t)th node. In case of radial networks, with on-line measurements available only at the primary substation, the variance moment represents the products of variances of the active and reactive power injections and their relevant electrical distance from the power system.

$$\overline{\mathbf{M}} = \begin{bmatrix} 1-\left|\mathbf{V}^*(1)\right| \\ \vdots \\ 1-\left|\mathbf{V}^*(N-N_V)\right| \end{bmatrix} \quad [MVA^2 \cdot \Omega] \tag{5}$$

Once the matrix M has been obtained, the algorithm calculates which node of the grid is characterized by the maximum variance moment. If the voltage estimation uncertainty exceeds the imposed limits, the algorithm allocates an additional measurement point in correspondence to the generator nearest to the node with the highest variance moment; the matrix M is re-calculated and the procedure continues until the voltage estimation uncertainty returns within the limits in every node of the monitored network.

3.2 DMS Operation

The DMS makes the system complying with the technical constraints at the minimum cost by resorting to:

- DG Generation Curtailment (GC),
- Reactive power exchange from DG (Ancillary Service-AS),
- Storage devices control,
- Demand Side Integration (DSI),
- OLTC control in the primary substation
- Network reconfiguration.

The DMS may be tripped only by constraint violations like voltage regulation problems (typically over-voltage caused by DG and voltage drops caused by high load) and thermal overload. DMS can also continuously strive to cut costs or reduce the exploitation of the existing assets (e.g. the network may be reconfigured to reduce the energy losses or to minimize the DG curtailed power).

In each time interval, the DMS finds the optimal combination of the available operation options and outputs the optimal set points to be sent to the local controllers of DERs and the open/closed status of the branches available for the network topology. Given one network topology, the problem to find the optimal combination is an OPF (Optimal Power Flow) problem. In the proposed DMS, load flow equations have been linearized to use linear programming in order to reduce the computing burden so that the algorithm can be applied to real time calculation. The DMS is also able to find the optimal network configuration within a range of optimal topologies to reduce the energy losses and to limit or avoid overloads in branches. In this case, the DMS also outputs the open/closed status of remote controlled breakers in the weakly meshed distribution network (Fig. 2). The external loop assesses all N_{config} feasible configurations and chooses the topology with the smallest calculated value of the Objective Function (O.F.). The proposed DMS architecture has been tested on real size networks and proved to be effective in allowing high shares of RES to be integrated in the system. The resort to flexible topologies allows reducing both energy losses and the variations to the production of RES. The research is now focused on the optimal control of storage devices.

Active networks require that accurate data about the network conditions are continuously available from the field. Large scale distributed measurement systems are necessary to carry out simultaneous measurements of electrical quantities in several monitored points on the system. Therefore, suitable techniques of DSE are necessary. The DSE provides a complete and consistent model of the operating conditions and it is essential for the DMS operation. The optimal placement of measurement devices is also covered.

Currently, the research is focused on the impact of uncertainties in state estimation on DMS.

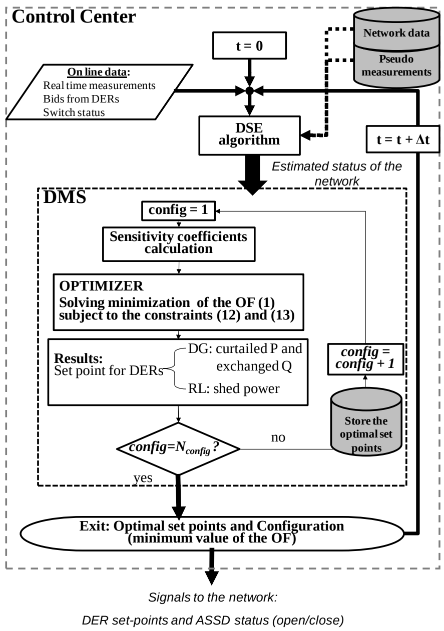


Fig. 2 - Flowchart of the proposed procedure.

The quality of DSE can be affected by different issues:

- the number and the position of the available measurement devices;
- the uncertainties introduced by the measurement devices;
- changes of the network topology (network reconfiguration) and deviations from their nominal value of the network parameters;
- partial lack of communication (emergency mode).

3.3 Optimal scheduling of Distributed Energy Resources

A possible solution for the optimal control of DER has been proposed in [16] where a two-stage scheduler has been proposed. It is composed by a day-ahead economic scheduler, that calculates the active power set points during the following day in order to minimize the overall costs, and an intra-day scheduler that, on the basis of measurements and short-term load and renewable production forecasts, updates the DERs and control set points every 15 minutes. The need for 24-hour horizon function is justified by the requirement for an optimal use of the available energy storage facilities and by inter-temporal operation constraints. The optimal intra-day scheduling of DERs is based on the use of a detailed three-phase load flow calculation and a MILP (mixed-integer linear programming) optimization algorithm. Fig. 3 shows the main concepts of the procedure in which the multi-objective function of the scheduling problem consists in the minimization of the voltage deviations with respect the rated value of the DERs production deviation with respect the maximum efficiency point calculated by the day-ahead scheduler. Also network losses are part of the objective function. The solution is based on the three-phase power flow calculation in which, in normal operating condition, the connection to the primary system is the slack bus. At each iteration, the initial values of the DERs control variables are modified by $\xi\Delta x$, where coefficient ξ is calculated so to minimize the value of objective function by means golden section method.

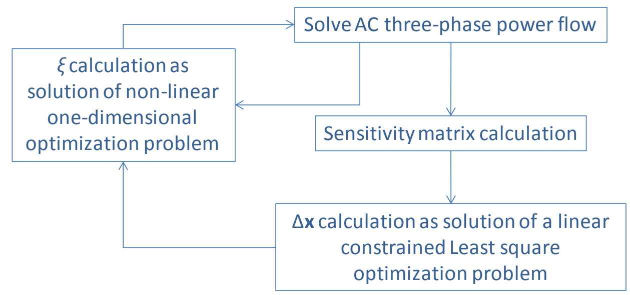


Fig. 3 - Scheme of the intra-day scheduling procedure.

3.4 Dynamic phasor monitoring of distribution systems

The evolution of power distribution networks from passive to active is determining major changes in their operational procedures: one of the main involved aspects is indeed the real-time monitoring of these networks. The above issue calls for a massive use of advanced and smarter monitoring tools that result into faster and reliable real-time state estimation of these networks. One of the most promising technologies in this field is certainly represented by the distributed monitoring based on the use of PMUs [17].

Synchrophasor estimation algorithms proposed in the literature are essentially based on the Discrete Fourier Transform (DFT) applied to quasi-steady state signals representing network node voltages and/or branch current waveforms. These DFT-based algorithms can be grouped into one-cycle DFT estimators, and fractional-cycle DFT estimators performing recursive and non-recursive updates [18]. Within this context, the algorithm presented by the Authors in [19], [20] belongs to the DFT algorithms; it has been conceived in order to: (i) allow the use of PMU in active distribution networks and (ii) keep the synchrophasor measurement accuracy within specific limits even in presence of distorted signal waveforms and electromechanical transients (namely, with frequency-varying signals). Concerning point (i), it is worth noting that, compared to transmission networks, active distribution networks are characterized by reduced line lengths and limited power flows. With reference to the use of bus voltage synchrophasors for the network state estimation, these characteristics result, in general, into very small phase differences between bus voltage phasors (generally in the order of tens of mrad or less). These characteristics calls for PMU devices characterized by synchrophasor phase uncertainty well below the limits provided by the IEEE Std. C37.118 [17]. Concerning point (ii), it is worth noting that distribution networks are characterized by much higher distortion levels than those of transmission networks.

Additionally, as active distribution networks are expected to operate even when islanded from the main transmission networks, PMUs appear a useful tool to support distribution system operators during the islanding and reconnection manoeuvres. In this respect, the application of PMU to monitor electromechanical transients, generally characterized by non-negligible deviations from the rated network frequency, could involve important bad estimation of the synchrophasors phases and frequencies.

3.5 Power Quality Issues in Active Distribution Networks

Regulatory scenarios give the framework to develop the DMS functions. For a fixed scenario, the evolution of the future distribution networks is influenced by the DMS itself. It is then important to define regulatory scenarios in which the DMS, its control and operation functions are expected to be tested and assessed. Among the various aspects regulated by National Authorities, like in Italy the Authority of Electric Energy and Gas (AEEG), the Power Quality (PQ) issues represent the constraints to be matched by the distributors in a liberalised market. In fact, regulators are increasingly utilizing performance based ratemaking (PBR) as a tool to promote efficiency in the provision of distribution services. Because of the cost pressures of deregulation and PBR regulation, most PBR schemes include minimum service quality requirements to ensure that utility service quality performance does not fall below desired levels. Most PBR regulatory schemes enforce the minimum standards for service quality performance through economic incentives for exceeding service quality targets or penalties for failing to meet those targets.

The presence of the DG globally influences the service quality that can be guaranteed to the final users, and, however, can be used also to compensate PQ disturbances exchanging with the network additional services, namely auxiliary services for PQ, besides the energy. This possibility requires: the presence of converters operating as distributed generation interface with the network and the measurability of the auxiliary services for PQ.

The measurability of the auxiliary services for PQ allows the centralised management verifies that the functions assigned to the active nodes have been performed, and then can be remunerated. In particular, it is worthwhile to evidence that the economical evaluation of PQ is one of the emerging issues in the National and International community of the research also for the push impressed by liberalised markets [21].

To develop a DMS able to implement functions for the management and the control of the network taking into account the PQ constraints, it is needed to ascertain and implement adequate methods and indices to estimate the PQ performance of the active networks.

First steps of the research have considered two main phenomena: the unbalances and the voltage dips.

Regarding the unbalances, the general problem of facing also the uncertainties that affect the evaluation of steady state operating conditions has been taken into account using the point estimate method. [22]. Moreover, since the point estimate method requires that the input random variables are uncorrelated, a suitable adjustment to take into account the correlation is applied. Different point estimate schemes (2m, 2m+ 1 and 4m+ 1 schemes) are presented and tested. The accuracy of the proposed techniques is tested on a three-phase unbalanced IEEE 34-bus test system; the results obtained applying the Monte Carlo simulation are assumed as reference. Both correlated and uncorrelated input random variables are considered, and multimodal probability density functions are tested. The final results evidenced that the 2m+1 scheme gives the best solution in terms of accuracy and computational efforts; in the case of correlated input random variables, an adequate procedure to take into account the correlation must be applied.

4 Conclusions

Major research activity is on-going on the subject of active distribution networks. The management and optimal use of distributed generation in distribution electrical networks is attracting several efforts and research funding.

In this context, the paper has presented the on mid-term results of a research national project that is involving five Italian Universities.

The architecture of a Distribution Management System (DMS) for active distribution network control has been presented focusing on its interface with remote monitoring and control systems of distribute resources including storage and loads.

Details and methodological insights on the DMS functions and operation such as state estimation, local controllers, communication systems and optimal scheduling strategies are provided.

The crucial importance of adopting new concepts for distribution network planning and operation, of using appropriate ICT technologies and equipment has been stressed. The necessity of identifying regulatory scenarios capable of permitting the full exploitation of distributed resources – generation, load and storage systems - has been put into evidence.

References

[1] European SmartGrids Technology Platform, "Vision and Strategy for Europe's Electricity Networks of the future", Document EUR 22040, 2006

[2] J.A.P. Lopes, N. Hatziargyriou, J. Mutale, P. Djapic, N. Jenkins, “Integrating distributed generation into electric power systems: A review of drivers, challenges and opportunities”, Electric Power Systems Research 77 (9), pp. 1189-1203, 2007.

[3] J. Northcote-Green, R. Wilson, Control and Automation of Electric Power Distribution Systems, Taylor & Francis Group, Boca Raton, 2007

[4] F. Pilo, G. Pisano, and G. G. Soma, 2011, "Optimal Coordination of Energy Resources with a Two-stage on-line Active Management", accepted for publication, IEEE Trans. on Industrial Electronics, 2011.

[5] C. Muscas, F. Pilo, G. Pisano. S. Sulis, "Optimal allocation of multi-channel measurement devices for Distribution State Estimation", IEEE Trans. Instrumentation and Measurement, June 2009, vol. 58, n.6.

[6] U. Häger, S. Lehnhoff, C. Rehtanz, H. F. Wedde, "Multi-Agent System for Coordinated Control of Facts Devices," 15th International Conference on In- telligent System Applications to Power Systems, 2009. ISAP '09, 8-12 Nov. 2009.

[7] A. Benigni, J. Liu, F. Ponci, A. Monti, G. Pisano, S. Sulis, "Decoupling Power System State Estimation by Means of Stochastic Collocation", IEEE Transactions on Instrumentation and Measurement, vol. 60 n.5, May 2011.

[8] A. De Bellis, S. Grillo, S. Massucco, S. Scalari, P. Scalera, F. Silvestro, "A Smart Grid Approach to Distribution Management Systems (DMS) for Electric Networks", CIRED Workshop, 7 – 8 giugno 2010, Lione, Francia, pp. 1–4

[9] CIGRE Brochure, "Communications technology fundamentals for the design of modern protection and control systems", Working Group D2.16, April 2007.

[10] A. Abur and A.G. Exposito, "Power system state estimation – Theory and implementation", New York, Dekker, 2004

[11] K. Li, "State estimation for power distribution system and measurement impacts", IEEE Trans. On Power Sys., 1995, vol. 11, 911-916

[12] F. Bignucolo, R. Caldon and M. Valente, "Probabilistic voltage estimation for the active control of distribution networks," Proc. 19th International Conference on Electricity Distribution, Vienna, 2007

[13] A. Shafiu, N. Jenkins, G. Strbac, 2005, "Measurement location for state estimation of distribution networks with generation", IEE Proc. Gener. Transm. Distrib. vol. 152/2, 240-246.

[14] F. Bignucolo, R. Caldon, "Optimizing the voltage measurements location for the management of active distribution networks", UPEC Universities Power Engineering Conference, Brighton, 2007;

[15] F. Bignucolo and R. Caldon, "Simplified method to assess the uncertainty degree of voltage estimation in active distribution networks", Proc. Powertech, Lousanne, 2007

[16] A. Borghetti, M. Bosetti, S. Grillo, S. Massucco, C.A. Nucci, M. Paolone, F. Silvestro, "Short Term Scheduling and Control of Active Distribution Systems with High Penetration of Renewable Resources", IEEE Systems Journal, Special Issue on "Identification and Control of Sustainable Energy Systems", Vol 4, Issue 3, September 2010, pp. 313-322.

[17] IEEE Standard for Synchrophasors for Power Systems, IEEE Std. C37.118, 2005.

[18] A.G. Phadke and J.S. Thorp, Synchronized Phasor Measurements and Their Application, Springer, New York, USA, 2008.

[19] A. Borghetti, C.A. Nucci, M. Paolone, G. Ciappi, A. Solari, "Synchronized Phasors Monitoring During the Islanding Maneuver of an Active Distribution Network", accepted for the publication on IEEE Trans. in Smart Grids.

[20] M. Paolone, A. Borghetti, C. A. Nucci, "A Synchrophasor Estimation Algorithm for the Monitoring of Active Distribution Networks in Steady State and Transient Conditions", Proc 17th Power Systems Computation Conference (PSCC'11) Stockholm, Sweden, August 22-26, 2011.

[21] JWG Cigrè Cired C4.107 : "Economic Framework for Power Quality", TB 467, June 2011.

[22] P. Caramia G. Carpinelli, P. Varilone: "Point Estimate Schemes for Probabilistic Three-Phase Load flow", Electric Power Systems Research – Elsevier Vol. 80, no. 2, February 2010, pp. 168-175.

Towards a Cooperative Life Cycle Documentation for Distributed Renewable Energy Power Plants

Johannes Schmidt
Institute for Applied Informatics (InfAI) e.V.
at the University of Leipzig
Leipzig, Germany
schmidt@infai.org

Antonius van Hoof
Baden-Wuerttemberg Cooperative State University
(DHBW) Stuttgart, Germany
a.vanhoof@hb.dhbw-stuttgart.de

Abstract – *Requirements analysis of IT-support for daily operation of power plants for renewable energy (viz. solar, wind, and biochemical) shows a need for new, primarily cooperation-supportive functions, especially in the area of maintenance and repair. One central consideration is the need for a so-called digital logbook, a complete and consistent central pool of all data and documents generated in the course of a power plant's life cycle. In this article, we describe an excerpt of a reference data model for such a logbook, showing how model concepts fit multi-party cooperative error detection and repair processes. Thereafter, we present an IT-architecture for logbooks as federated data service. In our work, we discuss and incorporate present results in international ISO and IEC standardization relevant to renewable energy power plants.*

Keywords: reference model, renewable energy, federated information systems, digital logbook.

1 Introduction

General ecological concerns and the wish to abandon nuclear energy technology have presently led to an increase of interest in renewable energy production, such as solar, wind, and bio-chemical energy production. In the context of the German "Energiewende" (energy transition), politics and large utility companies focus on reorganizing and extending the available infrastructure of electrical energy distribution, e.g. Smart Grid or Smart Metering. Although these considerations are of paramount importance, many experts are of the opinion that there is also an urgent need to optimize actual operation processes of renewable energy plants. Increasing cost pressure and increasing competition have to be met by an increase in process efficiency and better IT-support. In contrast to conventional power plants (e.g. coal, gas, nuclear), the geographic distribution and technological diversity of plant units amplified by the plurality of vendor-specific technologies pose great management challenges. Therefore, practical operation and life cycle management of renewable energy power plants call for new forms of organization and technical support.

The effective and efficient management of a renewable energy plant is a complex task. The actual degree of complexity depends on the power plants construction and its type(s) of power generation. The operator has to cope with several technical, organizational, and legal tasks and issues to ensure a technical and economical optimal energy production. Tasks are, for example, monitoring of the facility, coordination and control of maintenance and repair activities at set time intervals, fulfilment of statutory and legal obligations, and maintaining a complete documentation record of current and past technical configuration, condition and operation history of the facility. For all these tasks, operators depend on several third parties and institutions, like authorized experts, independent service providers (ISP), and public authorities.

The life cycle of a renewable power plant is defined as "series of stages through which an item goes, from its conception to disposal" [1, p. 14]. The operation phase of a power plant is the central stage. We focus on maintenance and repair which is a subtask of operation. During the entire period of operation, a complete and consistent documentation must be available for the power plant. The documents and data created in the construction phase form the basis for operation. In the disposal phase, the last operational configuration of the facility must be known for an optimal dismantling and disposal. Thus, the different phases depend on each other.

The formal operator of a power plant is legally responsible for maintaining a consistent and comprehensive documentation record on all relevant production data and KPIs, certificates and maintenance reports. As of today, relevant data and documents are still often exchanged between all involved business parties by email or even by fax or post. Thus, the formal operator is scarcely in a position to keep a consistent digital documentation record of the power plant life cycle. In line with terminology in the aviation industry, such a documentation record will be called a *logbook* here. In the following, we will outline our considerations regarding the design and implementation of such a logbook.

According to [2], documentation is an important task within the power plant life cycle, beginning from design and production and ending at disposal. The management of a consistent logbook is a cooperative issue, because the formal operator of the power plant has to demand documents and data from third parties involved in design and operational activities, has to update related information, e.g. master data, and provide information for other parties himself. The approach we outline is relevant both for operating companies and independent service providers.

Our investigations showed that manufactures of renewable power plants who offer technical operation as a service mainly use highly integrated application systems, e.g. based on SAP. Because they usually can staff most activities of plant operation with their own employees, they are in a position to define company-wide formats and systems for documents and data. However, the increasing success of vendor independent operation and service providers induces a need for a more cooperative, open and standardized implementation of a logbook.

The remainder of this paper is organized as follows. Section 2 shows exemplary the need for plant life cycle documentation by describing an abstract process for error detection and repair of a renewable energy power plant and by discussing the necessary information systems involved. In Section 3, the most important concepts for a digital logbook are explained. Section 4 introduces our approach for a virtual integrated logbook whereas section 5 compares our approach with other related work. Section 6 gives a short summary of the main findings.

2 Life Cycle Documentation

2.1 Error Detection and Repair Process

Figure 1 depicts a simplified process for error detection and recovery. Continuous monitoring of the power plant by the operator is a prerequisite for fault detection. A failure alarm can either be based on the analysis of recorded operational data or can be raised by the monitoring software of the power plant itself. Depending on the specific maintenance agreements, the error analysis and classification can be carried out by the operator itself, or by a contracted ISP or by the vendor of the power plant. In the latter two cases, the result of the analysis has to be reported to the operator who has to archive the resulting analysis document and link it with the original operational data or alarms that caused the analysis. After detection and classification, the operator can order a further maintenance company to remedy the defect. Such a scenario necessitates the coordination of data and documents for repair preparations between employees of the service provider(s) and the operator. Furthermore, the operator has to supervise the repair process. Afterwards, the maintenance service provider must provide a maintenance or modification log and will present billing documents. All this information must in principle be accessible to any further parties that will be involved in the future life cycle of the power plant.

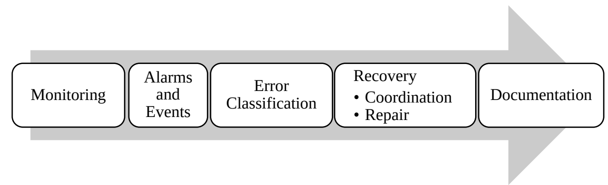


Figure 1. Simplified error detection and repair process

Even this overtly simplified process already gives an idea of the complex interactions between the involved actors. The power plant operator is responsible for the complete documentation of the power plant life cycle that is the basis for work performed by other parties. To increase the efficiency of this documentation process, IT-support is needed. An integrated and simplified access to all life cycle documentation by all parties involved can improve the quality of the maintenance and repair tasks and reduce maintenance costs. Furthermore, such IT-support by means of a holistic digital logbook has to take into account different types of renewable power plants, especially in the context of virtual multi-modal power plants, i.e. plants consisting of geographically disparate and technologically different units of power generation (e.g. combined wind and photovoltaic power generation systems within local public electricity supply utilities).

2.2 Information Systems

Today, there is a lack of standardized methods for the exchange of operational data and documents for technical operation. Although, several information systems (IS) can be purchased which support management of documents and data, they are not integrated and insufficiently support data exchange. As such, they do not provide immediate support for multi-party cooperative business processes in the context of maintenance and repair. However, such support is deemed necessary by experts to guarantee smooth operation of renewable energy power plants. In order to accomplish such an integration of information systems, an overview and classification of relevant IS and their corresponding business concepts including assignments to their owning actors are needed. Although first approaches have been published, no complete overview is available in the context of renewable energies yet.

Appelrath and González define three categories of information systems for utility companies: IS for integrated clearing and settlement, IS for mercantile administration and IS for technical operation [3]. For the purpose of our paper, mainly the latter category is relevant. Their approach is a good starting point to develop a specific classification of IS in the context of renewable energy, although some IS mentioned do not fit within our chosen context and some IS we address are missing. For that reason, we only describe selected IS in the following.

Information systems for technical operation form the basis of seamless life cycle documentation. These include Supervisory Control and Data Acquisition (SCADA) systems, Condition Monitoring Systems (CMS), Geographic Information System (GIS), and Control Stand Systems. SCADA systems and CMS are responsible for the acquisition and monitoring of operational data from the power plants. Several management systems support the operator to keep consistent master data, e.g. of the power plant park or the components of an asset. Furthermore, Maintenance Management Systems (MMS) or Workforce Management Systems (WMS) support the operator in supervising and dispatching scheduled maintenance activities. The latter two IS are often implemented as modules of an Enterprise Resource Planning (ERP) system.

In contrast to conventional power plants, weather forecast systems are highly relevant, especially for facilities that depend directly on weather conditions (like sun and wind). They enable the operator to forecast and react promptly on changing weather conditions to ensure optimal power generation performance. This kind of IS shall be classified to both mercantile administration and technical operation category. Based on weather forecasts, an operator can calculate the loss of revenue in the case of an infrastructure cut-off. On the technical side, the actual schedule of a maintenance activity depends on precise weather information. For example the maintenance schedule of a wind power plant (WPP) is oriented towards periods of minimal wind speed whereas maintenance activities for solar power plants (SPP) are oriented towards those days and hours in which high solar radiation is forecasted.

The present lack of integration of IS causes several problems of which two particular ones have to be addressed when designing a consistent digital logbook. Firstly, linking documents and data between different business units is not possible. Secondly, there are duplicates of data and documents, because each party tend to store their own data and documents, as well as data and documents provided by other parties. E.g. service provider both store their maintenance documents in internal applications and send them to the operator of the power plant. This kind of redundancy also makes it difficult to keep track of consistency and chronology of the stored information.

3 Towards a Reference Model

In this section, we describe an abstract reference model that forms the basis of a digital life cycle logbook for renewable energy plants. Figure 2 shows the most relevant domain concepts (bold font) in the context of maintenance and repair. The properties of the concepts are not shown. Furthermore, a concept can be an abstraction of a more complex partial model. Exemplary instances of the domain concepts are depicted in grey colour and font, and their relationships are labelled with “is a”.

3.1 Model Concepts

The digital logbook is a combination of different data and documents linking relevant information in the context of maintenance and repair. Every concept is related to a *power plant*. Different types of power plants are supported, like *WPP* or *SPP*. Their properties contain common information about the facility. Some of them are based on statutory specifications, but today, there is no mandatory standard. Each power plant consists of several *components* that form a tree. A component can contain *sensors* that provide different data, like status information, *alarms* or current *operational data*. The assignment of data to component forms the basis of the logbook.

As depicted in Figure 2, a *maintenance activity* is caused by different *triggers*. Firstly, an alarm can be raised by the power plant or analysis of operational data may indicate a distinctive feature (compare to Figure 1). Operational data include data provided by Condition Monitoring

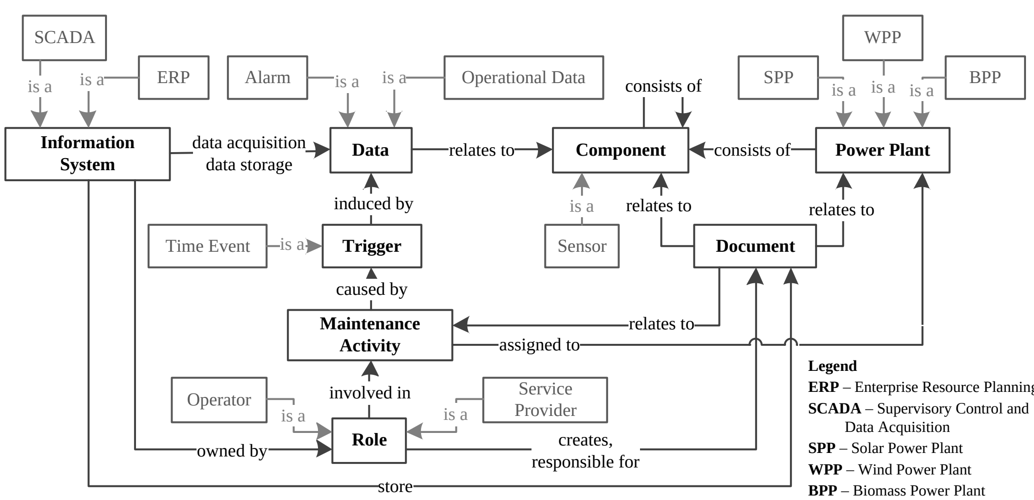


Figure 2. Reference model fragment of life cycle documentation for renewable energy plants

Systems (CMS), too. Secondly, there are periodic examinations, e.g. inspections. During and after completion of the activity, several *documents* have to be created, e.g. maintenance reports or certificates. The operator needs to relate these documents to the trigger and the log of error handling to achieve a consistent logbook. As far as inspection or exchange of parts is involved, the documents need to be related to the updated master data of the components.

Today, the data and information described in our reference model are stored and managed in different types of *information systems*, like *ERP* or *SCADA*. Every class of IS typically is related to a *role*. Thus, e.g. an *operator* uses a SCADA system to monitor the power plant whereas a maintenance service provider uses an ERP or Service Management Systems to schedule maintenance activities. A consistent model or portfolio for every role, including relationships between IS, actors, and the related business processes can ease system integration with the logbook. Thus, the logbook can be seen as a basis for a standardized exchange of documents and data for business-to-business integration.

3.2 Relevant Standards

The implementation of a digital logbook must base on established standards and technical guidelines to achieve a seamless integration of data and documents. Table 1 gives an overview of selected international standards related to the concepts from Figure 2. The concrete usage of standards depends on the type of the power plant.

Table 1. Related standards and technical guidelines

Standard	**Related Concepts**
IEC 61850, IEC 61400-25	Data, Component, Power Plant
ISO/TS 16952-10	Component (including Sensor), Power Plant, Document
IEC 61355, DIN EN 13460	Document
IEC 61968 IEC 61970	Information System, Data, Component, Power Plant

In the context of remote monitoring and data acquisition, IEC 61850 [4] and IEC 61400-25 [5] standards are used, whereas IEC 61400-25 is an extension of IEC 61850 and has been adapted for wind power generation. Both define special information models that refer to the structure of the power plant including abstract classification of components. Data properties and their semantics are determined and can be easily referenced in data analysis or maintenance reports.

The standard Reference Designation System for Power Plants (RDS-PP, ISO/TS 16952-10) [6] relates to different types of power plants (e.g. wind power [7]). Using RDS-PP, it is possible to uniquely designate every part or component of a power plant, elicit its specific function and functional dependencies, and determine its installation position and location on the plant site. The standard also offers means to link data, signals and documents with these parts. There are several document classifications available. The standard IEC 61355 defines a classification of technical documents, whereas the standard DIN EN 13460 especially focuses on maintenance documents. The so-called Zustands-Ereignis-Ursachen-Schlüssel (ZEUS) by FGW relates to RDS-PP and allows a standardized description of the current state of the power plant itself or one or more components [8]. We expect more standards and guidelines based on or related to RDS-PP to occur in the future.

The Common Information Model (CIM) (see IEC 61970 and IEC 61968) defines a reference model for data exchange between utility companies [9]. CIM defines concepts for SCADA, ERP, time series data, or the electrical infrastructure and forms the basis for application integration in the domain.

4 Logbook as Federated Data Service

We already hinted at the complexity of the task of keeping an up-to-date, complete and consistent digital logbook, comprising of all relevant data and documents throughout the complete life cycle of a power plant. This is due to multi-party involvement, with each party operating their own IS, adhering to their own data formats and providing their own rules concerning data access and modification. Our analyses show that almost every involved role, as a legal party, has concerns about data security and data privacy protection because of their acknowledged profound commercial value. Therefore, parties want to be in control of who has for what reason access to which particular data.

4.1 Architecture

The process of constructing a centralized special purpose data pool (in our case for storing and accessing information relevant for operation and maintenance) out of several, possibly technologically different, information sources is well understood. The software market provides ample tools to construct such so-called data warehouses, using semi-automatic ETL-Tools (Extraction, Transformation and Loading). However, the situation sketched above does not allow for the adoption of such pure data warehouse architectures.

Thus, we will design and implement the digital logbook as a virtual data pool – in the following named Federated Data Service (FDS). Users accessing the logbook will interact with it as though it contains all data (transparent data access), but all data are left unchanged at their original locations. All queries to the logbook will have to be transformed at query-runtime into a series of queries to relevant data sources. Returned results will have to be joined and condensed to one distinct result to be presented to the querying client. Systems adhering to such strategies are called Federated Information Systems [10].

A schematic view of the architecture of our system is shown in Figure 3. It consists of modules for query processing, registering data pools (handling access and data properties) and clients (handling authorization properties), mapping reference model concepts and relations to local data objects, planning and execution of local queries (i.e. transformation) and linking query results to produce a global transparent result to the clients. Underlying the FDS are all kinds of registered data pools, containing master data, operational data and documents.

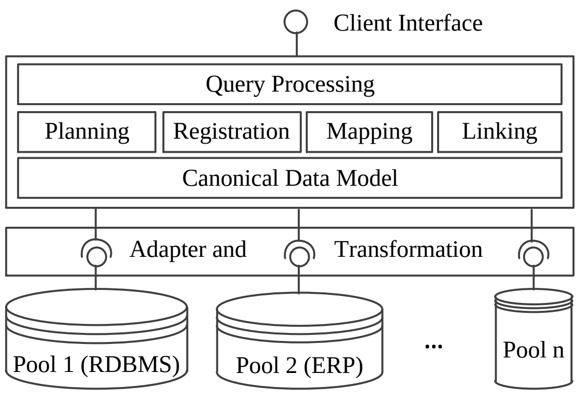


Figure 3. Schematic structure of FDS

The FDS provides through its mechanisms a functionality to deal with logbooks of several power plants of any kind. At the moment, we do not plan to implement a data/document caching on a per plant basis, which would provide a smooth transition from a virtual data structure to a real data warehouse for specific power plants. This might be a desirable feature for parties owning and or operating such plants as far as data protection and privacy issues allow for it.

Our digital logbook will be implemented as a web service. The same holds for the adapters which are responsible for the communication between the logbook and all data pools from which it will draw its data for query results. At the moment, we are evaluating several web service communication options: REST interfaces with JSON data object, Google Protocol Buffers and Web Intents [11]. A first experimental implementation of the FDS uses REST and JSON.

4.2 Semantics

A central concern in the construction of data federation systems is the issue of schema mapping. The FDS will publish to its clients a global data schema, also called canonical data model. Since there is no actual instantiation of this schema by means of a central data base, all elements (i.e. concepts, properties and relations) of the canonical model will have to be mapped to elements of the local schemata of the accessed data pools. Using this mapping, the FDS can translate a query relating to the canonical model into a series of local queries to data pools.

The reference data model in Figure 2, presented in the previous section of the paper, is the basis for the canonical data model of the logbook FDS. The largest part of that model is covered by concepts and relations standardized by RDS-PP (see previous section). The ISO-Standard offers a complete functional ontology and concept definition for all components of a plant and defines ways to link data, documents and signals/triggers to these components and thus to each other. The power plant itself as conjoint of subsystems and every component or part of the plant, documents, data and signals can be encoded as RDS-PP designation strings. Every part of such a string carries specific information on any of the above concepts.

It is possible to understand an RDS-PP string as a query to the FDS. For example the following string "#WPP.01 =G001 MDL10 BT012 -BT001 &MBB100/D00141" may be interpreted to mean "fetch document MBB100/D00141, being a document in the area of mechanical engineering, offering the 100th installation report, this particular one being concerned with a piece of equipment that is a temperature sensor having the function to measure the temperature of the yaw drive of the first wind turbine of power plant WPP.01".

Normal queries would leave certain parts of such strings unspecified and/or would add "wild cards". Thus, e.g. "#WPP.01 =MDL* &*DC *" would specify a query for any instruction manuals relating to yaw systems present on power plant WPP.01. In this way, RDS-PP can function as query language to link master data, operational data, incidents, and documents.

RDS-PP does not provide the concepts *Information System*, *Maintenance Activity* and *Role*, present in our reference model. These can be dealt with in following manner: *Information System* characterizes the particular data pools that are registered to provide information within the FDS. *Role* characterizes the particular data pool owners related to the maintenance activities they perform. FDS already has modules that take care of these. *Maintenance Activity* is a concept that will link roles, documents and data. Depending on the IS/data pools in use, maintenance activity records will be created and stored either by several local pools or be kept in a pool private to the FDS.

We plan the implementation of FDS in a way that parties can register any new information (data, documents, etc.) they offer with the FDS also through use of RDS-PP strings, together with information on access authorization for these data. Through this mechanism, power plant owner, operator and third parties in maintenance processes can collectively share all information necessary for dealing with specific maintenance activities.

5 Related Work

The "Fördergesellschaft Windenergie und andere Erneuerbare Energien" (FGW) presently works on a draft for a Global Service Protocol (GSP) [12]. They aim to support exchange of maintenance documents and data by defining a common data format based on XML and SOAP. In contrast to GSP, our approach includes querying capabilities and transparent data access for different roles involved. Furthermore, the logbook shall primarily be designed according to the processes and needs of the operator.

In the last years, some approaches concerning ontology based application integration in the energy domain have been published. They are based on CIM UML and derive messages for data exchange (e.g. [13]). It is possible to define custom CIM profiles that can contain constraints and selected elements of CIM [9]. This approach may be applicable in the context of a digital logbook, too.

6 Summary and Conclusion

Operation of renewable power plants is a complex task. Especially in the context of maintenance and repair, several different roles are involved, whose relationships depend on data exchange. To maintain a consistent digital logbook, a cooperative approach is needed. Based on a simplified domain reference model, we described a federated data service that is able to link documents and data from diverse data sources and query them based on the RDS-PP standard. This approach can be applied to other types of power plants, too. A concrete realization of the reference model must be based on at least the international standards listed in section 3, whereas standards for data acquisition form the basis of a consistent digital logbook. Today, there are interoperability problems between them. The IEC and other committees presently try to resolve them.

7 ACKNOWLEDGMENTS

This work is a result of the EUMONIS project, supported by the German National Academic Foundation (DFG) and the Federal Ministry of Education and Research (BMBF) as grant 01IS10033D.

References

[1] *Maintenance—Maintenance terminology; Trilingual version*, DIN EN 13306, 2010.

[2] *Technical documentation: Definitions and legal basics; Issue German/English*, VDI 4500 Part 1, 2006.

[3] H.-J. Appelrath and J. M. González, "Informationstechnik in der Energiewirtschaft," in 12/2010, *Handbuch Energiemanagement*, H.-P. Beck, A. Bauer, E. Meller, and C. Salander, Eds. 30th ed, Frankfurt am Main: EW Medien und Kongresse GmbH, 2010, pp. 10510–10534.

[4] *Communication networks and systems in substations—ALL PARTS*, IEC 61850-SER ed1.0, 2012.

[5] *Wind turbine generator systems—ALL PARTS*, IEC 61400-SER ed1.0, 2011.

[6] *Technical product documentation—Reference designation system—Part 10: Power plants*, ISO/TS 16952-10:2008, 2008.

[7] *RDS-PP Application specification Part 32: Wind energy*, VGB-S-823-T32;2012-04-EN, draft, 2012.

[8] *Technische Richtlinie für Erzeugungseinheiten—Teil 7: Instandhaltung von Kraftwerken für Erneuerbare Energien—Rubrik D2: Zustands-Ereignis-Ursachen-Schlüssel für Erzeugungseinheiten (ZEUS)*, FGW TR7 D2, 2012.

[9] M. Uslar, M. Specht, S. Rohjans, J. Trefke, and J. M. Vasquez Gonzalez, *The Common Information Model CIM: IEC 61968/61970 and 62325—A practical introduction to the CIM*, 2012nd ed. Berlin, Heidelberg: Springer 2012.

[10] U. Leser and F. Naumann, *Informationsintegration: Architekturen und Methoden zur Integration verteilter und heterogener Datenquellen*, 1st ed. Heidelberg: Dpunkt-Verlag, 2007.

[11] G. Billock, J. Hawkins, and P. Kinlan, "Web Intents: W3C Editor's Draft 17 May 2012," World Wide Web Consortium (W3C), 2012. [Online]. Available: W3C Website, http://dvcs.w3.org/hg/web-intents/raw-file/tip/spec/Overview.html [Accessed: 29 May 2012].

[12] P. Lyding, S. Faulstich, and K. Rafik, "Optimising flow of maintenance information to boost turbine efficiency," EWEA 2011, 14.-17.03.2011, Brussels, Belgium, 2011.

[13] M. Uslar, S. Rohjans, S. Schulte, and R. Steinmetz, "Building the Semantic Utility with Standards and Semantic Web Services," in *LNCS, vol. 5333, OTM 2008 Workshops*, R. Meersman, Z. Tari, and P. Herrero, Eds, Berlin, Heidelberg: Springer, 2008, pp. 1026–1035.

Analysis of Moroccan Wind and Solar Potential Using Artificial Neural Network Approach

Ahmed Ouammi
CNRST, TEER
Rabat, Morocco
ahmed.ouammi@cnrst.ma

Hanane Dagdougui
DIST Faculty of Engineering,
University of Genoa, Italy
hanane.dagdougui@unige.it

Driss Zejli
CNRST, TEER
Rabat, Morocco
zejli@cnrst.ma

Roberto Sacile
DIST Faculty of Engineering,
University of Genoa, Italy
Roberto.sacile@unige.it

Abstract - *An artificial neural network (ANN) model is used to forecast the annual wind speeds and solar irradiation in Morocco. Solar irradiation data are taken from the new Satellite Application Facility on Climate Monitoring (CM-SAF) - PVGIS database. The annual wind speed data are taken from (CDER, 2007). In this paper, the data are inferred using an ANN algorithm to establish a forward/reverse correspondence between the longitude, latitude, elevation, solar irradiation and wind speed. Specifically, for the ANN model, a three-layered, back-propagation standard ANN classifier is considered consisting of three layers: input, hidden and output layer. The learning set consists of the normalised longitude, latitude, elevation and the normalised mean annual wind speed of 20 sites and the normalised mean annual solar irradiation of 41 Moroccan sites. The testing set consists of patterns just represented by the input component, while the output component is left unknown and its value results from the ANN algorithm for that specific input. The results are given in the form of annual wind speed and solar irradiation maps. They indicate that the method could be used by researchers or engineers to provide helpful information for decision makers in terms of site selection, design and planning of new solar and/or wind power plants.*

Keywords: Solar Atlas; Wind Atlas; Artificial neural network; Geographic Information System; Morocco

1 Introduction

Renewable energy is an environmental friendly option which may be economically competitive with conventional power generation, where good resources are available. It represents an opportunity to enhance sustainable development in both countries which traditionally lacks fossil fuels and those who are constrained by some environmental policies and regulations. Renewable energy can play increasing role in countries that have an important environmental wealth and great renewable energy potential. Quality and accessibility of resource data will enable private investors and public policy-makers to access the technical, economical and environmental potential for large-scale investments in green technologies. For more accurate resource assessment, detailed site-specific micro-sitting analysis should be done considering existing transmission grid, accessibility, land availability, altitudes and topography [1, 2].

Renewable power generation has known a remarkably rapid growth in the past twenty years, and now it is a mature, reliable and efficient technology for electricity production [3]. In addition, in regions with proper resource characteristics, renewable energy may already be competitive with coal or nuclear power, especially when the cost of pollution is taken into account in the overall economic evaluation [4]. Any choice of renewable energy sources exploitation site must be based on the preliminary investigation of the average wind velocity, solar irradiation and potential, so that the accuracy of the resources data analysis is a crucial factor to be undertaken.

In the literature, many studies have been focused on providing a forecasting tool to predict and assess renewable energy sources and power production with good accuracy. From the wind speed predictions viewpoint, several authors [5-10] have used many approaches to predict wind speed and energy production.

2 Assessment of the available wind/solar energy potential

The meteorological measurement stations are usually insufficient to determine the potential of wind and solar potentials available in the whole territory. However, predictions methods are needed to estimate the wind and solar potentials in other locations where no measurement stations are available. Numbers of prediction method are available to construct an interpolated surface between available point's data measurement. Kriging is an interpolation technique that estimates unknown values from known sample values and semi-variograms. In recent years, the method has been widely applied in different research fields, among which the application in GIS in order to interpolate data.

The Kriging techniques have been adopted for the renewable energy predictions, among these Kriging techniques, good results have recently been obtained by the use of artificial neural network (ANN) techniques. In this paper, data have been inferred using an ANN algorithm to establish a forward/reverse correspondence between the longitude, latitude, elevation and the mean annual renewable energy.

Specifically, for the ANN model, a three-layered, back-propagation standard ANN classifier has been used consisting of three layers: input, hidden and output layer. The ANN input layer consists of 3 units which are associated to the longitude, the latitude and the elevation (linearly normalised between 0 and 1, taking into account, respectively, the maximum and minimum longitude, latitude and elevation of the territory) of a specific location.

In a back-propagation standard ANN learning phase, the characterising "weights" are defined on a given set of patterns. Specifically, the output yi of each unit i in the network is determined by:

$$y_i = \frac{1}{1+e^{-x_i}} \qquad (1)$$

This output has a real value between 0 and 1; xi is the total input to unit i given by:

$$x_i = \sum_j w_{i,j} y_j + b_i \qquad (2)$$

where wi,j is a real number, called weight, representing the strength of the connection from unit j to unit i.

The weighted sum of the inputs is adjusted by the bias characteristic of the unit i, bi. Network weights are initially assigned random values uniformly distributed in [-0.3, 0.3], in each back-propagation cycle, the weights are adjusted in the total output error. The learning ends either after a user-defined number of steps or when the total output error becomes asymptotic, where this error is defined as:

$$E = \sum_p \sum_j (O_{p,j} - D_{p,j})^2 \qquad (3)$$

where Op,j is the observed output on unit j for learning pattern p and Di,j is the desired output.

The ANN learning procedure is performed on learning set of patterns, where, in our model, each learning pattern p is represented by 3 parameters (input layer) and by 1 output parameter (output layer).

Table 1. Geographical coordinates and solar irradiation of the 41th sites

Sites	Longitude	Latitude	Altitude	Solar irradiation [kWh/m^2/day]
Agadir	9.6019	30.3963	0	5.44
Ain bnimathar	2.0377	34.0116	914	5.37
Safi	9.2447	32.3102	14	5.56
Azilal	6.5752	31.9705	1331	5.29
Bouarfa	1.9775	32.5283	1128	5.54
Boujdour	14.4238	26.1355	0	5.88
Dakhla	15.9463	23.6997	9	6.09
El jadida	8.5141	33.2338	24	5.6
Errachidia	4.4383	31.9519	1050	5.28
Esmara	11.6783	26.7455	178	5.84
Essaouira	9.7722	31.5175	0	5.68
Fes	5.0097	34.0344	410	5.21
Glyab	13.0844	21.2891	0	5.95
Imintanout	8.8602	31.1797	907	5.57
Kenitra	6.5861	34.2661	4	5.5
Khmisset	6.0644	33.7972	460	5.24
Khenifra	5.6741	32.9438	854	5.16
Khouribga	6.9047	32.8886	805	5.27
Laayoune	13.2108	27.1566	0	5.43
Lagouira	17.0725	21.0638	0	6.27
Larache	6.163	35.1827	40	5.58
Marrakech	8.0033	31.6344	455	5.25
Msaysat	15.8202	23.1605	135	5.35
Nador	2.9441	35.1647	29	5.54
Ouarzazat	6.8991	30.9208	1133	5.48
Oujda	1.9169	34.6963	547	5.25
Ousard	14.3261	22.553	306	6.19
Ouazane	5.5919	34.8	289	5.22
Chtoukane	14.7763	24.6369	129	5.33
Tanger	5.8005	35.7686	12	5.39
Tantan	11.1125	28.4347	48	5.07
Taounat	4.6525	34.538	582	5
Taourirt	2.9058	34.4202	389	5.21
Tarfaya	12.9308	27.9361	3	5.57
Taroudant	8.8877	30.4675	232	5.26
Tata	7.9813	29.7402	741	5.65
Taza	4.01	34.2252	487	5.14
Tetouan	5.3722	35.5766	172	5.19
Tinghir	5.548	31.5325	1314	5.4
Tiznit	9.7413	29.7069	231	5.37
Zagora	5.8555	30.3347	730	5.52

3 Results and discussion

In this paper, solar irradiation data of 41 Moroccan sites have been taken from the new Satellite Application Facility on Climate Monitoring (CM-SAF) - PVGIS database [11]. These data were in form of yearly solar irradiation on horizontal plane. Table 1 shows the geographical coordinates of the selected sites. From the analysis of the table, it can be seen that the mean annual solar irradiation is ranged between 5 and 6.3 kWh/m2/day. It can be concluded that the potential from solar of Morocco shows a quiet similar behaviour between all sites.

As regards wind data, 20 stations distributed over the whole Moroccan territory have been considered, these data have been collected from [12]. Table 2 shows the wind regime characteristics of the 20th sites. From the analysis of the table, it can be seen that the mean annual available wind power is ranged between about 50 and 1100 kW/m2. It can be concluded that the wind potential of Morocco shows an important variation between all sites in the whole territory.

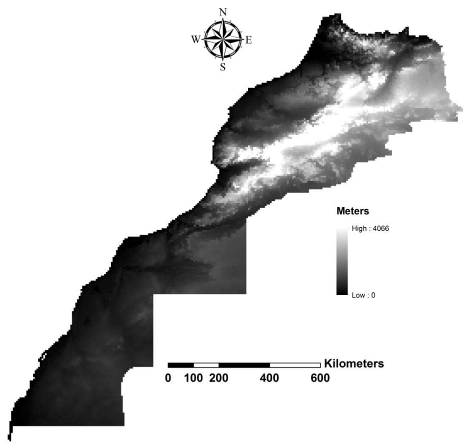


Figure 1. Digital elevation model of Morocco

Table 2. Wind regime characteristics of the 20th sites.

Sites	Wind speed [m/s]	C [m/s]	K [-]	Available power [kW/m2]
K. Elbaida1	10.94	12.1	2.75	1107.48
El fnidek	5.81	5.74	2	150.84
Houma	9.42	10.61	2.29	841.18
Fardiwa	9.1	10.22	2.51	703.85
Sendouk1	8.5	9.68	2.19	663.00
Tanger	6.8	7.64	2.28	315.17
Kenitra	3.88	4.35	1.66	83.30
Zagora	4.04	4.49	1.47	112.83
Tiznit	4.5	5.15	2.02	107.82
Tantan	5.12	5.75	2.15	141.26
Bouznika	3.83	4.35	2.25	58.80
Rabat	3.85	4.25	1.96	62.58
Dakhla	7.95	8.97	2.89	440.27
Nador	3.32	3.74	1.72	50.24
Tanger sadan	7.96	8.99	1.62	763.85
Essaouira cap	8.37	8.4	2.01	470.28
Essaouira tagant	5.55	6.41	1.65	269.03
Tarfaya	7.1	7.93	2.97	300.50
Taroudant	5.71	6.28	1.43	325.92
Safi had hrara	6.27	7.07	1.86	306.10

Starting from the DEM of Morocco reported in Figure.1, mean annual wind speed and solar irradiation maps of Morocco have been obtained. Figure. 2 displays the annual wind speed map obtained by ANN techniques of the whole Moroccan territory. This map shows, as it was expected and as it is also widely known from Moroccan decision makers and stakeholders and according to the wind resources that the most interesting sites in Morocco are in the northern, south-western and some internal territories from wind speed viewpoint. The Mediterranean coast and the southern Atlantic coast seem to be important from wind energy production viewpoint. But accurate monitoring campaign directly on the site is always highly recommended for a reliable assessment.

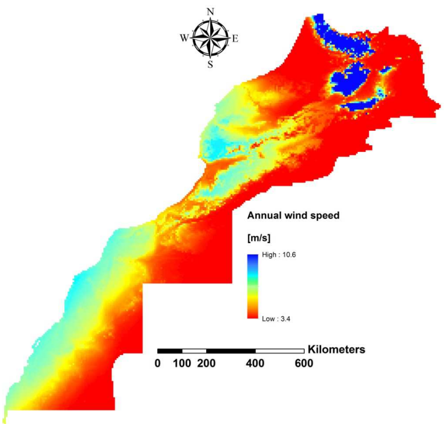


Figure 2. The annual wind speed map as obtained by ANN techniques

The solar irradiation map shows a quite different behaviour than the one observed for the wind. Figure. 3 displays the annual solar potential map from solar as obtained by ANN techniques of the whole Moroccan territory. It can be seen that the most interesting sites are in the southern and the eastern. Mediterranean coast, northern, and some internal territories are not of interest.

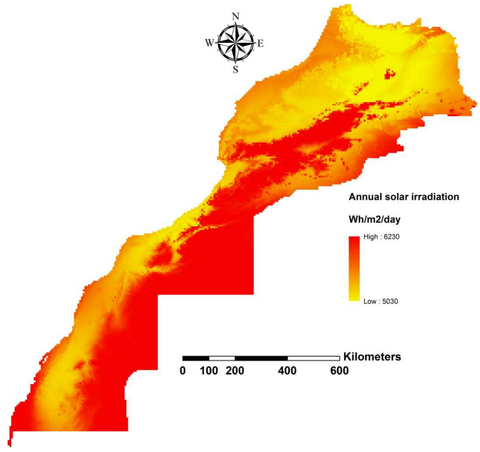


Figure 3. The annual solar irradiation map as obtained by ANN techniques

4. Conclusion

Renewable energy systems can be considered as an attractive option to conventional power generation as well to enhance sustainable development especially in developing countries like Morocco which traditionally lacks fossil fuels, but has an important environmental wealth, land availability and great solar potential. The obtained results indicate that the method could be used by researchers or engineers to provide helpful information for decision makers in terms of site selection, design and planning of new wind and/or solar power plants. Future research will be directed to the selection of the promising sites and the optimal technology to be installed under a multi-criteria approach. In particular, the focus will be dedicated to the design and planning of new renewable infrastructures.

References

[1] E.K. Akpinar, S. Akpinar, "An assessment on seasonal analysis of wind energy characteristics and wind characteristics". Energy Conversion and Management, Vol.46, pp.1848-1867, 2005.

[2] M.C. Mabel, E. Fernandez, "Growth and future trends of wind energy in India", Renewable and Sustainable Energy Reviews, Vol.12, pp.1745–1757, 2008.

[3] I. Fyrippis, J.P Axaopoulos, G. Panayiotou, "Wind energy potential assessment in Naxos Island Greece", Applied Energy, Vol.87, pp.577–586, 2010.

[4] K. Kaygusuz, "Renewable and sustainable energy use in Turkey: a review", Renewable and Sustainable Energy Reviews, Vol.6, pp.339–366, 2002.

[5] S. Rehman, "Wind energy resource assessment for Yanbo, Saudi Arabia", Energy Conversion and Management, Vol. 45, pp.2019–2032, 2004.

[6] A.S.A. Shata, R. Hanitsch, " Electricity generation and wind potential assessment at Hurghada", Renewable Energy, Vol.33,pp.141–148, 2008.

[7] Y. Himri,, A.B. Stambouli, B. Draoui, "Prospects of wind farm development in Algeria", Desalination, Vol. 23, pp.130–138, 2009.

[8] A.N. Celik, "A simplified model for estimating yearly wind fraction in hybrid-wind energy systems", Renewable Energy, Vol.31, pp.105–118, 2006.

[9] A. Ucar, F. Balo, "Evaluation of wind energy potential and electricity generation at six locations in Turkey", Applied Energy, Vol.86, pp.1864–1872, 2009.

[10] A. Conte, A. Pavone, C.F Ratto, "Numerical evaluation of the wind energy resource of Liguria". Journal of Wind Engineering and Industrial Aerodynamics, Vol.74,pp.355–64, 1998.

[11] http://re.jrc.ec.europa.eu/pvgis/

[12] CDER, Les ressources éoliennes du Maroc, 2007.

Supervision analysis and control system of photovoltaic power plants

Lambruschini P.
DYNATECH University of Genoa
Via Opera Pia, 11 A
16145 Genoa, Italy
lambruschini@dibe.unige.it

Raggio M.
Bajpai R.
Sharma A.
DYNATECH University of Genoa
Via Opera Pia, 11 A
16145 Genoa, Italy
{raggio, rajiv, abi}@dibe.unige.it

Abstract – *This paper describes a supervision system able to handling the data collection from photovoltaic implants, their analysis allowing providing prevision and control of energy production. The supervision system integrates different functionality usually demanded to single subsystems and advanced features devoted to the implants maintenance and security. The supervision system is able to interact with the PV plant varying its functionality.*

Keywords: Supervision, monitoring, smart-grid, photovoltaic, renewable energy, energy production

1 Introduction

In recent years the oil price increase, the need to reduce the environment pollution due to fossil combustible exploitation and the need to have low-cost energy sources due to the global economic crisis, have raised the interest in renewable energy sources [1].

Energy storage elements are nowadays expensive, breakable and allow low performances [2], hence the most efficient system to harvest energy from alternative sources is the injection into the grid [3]. This produces a change in the grid model [3] [4] until now known. The energy is no longer produced by a few large plants and consumed by many users, but is provided from a distributed network of sources. In this new scenario the energy producers do not have a direct control of the mains through their big power plants. This is especially true during specific hours of the day when the cost of energy produced from renewable sources is highly competitive and large thermal plants are exploited to a minimum. In this new model, become very important have the instruments for the control of the production implants in order to can manage production, according to consumer requirements and costs for each source. One solution that is being developed is to provide the electricity networks of intelligence going to make the smart grids. Under the name smart-grid usually means a series of technologies designed to build a new control system of the electricity network that makes use of distributed sensors, software for data collection [6] and analysis [7], and actuators [8]. All this with the idea of making a flexible network capable of transferring energy in real-time from the areas in which it is most convenient to produce it to those in which request comes. Our monitoring system described in this paper can be categorized in this light. It can be considered a supervision control system of one PV implant and a subcomponent of a wider and at higher level grid control system.

Another aspect a supervision system cannot ignore is the maximization of yield [1]. This can be achieved with a constant monitoring of the operating parameters of the devices that make up the system and through analysis of the data in a way to be able to compare the actual data with those expected. Assuming a typical lifetime for PV systems of about thirty years [2] [5], it is important to try to take prompt action on faults and constantly check the obsolescence of the panels in order to evaluate their replacement with newer, and more efficient.

In addition to the aspects of pure control of the implant parameters is fundamental and often overlooked that a supervisory system integrates a video surveillance system. The management of the surveillance of the site where the plant is located can be a key factor. Costs related to the management of safety, insurance contracts, and theft of equipment can bring down the productivity.

For the above exposed reasons, our supervision system integrates in one single tool the functionality of data logger, alarm states handling, communication on through several channels in order to guarantee the delivery of the message, data analysis, control of the energy production as a function of network request, control of implant functionality features following request of user or upper level control, and video surveillance management.

The paper is organized as follows. The system description section provides details about the system and its integration with other elements building a photovoltaic implant. In this section is shown even the integration with other elements needed to the supervision system in order to provide all its functionality. The system implementation section shows how the system has been designed and implemented. The tests and results section summarizes the condition under which the supervision system has been tested and the achieved results. In the last section some conclusions about the expected and achieved results are provided.

2 System description

The supervision system is integrated with the devices of the photovoltaic implant and with other elements needed for the implementation of all functionalities provided, as shown in Figure 1. In the centre of Figure 1, there is an embedded PC that is the hardware device where the supervision system core is implemented. Details about the implementation will be provided in next session of this paper, here we will be shown the integration with all devices around the core system.

Typically in the PV implants the base elements are inverters, string-boxes, and panel strings. The supervision system monitors all inverter and string-box supporting modbus communication protocol in TCP/IP or RTU versions, independently from device producers. The most important functionality of a supervision system devoted to PV implants is the control of the parameters provided by inverters and smart string-box. Data have to be analyzed to produce alarms status and control responses. Obviously each monitoring system must guarantee the communication with inverters, and string-box, while the panels usually do not have specific electronic for the communication [10] [11]. Sometimes panels can be equipped with surface thermal sensor and in this case the information is take in care by string box or additional electronic boars. The panels surface temperature is a key factor in the implant profitability, because the panels efficiency decreases when the temperature rises. A supervision system should know this important parameter mainly for two reasons. The first is the evaluation of panel malfunction when the panel temperature is anomalous respect to air temperature and solar irradiation level. The second is the calculation of the expected production respect the effective one. The production is even function of irradiation level and of whether condition. Our supervision system considers both parameters communicating with an irradiation sensor installed on the PV implant and with online climate data services [9]. The production is mainly influenced by irradiation, panel temperature and whether condition with particular attention to the presence of clouds. For this reason our supervision system can be integrated with all components needed to keep this information and can provide warning in case of significant difference between expected and achieved production. This information can be used to schedule supplementary maintenance.

To achieve a correct value of the energy produced our supervision system can be integrated with external counters. The integration can be done directly with device of mains operators and when this is not possible with devoted hardware. The system provides to compare data from the counters with the calculation producing itself, in order to avoid errors about this important parameter.

An important key factor of the monitoring systems is the integration with communication apparatus and the ability to produce information useful both for users and for systems automated controls. The communication should consider that the system could be installed in harsh environments or

Figure 1. System integration

in areas with difficult accessibility. Due this reasons, our supervision system can be integrated with GPRS or UMTS modem and LAN networks. This solution ensures that the system is always able to communicate alarm states, through mail or SMS or web and can always be reached by users, maintainers or upper level grid control systems. The capability of integration and communication with other systems became very important in case of inclusion inside a smart-grid. This scenario is shown in Figure 2, were it is shown a smart grid model with several control systems at different levels. Our supervision system can be considered the interface between the photovoltaic power plants and the local grid control systems being able to vary the operation of the plant.

Figure 2. Smart grid

In order to achieve a better management of the safety of the photovoltaic implant both from the point of view for the safety of workers who need access to the field both from the point of view of the integrity of the system itself, our supervision system is integrated with cameras. This functionality is important because insurance contract costs for safety of workers and for security of expensive installed apparatus can increase significantly in case of absence of surveillance systems. In some cases, this absence may prevent the issue of approvals required for the construction of the plant. Our surveillance systems provide the antitheft functionality and for the string panels can be integrated with antitheft systems provided by smart string-boxes based on measurement of impedance string. This functionality can be useful in nigh hours with low illumination.

3 System implementation

The supervision system is built on an embedded PC platform with a Linux operative system. This choice allows exploiting all well known Linux features in server field.

The PV implant in terms of its elements: inverter, string box, photovoltaic panel, production counter, is described with xml files allowing to the supervision-system the knowledge of all parameter to be monitored for each device. All other parameters, like rules about the response type in case of alarm or the list of the users allowed to access data or configuration thresholds, are described in xml language. In this way all document produced or processed by our system are in xml format.

The main elements of the supervision system are: a data logger to collect data from the devices of the PV plant, a xml database to store data, a control module to modify implant behaviour, four tools to perform the operation of alarm handling, data analysis, video surveillance, maintenance, and finally one module allowing the transmission of information to the system users. The whole system is schematized in the Figure 3.

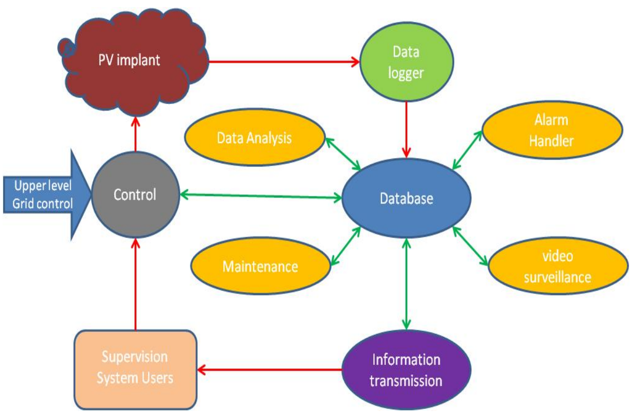


Figure 3. Supervision system

In the following each module of the above system will be described.

The data logger module handles the interconnection with the devices of the photovoltaic implant, in particular inverters and string box. The communication support MODBUS protocol in TCP/IP and RTU versions. The connection can be established via Ethernet wired or wireless in the first case or RS232/485 and USB in the second.

The database collects data from data logger and other modules in xml format. The usage of xml for all data exchanged among modules guarantee a standard format and the capability of integration with all tool supporting this technology. The xml format can be considered a key factor of our supervision system.

The data analysis module provides aggregate data from raw data: minimum, maximum, average, variance and instantaneous value for each parameter monitored. The data are compared with the expected one, considering the installation zone, the solar irradiation measured by sensors on site and data available from public database [9].

The alarm handler verifies if each parameter is inside constraints specified in xml configuration file. In case of alarm event produces an alarm status stored in database which is processed from other modules.

The maintenance modules check the presence of alarm status and verify if each procedure needed to silence the alarm has been performed or not.

The video surveillance module handles video cameras present on the implant site and provides functionality of antitheft. Its presence integrated inside the supervision system is a key factor because allows reducing insurance cost and avoids unauthorized presence on the implant site and possible thefts or damages to device and structure.

The information transmission module ensures all information from the supervision system arrives to the users. The information is published on web platform and in case of alarm users are worn by mail and sms and with colour changes on the web interface. In following figure the web interface developed to publish information and interact with users.

Figure 4. WEB interface.

The control module allows interaction with the photovoltaic implant through MODBUS protocol. It receives command from users or from an upper level grid control system and modify the functionality of the implants. Typical example could be the case of energy overproduction and the request from a grid control to reduce the amount of energy injected into the grid. Another example is when the grid handler request to inject current in the grid with a specific displacement different from the usual one. This condition can occur in the case where it is necessary to re-phase the network. The presence of a control module able to response at external command makes the system ready for integration with a smart grid.

4 Test and results

In order to test the system above exposed, a simulator of PV implants has been developed. This paper is not focusing on the simulator, but little details are here provided. The simulator is made by several processes running on one or more PC inside our departmental network simulating the condition of an implant with devices connected through a LAN. The simulator provides inverters, string box, and production counter behaviour and generates a large amount of malfunction conditions in order to produce alarm status that have to be processed from the supervision system. Could not be possible tests the supervision systems on a real solar power plant handling an onerous number of alarm states.

From our test the supervision system is able to handle correctly one alarm event each five seconds. The response time in case of events involving an action of the control module on the plant is 10 seconds. The aim of this supervision system is not to handle very fast events that could damage single device of the implant but only global malfunctions that have low evolution in the time. For event that should be detected in very short time, the single devices are provided with dedicated control.

5 Conclusion

The supervision system shown in this paper can be easily integrated with all devices usually present on photovoltaic power plants and with all communication network mostly exploited. The capability of monitoring all parameter of inverter and string box joined with advanced features on the energy production and the video surveillance cover all requirements typically present with a single system. The large interconnection possibility and the capability to response at external command make this supervision system suitable for the integration in smart-grid.

Acknowledgment

This work is supported by the European Commission in the context of the FP7 HEAP project (#247615).

References

[1] Suri, M.; Huld, T.; Dunlop, E.D.; Cebecauer, T.; "Geographic Aspects of Photovoltaics in Europe: Contribution of the PVGIS Website" *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 1, issue 1, pp. 34 – 41, March 2008

[2] Selvaraj, J.; Rahim, N.A., "Multilevel Inverter For Grid-Connected PV System Employing Digital PI Controller" *IEEE Trans. on Ind. Elecron.*, vol. 56, issue 1, pp. 149 – 158, Jan. 2009

[3] Dong-Jing Lee, Li Wang; "Small-Signal Stability Analysis of an Autonomous Hybrid Renewable Energy Power Generation/Energy Storage System Part I: Time-Domain Simulations", *IEEE Trans. on En. Conv.*, vol. 23 , issue: 1 pp. 311 – 320, March 2008

[4] Bialasiewicz, J.T.; "Renewable Energy Systems With Photovoltaic Power Generators: Operation and Modeling", *IEEE Trans. on Ind. Electron.*, vol.55, issue 7, pp. 2752 - 2758, July 2008

[5] Carrasco, J.M.; Franquelo, L.G.; Bialasiewicz, J.T.; Galvan, E.; Guisado, R.C.P.; Prats, Ma.A.M.; Leon, J.I.; Moreno-Alfonso, N.;"Power-Electronic Systems for the Grid Integration of Renewable Energy Sources: A Survey", IEEE *Trans. on Ind. Electron.*, vol.53, issue 4, pp. 1002 – 1016, June 2006

[6] CEI-IEC International Standard 61724- Photovoltaic system performance monitoring- Guidelines for measurement, data exchange and analysis, Ed. April 1998.

[7] S. Vergura, G. Acciani, V. Amoruso, G. Patrono, F. Vacca, "Descriptive and Inferential Statistics for Supervising and Monitoring the Operation of PV Plants", IEEE *Trans on INDUSTRIAL Electronics*, 2009, pp. 4456-4464.

[8] Edmundo Romàno, Ricardo Alonso, Pedro Ibanez et alt. "Intelligent PV module for grid-connected PV systems", in *IEEE Trans. On Industrial Electronics* , vol. 53, No. 3, Aug. 2006, pp. 1066–1073.

[9] http://www.ncdc.noaa.gov/cdo-web/
http://www.dicat.unige.it/meteo/
http://clisun.casaccia.enea.it/Pagine/Index.htm

[10] S. Vergura, E. Natangelo "Labview-Matlab integration for analyzing energy data of PV plants", *ICREPQ 2010*, March, 23-25, 2010, Granada, Spain (ISBN 978-84-613-7543-1).

[11] S. Vergura, E. Natangelo, "Labview Interface for Data Analysis of PV Plants", *IEEE-ICCEP 2009*, 9-11 June, 2009, Capri, Italy, pp. 236-241.

Developing an ICT-Enabled, Anti-Prophetic Approach to Sustainable Cities

Ellie Cosgrave
University of Bristol
Bristol, UK
ec5226@bristol.ac.uk

Theo Tryfonas
University of Bristol
Bristol, UK
theo.tryfonas@bristol.ac.uk

Kirsten Cater
University of Bristol
Bristol, UK
cater@compsci.bristol.ac.uk

Abstract - *Cities serious about achieving sustainability need to consider what the role of information might be for them, while acknowledging their municipalities as being more complex than the sum of the physical infrastructure. This paper argues that predicting the future is not a useful mechanism for city planning, and claims that this can actually be disruptive to sustainability because adaptability, differentiation and cultural development are effectively designed out. We argue that the ICT-enabled sustainable city will be an emergent consequence of a variety of factors, a System-of-Systems, including bottom-up design by creative citizens. The paper uses systems thinking to outline core principles for city leaders in dealing with the inherent complexity of delivering successful sustainable cities.*

Keywords: Cities, ICT, Sustainability, Systems thinking, Utopia.

1 Introduction

The last hundred years of human history have been characterised by unprecedented global population growth and urbanisation. In the majority of cases this has been achieved through the over-exploitation of natural resources, which is largely viewed as unsustainable [1]. If we are to continue to achieve beneficial economic development and the delivery of great urban centres, a new paradigm for urban design must be developed.

When considering how best to intervene in a city, one must first acknowledge that it comprises more than the sum of physical assets. Alongside the physical infrastructure that has traditionally characterised urban spaces, there are many more factors to consider – social interaction, cultural structure, and virtual communication channels are just a few. These sit within the physical, interwoven in space and time to form what we call 'a city'. If we are to progress in the delivery of sustainable cities, all of these dimensions must be taken into account and leveraged as important system components. Girardet supports this, saying "we need to look at cities as a whole – their economies, infrastructure, architecture, social networks, cultural realities and their environmental base – in order to grasp the full meaning of sustainable urban development" [2]. In modern cities, this must also include an acknowledgement of the digital and virtual aspects of the urban realm.

We are at a unique moment in history to creatively exploit Information Communications Technology (ICT) as a supporting force to bind the physical, social and virtual aspects of the city. As highlighted by Townsend, "(digital information technologies) are ready to become inextricably woven into everyday social and economic life of dwellers in every city on the planet" [3]. Recent advancements in ICT have seen the scale and role of information in every aspect of modern life expand almost exponentially. This information is transforming how we live our lives through better informed decision making that can be conscious to us or invisible to us (via automation/sensors). This opportunity is already being exploited by some forward thinking cities through implementation of e-governance, the roll out of smart grid infrastructure, the provision of open data portals etc.

Cities serious about achieving sustainability need to consider what the role of information might be for them, while acknowledging their municipalities as being more complex than the sum of the physical infrastructure. Taking this as a conceptual underpinning, this paper looks at the role of ICT in fostering sustainable communities that are not pre-planned, but that emerge from the creative leadership of the community and innovative companies (a non-'prophetic' approach). It investigates the implications of this for city leadership and decision-making around the concept of smart cities.

The paper takes a systems thinking perspective to highlight some of the fundamental principles that will support city leaders in navigating this field effectively. It starts by discussing the need for an anti-prophetic approach, investigates the unique role of ICT and goes on to present an overview of the 'city as a System-of-Systems'. Finally, the paper draws out some fundamental systems principles that will support city leaders in delivering sustainability through ICT investment.

2 The need for an anti-prophetic approach

The 20th century was peppered with visions of a future society (such as 'Garden Cities', 'Model Cities', 'Ecotopia' discussed below etc.), where imagined cities were scrutinized in meticulous detail. These often utopian visions offered alternative solutions to existing societal constructs. In the urban context, 'utopias' or 'ideal cities' describe harmonious places that are free of the turmoil of modern life. Put simply, city planners prophesied a vision of an ideal future society, and developed scheme designs in order to achieve them.

2.1 Background

Utopian visions of society first appeared during the sixteenth century, a time of significant social and technological change. These visions began with the political motivations of creating a communist or socialist society and a desire to overcome feudalism [4]. A prolific example of this is Thomas Moore's novel 'Utopia' (published in 1516), which charts the experiences of a traveller in an imagined world which functions within an ascribed (communist) political ideal. The short extract below (taken from Book II: Of Their Trades, and Manner of Life) clearly demonstrates how political ideals have permeated the vision of this 'Utopian' society.

"Besides agriculture, which is so common to them all, every man has some peculiar trade to which he applies himself, such as the manufacture of wool, or flax, masonry, smith's work, or carpenter's work; for there is no sort of trade that is not in great esteem among them. Throughout the island they wear the same sort of clothes without any other distinction, except what is necessary to distinguish the two sexes, and the married and unmarried."

In the aftermath of the industrial revolution, these utopian ideals began to reappear. Famously, Ebenezer Howard's vision of 'Garden Cities' offered a prescriptive solution to his societal concerns. Here concentrically planned zones of parks and open public spaces would be inhabited by a relatively small number of people and be entirely self-sufficient. This was developed largely as a reaction to the large-scale industrialisation and urbanisation that had characterised the late nineteenth century, and which had caused widespread poverty and poor living conditions [5].

Ernest Callenbach delivers another vision in his novel 'Ecotopia', set in 1999, which was at the time set 25 years in the future [6]. Here he catalogues the experiences of his protagonist William Weston in an imagined society that has rejected consumerism and that is striving for balance and harmony with nature. Callenbach conjures up vivid physical descriptions of the place, transporting the reader into this imagined future world. The following extract details his experience on leaving the main station:

"The 'street' itself, on which electric taxis, minibuses, and delivery carts purr along, had shrunk into a two-lane affair. The remaining space, which is huge, is occupied by bicycle lanes, fountains, sculptures, kiosks, and absurd little gardens surrounded by benches. Over it all hangs the most sinister quiet, punctuated by the whir of bicycles and cries of children".

2.2 Discussion

The reality of using utopian visions of society to guide decision making in cities, is entrenched in difficulty for many reasons. This section highlights some of our key concerns, and argues for a different approach.

1. At times, transformative technology (such as the internet) drives a fundamental shift in the way we operate. This cannot be captured by, or incorporated into utopian visions. Utopian visions cannot incorporate or capitalize upon future technological capability.

2. They allude to an 'end point' where society stops evolving, and a perfect equilibrium is reached. This goes against what we know a sustainable society to be like, where adaptability and continual development is at the core of what it means to be sustainable.

3. Utopian visions are founded on a single worldview, and are not able to incorporate a wide variety of perspectives.

4. They are allied to the particular social and political ideals of one generation.

5. The visions are prescriptive and take no account of the temporal, adaptive and complex nature of society. They are reductive and dictate a 'right' and 'wrong' way to exist. We know that there are many different ways in which to deliver great places, as demonstrated by the rich diversity of cities in the world today.

6. They seek to construct society through top-down design, rather than understanding that cities and cultures emerge from the collective actions of people and government.

A shift of focus is required from prophesying a physical construct to a 'process approach' that is able to adapt to and incorporate the changing needs of society. This process approach should acknowledge that the city is more than a sum of its physical elements, and that its emergent properties must be nurtured. Millat et al. explain "too many actors are involved to replicate solutions across communities, and parts of the value-network have to be recreated each time. Yet the skill of managing the network of actors and bringing a solution to the customer may emerge as the core competence in a market that has been neglected due to its complexity" [7]. So, while we cannot replicate solutions as the utopians allude to, we may be able to delineate some core principles for developing and managing sustainable cities.

This paper acknowledges that there is no one vision of a sustainable city and that attempting to create a generic vision for one is disruptive to moving forward because adaptability, differentiation and culture are restricted. Hence the paper is not intended to offer a design for society, but instead it looks at developing a process approach which will support a city to evolve in a way that delivers genuine value to citizens. There is merit in purposefully avoiding a prophetic vision of what cities should look like, and instead provide basic principles to support leaders in fostering successful, sustainable places.

3 The role of ICT in a sustainable city

There has been much discourse in recent years concerning the role of ICT in delivering successful and efficient cities. This dialogue has been largely captured in the 'smart city' debate which has involved large technology companies (such as IBM's smarter planet campaign, Siemens, and Cisco), technology consultants (Accenture, Arup), SMEs and entrepreneurs (Cosm etc.), and city leaders themselves (particularly Emer Coleman in London, Mayor Bloomberg in New York, Xavier Trias in Barcelona, Eberhard van der Laan in Amsterdam etc.). Kanter and Lintow argue that "someday soon, leaders will combine technological capabilities and social innovation to help produce a smarter world. That world will be seen on the ground in smarter cities composed of smarter communities that support the well-being of all citizens" [8].

From a review of the relevant literature as follows, and from having actively participated in this debate, three core areas of smart city value emerge:

1) Achieving operational efficiency;

2) Innovating service provision; and

3) Fostering the information marketplace.

3.1 Operational Efficiency

The use of ICT to optimise city services through ubiquitous sensing and actuation is the most common, and well argued aspect of the smart city.

When thinking about the role of data (or information) in the optimization of city services, it is useful to analogise it to the human body. Our bodies are constantly sensing the environment and responding to it, without us even being aware of it. It is constantly assessing temperature change, viral attacks, the light intensity and countless other things. Importantly, each of those sensors is coupled with a facility to respond and deal with the implications of what they measure. Sweating, white blood cell attacks, and pupil dilation all enable our body to run more efficiently, and respond to external events. Similarly, if we can sense what is going on in our cities, and are able to respond to the changing environment, or unexpected events, then we will be better equipped to run more efficiently. Increased amount of traffic data could allow us to respond better to traffic incidents by rerouting traffic in real time, or altering emergency services.

Sensing the city is becoming increasingly possible through the availability and variety of affordable sensors. These include light sensors, pollution sensors, noise sensors, location sensors, speed sensors and many more. We are at a point where we are able to distribute these sensors in our cities fairly easily. The challenge now is to work out how best to use this data, and couple it with response mechanisms that allow us to have meaningful impact on the efficiency of our systems. We need to think of this efficiency from a user perspective, a resource and cost perspective, and an environmental one.

Clearly, data can help us to understand what is going on in our cities, but without the ability to respond in real time, we are wasting a huge opportunity. For example, our bodies can sense temperature change, but that's not particularly useful without the ability to respond through e.g. sweating or shivering. Similarly, energy usage data in a city is not particularly useful without some ability to respond.

In that sense, we are coming to a point where the availability of data collected through ubiquitous sensing and novel technology is not the limiting factor for efficiency in cities. Instead the challenge will be learning how to compute and analyse this data, and respond to it in meaningful ways.

Rio, Brazil, have recently invested in a central control room that collates much of the city data in one place. This allows them to respond more effectively to problems, particularly in terms of emergency response [9].

3.2 Innovating Service Provision

There is a role for ICT beyond the much-discussed operational efficiency objective. ICT is actually driving a paradigm shift in the way that many traditional services within cities are delivered. New technologies are constantly appearing that enable us to provide basic services in different ways. For example, the move to the Oyster card in London has completely transformed how people interact with public transport in the city, facilitating ease of use, modal shift and flexibility [10]. This has real implications for the behaviour, quality of life, and welfare of citizens.

To refer back to our 'human body' analogy from above, service innovation can be thought of as an action we can take in order to achieve a certain goal. Our goal might be, for example, to lose weight and reduce our personal carbon emissions. To achieve these goals, we must change the way that we behave, and might choose to move from commuting in the car to cycling. This decision will be made based on a function of the affordability of the technology itself (in this case the bike), the safety of the roads (is there appropriate cycling infrastructure? e.g. cycle lanes etc.), as well as the impact it has on the original two objectives. In this way, new technologies (say, safer, cheaper bikes) can affect the way we choose to behave. Likewise, new technologies in cities are transforming the way they are able to deliver services to their citizens.

The key technology driving this change in cities is the Internet, which, coupled with the ever-increasing use of smart phones is transforming the way that cities are delivering services.

3.3 Fostering the Information Marketplace

There is a significant discourse around the role of information in cities delivering wider externalities than city service improvement. It has been argued that the availability of data across the entire city is fundamentally altering the economy. Information is available in many forms, and is creating opportunities for new products, services and business models in cities [9].

4 The City as a Complex System

Cities have a bewildering level of interrelationship and subtlety, whilst continually adjusting in response to stimuli. Sevtsuk and Beinart describe "a city is an enormously complex and open-ended system, with many intertwining force fields influencing its form simultaneously" [11]. It is this nature of cities that makes them robust to changing circumstances, but it also makes them extraordinarily challenging systems to mould and design.

It is this high level of complexity that makes systems thinking particularly relevant when trying to understand the nature and behaviour of cities. This section aims to illustrate some of the sources of complexity in cities.

4.1 Interconnectivity

Foth et al. argue that an acknowledgement of relationships between systems elements allows for an understanding of the hard-systems aspects of a city in its soft, human context [12]. Townsend sees an opportunity to deliver a bottom-up approach to urban design through these relationships, arguing that information technology can be harnessed to "re-engineer an infinite number of small-scale relationships", changing the overall urban form [3]. The current availability of ubiquitous sensing and actuation in cities is really making this opportunity a reality for city leaders.

4.2 Synergy

As Foth reports, "some of the fascination with human anatomy stems from the fact that a living body is more than the sum of its parts. Similarly the city is more than the sum of its physical elements" [13]. This synergy, evident in cities, offers a real opportunity for city leaders to foster sustainability through supporting emergence of solutions. Some view smart cities as places that will foster creativity, where citizens are generators of ideas, services and solutions, rather than subservient and passive recipients of them [14].

4.3 An Open System

Cities as aggregate entities are open, interacting with objects that lie outside their (ambiguously defined) boundaries. Planners and urbanists have "become aware of the limitations of viewing the city as a single 'closed' system" [12]. If the city is modelled as a self-contained entity, then arguably its true emergent nature can never be revealed. This is because supporting energy sources (e.g. international trade) that dictate a city's activities are being ignored. Thinking of a city as a collection of interacting systems enables us to capture the characteristically open nature of the city, making it an invaluable resource for understanding it.

4.4 Unintended Consequences

Efficiency is seen as a means of achieving sustainability and carbon reduction targets in cities. The principle is that if, for example, better access to traffic information leads to congestion reduction, there will be less traffic and therefore fewer emissions. However it is also conceivable that if there is less congestion then there will be more of an incentive to own and drive a car, thereby potentially increasing emissions. The complex nature of cities means there will always be a potential for unintended consequences. These should be actively sought out and dealt with by city leaders.

Systems concepts are an integral part of understanding the complexity of cities. If leaders are able to grasp this and understand the implications for investment in city systems, they will be better equipped to deliver sustainability and public value.

5 Implications For City Leaders

We have seen that there is an opportunity for ICT to support the delivery of sustainable cities. But we have also seen that the city is a complex entity, and that city leaders should avoid prophesying what the future will be like with the hope of being able to construct it. City leaders have a responsibility to invest in the city in ways that deliver value to their citizens and they must be able to deal with the complexity of the task. In that light, we explore some of the underlying principles of an approach to navigating this field.

5.1 A basic model

The promise of novel applications of ICT in cities drives the need to facilitate both top-down and bottom-up development within them. This will enable capitalising on the creativity and innovation of citizens, whilst fulfilling the obligation of city leaders to direct development in a way that delivers value to all. We also know that solutions cannot be cut and pasted from one city to another, but that cities must explore the most appropriate investment mechanisms for themselves. A useful first step in conceptualizing this problem is shown in Figure 1.

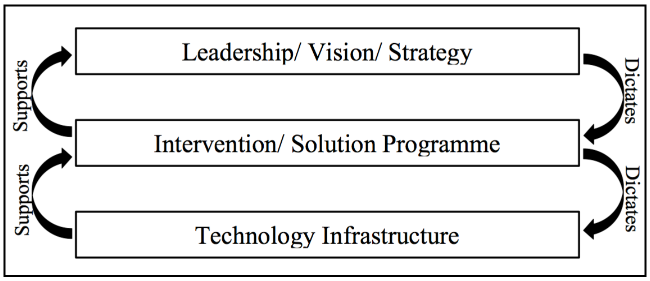


Figure 1. Layered model

The top layer is the high level aims of the city leadership. It requires them to consider, and explicitly articulate, the responsibilities that they have towards their citizens. What role must they play in directing the city, and what justification do they have for spending public money?

The middle layer is the most complex and intangible. It represents the complex network of processes and services in the city. These might be government actions, private companies operating core services (e.g. transport or utility companies), creative SMEs and innovators, or the actions of individual citizens. This layer represents the unique context in which each city operates.

The bottom layer is the collection of underlying technology infrastructures that support city operations (e.g. telecommunication infrastructure, sensor networks etc.). Government or private companies, or a combination of the two might invest in these.

Therefore, the whole structure can be thought of as a System-of-Systems, where social groups interests (i.e. stakeholder needs), complex projects, information technology, etc. can be though of as component systems of successive interactive layers. This model does not necessarily imply a hierarchical top-down approach to city management. It is evident that not all intervention programmes within cities are instigated as a result of an attempt to deliver on clearly defined strategic aims. Likewise not all technology provisions are implemented to deliver specific projects. For example, new technologies might be implemented for commercial purposes, which might then have such a significant impact on the city such that it must be incorporated into the strategy and vision. For example, some companies have taken the opportunity to deliver access to Wi-Fi in the public realm. This has had a significant impact on the city that city leaders are beginning to understand it as a key element of city life and have begun to incorporate it into city strategy. Jephson describes "a merged model of public policy formation that is neither top-down or bottom-up but a combination of both, one that recognizes that local actions are in need of some degree of centralized guidance that is conceptually consistent across jurisdictions" [15].

This approach therefore recognises the interdependent causal relationships between the three elements. It also incorporates the notion that the city strategy is dynamic and continually affected by the real-world system. There is no end point, but the city will always be defined to some extent by where is 'has been' and where it 'intends to go'.

5.2 A framework for smart cities

Using this model as a basic conceptualisation, we can overlay our understanding of the value delivered by smart cities through ICT as discussed in section 3, as shown in Figure 2.

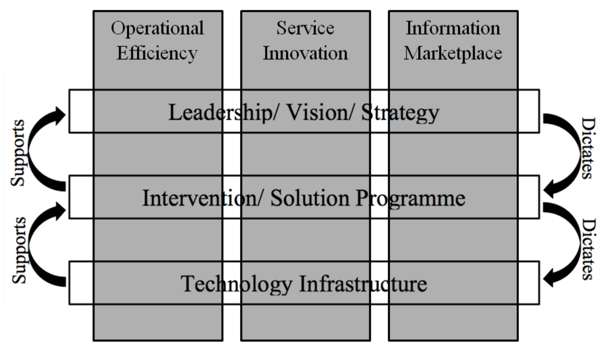


Figure 2. Smart city framework

Conceptualising the vision of a smart city in this way helps us to understand investment areas in manageable compartments. For example, if we wish to use ICT to deliver operational efficiency of public transport in the city, this might be a useful tool. It would compel us to consider the role of government, the existing operating context and the underlying technology infrastructure required. If these are understood, and their relationships investigated, then a clearer picture of the problem is built up. This empowers governments to think more creatively about where to intervene in the system. They may find that they need to invest in new infrastructure, or alternatively they might find that developing a closer relationship with network operators, or more clearly articulating their high level objectives would also deliver significant efficiency savings.

6 Conclusions and further work

Cities are complex multi-layered system-of-systems that are continually affected by a host of component systems and their relationships. As the global challenge of attaining sustainability in cities remains, if cities are to achieve their targets, they must avoid working towards utopian ideologies, bound by personal politics and aspiring to a prescriptive end-goal that can never succeed. Instead a process approach must be adopted that encourages leaders to acknowledge the value of emergence, and take account of the influences of time, space and people.

Of course our model as it stands is of limited use – it scopes however the context for an approach. We would like to develop this into a comprehensive model that can offer a detailed methodology for understanding appropriate investment in ICT and further work needs to be done. We plan to take further an action research approach, working directly with city leaders to understand the challenges and opportunities more clearly. This will culminate in the development of a methodology that can guide city leaders through effective smart city investment.

7 Acknowledgements

This work was supported by the Systems Centre at the University of Bristol, UK; the EPSRC funded Industrial Doctorate Centre in Systems (Grant EP/G037353/1); and Arup. The authors would like to thank Volker Buscher and John Davis for supporting this research.

8 References

[1] Bettencourt LMA, et al. Growth, innovation, scaling, and the pace of life in cities. Proc National Academy of Sciences USA. 104. 7301-7306.

[2] Girardet, H. Creating Sustainable Cities (Schumacher Briefing #2). Green Books Ltd. Foxhole, 1999.

[3] Townsend, A. Forward. in Foth, M ed. Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City. Information Science Reference, IGI Global, Hershey, 2009.

[4] Skroblies, H. 8 ideas and visions of 20th Century Planning, University of Technology Darmstadt.

[5] Howard, E. Garden Cities of To-Morrow, London, 1902.

[6] Callenbach, E. Ecotopia: The Notebooks and Reports of William Weston, Banyan Tree Books, Berkeley, 1975.

[7] Millat, AE. Haselmayer, S. Rasmussen, JH. Connected cities: your 256 billion euro dividend, Design London, London, 2010.

[8] Kanter, RM. Litow, SS. Informed and Interconnected: A Manifesto for Smarter Cities (working paper). Harvard Business School, 2009.

[9] Webb, M et al. "Information Marketplaces: the new economics of cities", The Climate Group, Arup, Accenture, Horizon, London, December 2011.

[10] Webb, F. 'Empowering Cities to survive and thrive'. June 2012. Available at <.http://theraconteur.co.uk/empowering-cities-to-survive-and-thrive/>, last accessed 09 June 2012.

[11] Sevtsuk, A. Beinart, J. The effects of ICT on City Form. MIT School of Architecture and Planning, Cambridge, 2005.

[12] Foth, Marcus. Odendaal, Nancy. Hearn, Gregory N. The View from Everywhere: Towards an Epistemology for Urbanites. in 4th International Conference on Intellectual Capital, Knowledge Management and Organisational Learning, (Cape Town, 2007), Academic Conferences International, 127-134

[13] Foth, M. Preface. in Foth, M ed. Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City. Information Science Reference, IGI Global, Hershey, 2009.

[14] Anonymous, Surely there's a smarter approach to smart cities? April 2012 Available at <http://www.wired.co.uk/news/archive/2012-04/17/potential-of-smarter-cities-beyond-ibm-and-cisco>, last accessed 09 June 2012.

[15] E.J. Jepson Jr. Sustainability and Planning: Diverse Concepts and Close Associations, Journal of Planning Literature, 15 (4), 499-510, 2001.

Towards an Intelligent System for Cognitive Behavioral Treatment

Panagiotis Karampelas
Hellenic American University
Manchester, NH, USA
pkarampelas@hauniv.us

Ioanna Laniti
Hellenic American University
Manchester, NH, USA
mlaniti@hauniv.us

Despina Konstas
Hellenic American University
Manchester, NH, USA
dkonstas@hauniv.us

Panagiotis Kalagiakos
Hellenic American University
Manchester, NH, USA
p_kalagiakos@yahoo.com

Abstract - *The paper describes an innovative intelligence information system of systems that can be used within the context of psychotherapy with individuals with depressive and bipolar disorders. The system achieves the following: (a) reveals psychological/ behavioral and physiological patterns of the disorders which will aid with diagnosing and long term prognosis; (b) provides information to therapists which will aid with treatment planning and improve treatment effectiveness and (c) provides both patients and therapists with services to improve their quality of life and enrich their professional experience respectively. The system through a closed-loop approach, further contributes in managing high-risk situations often associated with these disorders and significantly decreasing the potential risk for suicide.*

Keywords: intelligent system, personal health management system, bipolar disorders.

1 Introduction

Technology is considered an advancement for all fields, including treatment for medical conditions [1], [2], [4] as well as psychiatric disorders and is increasingly applied. The proposed interdisciplinary methodology empirically demonstrates the added value of incorporating technology into the established ways of conducting therapy producing a novel treatment model supported by an innovative wearable system which communicates with an intelligent monitoring system. Therefore, at the heart of the proposed methodology is the evolution of traditional cognitive behavioral treatment (CBT) taking advantage of the new developments in technology, such as the use of body sensor networks for monitoring behavioral and physiological patterns [15], which on one end allow the patient to self-monitor, predict and better manage behavioral maladaptive symptoms and on the other end, allow the mental health care provider to potentially better diagnose and provide a more accessible, informative and potentially efficient mode of psychological intervention.

The rest of the paper is organized as follows: Section 2 reviews the related work in the area. Section 3 describes the proposed approach and solution. Section 4 presents the system architecture and components and finally Section 5 discusses the current development and anticipated future work.

2 Related Work

The initial model of treatment is based on a well-established approach of conducting therapy for people with bipolar disorders. Cognitive behavioral therapy is a therapeutic approach that concentrates on monitoring and altering cognitions and/or behaviors to facilitate desired change in the patients' life. Studies have provided evidence for the efficacy of cognitive-behavioral interventions in treating bipolar disorders (e.g., [3]; [5]; [6]) as well as depression (e.g., [7]).

According to the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR, 2000) depression and bipolar are both mood disorders which means that patients having these diagnoses present primarily with mood disturbances. For a clinical diagnosis of any disorder, including the ones under investigation, the intensity of reported symptoms must cause a significant impairment in daily functioning (e.g., unable to cope with everyday demands within the social, professional or personal domain) or clinically significant distress. The proposed methodology focuses on depressive disorders including major depressive disorder (which may be secondary to a neurological or other Axis I or Axis II conditions). Major depressive disorder is characterized by one or more major depressive episodes (i.e., at least two weeks of depressed mood or loss of interest accompanied by at least four additional symptoms of depression, such as significant changes in appetite, sleep and social and psychomotor activity patterns, suicidal ideation etc.) (DSM-IV-TR, 2000).

The focus of the proposed approach also includes bipolar disorders, specifically bipolar I and bipolar II. Bipolar I disorder is characterized by one or more manic or mixed episodes (i.e., criteria are met for both manic and major depressive episodes except for duration), usually accompanied by major depressive episodes. Bipolar II disorder is characterized by one or more major depressive episodes accompanied by at least one hypomanic episode. A manic episode is the experience of an abnormally and persistently elevated, expansive or irritable mood lasting at least one week or any duration if hospitalization is necessary. Such episodes include such symptoms as inflated self-esteem or grandiosity, decreased need for sleep, pressured speech, flight of ideas, distractibility, or excessive involvement in pleasurable activities which may involve painful consequences (e.g., engaging in shopping sprees or sexual indiscretions). A hypomanic episode is a period of persistently elevated, expansive or irritable mood lasting at least four days but with symptoms not severe enough to cause significant impairment in daily functioning (DSM-IV-TR, 2000).

Overall, depressive and bipolar disorders are highly prevalent and carry with them a high cost in terms of damaged relationships, family suffering, and lost work productivity. Not all patients experience the same symptoms; the severity, frequency and duration of symptoms will vary depending on the individual and his or her particular condition. Yet, both depression and bipolar disorders may be treated and people with these illnesses can lead full and productive lives [8].

Regarding prognosis, patients have an increased frequency of recurrences as they age with later episodes occurring with greater intensity. Our proposed mode of treatment enhances compliance with treatment (psychotherapy and pharmacotherapy where applicable) and thus helps improving this progressive pattern. Additionally, with the population at hand, noncompliance with treatment appears to be the rule rather than the exception [9]. The proposed approach enables these patients to remain in treatment for a longer period of time so that their functioning is significantly improved and maintained at an acceptable level. New evidence further suggests that the long term course of bipolar disorder is chronic and involves a high prevalence of psychiatric comorbidity, and several symptom domains such as disturbances of circadian rhythms, mood instability, cognitive impairment as well as an increased mortality from medical disease [10]. The proposed system if systems consistently records such information and contributes towards a recognition of patterns for more accurate diagnosis as well as overall long-term prognosis. Although psychotherapies have been developed that enhance outcomes, reduce relapse, and aid adaptation to bipolar disorder and despite our greater understanding of the psychobiology of the disorder, we are only beginning to learn how to encourage patients to consistently use their pharmacological, psychosocial, and psychotherapeutic resources [11]. The proposed interdisciplinary approach is expected to aid in this direction as well.

3 Our Approach

The main objective of the proposed system of systems is to evolve the cognitive behavioral treatment (CBT) model for people diagnosed with depression and bipolar disorders, by incorporating multi-parametric measurements (physiological and behavioral) monitored by a mobile Personal Health Management System (PHMS), as well as enable therapists, through an intelligent Depression Bipolar Disorder Monitoring System (DBDMIS), optimize the treatment offered to such patients.

Patients, diagnosed with depression or bipolar disorders by their primary health care provider, are invited by their therapist to use the PHMS capable to monitor multiple behavioral and physiological parameters. Following a number of initial treatment sessions, they are provided with the PHMS which they are using to:

- record psychological/behavioral symptoms primarily based on the DSM-IV-TR (APA, 2000) as well as other symptoms (e.g., circadian rhythms, etc.) defined by the therapist
- monitor essential bodily and brain functions (e.g., unusual psychomotor activity, sleeping patterns, appetite changes, physiological and biochemical parameters)
- log the physical locations that reveal tendency for at-risk level behaviors (e.g., frequent visits at shops, banks or other places identified by the patient during the treatment sessions)
- review their health status by getting motivational feedback or warnings
- be alerted about prodromal symptoms and a possible depressive or manic episode in the near future
- be alerted for medication
- participate in a virtual community of patients with same diagnoses (i.e., depression or bipolar disorders) for patient-to-patient support
- send an alert to the primary health care provider in self-threatening situations
- psychoeducate themselves about depressive or manic episodes

On the other hand, primary health care providers using the DBDMIS which is the recipient of all the recordings by the patients' PHMS are able to:

- record patient clinical history as well as treatment session progress notes
- monitor patient mental health status in real-time
- receive, in high-risk situations (i.e., suicidal or homicidal intent), an alert including information on the location of the patient

- use stored data to provide patient with highly personalized treatment
- discover for each patient individual patterns of symptoms leading to new depressive or manic episodes through the use of data-mining techniques into behavioral and physiological data
- reach a treatment decision using the system recordings about the patient and system recommendations based on statistical analyses of the accumulated data
- better monitor treatment effectiveness through use of the accumulated data
- offer patients access to a service of patient-to-patient support by enrolling them in a virtual community of patients with similar symptoms
- access thorough reports generated by the system based on the patient's recorded information during the in-between-sessions intervals and the data recorded during the treatment sessions
- access informative reports on the targeted diagnoses (depressive and bipolar disorders) not only at an individual level (i.e., patient) but also at a country - (i.e., host country) level.
- participate in a virtual community of therapists exchanging their expertise on the field

4 System Architecture

The challenge of implementing such a system of systems is to combine the diverse devices and integrate the data from various sensors in a central system which will allow convenient access to the patient, the health care provider and emergency units. To address those challenges the PHMS has been designed in such a way to exist autonomously in an isolated environment when there is no connectivity and when connectivity is restored to communicate with the main DBDMIS sending the newly-acquired data. The mobile device connects with the sensors in order to collect the designated data and then analyzes them locally and sends them centrally for further analysis to the health care provider.

The implemented architecture of the system can be found in Fig. 1 where the various components of the system are visible. A three-tier approach for the DBDMIS has been adopted. The main components of the main system can be distinguished to those that serve the PHMS and those implementing the necessary functionality for the primary healthcare provider. More specifically, there is a common authentication mechanism that allows the legitimate users either doctors or patients to connect to the system.

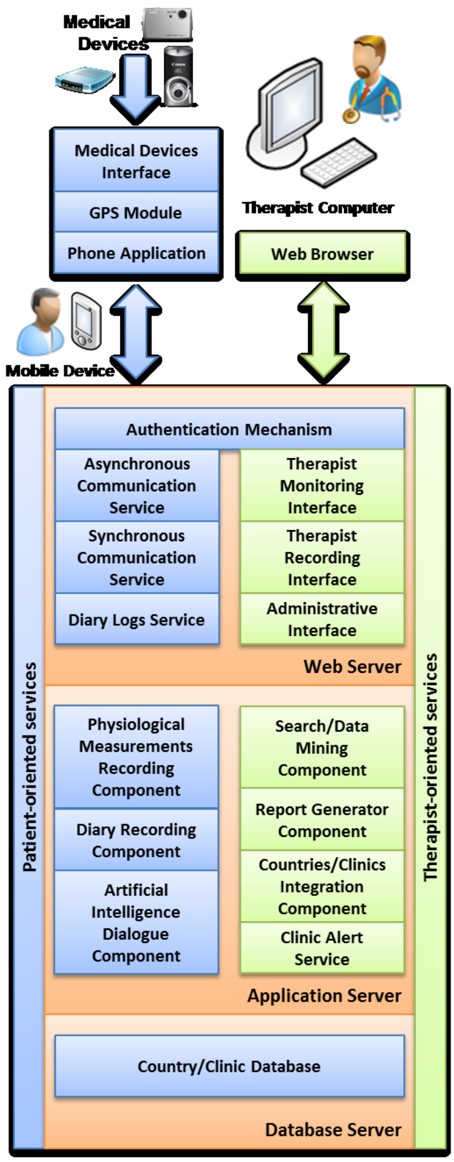


Figure 1. System Architecture

Once authenticated the mobile device can communicate either synchronously or asynchronously with the system to send the necessary data collected from the sensors and the journals completed by the patients. In the middle layer, the appropriate mechanisms to process and analyze the received data in order to respond to the user have been implemented. The specific component uses an artificial neural network (ANN) which has been trained with behavioral data from normal people and people suffering from depression and bipolar disorders covering the range of normal behaviors to manic episodes. Based on the training set the corresponding weights have been adjusted to match the cases and select the appropriate responses to the patients according to the identified stage of mood.

Concerning the therapist system, this is connected to the DBDMIS over the Internet using secure protocol. It offers three different services: patient monitoring, data recording and reporting, and administrative functionality. The middle layer of the therapist's system offers additionally a data mining component which is able to create association rules between the observed behaviors using the Apriori algorithm in order to find relations between the collected data. An integration component is also available that is responsible of data clustering when more than one clinic is connected and can provide comparative reports between the patients of the participating clinics from different countries with different culture, educational background, etc. Finally, an alert service is available to notify the primary health care provider in case of an emergency by processing the data collected from the PHMS.

Since the work is under development some of the components are not fully implemented yet such as the Medical Devices Interface which allows the communication between the mobile device and the sensors. The specific component depends on the licensing from a third party organization.

4.1 Mobile Device

Patients who are eligible to participate in the program after the screening process is complete they have to provide their full consent. Following a number of initial treatment sessions, they are provided with the Personal Health Management System which consists of a Mobile Computing Device (Fig. 2) and an integrated body sensor network compatible with the mobile device (not yet available). In the mobile device, the PHMS will be able to connect wirelessly (using Bluetooth) with the integrated body sensor network in order to gather physiological data monitored by the external devices when the implementation has been completed. There is evidence suggesting that physiological data such as heart rate, blood oxygen levels, brain signals, sleeping patterns, etc. may be indicators of depressive and bipolar disorders (e.g., [12]; [13]; [14]). Moreover, the mobile device is used to record psychological/behavioral symptoms based on the DSM-IV-TR (APA, 2000).

Patients monitor their own mood and behaviors but instead of writing daily logs, as commonly used in CBT, they enter specific information characteristic of these disorders such as, changes in appetite and sleep, alcohol and drug use, suicidal ideation etc. (in Fig. 2, Journal). These data, combined with the physiological measures (collected by the integrated body sensor network) are assessed by the PHMS. The system provides positive feedback or warnings or prompts on what to do to the patient depending on an analysis of the recorded data (personalized medical advice). The system, for instance, may prompt the patient to engage in specific, positive, health-promoting alternative behaviors in an attempt to decrease dysfunctional cognitions or maladaptive behaviors such as alcohol drinking (e.g., go to the gym or call a friend etc.) or reduce suicidal risk (e.g., call the hospital). More specifically, when the data recorded (i.e., symptoms) reaches predetermined at-risk levels, the system first provide immediate feedback or guidance to the patient, according to the severity of the situation, and then decide whether the patient needs immediate assistance by a primary mental health care provider. Such decision is made based on prior history of each individual patient and the patient patterns collected in the Depression Bipolar Disorders Monitoring Information System.

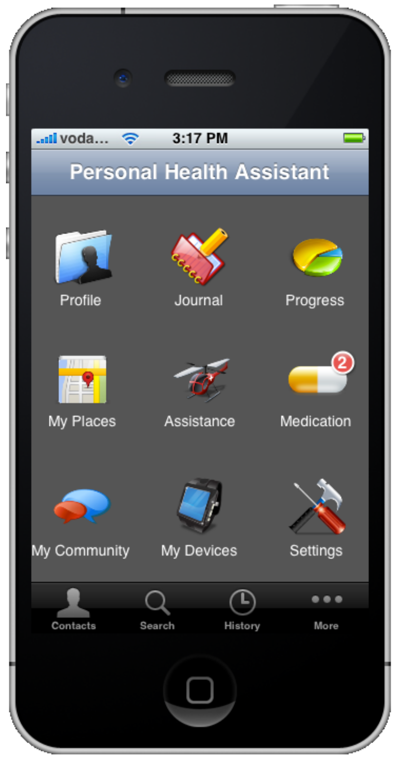


Figure 2. PHMS main page

In case of an emergency (i.e., when patients have surpassed the at-risk level of suicidal or homicidal ideation and have moved into intent levels or if the integrated body sensor network identify a critical situation through registration of vital signs, e.g., committing suicide), the PHMS sends an alert to the patient to contact one’s mental health provider. Subsequently and only with the patient’s consent to utilize this specific service, the system notifies/alerts the primary mental health care provider on duty. The PHMS alerts the monitoring center (DBDMIS) by sending the patient’s identification, recorded immediate maladaptive behaviors as well as the current position of the patient (taking into account limits of confidentiality). The latter is accomplished by taking advantage of the GPS and/or GSM functionality of the device- such a function is only activated in emergency situations (i.e., suicidal or homicidal intent) and only in cases in which the patient has consented. The closest available clinical setting (in case the patient’s primary mental health care provider is not available) is suggested to the patient and an alert is sent to the monitoring clinic. The operator of the DBDMIS system in the clinic then notifies the doctor on call who contacts the patient for further assessment and to decide on the appropriate course of action. In case the patient is at high risk for suicide or homicide but refuses to cooperate with the doctor, information is communicated to an ambulance (or an equivalent emergency service) for immediate response. Special provision is taken in order to preserve patient privacy and avoid revealing personal data to third parties without a life-threatening reason.

The PHMS, taking advantage of the location awareness (GPS and/or GSM), is also able to track behaviors such as repetitive visits to a specific place and associate visited places with potential at-risk related behaviors (as recorded by the patient during the initial treatment session or added later through the PHMS). For example, if a patient visits the bank three times on the same day, the system is able to identify such a behavior and advise the patient on how to handle the situation. In similar cases, e.g., when the patient visits several shops per day, the system also intervenes and suggests to the patient how to control his/her behavior. This is achieved by having the patient enter in the device- with the assistance of their therapist- those places possibly associated with bipolar/depressive behaviors such as banks, shops etc.

The PHMS has also the potential to provide the patient with access to a network of other individuals with similar disorders. This service can be used for daily patient-to-patient support. The system is able to support synchronous communication between the patients either through the mobile device or through website. Access to system is secure and private; username and passwords are used to enter the system. The conversations are treated as confidential and private, however when key words/terms are identified by the DBDMIS through text-mining techniques (such as, "I want to die"), it notifies the PHMS to start a dialogue with the user to help clarify the situation, and if necessary, prompt the patient to contact his/her mental health care provider.

All services provided by the PHMS are provided with the consent of the patient. It is up to the patient to accept or decline certain services provided. The PHMS is autonomous (close-looped system) and in cases that there is no connection with the respective DBDMIS of the health care provider is able to operate as usual and store the monitoring data and parameters until the connection is restored. In critical situations, the system tries to establish connection with all the possible alternatives. If this is not possible, then the system sounds an alarm to notify people in the perimeter. It is expected that the PHMS will support the independent living of the patient, increasing eventually the intervals that the patients will need psychological/ medical support. Furthermore, the PHMS will be available on a twenty-four/seven basis so the patient will feel secure since a preliminary support can be obtained directly from the system even at night time.

The proposed treatment protocol includes a thorough training of the patients on how to use the Personal Health Management System and the integrated body sensor network.

4.2 DBDMIS Platform

The Depression Bipolar Disorder Monitoring Information System is established in a participating clinic. The DBDMIS is able to run autonomously and it is scalable and able to connect with other DBDMIS in other clinics or countries to maximize data integration and thus enable better statistical analysis, pattern recognition and data-mining.

The DBDMIS running in the clinic enables the therapist to record the patient's clinical history at the beginning of treatment, record the therapist's progress notes for each session and log the services the therapist considers important for the patient's treatment. For example, the therapist may suggest to a patient the use of the PHMS to record specific behaviors, places identified as relevant to a manic state (e.g., neighborhood's ATM), as well as a series of physiological measurements. Alternatively or additionally, the therapist may suggest to a patient to become an active member of the virtual community group which aims at providing patient-to-patient support. Based on the clinical judgment of the therapist, the system's multiple capacities are used so as to provide patients with a highly personalized treatment. The protocol requires the therapist to acquire a signed informed consent form by the patient upon agreement of the two parties on the suggested services. The patient then receives the PHMS and the integrated body sensor network and actuators along with a thorough demonstration of the operation of these devices.

The DBDMIS is automatically updated with real-time data recorded by the patient through the PHMS and the integrated body sensor network. This way the therapist is able to have a real-time overview of the patient's status. Several research tools (pattern recognition, data-mining, neural networks) are available to the therapist, in a user-friendly format, for identifying behavioral symptom and/or physiological measurement patterns characterizing the status of each patient through time. Such data, combined also with clinical history and progress notes, are incorporated in a report which is also readily available to the therapist to use. Observed emerging/new behavioral patterns may be used for informed treatment decisions (e.g., changing current course of treatment) as well as treatment evaluation (i.e., effectiveness of current course of treatment).

Based on the data available the therapist is able to optimize the treatment provided to the patient taking into consideration statistical data from all the participating clinics and countries. In case of an emergency, e.g., the patient reveals at-risk levels behavior the DBDMIS system is able to send a warning to the therapist or the DBDMIS operator about the status of the patient, prompting for action. If the patient has accepted to be tracked by the PHMS, the DBDMIS system has the location of the patient available for immediate action e.g., ambulance intervention.

Finally, the therapist is also able to communicate with other experts in the field through the Virtual Community tools that are integrated to the DBDMIS. Relevant papers and publications, reports and other documents are available through the Virtual Community Platform along with a directory of experts. Communication tools such as Chat, Message Boards and Blogs are available for the therapists to create a network of knowledge in their field of expertise.

5 Conclusions

This paper presents work in progress towards the implementation of a treatment methodology for depressive and bipolar disorders using an innovative and intelligent information system of systems. The system upon completion is expected to aid with diagnosing and long term prognosis of patients with the aforementioned disorders by revealing psychological/behavioral and physiological patterns of the disorders. It also provides information to therapists which will aid with treatment planning and improve treatment effectiveness and provides both patients and therapists with services to improve their quality of life and enrich their professional experience respectively. The closed-loop approach adopted, further contributes in managing high-risk situations often associated with these disorders and significantly decreasing the potential risk for suicide.

Further areas of improvement that have been identified include the evolution of the data mining algorithms taking into consideration data entered by the patients, physiological data monitored by the integrated body sensor network and responses and feedback of the therapists to identify the positive and at-risk level behaviors. Furthermore, the development of the necessary tools to study associations between physiological measurements and bipolar/depressive behavior can enable the improvement of existing diagnostic and treatment methods for these disorders. Another important area of consideration is the compliance with the guidelines of the designated regulatory authorities for healthcare delivery in different countries in order to get the approval for usage and aid in the improvement of the quality of life of the targeted population.

References

[1] Kollmann A, Riedl M, Kastner P, Schreier G, Ludvik B. Feasibility of a mobile phone-based data service for functional insulin treatment of type 1 diabetes mellitus patients. J Med Internet Res. 2007;31;9(5):e36.

[2] Yoon KH, Kim HS. A short message service by cellular phone in type 2 diabetic patients for 12 months. Diabetes Res Clin Pract. 2008;79(2):256-261.

[3] Lam DH, Hayward P, Watkins ER, Wright K, Sham P. Relapse prevention in patients with bipolar disorder: Cognitive therapy outcome after 2 years. American Journal of Psychiatry. 2005;162(2):324-329.

[4] Pantelopoulos A and Bourbakis NG (2010) A Survey on Wearable Sensor-Based Systems for Health Monitoring and Prognosis. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 40(1), 1-12.

[5] Lam DH, Watkins ER, Hayward P, Bright J, Wright K, Kerr NG, et al. A randomized controlled study of cognitive therapy for relapse prevention for bipolar affective disorder: Outcome of the first year. Archives of General Psychiatry. 2003;60(2):145-152.

[6] Scott J, Paykel E, Morriss R, Bentall R, Kinderman P, Johnson T, et al. Cognitive-behavioural therapy for severe and recurrent bipolar disorders: Randomised controlled trial. British Journal of Psychiatry. 2006;188:313-320.

[7] Craigheart WE, Hart AB, Craighead LW, Ilardi SS. A guide to treatments that work (2nd ed.). Nathan PE, Gorman JM. New York, NY, US: Oxford University Press, 2002:245-261.

[8] Kessler RC, Chiu WT, Demler O, Walters EE. Prevalence, severity, and comorbidity of twelve-month DSM-IV disorders in the National Comorbidity Survey Replication (NCS-R). Archives of General Psychiatry. 2005;62(6):617-627

[9] Svarstad BL, Shireman TI, Sweeney JK. Using drug claims data to assess the relationship of medication adherence with hospitalization and costs. Psychiatric Services. 2001;52(6):805–811.

[10] Soreca I, Frank E, Kupfer DJ. The phenomenology of bipolar disorder: What drives the high rate of medical burden and determines long-term prognosis? Depression & Anxiety. 2009;26(1):73-82.

[11] Basco MR, Rush AJ. Cognitive behavioral treatment for bipolar disorder (2nd ed). New York, Guilford Press, 2005

[12] Salvadore G, Drevets WC, Henter ID, Zarate CA, Manji HK. Early intervention in bipolar disorder, part I: clinical and imaging findings. Early Intervention in Psychiatry. 2008;2(3):122-135.

[13] Lewis DA. Diagnosis-specific brain abnormalities. American Journal of Psychiatry. 2008;165(12):1511-1512.

[14] Miller A, Fox NA, Cohn JF, Forbes EE, Sherrill JT, Kovacs M. Regional patterns of brain activity in adults with a history of childhood-onset depression: Gender differences and clinical variability. American Journal of Psychiatry. 2002;159(6):934-940.

[15] Bonato P. Wearable Sensors and Systems. IEEE Engineering in Medicine and Biology Magazine, 2010:29(3), 25-36.

Socio-technical considerations for Enterprise System Interfaces in Systems of Systems.

C.E. Siemieniuch
ESoS Research Group
Loughborough University
United Kingdom, LE11 3TU

M.A. Sinclair
ESoS Research Group
Loughborough University
United Kingdom, LE11 3TU

Abstract - *The inclusion of humans in the systems comprising the System of Systems (SoS) mean that the boundaries of the SoS become fuzzy, and the interfaces become more complicated. Two particular issues are discussed; incompatibilities between organisations at the interfaces, requiring some changes to the ways companies operate; and incomplete information, requiring integrity of behaviour to enable trust to replace knowledge. As a consequence, some tool requirements for design and sustainment of SoS are outlined.*

Keywords: Sociotechnical systems, interface design, sparse information, incompatibilities, tools

1 Introduction

Much of the academic discussion about System of Systems (SoS) is at an abstract level, which means that some of the realities of SoS life are ignored. This paper discusses some of these, and some of the ways to ameliorate the undesirable behaviour of the SoS that may result.

For example, there is a particular difficulty in defining the boundaries of an SoS when a socio-technical perspective is adopted, arising from the fact that humans are now included. It is the capacity of humans to operate as thinking, decision-making, responsibility-accepting, goal-seeking entities that makes them of fundamental importance to enterprise systems, and it is their needs for a support network (sustenance, accommodation, training, recompense, understanding, etc.) which significantly widens the boundaries of any SoS within which they operate to include many more systems. In relation to boundaries, time is also of significance. For SoS of short duration (e.g. a month or less), many support systems can be ignored, because nothing much will change within that period. For longer-lasting systems, perhaps with an expected duration of 2+ years or more, issues of sustainability immediately emerge. It is also observable that most of the long-life SoS are essential to the continuity, safety, security and well-being of citizens in all the nations of the world, and that many of these SoS are global in extent. This paper considers some of the implications that emerge from this perspective.

We commence with some classifications, to bound the discussion below. In a paper that has become famous, Dahmann & Baldwin [1] listed four classes of SoS:

- Directed, typically owned and controlled within a single organisation
- Acknowledged, typically many system owners, but with a designated SoS manager
- Collaborative, typically many system owners, held together by contracts and with no central direction
- Virtual, typically many systems which may form a loose federation for a short period of time to accomplish some purpose, with no central direction. Systems and their owners may come and go during the time that the federation is in existence.

Unfortunately, most real SoS will have parts fitting the characteristics of different categories, both at any instance of time and during the lifecycle of the SoS. While the core of the SoS may fit the 'directed' category, as one moves towards the periphery of the SoS, where lurk the consultants and other specialist service organisations, the characteristics of the SoS may fit other classes better.

This brings to light the second part of the problem, that these different organisations may be structured and managed in different ways. In principle, this is not a problem (and abstract discussions in the literature bear this out by not discussing it), but in the real world where the SoS operates as an open super-system, issues of environmental noise and evolution will introduce emergent behaviour problems in the SoS that will need management. We give an example of these; the 2011 earthquake in Japan that has disrupted many supply chains. In the words of leMerle:

"*Recently, for example, Apple's supply of lithium-ion batteries, used in iPods, suddenly dried up. Unfortunately, as Apple quickly discovered, almost all its suppliers purchased a critical polymer used to make the batteries from the Kureha Corporation, a Japanese company whose operations were disrupted by the March 11 earthquake. In fact, Kureha's share of the global market for polyvinylidene fluoride, which is used as a binder in lithium-ion batteries, is 70 percent.*" *[2]*

The resultant behaviour of the erstwhile clients fits the category 'interesting'. ? not understand this bit

Emergent behaviour may affect different levels of interaction at the interfaces of the SoS. We may identify the levels developed by the Network-Centric Operations Industry Consortium [3]:

- Political or business objectives
- Harmonised strategy/doctrines
- Aligned operations
- Aligned procedures
- Knowledge/awareness of actions
- Semantic/information interoperability
- Data/object model interoperability
- Connectivity & network interoperability
- Physical interoperability

However, for this paper we consider only the top five layers, using a slightly different classification for ease of explication because the NCOIC layers are non-monotonic what does this mean:

- Government and regulatory (NCOIC political or business objectives) - At this level, the three main considerations are the legal framework that ensures ?fair competition, the governance of SoS, and the implications for society in general should any (or several) critical SoS fail.
- Strategy (NCOIC strategy) - This level is concerned with the strategy of the SoS itself. Its strategy will depend on its environment, the volatility of that environment, its geographic spread across different legislations in different regions, and the nature of its business domain and its perceived risks [4-12].
- Policy (NCOIC doctrines) - This general level is more concerned with how the SoS regulates and constrains what and how it executes its business. It is at this level where considerations such as inter-organisational trust, partnering, service level agreements, and contracts are important. Various fundamental issues such as the potential clash between lean operations and the need for resilience must be addressed, and plans for risk mitigations are prepared.
- Operations (NCOIC operations & procedures) - It is at this level where organisational and ITC considerations overlap. The ideal for the SoS is smooth operations within the established strategic and policy constraints of each of the individual organisations and of the SoS in which they are embedded. The quality of each interface between the organisations is critical in achieving this ideal, and we discuss some issues below.
- Transactions (NCOIC procedures)- At this level, IT&C plays a critical part, and the lower NCOIC layers are very significant.

The implication of this latter classification is that when the interface is specified between systems with different owners (or owned by different parts of the same organisation), the interaction between the two sides needs to be specified for these levels as well. We illustrate this with an example, albeit that this is hypothetical. Consider the different lifecycle states of organisations as identified in Fig 1 [13] below. As any new technology matures until it is commoditised, so the appropriate organisational structure changes, to optimise the benefits available from the technology at that stage. Thus, companies at the left-hand-side of the diagram will be start-ups; typically small or medium-sized, with flexible, project-oriented structures and perhaps led by a vision-based, obsessive leader, whereas companies at the right-hand-side will be fairly rigid, process-based organisations with defined procedures and minimal staffing, and a management focussed on cost-cutting and continuity of its money-earning processes.

Now consider an SoS in which two of these companies are partners, necessarily interfacing with each other – for example, the start-up provides a new module to interoperate with a standard module provided by the other company. The start-up has strengths in resilience and agility, whereas the other is much more rigid. Unless both companies choose to modify their behaviour at the interface, serious performance problems could emerge. In other words, the interactions at strategy, policy and operations levels will need to be addressed in addition to those at the lower NCOIC levels.

A second example is concerned with the characteristic of sparse knowledge and information in the SoS. Fig. 2 below captures some of the issues involved. The first implication of this diagram is that as change occurs in the SoS due to evolutionary pressures (including competition), organisational systems (OS) will be surprised by the changes, carrying the further implication that resilience is an important characteristic of SoS that needs to be engineered-in. The diagram carries the second implication that if one must operate without full knowledge, then one must trust the OS with which one must work; in turn, this has the implication that each OS must operate with integrity if this trust is to be established and maintained.

2 Minimising 'surprises' in the SoS

As a generalisation, the conditions of participation in a SoS between any two OS are established by contracts, and performance is set by service-level agreements (SLAs). However, as many managers and engineers involved in SoS will attest, when one must wave contracts and SLAs at another partner in the SoS in some argument about what happens at the interface, the co-operation is in well-established trouble. Fortunately, the conditions to avoid this state of affairs are fairly well-known [14-16] and we outline them briefly below, before discussing some of the engineering support that needs to be provided.

The two keys to an efficient socio-technical interface are firstly, performance (delivery to specification, on time, and in full), and secondly, trust (the result of keeping promises, and operational integrity). These are not independent; they have similar drivers. These are:

- The establishment of common goals and policies.
- Transparency about problems, and ways of working (including such notions as open-book accounting, and exposure of internal processes)
- A willingness to share benefits. (such as cost and time savings)
- A common understanding of terms and their usage
- Respect for confidentiality. This is a standard requirement within all organisations, strongly affected by the organisational structure and role design, and at a more mechanical level by the design and implementation of the IT infrastructure. It is a cornerstone of trust.
- Speedy and efficient execution of promises. The important determinants of this are organisational design; empowerment (responsibility, authority, and access to resources) of individuals and teams; effective control of processes; and access to timely, relevant, knowledge and information.
- Personal relationships, built up over time. This, together with the next point, comprise the other two cornerstones of trust. Knowing your peers' capabilities, freedoms, empowerment, ineptitudes, biases, and foibles provides the basis on which you can expect your peers to take the right actions. The quality of these personal relationships is affected by the organisation's human resource policies, and by its approach to empowerment of individuals.
- Recognition of the 'favour bank'. When surprises occur, there may be a need to call on others to provide help, resources, or alternatives. This may be accomplished by people going outside the normal procedures ('going the extra mile'; 'bending the rules') to resolve the surprise. In so doing, they 'bank a favour' to be redeemed when they encounter their surprise. This is directly affected by the level of empowerment of the individual person, and is an organisational design issue.

Assurance of these drivers at the interface requires attention to OS design, role design, and job support, as described, for example in [17]. However, in the context of SoS, the task of delivering these characteristics is not easy, and the provision of some appropriate tools would be highly beneficial. In the last part of this paper we outline some tools that, suitably tailored to a SoS environment, would be extremely helpful.

3 Tools for the process of designing SoS

For each of the categories below, tools do exist within the category. However, in many cases the tools require considerable tailoring to fit the needs of the stakeholders; too often they require experts to manipulate them, and often do not address issues of importance. There is a definite need for further work in this arena.

- Tools for Enterprise Architecture modelling & visualisation, to identify patterns and to explore the tradespace for the enterprise [18, 19].
- Simulation tools to run models separately and together. These models may be distributed, may only be available in 'black box' mode, and may need to be run almost continuously [20, 21].
- Dependability analysis tools, to enable the quality of the output from the tools above to be assessed, as an input to decision-making [22, 23]
- Mining tools to derive behaviour patterns, and thereby to provide an input to governance of the SoS [24].
- Governance metrics and measurement tools appropriate for the SoS environment
- Tools for cybersecurity assurance [25]
- Social network tools for organisations, for knowledge sharing and resolution of 'surprises' [26]

4 Including the Tools in SoS design

In this section, we outline briefly an approsach that enables socio-technical aspects to be addressed. The approach below is premised on large-scale SoS design; adaptation will be required for smaller-scale SoS.

Once it is decided that the design process must include consideration of social as well as technical aspects because of the essential roles of humans as agents of responsibility, authority, change management, control, operation, etc., necessarily one must consider the whole human agent, embedded within an organisational structure (which itself is embedded within a social order), capable of autonomous behaviour, and equipped with a personal history and set of attributes, knowledge, experience and skills. Essentially, this is a world of relationships, ethics, and perceived duties, in which physics and engineering principles and laws play little part, and constancy is an abstract concept. As one senior manager in a defence OEM remarked, "Our problem is not engineering - throw enough good engineers and money at the engineering issues, and they will be solved. Our big problem is management of our people."

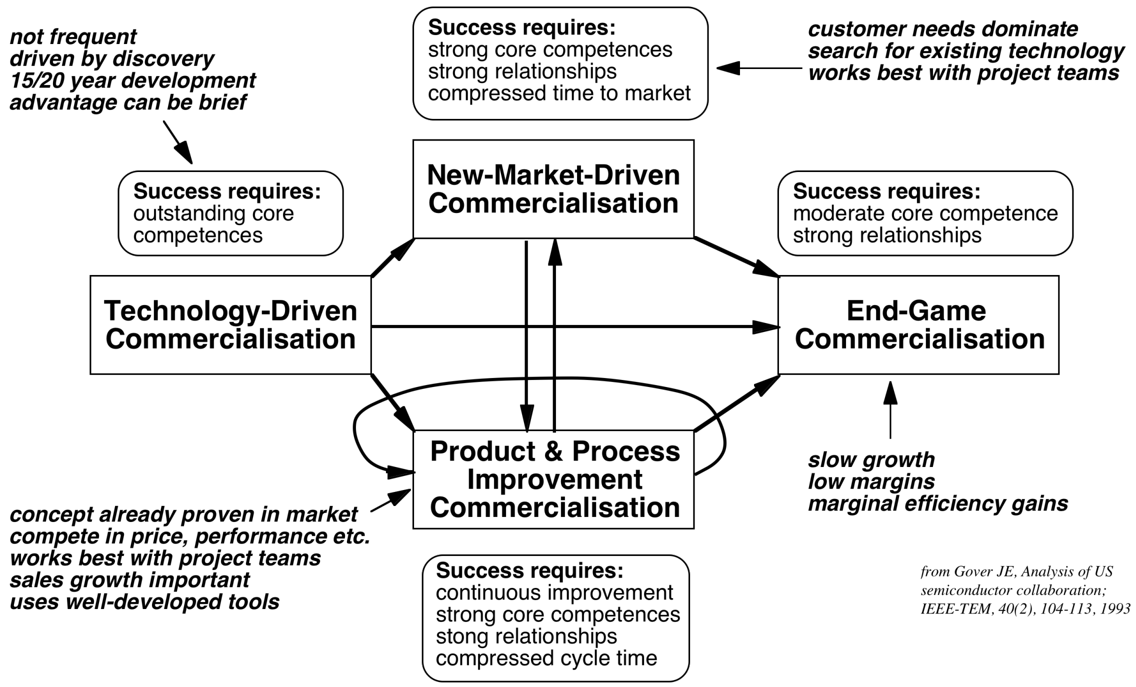


Fig. 1 As the technology evolves from novelty to commodity, so the kind of organisation that is best-suited to exploit it evolves from a start-up to a minimalised, lean, continuity-of-operations company.

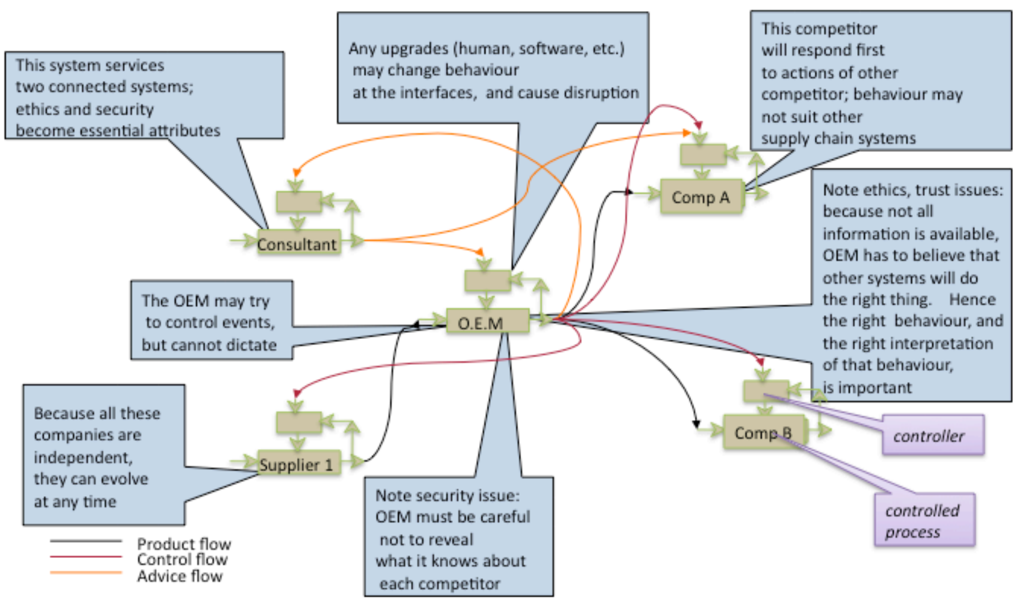


Fig. 2 Illustration of a supply chain SoS and the issues that might arise that will preclude full information transfer, and hence create 'surprises' in the SoS.

The implication of this is that SoS design becomes a 'wicked problem' [9, 27-29], with its attendant issues of indeterminacy, infinitude, and sub-optimal outcomes. This class of problems requires a different approach to their engineering [27] and management [9, 30], with the acceptance that their design is essentially a never-ending process predicated on continuous feedback of SoS performance as the SoS comes into being. It will be observed that the tools mentioned in Section 3 fit well into this ever-lasting process.

Initial phases of SoS design will be based on consortium-building, boundary identification (internal and external), and the establishment of trustworthy relationships (without which you might as well not start). As the capability engineering proceeds and the SoS as a socio-technical whole begins to take shape, so the enterprise modeling, architecture and simulation tools will be entrained; initially simple versions (in the absence of reliable data and information), mainly to ensure common understanding and to build collaboration. As the SoS enterprise develops, these simple versions will be replaced by more complex models of the enterprise, to ensure that it is and will continue to be resilient, to address the problems outlined earlier in this paper (and many others). Once the SoS enterprise is instantiated and is relatively stable in performance, the main function of the tools is to anticipate issues that will require further design; for example when external circumstances require adjustments to the architecture of the SoS. Since change takes time, especially where humans are the recipients of change, anticipation is vital; as one esteemed practitioner of change management memorably expressed it in a presentation, “if you don’t address the issues of change early and well, they will eat you for breakfast”.

At a lower level of SoS design, an approach that the authors have adopted [31] commences with an analysis of the decision-making structures and systems of the enterprise system (ES) under consideration to identify both efficiencies and inefficiencies, followed by an assessment of the cultural characteristics of the organisations and decision-makers of substance within the ES. This is then supplemented by the application of process modeling and role allocation and profiling techniques which provide an architectural analysis to assess authorities, responsibilities, control and communities of practice. The analysis then feeds into socio-technical performance tools which assess the likely performance of the new system that has been created. This approach could provide some of the input required for the tradespace and simulation tools outlined above. .

5 Conclusion

This paper has addressed a number of socio-technical issues to do with the real-world operations of SoS, with an emphasis on the civil domain. It has discussed a couple of well-known issues not often discussed in the SoS literature, and for which a socio-technical solution is required. The drivers of the solutions are fairly well-known and the organisational antecedents that must be in place for these drivers to be effective are also well known; the problem is that we lack appropriate tools to arrive at good solutions to the design and trouble-shooting of these. There is a short outline of the kinds of tools required, in order that the important stakeholders in the SoS can understand the performance and potential weaknesses of the SoS in order to deliver the resilience and efficiency that is expected now and will be necessary for future SoS.

Creation of dynamic short and long life-cycle SoS that comprise partnerships of enterprise systems will require changes to the individual enterprise system structures and processes. This is turn will require different people and resources perhaps from a broader spectrum of national, organisational and professional cultures, including a wider pool of available and accessible knowledge. These structures and processes must be accepted and integrated into existing/ legacy enterprise systems if they are to be effective. An integrated enterprise systems modelling and analysis tool could help facilitate such an integration. Considerations from agility suggest a paradigm shift is required, including elements notoriously difficult to change, such as culture and trust. Understanding how these things permeate and interact through an SoS may well help to avoid undesirable emergent behaviour than can have disastrous consequences.

References

[1] J. Dahmann & K. Baldwin, "Understanding the Current State of US Defense Systems of Systems and the Implications for Systems Engineering. ," 2nd IEEE Systems Conference, Montreal, 2008.

[2] M. leMerle. How to prepare for a black swan [Online].

[3] NCOIC, "NCOIC Interoperability Framework," Network Centric Operations Industry Consortium, Washington, DC 2011.

[4] M. Hammer, "The Getting and Keeping Of Wisdom ," Research Directorate, Public Service Commission of Canada, OttawaOctober 2002.

[5] D. Tranfield, D. Denyer, and J. Marcos, "Management and development of High Reliability Organisations," Cranfield University, Bedford, UK FOCUS 16-18-1, 2004.

[6] Y. Sheffi, *The resilient enterprise: overcoming vulnerability for competitive advantage*. Boston, MA: MIT Press, 2005.

[7] D. A. Fisher, "An emergent perspective on interoperation in systems of systems," Carnegie-Mellon University TECHNICAL REPORT CMU/SEI-2006-TR-003, 2006.

[8] D. D. Woods and E. Hollnagel, *Joint cognitive systems: patterns in cognitive systems engineering*. Basingstoke: Taylor & Francis, 2006.

[9] A. deMeyer, C. H. Loch, and M. T. Pich, "Managing project uncertainty: from variation to chaos," *Sloan Management Review,* vol. 43, pp. 60-67, 2002.

[10] C. E. Siemieniuch and M. A. Sinclair, "Using corporate governance to enhance `long-term situation awareness' and assist in the avoidance of organisation-induced disasters," *Applied Ergonomics,* vol. 39, pp. 229-240, 2008.

[11] M. A. Sinclair, "Ergonomics issues in future systems," *Ergonomics,* vol. 50, pp. 1957 - 1986, December 2007.

[12] N. N. Taleb, *The black swan: the impact of the highly improbable*. London: Penguin, 2008.

[13] J. E. Gover, "Analysis of US semiconductor collaboration.," *IEEE Transactions on Engineering Management,* vol. 40, pp. 104-113, 1993.

[14] A. B. Cherns, "Principles of socio-technical design revisited," *Human Relations,* vol. 40, pp. 153-162, 1987.

[15] C. W. Clegg, "Socio-technical principles for system design," *Applied Ergonomics,* vol. 31, pp. 463-477, 2000.

[16] T. Wafler, G. Grote, A. Windischer, and C. Ryser, "KOMPASS: a method for complementary system design," in *Handbook of cognitive task design*, E. Hollnagel, Ed., ed Mahwah NJ: Lawrence Erlbaum, 2003.

[17] C. E. Siemieniuch and M. A. Sinclair, "Implications of the supply chain for role definitions in concurrent engineering," *International Journal of Human Factors and Ergonomics in Manufacturing,* vol. 10, pp. 251-272, 2000.

[18] A. M. Ross, D. H. Rhodes, and D. E. Hastings, "Defining changeability: reconciling flexibility, adaptability, scalability, modifiability, and robustness for maintaining system lifecycle value," 2008.

[19] C. J. Roberts, M. G. Richards, A. M. Ross, D. H. Rhodes, and D. E. Hastings, "Scenario planning in dynamic multi-attribute tradespace exploration," presented at the SysCon2009 – IEEE International Systems Conference, Vancouver, BC, 2009.

[20] M. Kaâniche, P. Lollini, A. Bondavalli, and K. Kanoun, "Modeling the resilience of large and evolving systems," *International Journal of Performability Engineering,* vol. 4, pp. 153-168, April 2008.

[21] D. Daly, D. D. Deavours, J. M. Doyle, P. G. Webster, and W. H. Sanders, "Möbius: an extensible tool for performance and dependability modeling," in *Lecture Notes in Computer Science*. vol. 11th International Conference, TOOLS 2000, B. R. Haverkort, H. C. Bohnenkamp, and C. U. Smith, Eds., ed Schaumnurg, IL, 2000, pp. 332-336.

[22] A. Bondavalli, M. D. Cin, D. Latella, I. Majzik, A. Pataricza, and G. Savoia, "Dependability analysis in the early phases of UML based system design," *Int. Journal of Computer Systems - Science & Engineering,* vol. 16, pp. 265-275, 2001.

[23] H. Consortium, "HIDENETS - HIghly DEpendable ip-based NETworks and Services," IST-FP6-STREP- 26979 D1.3, 2006.

[24] J. F. Elder and D. W. Abbott, "A Comparison of Leading Data Mining Tools," presented at the Fourth International Conference on Knowledge Discovery & Data Mining, New York, NY, 1998.

[25] E. Guldentops, G. Hardy, J. Heschl, and R. Stroud, "Aligning COBIT, ITIL and ISO 17799 for Business Benefit," Office of Government Commerce, Norwich, UK2000.

[26] R. Shah, *Social networking for business - choosing the right tools and resources to fit your needs*: Wharton School Publishing, NJ, 2010.

[27] A. J. Daw, "Keynote: On the wicked problem of defence acquisition," presented at the 7th AIAA Aviation Technology, Integration and Operations Conference: Challenges in Systems Engineering for Advanced Technology Programmes, Belfast, N.I., 2007.

[28] H. W. J. Rittel and M. M. Webber, "Dilemmas in a general theory of planning," *Policy Sciences,* vol. 4, pp. 155-169, 1973.

[29] K. J. Vicente, C. M. Burns, and W. S. Pawlak, "Muddling through wicked design problems," *Ergonomics in Design,* vol. 5, pp. 25-30, 1997.

[30] J. Conklin, *Dialogue mapping: building shared understanding of wicked problems*. New York: J. Wiley & Sons, 2006.

[31] E.-M. Hubbard, C. E. Siemieniuch, M. A. Sinclair, and A. Hodgson, "Working towards a holistic organisational systems model" presented at the 5th Int. Conf. Systems of Systems Engineering (SoSE), Loughborough, UK, 2010.

1.

Astronomic Sun Tracker Performance and Solar Energy Collection Comparison for Different Italian Sites

Fossa M.
Dime, University of Genova, Via Opera Pia 15a, Genova, Italy
marco.fossa@unige.it

De Domenico C.
Dime, University of Genova, Via Opera Pia 15a, Genova, Italy
c.dedo@email.it

Abstract - *In many Mediterranean Countries the level of solar insolation on horizontal varies from 4 to 5.5 MJ/(m day), as annual average. When the beam fraction of the solar energy is predominant, tracking the solar disk is a possible strategy to increase the energy yield in PV and thermal solar systems. This paper is addressed to the comparison (in terms of collectable energy) among feedback, double axis, trackers and a particular single axis one, driven by astronomic information. The comparison is made in terms of hourly simulations, based on TMY solar data. The reference tracker is assumed to be able to minimize the surface zenith angle and even to orient the collectors horizontally under cloudy conditions. The results show that the differences (reference vs astronomic) in yearly collectable energy are within few percent, thus demonstrating that non concentrating solar systems can be successfully applied with simple, astronomic tracking devices.*

Keywords: solar energy, tracking systems, hourly simulations.

1 Introduction

Solar energy exploitation is gaining more and more attention due to a number of reasons related to primary energy demand, costs, availability, environmental issues. In Europe, furthermore, the 2020 objectives force the member Countries to address their efforts to the green technologies and strategies like the renewable resources utilization, energy saving interventions (especially devoted to buildings), enhancement of power plant efficiency.

Solar energy is a widespread available renewable resource. As it is well known, the solar energy can be converted in different ways, from low enthalpy thermal applications, to photovoltaic electricity production, passing trough the high enthalpy thermodynamic conversion into mechanical energy.

In many Mediterranean Countries the level of solar insolation on horizontal varies from 4 to 5.5 MJ/(m day), as annual average. Global solar irradiation is mainly related to latitude, even if sites at the same latitude can show relevant differences in available beam insolation, due to the peculiar meteorological conditions. When the beam fraction of the solar energy is predominant, tracking the solar disk is a possible strategy to increase the energy yield in PV and thermal solar systems, either in case of concentration system (heliostats, parabolic mirrors, compound parabolic collectors, Fresnel lenses and mirrors) or non concentrating panels.

Concentrating the light at high suns, involves the need of high aiming accuracies, in order to focus on the target within the acceptance angles. These angles, as it is well known, become smaller and smaller, as the concentration ratio increases. On the other hand, non concentrating, tracking systems have weaker requirements in terms of tracking accuracy.

There are a great number of different tracking systems, classifiable in terms of orientation and number of axes of rotation [1]. Many of them, even those employed in non concentrating solar systems, are based on double axis mechanisms, with feedback controls devoted to correct the surface position and even to check for the presence of the solar disk, whose presence depends on the cloud coverage. One possible type of driving the system makes use of a light sensing device: a system of sensors detects the brightest point of the sky, transmitting the data to the control that properly actuates servomechanisms to point the surface. In case of overcast sky, the (feedback) tracker can choose to orient the solar panel (PV or thermal) horizontally, in order to collect the maximum diffuse radiation from the sky dome: Kelly and Gibson [2] show that during cloudy periods a horizontal surface orientation increases the solar energy capture by nearly 50% compared to a surface pointed directly toward the sun disk.

A different type of driving system is the so-called chronological (or astronomic) guide: there are no light sensors or feedback circuits and the servo mechanisms work on the basis of an algorithm which is able to predict the position of the sun in the sky in any moment of the year. As it is well known, the path of the sun can also be described with good approximation periodic on an annual basis. Moreover, whatever the considered site and the time measurement (real solar or conventional time), the sun will always be within a well assigned and fairly narrow area in the sky dome, so that, losing part of the tracking accuracy, its motion can be tracked through the use of very simple systems. Based on the above considerations, it is possible to conceive and build a chronological tracker whose inaccuracies (in aiming the sun disk) are more than acceptable for non concentrating solar systems. These systems can be based on mechanisms (cams and shafts) whose kinematics is appropriate for the task of maximizing the yearly collected energy while dramatically limiting the use of electronics and hence reducing the cost of the installation.

This paper is addressed to the comparison (in terms of collectable energy) among feedback, double axis, trackers and a particular single axis one, driven by astronomic information system, stored on board of the tracking device.

The comparison is made in terms of hourly simulations, based on TMY solar data. The reference, double axis tracker is assumed to be able to minimize the surface zenith angle in all the circumstances and even to orient the collectors horizontally under cloudy conditions. The astronomic (single axis) tracker under consideration is driven by a peculiar but simple kinematics and it “knows” the position of the solar disk during the year, but is not able to detect its presence. Furthermore the astronomic tracker, due to its simplified tracking mechanism, is characterized by inaccuracies in aiming the sun disk.

To compare the solar trackers performances is necessary to know the distribution of diffuse irradiance from sky dome, but in most cases only global irradiance on horizontal surface is known, so the accuracy of the tilted irradiance predictions degrades significantly. In 2008 Gueymard demonstrated [3] that this is mainly conditioned by the local performance of the direct/diffuse separation method; in this paper beam an diffuse irradiation are estimated from horizontal global irradiation using the Perez irradiance model, since it has been shown [4] that this is one of the best performing model. (*Soga*, 1999).

On these basis, the analysis of both tracking systems is performed taking into consideration different models for sky radiance, starting from the simplest model of uniform sky proposed by Liu and Jordan in 1960 [5]. Tilted surface insolation is estimated from beam and diffuse irradiation also using two different anisotropic sky radiance models: the Perez diffuse irradiance model [6] [7] and the HDKR (Hay, Davies, Klucher, Reindl) model [8].

The results show that the differences (reference vs astronomic) in yearly collectable energy are within few percent, thus demonstrating that non concentrating solar systems, driven for pointing the sun, can be successfully applied with simple, single axis tracking devices.

2 Theoretical background

The aim of a tracking system for solar purposes is to orient a solar panel in order to capture the highest fraction of radiant energy. Concerning beam radiation, the efficiency of a solar collector increases as the sun rays angle of incidence on the irradiated surface is reduced. This is expressed by the equation:

$$H_{b,t} = H_{b,n} \cdot \cos(\theta_w) \quad (1)$$

where $H_{b,t}$ and $H_{b,n}$ are the beam insolation expressed as J / (m^2 day) on the tilted and on the normal surfaces and θ_w is the incidence angle, respectively.

The tracker goal is hence to try to keep the sun rays as much as possible perpendicular to the surface of the solar panel, which in turns has to be displaced in order to trace as well as possible the sun path in the sky during the day. The reflected insolation (due to the surrounding surfaces, e.g. the ground) on a tilted surface can be expressed as:

$$H_{r,t} = H_{g,h} \cdot F_r \cdot \rho \quad (2)$$

where $H_{g,h}$ is the global insolation on the horizontal surface, ρ of the surrounding surfaces and F_r is the isotropic reflected view factor. In this paper ρ is assumed to be equal to 0.2.

To estimate the insolation on a tilted surface must also be taken into account the diffuse radiance and its distribution in the sky. The simplest description is the uniform sky model proposed by Liu and Jordan, often used as a reference. Eq. (3) describes the relationship between diffuse radiation on horizontal and tilted surfaces as:

$$H_{d,t} = H_{d,h} \cdot F_d \quad (3)$$

where $H_{d,h}$ is the diffuse insolation on the horizontal surface, and F_d is the isotropic diffuse view factor.

This isotropic model is too inaccurate and conservative as it underestimates the diffuse insolation. More accurate sky radiance models are composed of three parts, that are the isotropic from sky dome, the circumsolar and the horizon brightening. In the model proposed by Perez et al. the expression that combines these three components of the diffuse radiation is:

$$H_{d,t} = H_{d,h} \cdot (1 - F_1) F_d + H_{d,h} \cdot F_1 \frac{a}{b} + H_{d,h} \cdot F_2 \sin(\beta) \quad (4)$$

where F_1 and F_2 are coefficients expressing the degree of circumsolar anisotropy and the anisotropy at the horizon respectively. The terms a and b are finally defined as:

$a = max(0, cos\theta_w)$ and $b = max(0.087, cos\theta_z)$

where θ_z is the zenith angle. The coefficients F_1 and F_2 represent the brightness of the sky and they are function of parameters that describe the sky conditions. Statistically derived coefficients that can be found on tabulated data.

Since sometimes it is difficult to evaluate the F_1 and F_2 terms correctly, a number of more simple sky models have been proposed in literature. Among those models, the HDKR one has the following expression:

$$H_{d,t} = H_{d,h}\{(1-A_i)\cdot F_d[1+f(\sin\frac{\beta}{2})^3]+A_i R_b\} \quad (5)$$

where β is the tilt angle of the surface, A_i an anisotropy index function of the transmittance of the atmosphere for beam radiation and the factor f accounts for cloudiness.

In the following, the above models (i.e. Perez and HDKR) are employed to infer the hourly values of diffuse insolation, in order to carry out a performance analysis between the two different tracking systems, a reference double axis feedback tracker and an astronomic single axis tracker.

The investigation is referred to the global energy captured in one year, and it is based on TMY solar hourly data, where the TMY diffuse radiation on horizontal $H_{d,h}$ is converted into diffuse radiation on tilted surface by means of the Perez and HDKR models.

3 Present tracker model and analysis

The reference solar tracker is assumed to be a perfect system capable to exactly point the solar panel surface toward the center of the sun disk and to orient it horizontally in cloudy days in order to collect the 100% of the available solar energy under any situation.

On the other hand, the kinematics of the astronomical tracker is supposed to be affected by a number of pointing errors, as well as the inability to detect the presence (and absence) of the sun disk, which is in principle its main drawback.

The inaccuracies of the astronomic driving system are due to:

- the law of azimuthal motion which does not change every day of the year (maximum error on azimuthal angle: 19°),
- the law of daily tilt motion which does not change every day of the year (maximum error on tilt angle: 5°),
- the simplified law of seasonal tilt motion (maximum error on tilt angle: 3°),
- the effects of the Equation of Time and other minor variables neglected (maximum error on azimuthal and tilt angles: 5°),
- the inability to orient the solar panel horizontally in cloudy days (variable penalization according to different sky conditions).

All the above inaccuracies affect the collection of both beam and diffuse radiation, except for the last one that only involves the diffuse radiation.

Once derived the laws of motion of the astronomical tracker (not described here for the sake of brevity), the procedure that has been implemented employs TMY solar data related to a number of selected sites to calculate the hourly insolation values according to two sky models.

The TMY hourly data employs the Perez irradiance model for generating direct and diffuse data. Other parameter available in TMY series are: solar angles; beam, diffuse and global insolation on the horizontal surface; beam and global insolation on the normal surface.

4 Results and discussion

The mathematical model that allows the motion of the astronomical tracker to be described, is based on the assumption that the mechanical connections can be based on the use of a proper cam, whose shape and position at ground is able to reproduce the trajectory of the celestial equator. Since the celestial equator is crossed by the sun twice a year, at the equinox days,the cam has been conceived in order to track the solar altitude angle in those two particular days (Figure 1).

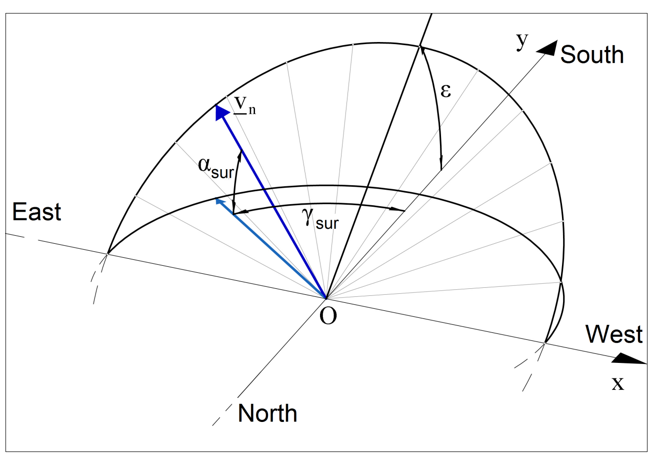


Figure 1. Synthetic sketch describing the laws of motion of the present astronomic tracker

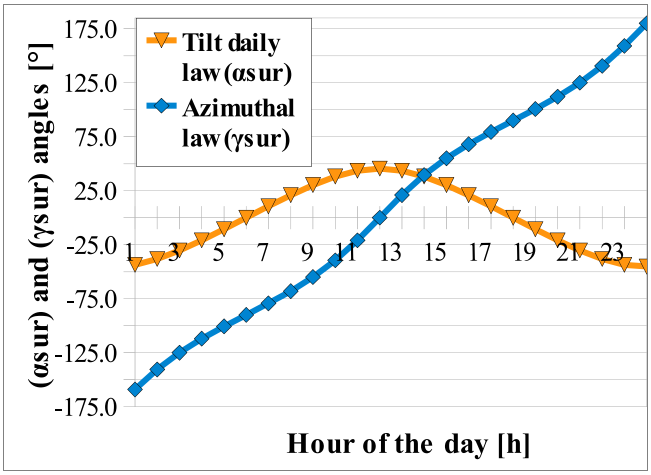


Figure 2. Synthetic description of the laws of motion of the proposed tracking systems

The selected cam has in particular a plane circumferential shape and it is tilted with respect to the horizontal plane by the angle ε. The two planes intersect at the diameter of the circle, creating a line from east to west. The angle ε is equal to the co-latitude of the site. The radius of the circle represents the unit vector v_n, that sweeps the space (Figure 1) as it is moving with constant angular velocity of 15°/h (following the hour angle of the sun). The inclination of the vector is $\alpha_{sur} = 90° - \beta$, while the angle that its projection on the horizontal plane forms with the y axis is the surface azimuth angle γ_{sur}. The cam is fixed (the seasonal variations of tilt can be managed by another mechanical device) and hence the law of the daily tilt variation and the azimuthal movement are always the same during the whole year.

The laws of motion of the proposed tracking systems, in terms of α_{sur} and γ_{sur} angles are graphically described, in a simplified version, in Figure 2: it shows that the azimuthal and tilt movements of the solar panel are both contained within the corresponding sun path curves during solstices (June and December).

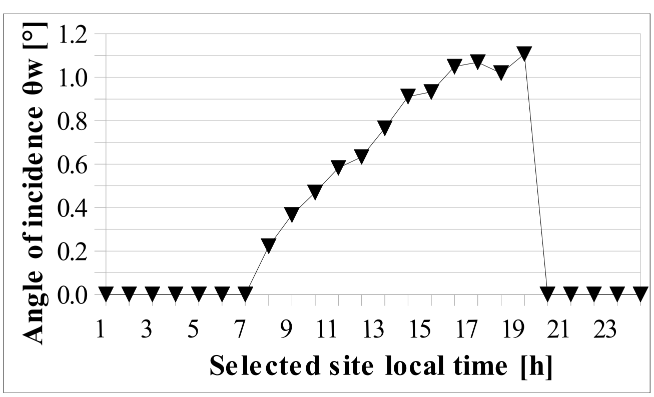


Figure 3. Accuracy of the astronomical tracker in aiming the sun disk during the spring equinox (selected site is Genova, Italy)

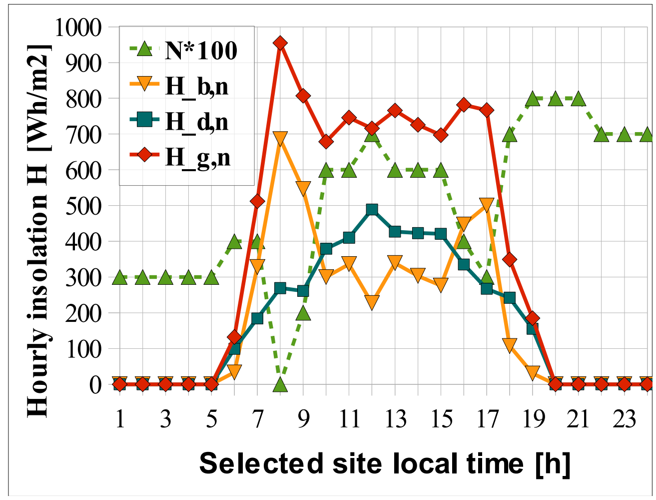


Figure 4. TMY hourly insolation data during a variable cloudy day

The law of motion of α_{sur}, and then the tilt angle β, is composed by two parts: the daily tilt law, which is invariant in time, and the tilt seasonal law. For the sake of brevity Figure 2 shows the daily tilt law of motion, but not its seasonal adjustments (which is not managed by the cam). To represent the seasonal adjustment of tilt on the graph, the relative curve should be moved up or down depending on the day of the year: in this way the tilt function keeps the same shape depicted in Figure 2.

Since the cam has been designed with reference to the celestial equator, the astronomical tracker follows the sun disk with its best accuracy during the equinox days. This is shown in Figure 3 (where the selected site is here Genova, Italy), where it can be observed that the angle of incidence θ_w assumes very small values during the daylight hours. Worth noticing from equation (1) the beam insolation collected by the solar panel is directly in proportional to $\cos\theta_w$: so for very small values of the angle of incidence, its cosine is nearly equal to 1 and the system collects almost all the available beam energy.

Figure 4 refers to a surface normal to the sun rays in a highly variable partially cloudy day. The Figure show the profiles of diffuse, beam and global insolation as well as the cloudiness index N, which ranges between 0 (completely clear sky) and 8 (fully overcast sky).

Under the cloudy conditions depicted in Figure 4, the diffuse insolation $H_{d,n}$ reaches its maximum values, sometimes even higher than the beam insolation. The latter is never nil due to the presence of bright clouds at sun disk position, that reflect and diffuse toward the ground the sun radiation without screening it completely. Usually in these cases the insolation on a normal surface is greater than the one on the horizontal plane.

During days like the one described in Figure 4, the electronics that drives a feedback tracker could be affected by pointing errors due to the continuous changes in intensity and direction of the radiation. This is not the case in the present reference solar tracker, which is assumed to be always able to exactly detect the brightest point of the sky and to consequently orient the solar panel accordingly.

Figure 5 presents in abscissa the months of the year and in ordinate the ratio of the diffuse (and reflected) insolation collected by the "astronomical" solar panel with respect to that collected by the reference solar tracker.

The Figure also shows the effects of considering two different sky models, namely the Perez and HDKR ones.

Among the limits of the proposed single-axis tracker, the one that mostly affects the diffuse radiation collection is lack of possibility to place the solar panel horizontally on overcast days. This reason is the explanation of the lower values of the diffuse radiation ratio in Figure 5 during the winter months, when the probability of overcast days during the month is higher. Furthermore, the graph shows that the difference between the two systems clearly follows the trend of the angle of solar declination.

This second effect can be ascribed to the fact during the winter the average tilt of the surface is high. In these conditions the astronomic driven system works (during the cloudy hours) in the worst conditions (non collected energy up to 5-7%) with respect to the perfect tracker, which move the collector parallel to the horizontal plane. Conversely, in July or June the difference in collected diffuse energy is reduced to few percents. The overall effect, since there is much more insolation during the summer than during the winter, the yearly loss of diffuse radiation of the astronomic system with respect to the reference tracker is contained between 2.3% and 3.9%, depending on the sky model considered.

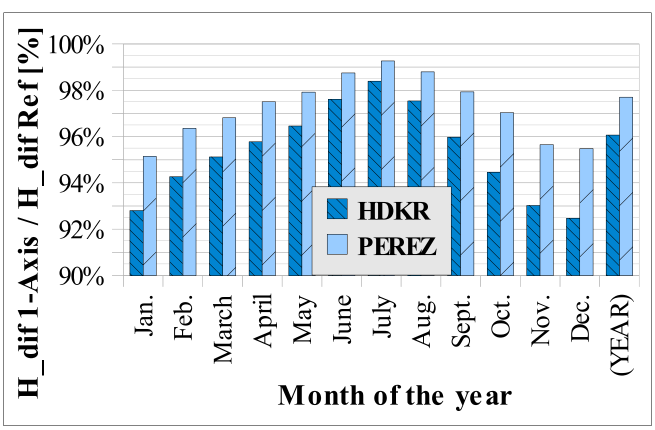


Figure 5. Diffuse insolation collection during the year according to Perez and HDKR sky models. Astronomic tracker is here compared to a perfect feedback tracker

These above figures referred to the location of Genova, Italy would reduce once a different location, having a higher H_b / H_d ratio on an annual basis, were taken into account.

Figure 6 is similar to Figure 4 but it refers to the global radiation: the performance of the astronomic tracker are compared to those of the reference double axis feedback tracker, which, in addition to exploiting the horizontal positioning for the collection of diffuse radiation, has no pointing error ($\theta_w = 0$) for the collection of direct radiation.

The percentage difference referred to global radiation is on average lower than the corresponding values referred to diffuse radiation, because the astronomic tracker is able to collect the direct radiation with good accuracy, and direct sunlight is the major fraction of the global insolation.

As already discussed, the present single-axis tracker is very accurate close to the equinoxes and much less accurate at the solstices, a feature that is described by the oscillatory trend of the H_{glob} profiles of Figure 6. Figure 6 introduces a new parameter which is the site location. Figure 6 in particular shows the results related to the HDKR model in terms of collected global energy in different sites in Italy. These sites are namely Genova (latitude 44.45°), Rome (latitude 41.88°) and Siracusa (latitude 37.07°), each of them characterised by proper TMY series.

Although the absolute values of solar energy were very different in the set of locations here presented, the relative difference between the two tracking systems is comparable. The disadvantage of the tracker Astronomic tracker compared to the reference one is reduced in those localities that have a dry climate and long sunshine days, typical features of southern areas of Italy

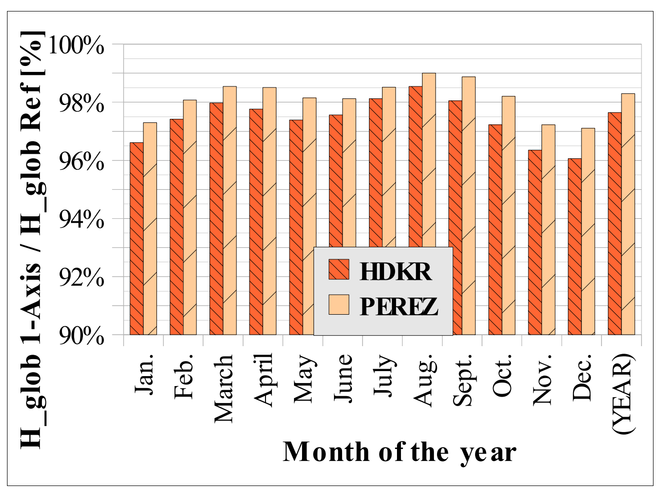


Figure 6. Global insolation collection during the year according to Perez and HDKR sky models. Astronomic tracker is here compared to a perfect feedback tracker

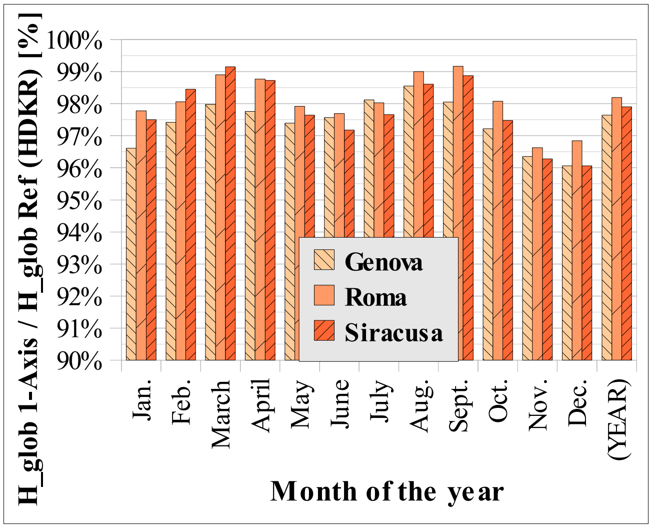


Figure 7. Global insolation collection during the year according to Perez sky model. Astronomic tracker is here compared to a perfect feedback tracker in different locations in Italy

5 Conclusions

In this paper a comparison in terms of collectable solar energy among has been made with reference to two tracking strategies. The reference tracker is a "perfect" feedback tracker able to minimize the surface zenith angle and even to orient the solar collector horizontally during overcast days. The second tracker system, proposed in this paper, is a particular single axis one, driven by astronomic information system, whose mechanics is based on the use of a simple cam rod. The performance comparison is made in terms of hourly simulations, based on TMY solar data and two different diffuse radiation models have been considered.

The results shows that the HDKR and Perez sky models give almost the same results, at least for the present investigation purposes. Regarding the performance of the proposed astronomic tracker, due its intrinsic design, has its best performance (in pointing the sun disk) at equinoxes, while especially close to the winter solstice its behaviour is some 4-5% far from the performance (global insolation collection) of the reference tracker and some 7% if only the diffuse radiation is considered.

A comparison referred to different Italian sites (north, centre and south Italy), revealed that an astronomic tracker like the one here presented is able to collect 97-98% of the global solar energy with respect to a perfect tracking systems, thus demonstrating that non concentrating solar systems, driven for just pointing the sun, can be successfully applied with simple, single axis tracking devices.

References

[1] Mousazadeh H., Keyhani A., Javadi A., Mobli H., Abrinia K. and Sharifi A., 2009, A review of principle and sun-tracking methods for maximizing solar systems output, Renewable and Sustainable Energy Reviews, vol. 13, issue 8, pp. 1800-1818

[2] Kelly N. A., Gibson T. L., 2009, Improved photovoltaic energy output for cloudy conditions with a solar tracking system, Solar Energy, vol 83, pp. 2092-2102.

[3] Gueymard C. A., 2008, Direct and indirect uncertainties in the prediction of tilted irradiance for solar engineering applications, Solar Energy, vol 83, pp. 432-444.

[4] Soga K., Akasaka H. and Nimiya H., 1999, A comparison of methods to estimate hourly total irradiation on tilted surfaces from hourly global irradiation on a horizontal surface, Journal of Architecture, Planning and Environmental Engineering, vol. 519, pp. 31-38

[5] Liu B. Y. H. and Jordan R. C., 1960, The Interrelationship and Characteristic Distribution of Direct, Diffuse and Total Solar Radiation, Solar Energy, 4.

[6] Perez R., Seals R., Ineichen P., Stewartt R. e Menicucci D., 1987, A new simplified version of the Perez diffuse irradiance model for tilted surfaces, Solar Energy, 39/3: 221-231.

[7] Perez R., Stewartt R., Arbogast C., Seals R. e Scott J., 1985, An anisotropic hourly diffuse radiation model for slopimg surfaces: description, performace validation, site dependency evaluation, Solar Energy, 36/6: 481-497.

[8] Duffie J. e Beckman W., 2006, *Solar Engineering of Thermal Processes*, John Wiley & Sons, New York.

The Viewband Concept: Introducing Life-cycle Modeling in Enterprise Architectural Frameworks

Daniele Gianni
Freelance Researcher
The Netherlands
daniele.gianni@incose.org

Andrea D'Ambrogio
Dept. of Enterprise Engineering
University of Rome TorVergata
Rome, Italy
dambro@uniroma2.it

Abstract - *Enterprise architecture (EA) frameworks have been successfully used to support enterprise engineering activities by providing graphical and machine-processable constructs to represent enterprise systems. However, existing EA frameworks lack modeling approaches to represent the increasing complexity and coupling of modern enterprises, as these frameworks do not distinguish architectural blocks supporting the different enterprise lifecycle processes, such as operation, maintenance and governance. In this paper, we propose the Viewband concept, which can be used to explicitly introduce lifecycle modeling, thus overcoming the limitations of existing EA frameworks. We also present an overlay methodology to implement the Viewband concept by use of standard modeling structures in existing EA frameworks, thus gaining compliance with currently available technologies and documentation standards. We show the application of the Viewband concept, and the related implementation methodology, through a simplified example enterprise model in UPDM (Unified Profile for DoDAF and MODAF).*

Keywords: Life-cycle, Modeling, UML, UPDM, Framework.

1 Introduction

Enterprise modeling is central to most of the enterprise engineering activities as it provides the foundation upon which technical and strategic architectural decisions can be evaluated. Enterprise modeling is generally supported by Enterprise Architecture (EA) frameworks (e.g., *UPDM* [1][2], MoDAF [3][4], DoDAF [5] or ESAAF [6]) which provide graphical and technical means to guide the representation of diverse facets characterizing an enterprise. In particular, EA frameworks are based on the concepts of *View* and *Viewpoint*, according to which enterprise projections can be consistently represented. Each projection specifically concerns a pre-identified type of issue and links to other projections through shared model blocks, providing a multidimensional enterprise representation. However, these concepts are not fully suitable to represent modern enterprises which present an increasingly complexity and interconnectivity with other enterprises, when participating in Systems of Systems (SoS) configurations for the provision of new services [7]. These configurations often rely on interactions spanning various enterprise life-cycle processes (e.g. operation, maintenance and governance) of the composing enterprises. Existing frameworks do not provide modelling structures to identify the specific life-cycle processes to which a model block pertains. As a consequence, these EA frameworks cannot be used to produce enterprise models that can offer a higher degree of accuracy, fully representing the interdependency characteristics of the enterprise life-cycle processes across the enterprises in a SoS configuration. For example, *UPDM* (Unified Profile for DoDAF and MODAF) does not explicit distinguish the type of interactions occurring in the context of an operational scenario. Thus, the information related to the enterprises life-cycle processes is not represented, causing the respective lack of information concerning the enterprise interdependencies, such as interactions sustaining the enterprise operation rather than the enterprise maintenance.

In this paper, we address the problem of distinguishing life-cycle processes by introducing the *Viewband* concept. Exploiting the analogy with the visible domain, we note that the perception capability of a wavelength (color) can be used, along with *Viewpoint* and *View*, to represent the life-cycle processes by distinguishing which model blocks and which relationships support an identified process. The *Viewband* concept can be used to enrich enterprise models with such information and to provide the opportunity of a more accurate and faithful enterprise representation. In addition, we show how existing EA frameworks can incorporate the *Viewband* concept using two simple overlay methodologies, avoiding any structural modification while leveraging on the availability of standards technologies and documentation.

The paper is organized as follows. The background section recalls the main concepts underlying an enterprise architectural model. The *Viewband* Concept section introduces the methodology to distinguish the life-cycle processes within an enterprise model. The Concept Implementation section presents three possible implementation strategies, discussing also a preliminary evaluation of the strategy impact and effectiveness on existing EA frameworks. The Example section presents an application of the methodology to a simplified enterprise model in *UPDM*.

2 Background

The *Viewband* concept and the demonstrative example are based on the following architectural concepts and EA frameworks.

2.1 Architectural Concepts

IEEE defines architecture as "*the fundamental organization of a system embodied in its components, their relationships to each other, and to the environment, and the principles guiding its design and evolution*" [8]. The standard also specifies that an architecture can be represented by one or more architectural descriptions, depending on specific rationales. In turn, an architectural description addresses the needs of an identified set of stakeholders and consists of *Views* and *Viewpoints*.

The standard defines *View* as "*a representation of a whole system from the perspective of a related set of concerns*". This representation conforms to a *Viewpoint*, i.e., "*a specification of the conventions for constructing and using a view. A pattern or template from which to develop individual views by establishing the purposes and audience [i.e., stakeholders] for a view and the techniques for its creation and analysis.*"

When addressing physical systems, a figurative analogy is also often used to better communicate and help to internalize the concepts of *View* and *Viewpoint*. For example, a common analogy is to represent the system as an object that can be observed by a human, as shown in **Error! Reference source not found.** [9]. In this analogy, a *Viewpoint* can be defined as the perspective from which the object is observed. Differently, the *View* can be imagined what is actually perceived.

2.2 Enterprise Architectural Frameworks

The definition and maintenance of enterprise architecture models can be a complex task for the inherent complexity in terms of enterprise scope, number of socio-technical concepts, and number and variety of stakeholders. Enterprise architectural frameworks have therefore been introduced to provide guidelines and contextual system modeling schemas to support the creation and maintenance of enterprise architectural models. Although EA frameworks share common purposes and the above introduced concepts, these frameworks have often been tailored to the specific characteristics of the enterprise domains to further facilitating framework users to represents architectures of their interest. Several frameworks have been introduced, each addressing a specific domain and community. For example, EA framework TOGAF is of general purpose and can be applied to many business contexts [10]. Differently, the ESA-AF is for European Space-based SoS [6], while DoDAF and MoDAF primarily concern the military domain. DoDAF and MODAF have become popular and their wide adoption has motivated the definition of *UPDM*, a standard interoperable UML-based profile for the exchange of DoDAF and MODAF models in XMI coding format. With the objective of reusing and integrating with available standards, *UPDM* also encapsulates UML/SySML specifications [11] along with the SoaML profile [12]. In addition, *UPDM* is organized into logical units, avoiding circular dependencies and offering several levels of compliances. Currently, *UPDM* offers two levels of compliance, Level 0 and Level 1.

UPDM Level 0 consists of three sub-profiles: Core, DoDAF and MODAF. Core profile contains the definition of elements shared by DoDAF and MODAF. Differently, DoDAF and MODAF profiles contain the definition of the respective non-shared elements. In this paper, we are interested mainly in the generic characteristics of *UPDM*/MODAF, and therefore we omit the description of *UPDM* Level 1 and of DoDAF specific aspects. The Core consists in turn of a set of architectural *Viewpoints*, each implemented as an UML profile. Specifically, such profiles are: All—for descriptive/metadata views on the EA model; Acquisition—for views on related to the acquisition of existing and procured systems; External types—for elements surrounding the archicture; Operational (OV)—for views on the functional architecture; Service (SV)—for views on service definition and consumption; Strategic—for views on high level capabilities, enterprise goal and mission; System—for views on the physical architecture; Technical—for views on standard and technologies used in the enterprise architecture. For each of these *Viewpoints*, *UPDM* also defines a set of *Views*. For example, Operational Viewpoint defines: High Level Operational Concept Graphic View (OV-1), Operational Node Relationships Description View (OV-2) and Operational Activity Model (OV-5). OV-1 provides an overall view of the scenario in which the enterprise operates. The scenario is mainly identified by the graphical or model description of the involved roles or actors. The interactions among the actors are detailed in OV-2, which formally defines the interaction directions and the interaction content. The content can be of several types: Information Element—to indicate message exchange; Resource Artifact—to indicate physical object exchange; Movement of People—to indicate the exchange of humans; and Capability Configuration—to indicate the combined exchange of Resource Artifact and Movement of People.

3 Viewband Concept

The *View* and *Viewpoint* concepts cannot be used to distinguish the interleaving roles of architectural elements and interactions that characterize enterprise operation, maintenance and governance chains of enterprises in SoS configurations. To overcome this limitation, we introduce the *Viewband* concept, an abstraction primitive that can be used to support the distinction of elements and interactions roles. In line with the above description of *View* and

Viewpoint concepts through the above visual metaphor, we exploit the visual analogy noting that:

- An actor's eye can distinguish the color of an object
- The color is often composed by a mixture of primitive colors
- An actor can filter colors using special devices, such as sun glasses or an intense external light of a complementary color.

Underlying these observations, there is the fact that the human eye and our brain can distinguish colors, basing on the perceived wave length that is refracted by objects (1). Using a prism, the white light can be separated in beams of different colors, which combination can produce any other color (2). In addition, using sunglasses or other glasses filtering the electromagnetic waves, we can alter our perception of the colors (3)—sometimes to stress the differences among objects or to minimize the visual impact of objects of given colors.

Similarly, an enterprise model is an object that is characterized by a set of related blocks, each representing an enterprise element participating in one or more life-cycle processes, i.e., operation, governance or maintenance. Aside from the structure in *View* and *Viewpoints*, the blocks need to be "colored" according to their roles in the enterprise life-cycle. This is particularly important to describe and identify interconnections among life-cycle processes in SoS configurations, contributing to highlight the critical operation-maintenance-governance chains sustaining the provisioning of new services. For this reason, we introduce the architectural concept of *Viewband* as:

"a distinguishing characteristic that can be used to indicate, and subsequently identify, the enterprise life-cycle in which an architectural element is participating".

Applying this definition, enterprise modelers will be able to distinguish the role of a model block by "coloring" the block according to their role in an enterprise life-cycle. As a result, the blocks can be filtered to visualize only those supporting each life-cycle process. In addition, this coloring will enable enterprise architects to visually identify blocks which will be needed in the various life-cycle processes of operation, maintenance, and governance, in particular those blocks that are part of independent enterprises. This distinction is also an important point for the definition of mathematical models for SoS performance models, as failures in the maintenance or governance chains might indirectly affect the operation performance.

The *Viewband* concept adds additional information to the enterprise model blocks, and a non-intrusive implementation is essential to promote the application of this concept within the standard technologies of existing EA frameworks, specifically the *UPDM* one.

3.1 Concept Implementation

The *Viewband* concept introduces information concerning the enterprise operation type in which an architectural building block is participating. Using existing EA frameworks, this extra piece of information can be represented in several ways, from block packaging to block coloring, both physically (i.e., using a distinguished color palette), and semantically (i.e., introducing attributes). In any case, it is essential the impact that the concept implementation may produce on the conformance to the *UPDM* standard. A possible and compliant implementation is based on the combined use of UML packages, block physical coloring and block attribute marking.

Package structuring

The model can be structured to implicitly convey information related to the model blocks stored within pre-designed packages. Before detailing these considerations, we implicitly assumed that any architectural diagrams use the following *storing criterion*: diagrams are consistently stored in UML packages conforming to the convention <containingPackage>/<viewpointName>/<view>, where *Viewpoint* and *View* uniquely identify the diagram type. For example, at root level, instances of the OV-2 view are stored within the packages Operational and OV-2. Assuming the diagram is named "Example_OV-2", the diagram would be uniquely identified by the relative URL: Operational/OV-2/Example_OV-2.

Next, we have formulated the *root criterion* for the package structure of the enterprise model:

"at root level, the enterprise model shall contain any view diagrams of general interest to the entire architecture (i.e., not specifically related to any enterprise life-cycle) as well as the newly created packages: architecturalBlocks, operation, maintenance, governance, mixed".

Concerning the *Viewpoints*, basing on the above definitions, All and Strategic *Viewpoints* can be identified as not referring to any specific life-cycle process. The All viewpoint is for an overarching and life-cycle independent summary representation of the enterprise model. Differently, Strategic concerns the general motivations for the enterprise, including its capabilities, mission and goals, and must be similarly stored at root level. Concerning individual views, High Level Operational Concept Graphic View (OV-1) is the only one that can be identified as specifying the scope of the entire architectural model, and thus not addressing a specific life-cycle process.

Once created an enterprise model, the root structure should display as follows:

- All, containing instances of the respective *Views*
- ArchitecturalBlocks, containing model blocks composing any diagram
- Governance, containing only instances of *Views* on enterprise governance

- Maintenance, containing only instances of *Views* on enterprise maintenance
- Mixed, containing only instances of views involving two or more enterprise life-cycle processes
- Operation, containing only instances of *Views* on enterprise maintenance
- Operational, containing instances of OV-1
- Strategic, containing instances of the respective *Views*.

ArchitecturalBlocks can be further structured in Governance, Maintenance, Mixed, and Operation sub-packages, though this sub-organization can be seen only as further model consistency constraint rather than part of the *Viewband* concept implementation. It is important to remark that views instances are stored in the Governance, Maintenance, Mixed, and Operation packages according to the storing criterion above defined.

Although the Model Structure is the first step in distinguish elements involved in the different life-cycle processes, the structure itself does not cover all the possible cases. For example, in the Mixed package, views instances might contains model blocks interacting in different life-cycle processes, for example to graphically highlight how the enterprise operation is affected by the maintenance. This type of disambiguation is offered by the coloring criterion.

Physical Coloring

Physical color coding is the most affine method with the analogy of the visual domain above introduced. The method consists in defining a color coding for elements in the different processes, and subsequently applying these to the blocks in an architectural model. In this way, enterprise architects can visually discern the elements supporting operations, maintenance and governance. In theory, the method can also consider the mixture of two or more colors, if a model block participates respectively in two or more life-cycle processes in a view instance, respectively.

This solution is certainly non-intrusive on existing EA standards as it does not require any change to the underlying metamodels. However, this solution leaves the visual ambiguity of the color graduation and prevents from a systematic and standard computer-based approach to the processing of viewband information. Indeed, the diagramming information is often tool specific and no warrantee is generally provided on the format compliance. For this reason too, it is important to restrict the use this method only for human processing of the enterprise model, in particular for those view instances in the Mixed package. In this way, other internal color coding will remain unaffected by the concept implementation.

A more formal approach, with less visual impact, is offered by the attribute marking method.

Attribute Marking

The attribute marking aims to overcome the limitations of the physical coloring, while offering a machine-processable identification of the *Viewband* concept. This method consists in marking model interactions with one or more Boolean attributes, one per each of the life-cycle processes in which the interaction is occurring. The attribute presence, and its value set to true, indicate that the element participated in the respective life-cycle process. For example, a Needline can be associated to the attribute governance, which must be set to true, to indicate that the Needline is part of enterprise model concerning the governance. In our analysis, we observed that a model block can participate in different scenarios covering diverse life-cycle processes. While the block structural composition remains uniform over the possible block roles, the interactions with other model blocks may change, depending also on the contextual scenario and conditions. Considering the possible multiple roles of a model block, the roles played in a pre-chosen scenario can only be determined through the interactions taking place in the scenario. Basing on these observations, we have identified the interactions as the model block to be enriched with the *Viewband* information. The block role will remain derivable mechanically by observing the incoming and outgoing interaction in a pre-chosen scenario.

Below, we show how the above methods can be used to implement the *Viewband* concept for a simplified enterprise model.

4 Example

We consider a simple example concerning a National postal enterprise for the delivery of envelops over the national territory. For conciseness of presentation, we only present the views concerning the High Level Operational Description (OV-1), Operational Node Relationship Description (OV-2), and Operational Activity Model (OV-5).

Error! Reference source not found. shows the OV-1 view of the overall scenario, defining the enterprise operation as consisting of two customers, two local branches, and a national hub for the regional mail dispatching.

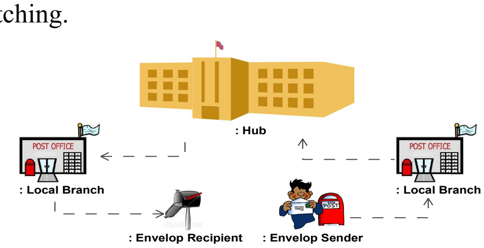


Figure 1 High Level Operational Scenario (OV-1) for a National Postal Enterprise

In the operation scenario, we assume that customers must visit the local post office for posting a mail. In addition, post vans are used to deliver mails between the hub and any of the local branches. Finally, riding a scooter, a postman delivers mails to the local customers. Aside from the operation process, we identify a simplified set of maintenance and governance processes. Concerning the enterprise maintenance, we assume that the vans and the scooters need occasional servicing for mechanical repairs and for refueling. For conciseness of discussion, we limit the representation of the enterprise governance to the payments for the servicing and refueling.

Using the *Viewband* concept, we can model the enterprise distinguishing among operation, maintenance and governance. Using the above package method, the *Viewband* concept can be implemented by structuring the model packages as shown in Figure 2.

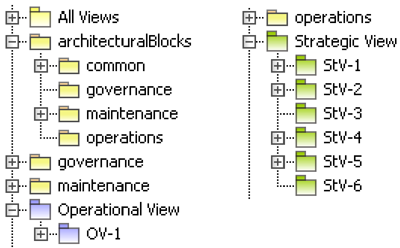


Figure 2 Package Structure

Before proceeding with the presentation of the OV-2 diagrams, the exchanged data and objects need to be identified. From the above scenario description, we have identified the entities: “Envelops”, “Envelop”, “PostVan”, “PostScooter”, “Invoice”, and “Money”. “Envelops” is a “Resource Artifact” grouping two or more “Envelop” entities, both “Resource Artifacts”. “PostVan” is a “Capability Configuration” consisting of a “Driver” and a “Van”. “PostScooter” consists of a “Postman” driving a “Scooter”. “Invoice” and “Money” are “Resource Artifacts”.

Figure 3 shows the OV-2 diagram supporting the enterprise operation. The diagram can read easily and congruently with the above definition of the operation scenario.

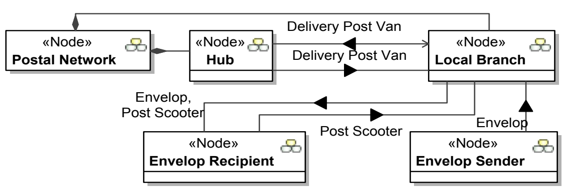


Figure 3 Operation OV-2 View

Figure 4 shows the OV-5 diagram defining the “Local Branch” activities for the enterprise operation. These activities are also related to OV-2 in Figure 3, linking the object exchanges to the consuming and producing activities of each node, specifically “Local Branch” in this case. For example, “Envelops Collection from Customers” is the consuming activities for the incoming “Envelop” exchange with “Envelop Sender”. Similarly, “Envelop Dispatching to Headquarters” is the producing activity of the “Delivery Post Van” from the “Local Branch” to the “Hub”. Analogous considerations hold also for the “Hub”, “Envelop Recipient” and “Envelop Sender”.

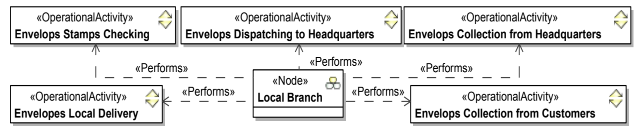


Figure 4 Operation OV-5 for Local Branch

From the above description of the operational scenario, we can infer the external nodes, interactions, and activities that support the enterprise maintenance. In particular, Figure 5 shows the OV-2 diagram supporting the enterprise maintenance. This process involves the “Hub” and “Local Branch” nodes, which constitute the enterprise core elements implementing the operations, and the two external nodes “Garage” and “Petrol Station”. These nodes provide servicing for mechanical repairs and for refueling, respectively.

Figure 6 shows the OV-5 for defining “Hub” and “Local Branch”’s operational activities that support the enterprise maintenance. The enterprise routinely performs these activities, which will also be interleaved with operation and governance activities. “Check Fuel Level” is a periodic test activity that the van drivers and the postman will perform before any journey. “Perform Vehicle User Test” and “Perform Scooter User Test” are respectively the mechanical check that drivers and postmen will perform periodically, once a month, or upon failure to start the respective engines.

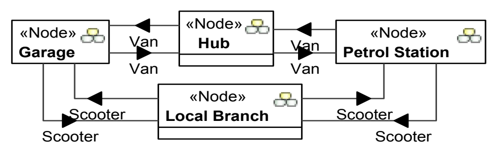


Figure 5 Maintenance OV-2

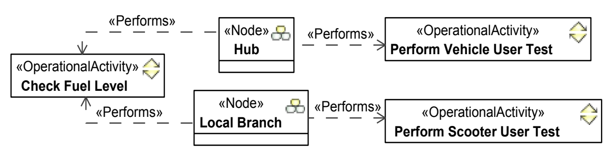


Figure 6 Maintenance OV-5 for Local Branch and Hub

From the above description of the operational scenario, we can infer that external nodes, interactions, and activities support the enterprise governance. Figure 7 shows the OV-2 diagram supporting the enterprise maintenance. This process involves the exchange of the servicing invoices from the “Garage” and “Petrol Station” to the “Hub” and “Local Branch”. The process also includes the payment of the respective invoice, in the opposite direction.

For completeness, Figure 8 shows the OV-5 for defining “Hub” and “Local Branch”’s operational activities that support the enterprise governance. Analogously to the above OV-5s, these activities are linked to the production and consumptions of the Governance OV-2 exchange objects. For example, “Collect and Archive Invoice” consumes the “Repair Invoice” coming from the “Garage”.

Similar observation holds for the “Refueling Invoice from the Petrol Station”.

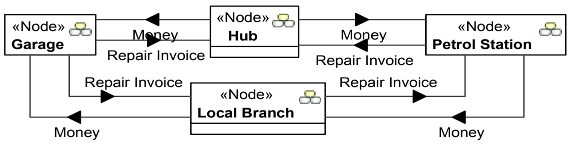


Figure 7 Governance OV-2

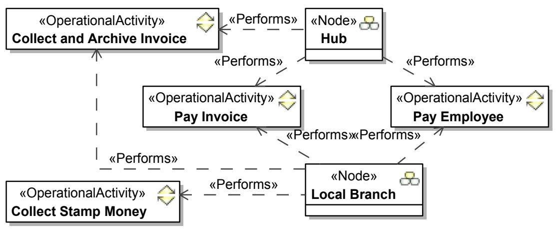


Figure 8 Governance OV-5 for Local Branch and Hub

With the objective of supporting an engineering process from a system thinking perspective, we use the physical coloring method to derive a combined OV-2 diagram illustrating the interleaving of the enterprise operation, maintenance and governance interactions. The resulting OV-2 is shown in Figure 9, where red indicates the operation, blue indicates the maintenance and green indicates the governance.

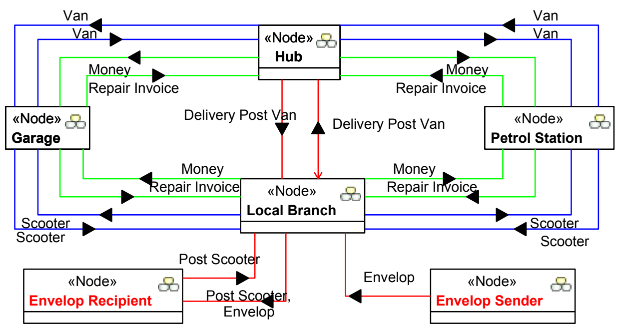


Figure 9 Combined Operation-Maintenance-Governance OV-2

In particular, using the *Viewband* concept and the above introduced approach, it is possible to highlight the maintenance and governance interactions that sustain the enterprise operation of mail collection, dispatching and delivery. In addition, browsing the enterprise model from this diagram, more punctual information can be inferred, concerning the enterprises integrations across these life-cycle processes.

5 Conclusions

Enterprises are increasingly interconnected and enterprise models need to accurately represent the interleaving interactions of enterprise elements participating in different life-cycle processes, such as operation, maintenance and governance. However, existing enterprise architectural frameworks, such as *UPDM*, lack modeling structures to distinguish these processes and the roles played by architectural elements. As consequence, enterprises models cannot explicitly represents these aspects and potentially fail in supporting the enterprise engineering. Exploiting the analogy with the visual domain, we have introduced the *Viewband* concept, which can be used to distinguish architectural elements supporting different life-cycle processes. We have also presented an overlay implementation approach on standard architectural frameworks, such as *UPDM*. The method consists in three main parts: package structuring, physical coloring and attribute marking. In addition, we have shown an example application of the *Viewband* concept and of the respective implementation approach, using a simplified example scenario concerning an enterprise for the mail distribution over the national territory.

References

[1] Hause, M., “The Unified Profile for DoDAF/MODAF (UPDM), enabling systems of systems on many levels”, *Proceedings of the 4th Annual IEEE Systems Conference*, 2010.

[2] OMG, UPDM Specification 11-05-01, http://www.omg.org/spec/UPDM/1.1/PDF/, last access January 2012.

[3] MoD: UK Ministry of Defence Architectural Framework (MODAF) - detailed guidance, http://goo.gl/HntQj (last access June 2012).

[4] Bailey, I.: Brief Introduction to MODAF with v1.2 Updates. IET Seminar on Enterprise Architecture Framework, London (2008), 1-18.

[5] US DoD, DoDAF V2.0 Volume 1: Introduction, Overview, and Concepts - Manager’s Guide, http://goo.gl/bm6bM (last access June 2012).

[6] Gianni, D., Lindman, N., Fuchs, J., Suzic, R., “Introducing the European Space Agency Architectural Framework for Space-based Systems of Systems Engineering,” In: Hammami, O., Krob, D., Voirin, J.-L. (eds.): *Proceedings of the 2nd International Conference on Compelx Systems Design & Management (CSDM 2011)*. Springer Verlag, Paris (2011) 335-346.

[7] Jamshidi, M.: Systems of Systems Engineering: Innovation for the 21st Century. Wiley (2009).

[8] IEEE Std 1471 Recommended Practice for Architectural Description of Software-Intensive Systems. IEEE (2011).

[9] Hilliard, R.: All About IEEE Std 1471. (2007).

[10] The Open Group, Welcome to TOGAF Version 9 -- The Open Group Architecture Framework (2011).

[11] Friedenthal, S., Moore, A., Steiner, R.: A Practical Guide to SysML: The Systems Modeling Language. Morgan Kaufmann (2008).

[12] OMG: Service oriented architecture (SOA) Modeling Language (SoaML), http://www.omg.org/spec/SoaML/ (last access January 2012).

Human Performance Modeling in System of Systems Analytics

Craig R. Lawton
Distinguished Member of Technical Staff
Sandia National Laboratories
Albuquerque, NM, USA
crlawto@sandia.gov

John H. Gauthier
Principal Member of Technical Staff
Sandia National Laboratories
Albuquerque, NM, U.S.A.
jhgauth@sandia.gov

Abstract - *The Department of Defense has identified that integrating the human element into large scale System of Systems (SoS) models is a significant challenge that remains unaddressed. Failure in doing so leads to significant limitations in our SoS analytical capabilities as human performance is a large contributor to the performance of a SoS. The primary challenge is that, in most SoS domains, the problems being analyzed are large in scale. Conversely, most Human Performance Modeling (HPM) initiatives look at integrating detailed cognitive models that capture fine grained details of human perception, decision making, and response with detailed systems models and simulations (e.g., Lebiere et al., 2003). It is not feasible to integrate such fine grained cognitive models with systems models and perform SoS scale analysis. This paper documents a capability that integrates HPM into a large scale SoS simulation toolset and demonstrates the utility of the toolset.*

Keywords: Systems Engineering, System of systems, modeling and simulation, Human Performance Modeling.

1 Introduction

As early as 1960, an Air Force report stated that humans contributed to 20-53% of system failures (Shapero et al., 1960). The military has subsequently identified Human Performance Modeling (HPM) as a significant requirement and challenge as can be seen in the Department of Defense's (DoD) Defense Modeling and Simulation Office's (DMSO) Master Plan (DoD 5000.59-P 1995). In the plan, the DMSO identified the capability to robustly represent individuals and group behaviors as a critical need. The DMSO officially recognized human behavior representation as a prominent technology challenge. Specifically, the challenge is to develop and integrate accurate sub-models of human behavior in a manner that accounts for human perceptual, cognitive, and motor output in the task being modeled. To this goal, the military is currently engaged in programs devoted to HPM in various military contexts. Examples include the Human Performance Modeling Integration (HPMI) Program within the Air Force Research Laboratory focused on integrating HPMs with constructive models of systems (e.g. cockpit simulations) and the Navy's Human Performance Center (HPC) established in September 2003.

The primary challenge is that, in most SoS domains, the problems being analyzed are large in scale. Most HPM initiatives look at integrating detailed cognitive models that capture fine grained details of human perception, decision making, and response with detailed systems models and simulations (e.g., Lebiere et al., 2003). It is not feasible to integrate such fine grained cognitive models with systems models and perform SoS scale analysis.

Sandia National Laboratories (SNL) has been in the forefront of modeling System of Systems (SoS) with the SoS Analysis Toolset (SoSAT). SoSAT modeling, and most SoS modeling efforts to date, have focused on military hardware systems, analyzing performance attributes such as mobility, lethality, availability, etc.—the human element has been ignored. However, humans are the driving force behind all military SoS operations. Sustained military operations produce deficits in soldiers' cognitive/physical performance abilities that can result in inefficiencies and errors, negatively affecting overall SoS performance. Omitting them from an analysis might cause the largest performance factor of the SoS to be missed and produce results that can be misleadingly optimistic.

This paper presents an approach to integrating HPM into a large scale SoS simulation and modeling environment and a demonstration of how this capability can be applied.

2 Overview of SoSAT

In the pursuit of modeling and analyzing complex system of systems (SoS) capabilities,Sandia National Laboratories developed a multi-system time simulation capability called System of Systems Analysis Toolset (SoSAT).

SoSAT development was driven by the need to support Military modernization decisions and particularly the performance of Brigade Combat Teams (BCTs);

however SoSAT can be applied to various systems of systems problems. A significant engineering and integration challenge exists in the Army's modernization programs which consist of a families of manned and unmanned platform systems, connected by a common network, enabling a modern modular force, providing our soldiers and leaders with leading edge technologies and capabilities allowing them to dominate in complex environments.

SoSAT is a time-step stochastic simulation tool designed to model and simulate the multi-echelon operation and support activities projected to be conducted by future Brigade Combat Teams. It provides logistics analysts with the ability to define operational and support environments and ascertain measures of its performance effectiveness based on multiple trials. SoSAT characterizes sensitivity changes to all platforms, support systems, processes and decision rules as well as vehicle reliability and maintainability (R&M) characteristics. It is designed to be a robust decision-support tool for evaluating the readiness and sustainment of the modernized BCTs to include fuel, water, ammunition and maintenance operations. SoSAT can also take into account external conditions (e.g., storms or extreme terrain) and combat damage. Simulation output results assist the user in identifying platform, as well as SoS level performance and logistics support issues.

Key to the multi-system simulation capability has been the development of a State Model Object (SMO) that enables a system, its elements, and its functionality to be encapsulated for use in the simulation. Every system in the simulation is represented by an SMO which has a defined composition of items that help define the system's functionality. SMOs can represent air vehicles, ground vehicles, manufacturing equipment, etc. The systems are the central objects of the model and are the entities that march through the simulation.

The basic structure for modeling a system as an SMO in SoSAT is as follows. A system performs functions (e.g., mobility, communications, sensing, lethality, etc.). Functions are supported by elements of the system, including primary elements (engine, instrumentation, sensors, etc.) and consumables (fuel, ammo, etc.). Elements can fail through normal reliability processes, external conditions (combat damage, external elements—e.g. severe weather, hilly terrain, etc.), and the failure of other systems where there are dependencies (e.g., logistics). Failure of an element affects system function. Failure of a function can affect other systems and system availability.

Implementation of probability-based HPMs in SoSAT can occur in two ways: (1) a human can be defined as a system, and (2) a human can be defined as an element of a system. The first way is applicable to humans that are not assigned to a single system, and whose activities might have influence over a number of systems, e.g., a platoon leader. The second way is applicable to humans that are assigned to or do not have direct effects on a single system, e.g., a truck driver. The focus of this paper will be on implementation of the second type of HPM.

3 Human Performance Modeling

In SoSAT we base our representation of Human Performance in part on probability based human performance. Swain (1963) laid out the basic concept of probability based human performance (referred to as Human Reliability Analysis in Swain's context). Briefly, probability based performance modeling is based on the ideas that a human job can be divided into individual tasks, failure of successful completion of an individual task can be quantified by a probability (called a Human Error Probability or HEP), and failure of successful completion of the job can be determined by combining the probabilities (often with the assumption that the tasks are independent).

In SoSAT, parameters are set for expected error rates and times to recover from errors (called time-to-reset) for humans under nominal conditions. For example, a repairman will make a mistake a certain fraction of the time when replacing an alternator or performing any repair, so that the time it takes to perform the repair will be the nominal time to do the job plus the time to recover from the mistake. In addition, SoSAT allows for the variability of performance – due to internal or external conditions - based on performance shaping factors (PSF) such as fatigue level. Performance degrades or improves when a PSF is applied (e.g., increasing or decreasing the failure rate).

SoSAT is designed around systems and their elements. The elements that affect a system can be primary elements (parts or subsystems), consumables, external elements, and reference elements (to functions of other systems). Humans have been included as a fifth element type. For humans to affect an analysis, they must be defined as an element and be assigned to a SoSAT system. For example, a driver is assigned to an infantry carrier vehicle (ICV) or a repairman is assigned to a mechanics pool.

At any point in time an element is either up or down in SoSAT. Humans are down if they are recovering from a mistake (reliability error) or if they have suffered combat damage. The input and implementation for reliability errors are described in Section 3.1 Human Error Rates. The duration of time that elapses when humans are down depends on the reason for being down.

Human mistakes and human absence adversely impact the performance of the SoS according to the duties they are assigned. In SoSAT, human involvement fits one of three categories:

- Direct impact – the performance of a driver affects the mobility of a vehicle. The driver is included as an element in the Boolean expression (union of cutsets) that describes the mobility function for the system.
- Logistics authority – the failure of an individual to authorize replacement of parts and consumables in a timely fashion can cause or extend vehicle downtime. Authorities are granted when the human is first defined.
- Repair – a mistake made by a mechanic during a repair job extends repair time for primary elements. Repair personnel are identified when humans are first defined.

3.1 Modeling Human Error Rate in SoSAT

When a human (element) is assigned to a SoSAT system type, properties include a probability distribution to describe basic error rate for the human. SoSAT assumes that in the absence of any PSFs the error rate for an individual would be constant in time. A distribution models the uncertainty and the human-to-human variability of the constant error rate. Prior to the first simulation, SoSAT samples the distribution N times, where N is the number of instances of the system, hence the number of human elements, using stratified median sampling. Each of the N humans is then assigned one of the sampled error rates.

In keeping with the treatment of parts and subsystems, SoSAT requires a time-to-failure (TTF) distribution for humans. The appropriate distribution for a constant failure rate model is the exponential distribution. Thus, the sampled error rate, λ, defines the required parameter for the density function, f, and the cumulative distribution function F,

$$f(t) = \lambda e^{-\lambda t} \quad \text{and} \quad F(t) = 1 - e^{-\lambda t} \quad \lambda > 0, t \geq 0 \qquad (3\text{-}1)$$

At each time step in a simulation SoSAT samples the TTF cumulative distribution function to determine if the human fails during that time step. The sampling is random, but conditioned on survival to the current simulation time, S. With a constant failure rate model and in the absence of PSFs the calculation is effectively unconditional. That is, it is equivalent to unconditionally sampling the exponential distribution and adding S to the result. This time-independent model would apply throughout a simulation if not for the presence of PSFs.

SoSAT distinguishes human errors as recoverable and catastrophic. The human resets after a prescribed time following a recoverable error. If the human commits a catastrophic error to a system, the system is disabled for the remainder of the simulation (nonhuman system) or until a replacement system arrives (human system).

The probability of catastrophic error is conditional. It the human fails during a time step in SoSAT the possibility of catastrophic error is checked. Specifically, if a uniformly generated random value in the interval (0, 1) is less than the catastrophic error probability, the error that the human has committed is catastrophic. If the error is not catastrophic, the error is recoverable.

Reset time is the time required to recover from a non-catastrophic mistake. The time is defined by a probability distribution when the user assigns a human element to a SoSAT system type. Each instance of the system type means another instance of that human element. Every such human element uses the same reset time distribution.

When the human makes a reliability error, the reset time distribution is randomly sampled. The sampled value is used as follows.

- If human performance directly affects performance of the system (e.g., driver), it is the time required for the human to reset. So, the human (and possibly the system) is down for that time.
- If the human has authority to order parts or consumables for systems, one of those systems requests that the human place an order, and the human is experiencing reset time, the human does not place the order until the reset time has elapsed.
- If a mechanic makes an error during a repair job, the time required to finish the repair is extended by the reset time.

3.2 Performance Shaping Factors

As mentioned in addition to basic human error rates, SoSAT applies Performance Shaping Factors (PSF) to error rates to capture time independent and time dependent factors that might differentiate the performance of specific individual humans or human types that would be part of the SoS. SoSAT models Training, Experience Level, Perceptual Skills, and Fatigue as PSFs as time independent and are defined prior to simulation of the SoS mission. The fourth PSF, fatigue, is modeled in SoSAT as time dependent.

For the time independent PSFs, the idea is that an individual with increased training, with longer experience, or with greater perceptual skills is less likely to commit an error than someone with less of these attributes. The user inputs an adjustment factor for each of the three PSFs.

SoSAT assumes that the factors are multiplicative and takes the product of the three factors.

The adjustment factors can be defined for the entire force or for human types. To calculate the final adjustment factor for a human type, SoSAT multiplies the product defined for the force by the product defined for the human type.

SoSAT divides the adjustment factor into the basic error rate for the affected humans. (e.g., PSFs that improve performance should have adjustment factors greater than 1.0 to decrease the error rate.)

The effect of Fatigue is to increase the error rate with time and SoSAT assumes that the increase is linear with time. Rather than prompt the user for the rate of increase (slope), SoSAT prompts for an error-rate-multiplier that is valid at a future point in time. We have chosen 24 hours for that point in time and the user supplies a distribution for the multiplier.

The information is used as follows. First, the sampled error rate assigned to the human is treated as an initial rate ($\lambda_0 = \lambda(0)$). Second, the fatigue multiplier distribution is sampled and the sampled value (v) is assumed to be a multiplier of the initial rate that is valid after 24 hours of operating time ($\lambda(24) = v \cdot \lambda_0$). SoSAT uses this information to define a constant slope m.

$$m = \lambda_0 (v - 1)/24 \qquad (3\text{-}2)$$

The error rate then increases linearly with time according to $\lambda(t) = \lambda_0 + m \cdot t$. If the fatigue multiplier is fixed at 1.0, then the slope m = 0.0 and the constant error rate model from Equation (3-1) applies.

For the general case (m > 0) the effects of fatigue are used to define a modified TTF distribution function

$$F(t) = 1 - e^{-(\lambda_0 + mt)t} \qquad \lambda_0 > 0, m > 0, t \geq 0 \qquad (3\text{-}3)$$

An example comparison of cumulative distribution functions (CDFs) with $\lambda_0 = 0.04$ is shown in Figure 3.1. The no fatigue case uses Equation (3-1). The fatigue case uses Equation (2-3) with m = 0.000833 derived from multiplier v = 1.5 in Equation (3-2). At any given time the cumulative probability of error to that time is greater with fatigue. Of greater interest note that time differences increase with increasing probability. For example at the median there is a 50% chance that the human has made an error by 17.4 hours with no fatigue, but with fatigue there is a 50% chance by 13.6 hours, or 3.8 hours earlier. The difference grows to 14.2 hours (40.3 – 26.1) at the 80th percentile.

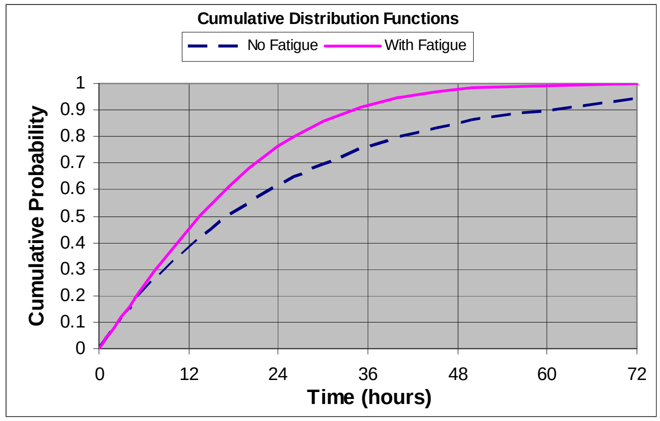


Figure 3.1. Fatigue Distribution Example.

The modified TTF distribution is referred to as the linear rate exponential in SoSAT. For m > 0, its mean (μ) and variance (σ^2) are stated in terms of the complementary error function (erfc).

$$\mu = \frac{e^{\lambda_0^2/4m}}{2}\sqrt{\frac{\pi}{m}}\ \mathrm{erfc}\left(\frac{\lambda_0}{2\sqrt{m}}\right) \quad \text{and} \quad \sigma^2 = \frac{1-(\lambda_0 + m\mu)\mu}{m} \qquad (3\text{-}4)$$

$$\text{where} \quad \mathrm{erfc}(x) = \int_x^\infty \frac{2}{\sqrt{\pi}} e^{-S^2} dS$$

For m = 0 (Equation 3-1) they are $\mu = 1/\lambda$ and $\sigma^2 = 1/\lambda^2$.

4 Example Model of HPM in a SoS

The following example will serve as the basis for demonstrating the utility of modeling HPM in a SoS. The example problem models a military combat unit that contains a company that has two platoons and three separate support vehicles. The force also has a repair facility and a consumables depot that are independent of the company. Key systems of the platoons are

- Their own water supply (CAMEL)
- Command vehicles (CV)
- Sandia all purpose vehicles (SAPV)
- Platoon leaders (Platoon Leader)
- Senior NCOs (NCO)

The company's three support vehicles are trucks for supplying fuel, ammunition, and spare parts. The repair facility has a parts depot and a pool of mechanics.

Personnel include the aforementioned platoon leaders and NCOs. Also, the mechanics pool consists of three repair personnel. These are generic mechanics that are qualified to make any kind of repair that systems may require. The only other humans modeled in this problem are the drivers of the vehicles.

Systems have primary elements that can fail and require spare parts and mechanics for their repair. Systems can also fail if they run out of consumables. Systems must access someone with authority to request that spares and consumables be ordered. The persons of authority for the systems of a platoon are their platoon leader and NCO.

All mechanics are initially assigned to the mechanics pool. They are temporarily reassigned as needed to systems that require repair. Drivers are permanently assigned to systems. Their performance directly impacts the performance of their system.

The mission for the company includes a 72-hr stretch in the field, a 24-hr break at the repair facility, and another 72-hr stretch in the field. During each period in the field there could be a 12-hr battle. The 3-day stretches in the field cause fatigue for the humans, which increases their error rates. The battle causes stress which also causes humans to make errors more frequently. In addition the probability of combat damage to humans increases with battle.

Each human is assigned a basic error rate distribution. The sampled error rate can be influenced by several performance shaping factors (PSFs). This example problem deals with fatigue. Capability of dealing with time-independent PSFs (inadequate training, inadequate experience, inadequate perceptual skills) is not exercised.

The combat unit force structure for our example is shown in Figure 4.1 There is one instance of each system type for most systems. The exceptions are SAPV (SAPV-001 and SAPV-002 in Platoon A and SAPV-003, SAPV-004, and SAPV-005 in Platoon B) and CV (CV-002 and CV-003 in Platoon B).

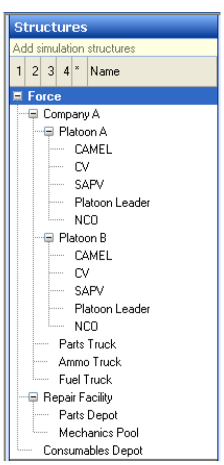


Figure 4.1. Military Unit Force Structure System Types.

Systems can endure combat damage based on a pre-specified damage rate profiles. In addition to combat systems, systems can be service providers, spare parts providers, and consumable providers. In addtion, systems are assigned scenarios that define how the systems are utilized (e.g. time on, miles traveled, etc.). SoSAT simulates the process of the systems performing their missions and stepping through time incurring failures, consuming supplies, and being resupplied.

Human elements are assigned to systems. Figure 4.2, shows a supplier driver being assigned to an ammo truck. In addition, human element parameters are displayed such as basic error rate, time to reset, and fatigue PSF settings for this human element.

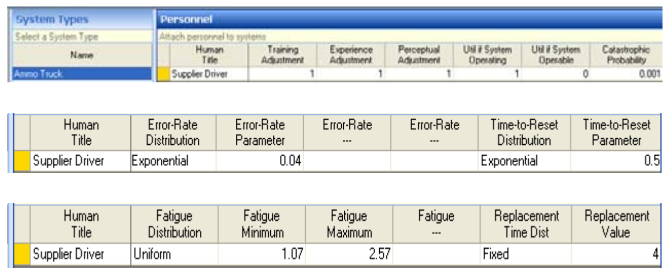


Figure 4.2. Personnel for Ammo Truck System

The problem is run for 200 trials and selected results are tabulated.

5 Results & Conclusions

The element details result, displayed in Table 5.1 shows six measures of performance: uptime, downtime, random failures, combat damage failures, total failures, and system availability for the 20 human elements Table 5.1. Only the mean results, across the 200 trials, are shown.

The drivers and supplier drivers are elements of simple cutsets for the "operability" function for their system. Thus, their mean downtime is the mean time their system is down due to their failure. These downtimes range from 1.1 hours (SAPV-002) to 20.3 hours (CAMEL-001). Judging from the ratio of random (reliability) failures to combat damage failures, the downtime primarily arises from reliability failures.

Note that the sum of up time and downtime is mission time (168 hours) for the NCOs, platoon leaders, and mechanics. This is not the case for either type of driver. The reason for this is that SoSAT discontinues time accounting for elements when a system becomes disabled. So by subtracting the sum of up time and downtime from mission time we can know the average time lost due to system disabling.

Doing the subtraction the disabled time for Ammo Truck-001 is 6.7 hours on the average. The worst system for average disabled time is CV-002 (31.6 hours). One metric of interest in military SoS is system availability. System availabilities are inherently interdependent as a combat system's availability will be dependent on a support or supply system's availability. The final column in table 5.1 include human element availability. Note that the human element availabilities capture the amount of system availability that is not accounted for if HPM were not to be modeled

System	Name	Up Time	Down Time	Random Failures	Combat Damage Failures	Total Number of Failures	Availability
Ammo Truck-001	Supplier Driver	152.3	9.0	16.1	0.03	16.1	0.944
CAMEL-001	Driver	125.0	20.3	36.0	0.05	36.1	0.863
CAMEL-002	Driver	148.5	2.8	4.7	0.04	4.7	0.982
CV-001	Driver	150.7	4.4	8.0	0.05	8.0	0.971
CV-002	Driver	120.4	16.0	29.3	0.04	29.4	0.884
CV-003	Driver	146.8	3.2	5.6	0.05	5.6	0.979
Fuel Truck-001	Supplier Driver	153.9	7.7	14.0	0.03	14.0	0.952
Mechanics Pool-001	Mechanic1	164.6	3.4	0.5	0.00	0.5	0.980
Mechanics Pool-001	Mechanic2	164.6	3.4	0.5	0.00	0.5	0.980
Mechanics Pool-001	Mechanic3	163.5	4.5	0.6	0.01	0.6	0.973
NCO-001	NCO	160.4	7.6	5.1	0.61	5.7	0.955
NCO-002	NCO	141.2	26.8	41.7	0.45	42.1	0.840
Parts Truck-001	Supplier Driver	150.1	8.7	15.6	0.01	15.7	0.946
Platoon Leader-001	Platoon Leader	160.7	7.3	5.3	0.56	5.8	0.957
Platoon Leader-002	Platoon Leader	140.8	27.2	41.3	0.49	41.8	0.838
SAPV-001	Driver	128.1	18.8	33.8	0.04	33.9	0.869
SAPV-002	Driver	154.0	1.1	1.8	0.03	1.8	0.993
SAPV-003	Driver	154.2	2.0	3.5	0.06	3.6	0.987
SAPV-004	Driver	141.8	6.8	12.4	0.04	12.4	0.954
SAPV-005	Driver	128.4	14.8	26.6	0.04	26.6	0.896

Table 5.1. Element Details for Humans.

Results compiling system availabilities with and without the inclusion of HPM in the SoS model have been tabulated. Figure 2.27 shows the performance of the force in terms of operational availability. The impact of human performance on system performance is not evident here, so we modify the input and make a second run without the influence of humans.

Functional Availability Rollup by Structure

System	Operability Mean	5th	95th
Force	0.857	0.744	0.944
Company A	0.828	0.691	0.933
Platoon A	0.827	0.652	0.991
CAMEL	0.857	0.261	1.000
CV	0.727	0.102	1.000
SAPV	0.689	0.285	0.986
Platoon Leader	1.000	1.000	1.000
NCO	1.000	1.000	1.000
Platoon B	0.809	0.650	0.966
CAMEL	0.859	0.173	1.000
CV	0.750	0.317	0.995
SAPV	0.705	0.393	0.972
Platoon Leader	1.000	1.000	1.000
NCO	1.000	1.000	1.000
Parts Truck	0.843	0.153	0.999
Ammo Truck	0.845	0.205	0.993
Fuel Truck	0.948	0.075	1.000
Repair Facility	1.000	1.000	1.000
Parts Depot	1.000	1.000	1.000
Mechanics Pool	1.000	1.000	1.000
Consumables Depot	1.000	1.000	1.000

A

Functional Availability Rollup by Structure

System	Operability Mean	5th	95th
Force	0.800	0.671	0.899
Company A	0.759	0.606	0.878
Platoon A	0.782	0.607	0.909
CAMEL	0.687	0.102	0.864
CV	0.755	0.306	0.966
SAPV	0.689	0.315	0.900
Platoon Leader	0.952	0.082	0.991
NCO	0.949	0.872	0.995
Platoon B	0.725	0.556	0.879
CAMEL	0.831	0.209	0.988
CV	0.659	0.371	0.806
SAPV	0.684	0.415	0.899
Platoon Leader	0.813	0.744	0.887
NCO	0.816	0.744	0.894
Parts Truck	0.824	0.292	0.939
Ammo Truck	0.843	0.547	0.935
Fuel Truck	0.773	0.256	0.954
Repair Facility	1.000	1.000	1.000
Parts Depot	1.000	1.000	1.000
Mechanics Pool	1.000	1.000	1.000
Consumables Depot	1.000	1.000	1.000

B

Figure 5.1. System Availability with (A) and without (B) HPM.

Figure 5.1. compares availability when humans are included versus not included. There are instances, however, where availability improves with the addition of humans. This is an artifact of the statistics. Presumably if more trials were run, this behavior would not appear.

To facilitate comparison at the command level we copied each grid into a spreadsheet and generated the plot shown in Figure 5.2. The decrease in availability due to the actions of humans is evident for the force, the company, and both platoons.

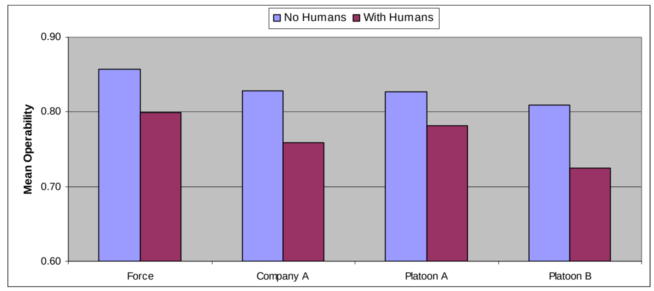


Figure 5.2. System Availability Comparison.

It is clear that including HPM in SoS analysis is quite important as human performance is a significant contributor to SoS performance. The capability described and demonstrated in this paper constitutes a significant advance in SoS analysis and opens up SOS analysis to areas not heretofore possible.

6 References

[1] Shapero, A., T.I. Cooper, M. Rappaport, K.H. Schaefy, C. Bates Jr., *Human Engineering Testing and Malfunction Data Collection in Weapon system Test Programs*, WADD-TR-60, Wright-Patterson AFB, Ohio, 1960

[2] Swain, A.D., *A Method for Performing a Human-Factors Reliability Analysis*, SCR-685, Sandia Corp., August 1963.

[3] Lebiere, C., R. Archer, D. Schunk, E. Biefeld, T. Kelly, and L. Allender, Using an integrated task network model with a cognitive architecture to assess the impact of technological aids on pilot performance, *Proceedings of the 2003 Conference on Human Performance Modeling of Approach and Landing with Augmented Displays*, NASA/CP-2003-212267, 165–187, 2003

[4] Miller, D.P., and C.R. Lawton, *Human Performance Modeling for System of Systems Analytics: Combat Performance-Shaping Factors*, SAND2006-7911, Sandia National Laboratories, Albuquerque, NM, 2007

Efficient Systems Analysis by combining SysML and Coevolution

David Morgan, Antony Waldock
BAE Systems Advanced Technology Centre
Filton, Bristol, BS34 7 QW, UK
david.morgan11, antony.waldock@baesystems.com

David Corne
School of MACS, Heriot-Watt University,
Edinburgh, UK
david.corne@macs.hw.ac.uk

Abstract – *The engineering of a complex and large scale system with hundreds of competing requirements is a time consuming and costly process. In recent years, Model Based Systems Engineering has been adopted as a means of moving from a document-centric approach to a model based approach where reusable models can be used to analyse the proposed system. The application of multi-objective optimisation algorithms to generate a set of designs that represent the trade-off between competing system requirements would be highly desirable. In this paper, the authors apply different optimisation strategies, inspired by coevolution, to efficiently generate a set of solutions by identifying and exploiting the structure within the design. The preliminary results show that using the structure inherent in a SysML design has significant benefits in terms of the number of evaluations needed to generate the solutions.*

Keywords: Model Based Systems Engineering, Multi-Objective Optimisation, SysML, Coevolution.

1 Introduction

Typically, the engineering of a complex system is decomposed into a set of subproblems, which are then developed, in relative isolation, by a set of specialised teams (training, logistics, engineering costs, performance requirements, maintenance etc.). Each team produces design documentation using an appropriate toolset (Matlab, Excel, finite element simulation etc.) to model and understand the trade-offs between a subset of the derived system requirements. Understanding the trade-off between competing requirements across the entire system, by collating and reviewing the individual models or design documentation, can be a time consuming and error prone process.

The first step in understanding the trade-off between competing requirements at a systems level is the ability to integrate the different views from the specialised teams. Model Based Systems Engineering (MBSE) [1] is the application of modelling to support the development and later life cycle phases of a system using a model centric approach rather than a document centric approach to reduce system design time [2]. The model represents selected aspects of the structure, behaviour, operation or other characteristics of a system. The goal is to develop a single model that can be viewed and manipulated by a diverse range of teams enabling a common understanding, consistency and eventually reuse. In recent years, SysML has become the standard modelling language, developed by the Object Modelling Group (OMG), to support MBSE.

SysML is an extension of the Unified Modelling Lanuage (UML) [3] and in particular, provides two new views or diagrams, Requirements and Parametric, with which to support systems engineering. A Requirements Diagram (RD) visually represents system requirements and the interconnections between them. The Parametric Diagram (PD) represents the constraints between different system components. The PD can be populated from a library of parametric models that represent the performance of different system components and can be written in a range of specialised toolsets.

As well as graphically representing the system, toolsets such as MagicDraw and IBM Rhapsody support analysis of the constraints by instantiating the graphical models. For example, ParaMagic is capable of calculating unknown system parameters given input parameters and constraint models using Matlab and Mathematica. Some toolsets are capable of performing limited trade-off studies by evaluating a range of input parameters and plotting the resultant output parameters. The development of an approach that enables a systems engineer to automatically view and explore the designs that represent the trade-off between competing system requirements would be highly desirable. Although for complex and large systems with hundreds of requirements, performing a black-box multi-objective optimisation is likely to be extremely time-consuming to yield a reasonable set of designs.

Fortunately, the SysML language means that the multi-objective optimisation does not have to be a completely blind black-box optimisation. The structure inherent in the SysML representation gives the multi-objective optimisation algorithm some insight into how the problem is structured and therefore could enable the algorithm to be automatically adapted to exploit these properties. In this paper, the authors present a preliminary study into automated methods of efficiently optimising competing system requirements across an entire system by combining a Model Based System Engineering (MBSE) approach with multi-objective optimisation algorithms based on coevolution.

Section 2 of the paper firstly introduces a case study problem to evaluate the performance of the different optimisation approaches and shows the types of structure that can be identified and exploited. Section 3 then outlines the different approaches that could be applied to the multi-objective optimisation algorithms. Section 4 presents the experimental setup and results from this preliminary study with conclusions and future work presented in Section 5.

2 Case Study

In this work an Air Traffic Control (ATC) case study was used to evaluate the approach. The goal of the case study is the hypothetical design of an ATC system for U.K. airspace. The problem is to choose the optimal set of radar types and associated configurations at 17 different sites, such that the system as a whole is able to provide sufficient coverage for minimal operating costs. The approach taken in this paper can be applied to any system developed within SysML and this case study was selected because of the authors' familiarity with the problem and the suitability of the models used.

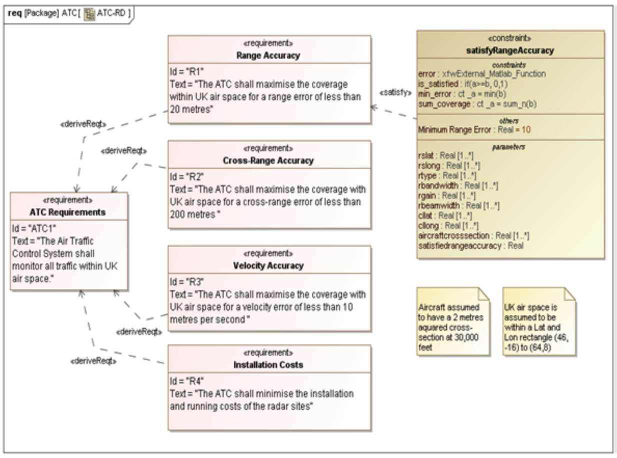


Figure 1 - ATC Requirements Diagram

The first diagram considered for the ATC case study is the RD. In this case study, the RD (shown in *Figure 1*) shows four derived requirements; three focusing on the tracking performance of the radar system and one on the costs associated with upgrading and operating the radars. Each of the derived requirements has an associated constraint that verifies an instance of the system based on the requirement. The second diagram considered for the ATC case study is the BDD which shows the structure of the ATC system (see *Figure 2*). The ATC System consists of two parts; the set of radar sites and the set of coverage locations that the system is required to cover. Associated with each radar site is a spatial location and a corresponding radar with the parameters; type, gain, bandwidth and beamwidth. The set of coverage locations has a position and the radar cross section of the size of aircraft that should be detected. The ATC Analysis block has a number of constraints associated that represent the derived requirements seen in the RD. The third diagram is the PD, which glues the blocks in the BDD diagram with the constraints. The PD is shown in *Figure 5* and shows a mapping between the component parameters for each component and the constraint: range accuracy. The output of the constraint is the number of satisfied coverage locations i.e. the number of locations with a range error below 20m. The constraint can be represented as a single element or can be broken down into more detail as shown in *Figure 6* where the sum and min operators are displayed. Exposing the operators involved in a constraint could enable the optimisation algorithm to find the set of solutions more efficiently.

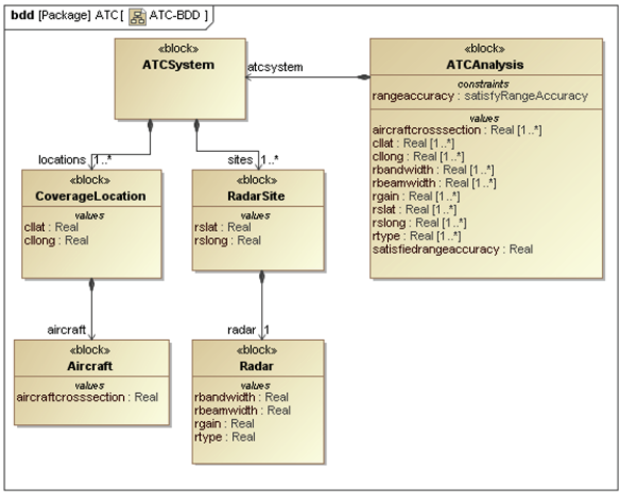


Figure 2 - Block Definition Diagram for an ATC

2.1 Constraint Functions

The constraint functions used to evaluate the requirements are split into two categories; radar performance and cost modelling. For radar performance, three constraint functions are f_1, f_2 and f_3 represent the percentage of coverage locations that have sufficiently small range error σ_1, the cross-range error σ_2 and the velocity error σ_3 as defined in [12]. In this paper, each radar station is assumed to process its data independently, and the error at a particular coverage location is taken to be the smallest error obtained by any single site. It is also assumed that the same antenna dish transmits and receives the signal. The range error is assumed to be satisfactory if σ_1 is less than 20m, the cross-range error less than 200m and the velocity error less than $10ms^{-1}$.

A fourth constraint is defined as cost, the function of which is filled by the cost metric (1) in this analysis. It is a sum over the N_i different radar sites and is composed of three terms; the installation and commissioning of a particular type of radar (C_{fix}), the costs of operating the system over a period of α years (C_{op}) with a reduction in costs when sites are of a similar type (N_s) and the final term is a configuration based cost (C_{con}) and is a function of the gain of a particular site. This reflects the fact that installing a large, high performance antenna dish will come at an increased cost.

$$f_4 = \sum_{i=1}^{N_i} C_{fix}^i(T) + \frac{\alpha}{N_s^i} C_{op}^i(T) + C_{con}^i(G,T) \quad (1)$$

In this paper, we consider five types of radar in the trade-off analysis, the first two of which represent the extremes of high performance and high cost (T1), and low performance for relatively little cost (T2). The remaining three models in general fall between these extremes (See *Table 1*). Another allowable option is to leave a radar site unoccupied, in which circumstance it is taken to have infinite error and zero cost. The goal is to select the four radar parameters to maximise coverage and minimise costs.

	T1	T2	T3	T4	T5
P (kW)	2000	500	1200	1200	1200
P_C	100	25	50	50	50
F (dB)	40	30	30	30	30
PRF (Hz)	400	1000	400	400	400
T_{scan} (s)	3	2	3	60	4
λ (cm)	10	30	15	15	15
G (dB)	27-32	15-22	23-26	23-26	22-26
θ (°)	0.8-1.2	2.0-3.0	2.6-3.5	0.1-0.5	1.4-2.2
B (MHz)	0.3-0.6	0.1-0.2	1.5-8.0	0.3-1.0	0.02-0.3
C_{fix} (£M)	100	10	35	42	30
C_{op} (£M)	1.5	0.5	1	1.25	1.5

Table 1 - Radar Types

3 Coevolution

The case study defined in the previous section represents a problem with 68 variables (4 for each of the 17 radar stations) with four objective functions (constraints that satisfy requirements). Although this represents a sufficiently complex and large scale problem for this study, the engineering of real-world systems is likely to be far more complex and include hundreds of requirements and variables. Finding the set of solutions that represent the trade-off between the competing objectives in such a high-dimensional space with many objective functions without using problem decomposition is likely to be intractable for this scale of problems. Fortunately, recent promising methods of problem decomposition have emerged from the field of multi-objective optimisation [4][5].

Problem decomposition deals with approaches to large-scale problems that involve breaking the problem down into a collection of smaller subproblems. If a problem is fully decomposable, in the sense that there is no interaction between the subproblems, then problem decomposition is clearly a simple and successful approach, since we can simply solve each problem independently and then concatenate the solutions to the subproblems and obtain a good solution to the larger problem. However, real world cases are invariably not decomposable in this way. It is always possible to decompose a larger problem into arbitrary smaller ones, however this cannot be done without significant interactions between the subproblems; this means that the concatenation of solutions to the subproblems will generally be a poor solution to the global problem. For example, if we combine the most mass-efficient wing design with the most energy-efficient engine, the wing may be structurally unstable beyond 200mph, and so forth. Problem decomposition methods therefore operate by combining independent optimization of subproblems with communication and monitoring processes, which convey information between the processes that are optimizing the subproblems, and continually monitor and steer them on the basis of the current ‘global’ solution.

Although problem decomposition is a highly generic idea, precise methods and approaches tend to be *ad hoc*, and closely tied to a particular optimisation process. For example, ‘Benders’ decomposition’ approach [6], and variants thereof, are common approaches when the problem can be stated as a mathematical programming problem, and exploits a particular way of grouping the variables and stages of optimisation within the linear or stochastic programming formulation. Meanwhile, a range of ‘divide and conquer’ approaches to problem decomposition have been used in several studies in the evolutionary computation community. For example, in [7] the authors use an evolutionary algorithm first to find suitable problem decompositions; then, a fast heuristic solves each subproblem, and a ‘patching’ heuristic combines the subproblem solutions. This approach is appropriate when suitable fast heuristics exist, including one that helps evaluate the suitability of any given decomposition. An alternative approach was introduced in each of [8][9] where, broadly speaking, a problem was randomly partitioned into subproblems that were solved independently, but where the quality of a subproblem’s solution was calculated dynamically based on combining it with a random selection of the current solutions from other subproblems. This type of approach was termed ‘co-operative coevolution’ in [10], and there has been growing recent interest following a number of recent studies [10], which have shown it to work very well across a wide range of large-scale benchmark optimization problems (e.g. including 500 and 1,000 dimensional versions of the benchmark suite from [11]).

The most recent and effective approaches to co-evolution are those discussed in [10], and the approach we describe in the next section is based on this. Essentially, coevolution works by first partitioning a problem into a group of smaller subproblems. Optimization then progresses by combining independent optimization of these subproblems with intermittent communication of information between them. A partition into subproblems is essentially a partition into groups of parameters. While an individual subproblem is being optimized, we need to continually estimate the solution quality of our current solution to this subproblem; this means that we need to assume a set of instantiations for the parameters that are *not* involved in that subproblem. So, each independent optimization of a subproblem proceeds by assuming that the current solutions to the other subproblems are fixed – the communication between subproblems takes the form of occasionally updating and resetting each subproblem’s view of the other subproblems’ latest parameter settings. Hence, at intervals, each subproblem receives a global steering signal, while in between these times they make fast progress on their own part of the problem, Recent studies, as summarized in part in , invariably show that this approach can significantly improve performance on large-scale problems, either achieving similar results to a global optimization, but much faster, or by achieving results that are both faster and better quality.

Almost all studies of co-evolution so far, however, have involved benchmark function optimization problems or, in the earlier studies such as [7][8], classical combinatorial problems such as the travelling salesperson problem or set-covering. The potential for using this approach in large scale systems engineering seems clear, and open for investigation. This paper represents an initial investigation of the approach in a systems engineering context. In particular, noting that the ability to use co-evolution relies on the ability to partition the problem into subproblems, we describe how to exploit the structure inherent within the SysML diagrams to guide the co-evolution method.

4 Experimental Study

The MOEA/D [4] Genetic Algorithm (GA) is known to perform well over a wide range of high-dimensional optimisation problems. In particular, decomposition based approaches are known to be effective when there are several objective functions. For these reasons we use the version of the algorithm described in [5]. In this study the population size is 1200, each associated with a different scalar aggregation function (subproblem). Each subproblem has an optimal solution that is located at a specific point on the Pareto front, and together these single objective problems characterise the multi-objective problem. The part of the Pareto optimal front containing non-dominated solutions with respect to the performance metrics f_1, f_2 and f_3 is shown in *Figure 3*, with the cost metric f_4 shown on the colour axis.

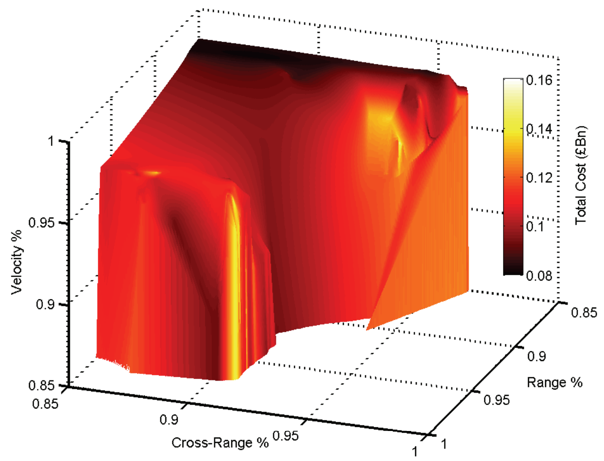


Figure 3 - The known Pareto optimal front for ATC

Better performing sets enclose larger hyper-volumes in the objective space and have darker surfaces (indicating lower cost). The Pareto set contains many different types of solution vectors. Common features were that some locations were always empty (indicating redundancy of particular sites), and that the higher performing solutions tended to contain mixtures of types 3, 4 and 5 rather than type 1 (suggesting the benefits of mixing stations with complementary strengths). The cheaper type 2 model also features in many solutions in the Pareto set. Typically these are seen later in the optimisation, i.e. high performing solutions are often found quickly, but similarly performing cheaper solutions take longer to emerge. To measure the convergence rate under several different coevolution strategies, we use the standard IGD performance metric [5]. It measures the average minimum distance between a uniformly spaced set of points on the Pareto front and its current approximation. In determining the rate of convergence, one evaluation is defined as the evaluation of all of the performance metrics for one radar site *i* at one coverage location *j*. We use a uniformly spaced grid of 256 evaluation points across the coverage area, so a single evaluation of $f_{1\text{-}4}$ for a given state vector will by default require 4352 evaluations. For all of the strategies tested, the system was found to exhibit logarithmic convergence. The convergence rate β was determined by least squares fitting to $evaluations^{-\beta}$ in each of 20 independent runs.

Firstly, three baseline optimisation approaches were tested on the problem: ‘black-box’ optimization, coevolution using random partitions and coevolution with partitioning based on information from a human designer. In the first, the GA is free to mutate every element of the solution vector. The second random approach, involves restricting mutation to only 17 indices randomly selected from the 68 in the problem. The resulting four replica populations are then evolved independently for 20 generations, after which time the best solution for each subproblem is selected. This type of stochastic approach often performs well in optimization problems with large numbers of local minima. For comparison, the third baseline approach used a ‘model’ system engineer’s knowledge of the problem to cluster the radar sites and their associated variables into four fixed coevolution groups based on geospatial location.

In this preliminary study the authors proposed two approaches that are based on utilising the structure in the SysML. The first approach stems from the BDD SysML diagram where the structure of the parameters can easily be extracted. Instead of randomly choosing 17 elements as in the Random baseline method, this approach randomly selects 10 elements from each of the four groups in the diagram (i.e. type, gain, bandwidth or beam-width) to coevolve.

The second approach developed by the authors utilises information about the operators defined in the SysML diagram. The ‘min’ operator, defined in *Figure 6*, may mean that some of the inputs have little effect upon the output. For example, it is unlikely that the radar station in Belgium will ever provide the best result for a coverage location in the Atlantic Sea. To exploit this knowledge, a probability distribution can be learnt based on the frequency with which a particular radar site gives the accepted result for a particular coverage location. In the optimization, the links to evaluate are chosen from this distribution (without replacement) at each generation until 60% are selected. The results for the fitted convergence curves of the five methods are shown in *Figure 4*.

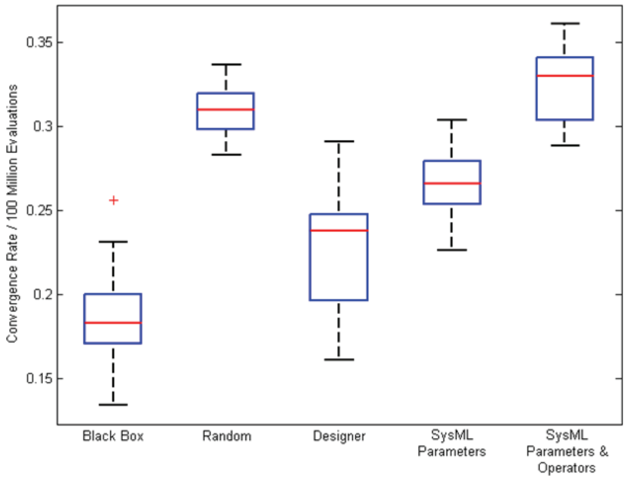


Figure 4 - Experimental results for the baseline methods (Black Box, Random, and Designer) with the SysML.

As was expected, it is possible to significantly speed up convergence by using coevolution. In particular, this problem responds very well to the random baseline approach. SysML exhibits a faster convergence rate than the 'black box' and designer but is outperformed by the coevolution with random partitioning. Introducing the link strength probability distribution significantly augments SysML, with a speedup not far short of what would be expected if all of the removed links were completely redundant. By using all of the structure available within the SysML diagrams, it is possible to improve upon the convergence rate of all the baseline approaches. In more highly correlated problems, it is likely that exploiting structure in this way would outperform the stochastic approach by a wider margin.

5 Conclusions and Future Work

The development of an approach that enables a systems engineer to automatically view and explore the designs that represent the trade-off between competing system requirements would be highly desirable. In this paper, the authors have shown that using coevolution strategies can provide an improved convergence rate over a 'black box' optimisation, where utilising the structure inherent in a SysML design can improve the convergence rate further. The ability to exploit structure within the SysML diagram will be extremely important when this approach is applied to real-world design problems with large numbers of variables and hundreds of system requirements.

Future work could be improved in three possible directions; integration, optimisation and visualisation. The methods outlined in the paper would require integration with current toolsets. Means of identifying and exploiting structure could be extended much further than the initial developments. The visualisation of the resulting optimal designs for human assessment is another interesting and related area.

References

[1] A. Wayne Wymore , "Model-Based Systems Engineering: An Introduction to the Mathematical Theory of Discrete Systems and to the Tricotyledon Theory of System Design", CRC Press, Boca Raton FL 33431, 1993 (1 800 272 7737), xiii+710 pages, ISBN 0-8493-8012-X.
[2] http://www-01.ibm.com/software/success/cssdb.nsf/CS/KJON-8SJU3L?OpenDocument&Site=default&cty=en_us
[3] OMG SysML 1.2 specification - http://www.sysml.org/docs/specs/OMGSysML-v1.2-10-06-02.pdf
[4] Q. Zhang, H. Li, "MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition", IEEE Transactions on Evolutionary Computation, Vol. 11:6 pp.712-731, 2007
[5] Q. Zhang et al., "The Performance of a New Version of MOEA/D on CEC09 Unconstrained MOP Test Instances", IEEE Congress on Evolutionary Computation, pp.203-208, 2009 (CEC)
[6] J. F. Benders, "Partitioning procedures for solving mixed-variables programming problems," *Numer. Math.* 4, 3 (Sept. 1962), pp. 238–252
[7] Christine L. Valenzuela and Antonia J. Jones. 1993. "Evolutionary divide and conquer (i): A novel genetic approach to the tsp". *Evol. Comput.* 1, 4 (December 1993), 313-333.
[8] Luis F. Gonzalez Hernandez and David Corne. 1996. "Evolutionary Divide and Conquer for the Set-Covering Problem". In *Selected Papers from AISB Workshop on Evolutionary Computing*, Terence C. Fogarty (Ed.). Springer-Verlag, London, UK, 198-208.
[9] M. Potter, K. De Jong, "A cooperative coevolutionary approach to function optimization", in Proceedings of the Third Conference on Parallel Problem Solving from Nature, vol. 2, 1994, pp. 249–257.
[10] Z. Y. Yang, E. K. Tang, and X. Yao:, "Large scale evolutionary optimization using cooperative coevolution," *Information Sciences* 178 (2008) 2985–2999
[11] P.N. Suganthan, N. Hansen, J.J. Liang, K. Deb, Y.P. Chen, A. Auger, S. Tiwari, "Problem definitions and evaluation criteria for the CEC 2005 special session on real-parameter optimization", Technical Report, Nanyang Technological University, Singapore, 2005, http://www.ntu.edu.sg/home/EPNSugan
[12] G.R. Curry, "Radar System Performance Modeling (Second Edition)", Artech House Publishing, 2004

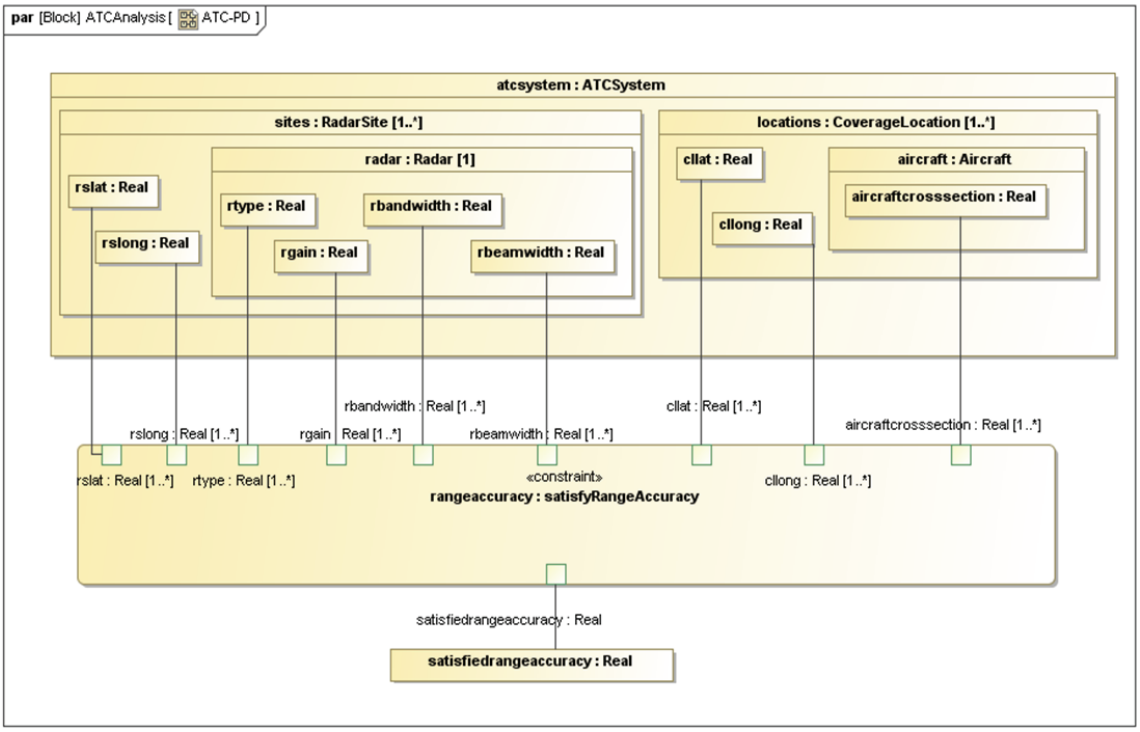


Figure 5 - Parametric Diagram linking the ATC System with the Range Accuracy Constraint

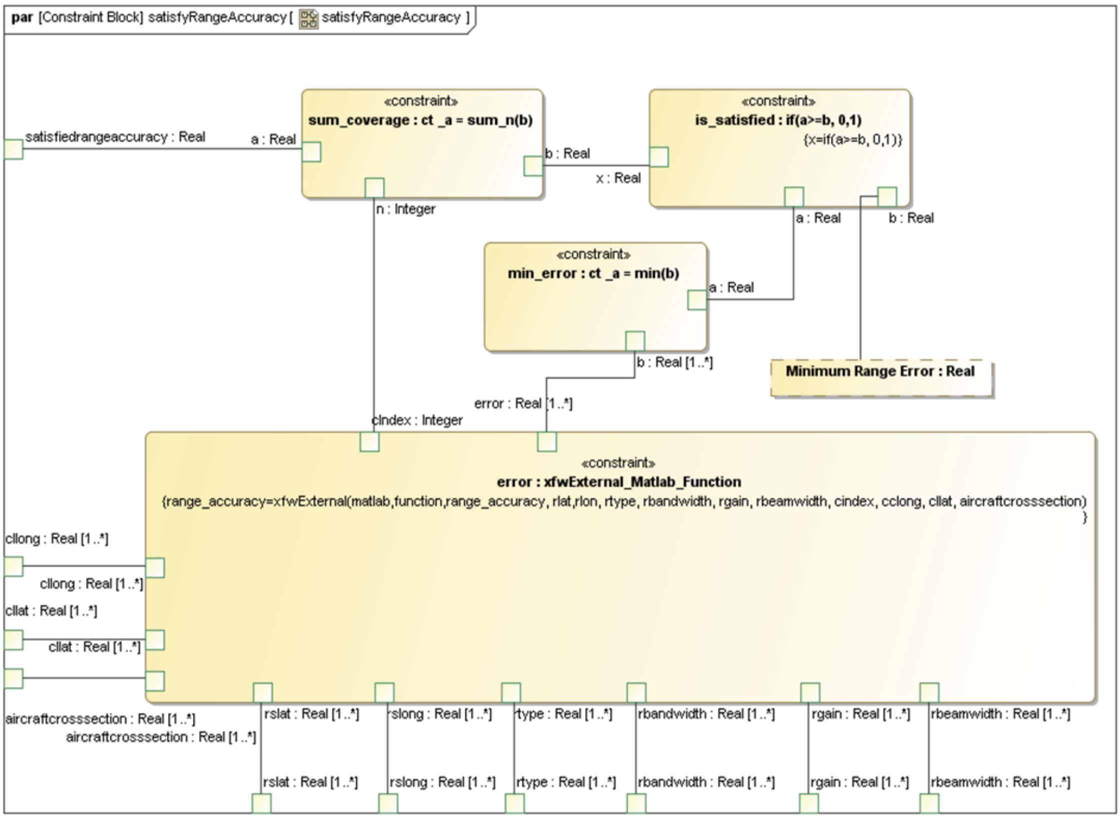


Figure 6 - Parametric Diagram for the Constraint - Range Accuracy

A New Automated Procedure for Estimation of Evapotranspiration for Universal Acceptance

Rai Sachindra Prasad,
Department of EEE,
Graphic Era University,
Dehradun, India.
drsachpd@yahoo.com

Abstract - *Evapotranspiration (ET), a complex dynamic, nonlinear phenomenon, is composed of evaporation and transpiration. Numerous models for estimation of ET exist but none has performed equally well over all regions of the world. Besides, complex calculations, local calibrations and adjustments of parameters in some popular models render them unfit for field applications from non-expert users' point-of-view. This paper introduces, for the first time, a fully automated user-friendly procedure, requiring as input only climate data of the location. The procedure identifies the most dominant variables which influence ET for the location, recommends application of Artificial Neural Network (ANN)/ Ridge Regression (RR), whichever is judged appropriate on a set of developed rules. It eliminates the need of several trials required in the ANN applications for choosing the best sub-set of variables. The procedure, tested on climate data of several regions of the world, is found to be worthy of applications in field.*

Keywords: *Evapotranspiration, Multicollinearity, Singular Value Decomposition, Variance Decomposition Proportions, Artificial Neural Network, Ridge Regression.*

1 Introduction

Evapotranspiration (ET) is a very important parameter in irrigation and hydrology. The physics and methodology of the ET process has been described in several papers [1]. However, questions of great relevance still remain unanswered. These relate to the exact manner in which soil and plant factors affect ET, the definite nature of interactions among the micrometeorological effects on ET, and interactions in the soil-water-plant-atmosphere continuum. Several well-known ET models on all possible approaches have been proposed in [2]. The issues associated with most of the models are the requirements of calibrations with change in location. These situations have forced the researchers, in the recent past to apply artificial neural networks (ANN) [3], [4]. ANN requires large size of data for training [5]. In the ANN approach it is unavoidable not to try all possible combinations of input variables in the quest for optimal performance. Again since the choice of dominant variables influencing ET is not known uniquely, it is also confusing which architecture would perform best and which are likely to behave poorly. Another important issue in estimation of ET is the aspect of user-friendliness for practical use [6]. This paper addresses all these issues in developing a fully automated procedure worthy of universal applications. The tools used are: singular value decomposition (SVD), variance decomposition proportions (VDP), for identification of dominant variables; and use of ANN or ridge regression (RR), which is identified on-line. The rest of the paper is arranged as follows: in section 2, we introduce briefly SVD, VDP, RR, ANN and its algorithm. Section 3 explains the algorithm for the composite procedure, 'RS'. Section 4 provides results and comments before finally concluding. Due to space constraints emphasis is on the algorithms and results, only briefly describing the application parts of the theory of various tools used. Since Radial Basis Function Networks (RBFN) have been used in estimation of ET more than any other architecture of ANN, we have used this in developing the procedure RS. The main objective of the paper is to develop a user-friendly computer program with permissible accuracy of less than fifteen percent [7]. The limitations of ANN requiring large size of data and high computation cost have been offset to an extent, first by choosing RBFN, and secondly, to a large extent, by identifying on-line the dominant inputs among a set of possible inputs. Again, the application of RR in the procedure is for situations where input variables are identified to be highly interrelated, which all climatic variables (CV), normally, are expected to varying degrees of multicollinearity (MC). The motivation for this work comes from the comment [8] that it is difficult, if not impossible, to predict transpiration, which is an integral part of the composition of ET. Sources of data are reported in [14]-[17].

2 Singular Value Decomposition (SVD)

SVD has been suggested as remedial measure for ill-conditioned data matrices [9]. The essential features of SVD are contained in the following theorem which is stated without proof:

Theorem Let A be any n x p matrix, U (n x p) and V (p x p) be the orthogonal matrices, then A can be decomposed as

$$A = U\,D\,V' \qquad (2.1),$$

where, D is a diagonal matrix, of the order p x p, containing non-negative diagonal elements, μ_j, j = 1,2,..,p; $\mu_1 > \mu_2 > .. \mu_p \geq 0$. The diagonal elements, μ_j, are called singular values of A. These elements are crucial in determining MC among the columns of A. Any practical situation, where μ_j approaches zero, is of considerable interest in the analysis of climate data (CD). This very feature has been used in the analysis of CD forming rectangular matrix A, whose columns represent the CV. Standardization of CD is essential for any meaningful interpretation of SVD application results.

2.1 Variance Decomposition Proportions

Let Z represent standardized form of A. Then, condition number, k(Z), of Z is defined as $k(Z) = \|Z\| \times \|Z^{-1}\|$ (2.2), where, $\|.\|$ denotes spectral norm. Using matrix algebra, it can be shown that

$$\|Z\| = \mu_{max} \text{ and } \|Z^{-1}\| = 1/\mu_{min} \quad (2.3),$$

and from (2.2), $k(Z) = \mu_{max}/\mu_{min}$ (2.4),
where, μ_{max} and μ_{min} are maximum and minimum singular values respectively, and $1 \leq k(Z) \leq \infty$. When k(Z) = 1, the predictors are perfectly orthogonal, whereas, k(Z) > 1 indicates existence of MC. A large value of k(Z) indicates severity of MC. The VDP of the estimated regression coefficients are obtained as follows: Variance-Covariance matrix is as in (2.5)

$$V(\hat{\beta}) = \hat{\sigma}^2 \ (Z'Z)^{-}1 = \hat{\sigma}^2 = \hat{\sigma}^2 \ (V \ D^{-2} \ V') \quad (2.5).$$

For the sth regression coefficient,

$$\text{var}(\hat{\beta}) = \hat{\sigma}^2 \sum (v_{sj})^2 / (\mu_j)^2, \ j = 1,2,..,p \quad (2.6),$$

where $V = (v_{sj})$. The var($\hat{\beta}$) is the sum of the decomposed components of the sth regression coefficient, each one of which is associated with one particular singular value. Define $p_{sj} = (v_{sj})^2/(\mu_j)^2$, $p_s = \sum p_{sj}$, and $\pi_{js} = p_{sj}/p_s$; (2.7), s,j = 1`,2,..,p, and p_{sj} is the variance of the jth component of decomposition for the sth regression coefficient and π_{js} denotes the proportion of the variance of the sth regression coefficient associated with the jth component of its decomposition. These are reflected in VDP table 1

TABLE 1 VDP TABLE

μ	Proportions of var ($\hat{\beta}_1^*$)	var ($\hat{\beta}_2^*$)..	.var (($\hat{\beta}_p^*$)	k(Z)
μ_1	π_{11}	π_{12}	.. π_{1p}	$k(z)_1$
μ_2	π_{21}	π_{22}	.. π_{2p}	. $k(z)_2$
---	---	---	---	---
μ_p	π_{p1}	π_{p2}	.. π_{pp}	. $k(z)_p$

2.2 Ridge Regression (RR)

Ridge Regression has been fully explained in [10a,b]. Ridge estimated regression coefficient is given as in (2.8),

$$\hat{\beta}_K = (z'z + kI)^{-1} z' \ y \quad (2.8),$$

k > 0, y is the response (ET), z'z is in correlation form, and I is the identity matrix. RR is justified when

$$MSE(\hat{\beta}_K) < MSE(\hat{\beta}^{*}) \quad (2.9),$$

$\hat{\beta}^*$ is the ordinary least squares (OLS) estimated coefficient standardized form.

2.3 Ridge Trace: Stability Region

The ridge trace is a plot of ridge estimated regression coefficients, $\hat{\beta}_{ki}$, i = 1,2,..p against k, o < k < 1. Ridge trace furnishes a visual representation of instability due to collinearity and provides a basis for selecting a value of k. The problem with practical use of trace is with regard to selection of unique k by visual estimation which is prone to error. This weakness has recently been removed by an algorithm, 'modified trace' (MT) [11], where, k_I denotes the commencement point of the stability region, k_F denotes the end point of the region such that $MSE(\hat{\beta}_k) = MSE(\hat{\beta}^*)$, k_{OPT} represents minimum $MSE(\hat{\beta}_k)$, and k_{RS} is final choice such that (2.9) is not violated.

2.4 Artificial Neural Network (ANN)

The theory and applications of ANN have been adequately dealt with in [5]. For prediction, most often used networks are back-propagation and RBFN. RBFN has the merits of simple architecture, less computation cost and memory requirement, and is particularly suitable for including in a computer program. It consists of three layers: input, hidden and output. The hidden layer is non-linear while the output layer is linear. The jth input vector, $Zj = [z_{j1}, z_{j2},.,z_{ji},..,z_{jn}]$, i = 1,2,..,n; j = 1,2,..,m; m being the number of input vectors. The number of output nodes, y_j, is equal to the number of training input-output data pairs, where, input is in matrix form and output is a vector. The jth hidden nodes vector is: $H_j = [h_{j1}, h_{j2},..,h_{ji},.., h_{jk}]$, j = 1,2,..,m; k= 1,2,.., q; q being the number of centers chosen. The weights w_{jk} connect the kth hidden node with the jth output node. In the hidden layer $C_j = [c_{j1}, c_{j2},.,c_{jq},..c_{jq}]$, is called the jth center for RBFN, and $\|z_{ji} - c_{jq}\|$ is the Euclidean distance obtained in (2.10)

$$\|z_{ji} - c_{jq}\| = \sqrt{\sum} (z_{ji} - c_{jq})2 \quad (2.10),$$

$$h_{jk} = \Phi_{jk} (\sqrt{\sum} (z_{ji} - c_{jq})2 \quad (2.11),$$

$$\Phi(z) = e^{-(z)2/(\alpha)2} \quad (2.12),$$

Φ( ) is the Gaussian function, and α is the width parameter. In the output layer, w_{jk} is the weight between the hidden node h_{jk} and the output node y_j. The jth output of the output layer is given in (2.13) $yj = \sum w_{jk} h_{jk}$ (2.13),

Optimization criterion mostly used is root mean square error (RMSE),

$$RMSE = \sqrt{\sum} (1/m) * (error_j)^2 \quad (2.14),$$

$$(error)_j(ep) = (t_j - y_j(ep)) \quad (2.15),$$

$(error)_j$ (ep) and y_j (ep) are the jth error and the jth output of the epth epoch and t_j is the jth true value. The elements of the three learning matrices, w_{jk}, c_{jk}, α_{jk} are determined using various rules, some of them heuristics. In this paper, these are random positive values between 0 and 1, and updated dynamically during training following [12]. The three learning parameters, μw, μc, μα, in (2.16)-(2.18) are estimated, for the first time, using a simple and effective error analysis technique following [13]. This is explained in ANN algorithm which is described next.

$w_{jk\mu}$ (ep+1) = w_{jk} (ep) – μw_j (δ ⁄ δw_{jk}) $error_j$ (ep) (2.16).

c_{jk} (ep+1) = c_{jk} (ep) − μc_j (δ ⁄ δc_{jk}) $error_j$ (ep) (2.17).

α_{jk} (ep) = α_{jk} (ep) − $\mu\alpha_j$ (δ ⁄ δα_{jk}) $error_j$ (ep) (2.18).

2.5 Algorithm ANN

Step 1 Choose the no of centers (C), say q

Step 2 Initialize elements of matrices W_{jk}, C_{jk} and α_{jk} randomly, j=1,2,..,m; k=1,2,..,q

Step 3 Present input data vectors, z_{ji}, j= 1,2,..,m; i=1,2,..,n

Step 4 Conduct error analysis: for each j, compute $(error)_j$ using (2.15), outlined in steps 5-9.

Step 5 Generate an array of equally spaced positive Numbers ranging from say, 0.0050 to 1.0000 in m steps for all the three learning parameters.

Step 6 Keeping two of the three learning parameters constant at any arbitrary value, say 0.5, compute $(error)_j$, for j=1,2,..,m, of the third learning parameter vector using (2.16)-(2.18).

Step 7 Repeat step 6 for the other two learning parameters.

Step 8 For each error vector in steps 6 and 7, determine the element values of each of the three learning parameter for which the $(error)_j$ is within ± 10% deviation from zero error. (If needed, steps 5-8 may be repeated for a new set of values in each of the three learning parameter's array to estimate a near optimal value by progressively narrowing the difference in element values so that its elements have nearly the same value, and as close to zero error as possible.)

Step 9 Use the values determined in step 8 for μw, μc and μα to train the network, meeting the criterion of small RMSE.

Step 10 Present test data for the desired goal.

Step 11 STOP.

3 Algorithm ('RS')

The algorithm of the entire procedure, 'RS', is composed of six modules, (i) normalization of data, (ii) SVD and VDP analysis, (iii) identification of climatic variables (CV), (iv) estimation of optimal ridge parameter, k, using algorithm 'MT', (v) ANN algorithm, and (vi) performance evaluation

Step 1 Normalize data, A→Z.

Step2 Compute OLS estimated coefficients of normalized data, $\hat{\beta}^*$.

Step 3 Compute SVD of Z.

Step 4 Form VDP Table.

Step 5 Using rules from Table 5, check if any CV requires elimination (k(Z) & π_{js} must be the joint condition).

Step 6 If rules permit, form a new data matrix, $Z_{new.}$

Step 7 Repeat steps 3-6 if permitted, else go to next step.

Step 8 Compute ridge estimated coefficients, $\hat{\beta}_{ki}$, i = 1,2,.., p, against 0 < k < 1.

Step 9 Compute MSE ($\hat{\beta}_{ki}$), i = 1,2,..,p.

Step 10 If k_F found, fix it, else go to step 20.

Step 11 If $k_I < k_{OPT} < k_{F,}$ then set $k_{RS} = k_{OPT}$, else set $k_{RS} = k_{I.}$

Step 12 Compute $\hat{\beta}_{kRS}$.

Step 13 Compute MSE ($\hat{\beta}_{kRS}$).

Step 14 Select k_{RS} on minimum value of MSE in step 13, $k_I < k_{RS} < k_{F,}$; $k_{OPT} < k_I$.

Step 15 Compute ($\hat{\beta}_{kRS}$)$_i$, i = 1,2,..,p.

Step 16 Evaluate performance parameters (RMSE, MAE).

Step 17 If performance is satisfactory, go to next step, else to ANN module.

Step 18 Print model's input (predictors), generate y_j and fit a model.

Step 19 STOP.

Step 20 If in step 10, k_F is not found, then test if MSE ($\hat{\beta}_k$) < MSE ($\hat{\beta}^*$).

Step 21 If condition in step 20 is true, then check if k(Z) > 35, if yes, go to next step else to ANN module.

Step 22 From table 5 check the variables which have $\pi_{js} \geq 0.95$, for $p \geq 2$, if true, delete the variable which has the highest value of $\pi_{js,}$ and go to next step else to step 24.

Step 23 Go to step 4.

Step 24 Apply ANN module.

Step 25 Conduct training of the network, and performance test until desired parameters (RMSE, MAE) are met.

Step 26 Go to step 18.

4 Results and Discussions

Due to space limitations, results are shown only for five locations. Table 2 presents singular values, condition number, and records of CD used, Tables 3, 3A and 4A-4D show the VDP analysis for the locations. Based on experiments, rules for elimination of a variable, and for applicability of the procedure in a given situation, have been shown in Tables 5 and 6 respectively. If the data size is small (< 200) and has 'mild' or 'nil' MC, then any popular model (P) reported in [1]-[3] may be used.

The CD of the five locations were subjected to VDP analysis, but ANN applications have been shown only for locations 2, 4 and 5.

TABLE 2 SINGULAR VALUES AND CONDITION NUMBERS

Location 1	Santa Monica (USA)
No of record	4 years (2001-04) data of daily observations.

Predictors(p)-4	AT RH W SR
Singular Values, k(Z)	1.96 0.34 0.17 0.11, 17.7
Location 2	Davis (USA)
No of records	4 years (2001-04) data of daily observations.
Predictors (p) -4	AT, RH, W, SR
Singular values, k(Z)	1.90 0.47 0.38 0.13, 14.16
Predictors -6	ATmax, Atmin, W, SR, RHmax, RHmin
Singular values, K(Z)	2.329 0.634 0.356 0.175 0.124 0.04, 55.45
Location 3	Dirab (Saudi Arabia)
No of records	26 Monthly averages of data (1981-83).
Predictors 4	AT, RH, SR,W
Singular values, K(Z)	1.93 0.477 0.202 0.047, 41.06
Location 4	Yuma Mesa (USA)
No nd 4 records	319 (2009 data of daily observations)
Predictors 4	AT, RH, W, SR
Singular values K(Z)	1.93 0.37 0.32 0.12 16.08
Location 5	Pomona (USA)
No of records	4 years(2001-04 data of daily observations
Predictors 4	AT, RH, W, SR
Singular values, k(Z)	1.96 0.34 0.17 0.11, 17.8

TABLE 3 VDP TABLE OF LOCATION 1

μ	VDP Elements				k(Z)
	AT	RH	W	SR	
1.96	0.002	0.000	0.001	0.014	1.00
0.34	0.068	0.047	0.048	0.222	53.81
0.17	0.308	0.027	0.933	0.011	11.60
0.11	0.621	0.926	0.018	0.753	17.70

TABLE 3A VDP TABLE OF LOCATION 1

Six variables

μ	VDP Elements						k(Z)
	ATmx	ATmn	RHmx	RHmn	W	SR	
2.40	0.00	0.00	0.00	0.00	0.00	0.00	1.0
0.35	0.00	0.00	0.00	0.00	0.00	0.00	6.8
0.26	0.00	0.00	0.02	0.04	0.00	0.00	9.3
0.17	0.03	0.08	0.01	0.00	0.00	0.00	14.2
0.10	0.08	0.02	0.00	0.00	0.36	0.07	22.9
0.03	0.87	0.89	0.96	0.94	0.63	0.67	77.4

TABLE 4A VDP TABLE OF LOCATION 2

μ	VDP Elements				k(Z)
AT	RH	W	SR		
1.90	0.002	0.006	0.014	0.005	1.00
0.48	0.004	0.083	0.176	0.017	3.95
0.38	0.009	0.214	0.810	0.005	5.00
0.13	0.984	0.697	0.000	0.829	14.6

With 168 data points, k(z)=11.0

TABLE 4B VDP TABLE OF LOCATION 3

μ	VDP Elements				k(Z)
	AT	RH	W	SR	
1.94	0.001	0.001	0.000	0.001	1.00
0.44	0.005	0.023	0.002	0.016	4.40
0.12	0.147	0.017	0.077	0.023	10.20
0.05	0.846	0.959	0.921	0.960	40.0

Among the five locations, CD of locations 1 and 5 have also been used for ANN application in [3], where two very significant results were highlighted: (i) ANN performs, overall, equal to, or better than, the best among popular models (P), (ii) data of a nearby station (Santa Monica) used for predicting ET for another station (Pomona) yielded good results.

TABLE 4C VDP TABLE OF LOCATION 4

μ	VDP Elements				k(Z)
	AT	RH	W	SR	
1.93	0.003	0.000	0.001	0.006	1.00
0.37	0.000	0.000	0.001	0.006	5.21
0.37	0.115	0.696	0.051	0.051	6.03
0.12	0. 799	0.011	0.853	0.987	16.8

TABLE 4D VDP TABLE OF LOCATION 5

μ	VDP Elements				k(Z)
	AT	RH	W	SR	
1.96	0.000	0.009	0.009	0.000	1.00
0.34	0.003	0.023	0.848	0.001	5.76
0.17	0.003	0.023	0.000	0.998	11.52
0.11	0.984	0.886	0.143	0.001	17.81

TABLE 5 RULES FOR ELIMINATION

K(Z)	&	π_{js}	MC LEVEL	Eliminate?
< 20	&	0.70	NIL	No
20 < k(Z) ≤ 35	&	0.80 < π_{js} <0.95	MILD	No
k(Z) > 35	&	0.95 < π_{js} ≤1.00	HIGH	Yes

TABLE 6 RULES FOR CHOICE OF PROCEDURE

Level of MC	Choice of Procedure	
	Size of Data	
	Small	Large
NIL	P	ANN
MILD	RR/P	ANN
HIGH	RR	ANN/ RR

TABLE 7 RESULTS OF PROCEDURE APPLIED ON TEST DATA

Location	2	3	4
Training Data	200	12*	180
Test Data	143	14*	128
RMSE	0.0014	4.79	0.034
MAE	1.84e-004	0.84	0.003
Learning Parameters	μ_α = 0.68	k_I=0.007*	μ_α = 0.72
	μ_w =0.60	k_{OPT}= 0.003*	μ_w =0.78
	μ_c =0.72	k_{RS}= 0.007*	μ_c =0.58
Centers	C = 20	----	C = 20

Number of epochs = 350-400 (Training of RBFN)

* RR application figures on ROM

Both the results were obtained in [3] after seven combinations of physical variables, (i) W, (ii) RH, (iii) AT, (iv) SR, (v) SR and AT, (vi) SR, AT and RH, (vii) AT, RH, W, and SR, in trial runs during training of the network, involving huge memory and high computation cost. Against this background, VDP analysis and RBFN were used in the new procedure to yield the same results without any exercise in combinations and trials. A look at the SVD and VDP results for the two locations in Tables 2, 3 and 4D demonstrate exactly the same result, since k(Z) is nearly the same (17.71) and (17.81) and the VDP figures at the bottom of the table indicate the effect of MC as 'nil', indicating that all four model inputs are significant contributors to the response (ET), also claimed in [3]. Except for location 3, all others (4 predictors) do not exhibit even moderate condition of collinearity. For locations 1 and 2, when the number of predictors was increased to six in an experiment, k(Z) shot up to 77 for location 1 (Table 3A),, while in location 2 it increased to 55 (Table 2), indicating severe condition of MC. For location 1, the VDP Table 3A, on application of rules, shows that there are two possible variables which have very little predicting power. Removing one variable, RHmax, would form another VDP table which will further guide the status of variable elimination, following rules of elimination.

Location 3 with the same four predictors (AT, RH, W, SR) shows severe MC, indicating possible elimination of one variable as per the rules. Since the data size for this location is only twenty six, use of ANN is out of consideration. RR is, therefore, applied using twelve data points (December 1981 to November 1982) for model building, and the rest fourteen used for prediction. Since k(z) and π_{js} values in the VDP table of the location reflects severe collinearity between RH and SR, one of the four variables, SR, was eliminated. Resultant VDP of the three predictors did not indicate further elimination (not shown). The reduced order model (ROM) was then obtained in (2.19),

$$ET = 0.025 + 0.348\ (AT) - 0.087\ (RH) + 2.108(W) \quad (2.19)$$

where AT, RH and W are in raw form. A comparison of the RR procedure with true values, and popular models, is shown in Figures 2 and 3 respectively. Table 8 indicates comparison on RMSE values.

RR procedure on comparison with true values, and popular models, is shown in Figures 2 and 3 respectively. Table 8 indicates comparison on RMSE values.

TABLE 8 COMPARISON OF 'RS' PROCEDURE WITH POPULAR MODELS (P) (Location 3)

Models	H	J-H	MJ-H	T	B-C FAO	PM	RS (ROM)
RMSE	4.46	11.	5.56	4.03	6.97	14.2	4.79

H=Hargreaves, J-H= Jensen-Haise, MJ-H=Modified J-H, [1]-[3]
T=Turc, B-C (FAO)=Blaney-Criddle (FAO), PM= Penman

The RR procedure competes very well with them, in fact almost as good as the best models (H, T). Since application of ANN for prediction of ET for location 1 has already been reported in [3], this is not considered again, but was included for the specific purpose to show that VDP analysis is a reliable technique to determine which variables are significant, thereby avoiding unnecessary computations involved in combinations of variables, and trials during training. Similar results were also obtained for location Pomona, which also figured in the ANN application [3]. Results on test data for ANN application and on prediction data for RR application are shown in Table 7. Since experiments (Table 4A bottom) indicate that k(Z) varies little with variation in data size over the minimum around twenty, hence, irrespective of data size above the minimum, SVD and VDP together can be relied upon to investigate MC among predictors, also reported in [18]. Results of ANN applications are evaluated on root mean square error, RMSE, and mean absolute error, MAE, defined as: $RMSE=\sqrt{(1/m \sum (ET_{i\ observed} - ET_{i\ predicted})^2)}$ (2.20)
$MAE=1/m \sum (ET_{i\ observed} - ET_{i\ predicted})$. (2.21).

Figures 1-6 represent the performance of the ANN algorithm. All the graphs were obtained on the MATLAB platform. Total steps involved for realization of the RS algorithm are nearly two hundred. No attempt was, however, made to optimize computer program.

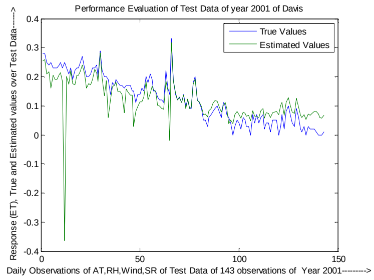


Fig. 1 Performance of RBFN of location 2

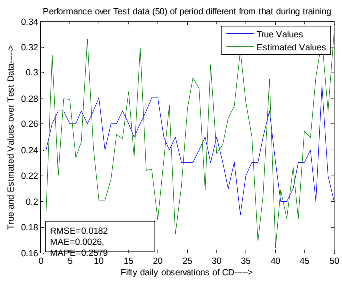


Fig. 2 Perferformance over test data of year 2002, different from the period of training data (2001), location 2.

5 Conclusions

The automated procedure holds the prospects of universal applicability, identifying on-line the physical variables relevant for the most appropriate model of a location, ensuring acceptable accuracy. Besides, the procedure, being user-friendly, is expected to fulfill a great need of the users. However, for RR application, since the procedure has been designed for severe cases of MC, it would not be applicable for mild MC with smaller size of data

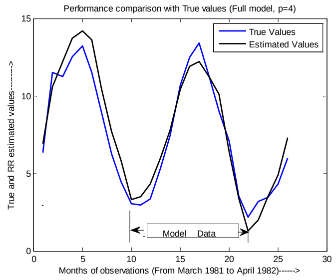


Fig. 3 Comparison of RR with true values, location 3 (ROM).

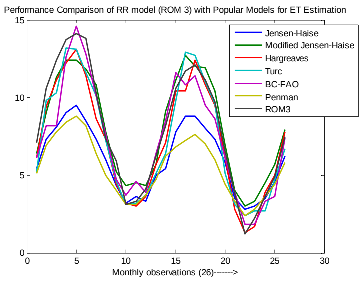


Fig. 4 Comparison of RR with popular models

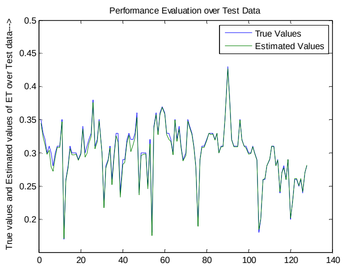


Fig. 5 Performance of RBFN on test data, location 4

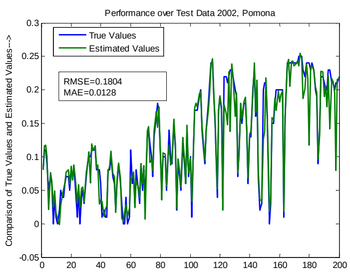


Fig. 6 Performance of RBFN on test data, Location 5

References

[1] ASCE, “Methods for estimating evapotranspiration”, Proce Irrig. And Drainage, Special Conference, Las Vegas, Nov 2-4, 1966, ASCE.

[2] M.E.Jensen, R.D.Burman, R.G.Allen (eds), “Methods for estimating evapotranspiration, and Irrigation Water Requirements”, ASCE manuals and reports on Engineering practices No. 70, ASCE; 1990

[3] O. Kisi, ”Evapotranspiration modeling from climatic data using a neural computing technique”, Hydrologic Process, vol 21, 2007, pp 1925-1934.

[4] S.Trajkovic, B. Todorovic, M. Satanical, ”Forecasting reference evapotranspiration by artificial neural networks”, ASCE Jour of Irrigation and Drainage Engineering, vol 129, 2003, pp 454-457.

[5] S. Haykin “Neural Networks, A Comprehemsive Foundation”, New York, 1994, MacMillan College Publishing Company.

[6] S. Pleban and I. Israeli,” Improved approach to irrigation scheduling programs”, Journal of Irrigation and Drainage, vol 115, ,No 4, 1989, pp 577-587.

[7] M.L.Sharma,”Estimating evapotranspiration”, In “Advances in Irrigation”, (Edit D.Hillel), Vol 13, 1985, pp 213-281, Acad Press, N.Y.

[8] S. Naoum and I. K. Tsanis,” Hydroinformatics in evapotranspiration estimation”, Environmental Modeling and Software, vol 18, 2003, pp 261-271

[9] V.C. Klema and A.J. Laub “ The Singular Value decomposition: its computation and some applications”, IEEE Trans Automatic Control AC-25 No2, 1986, pp 164-176.

[10] A.E. Hoerl and R.W. Kennard, “Ridge Regression: Biased estimation for non-orthogonal problems”, Technometrics, Vol 12, No 1, 1970a,b, pp55-67, 68-82.

[11] R.S. Prasad, “An algorithm for estimation of optimal ‘k’ in Ridge Regression applied in modeling Evapotranspiration”, INDICON, Annual IEEE Conf, Hyderabad, Dec 16-18, 2011, pp 1-4, (IEEE Xplore).

[12] T. Poggio and F.Girossi, ”Networks for approximations and learning”, Proc IEEE, Vol 78 (9), 1990, pp 1481-1497

[13] E.O.Doebelin, “Measurement Systems: Application and Design”, Third Edition, International Student Edition, McGraw Hill International Book Company, New Delhi.

[14] Climate Data of Santa Monica, Pomona and Davis locations, website of California Irrigation Management Information Service (CIMIS).

[15] Climate Data of Yuma Mesa, website of Arizona Meteorological Network, University of Arizona.

[16] M. Saeed, ”The estimation of evapotranspiration by some equations under hot and arid conditions”,Trans ASAE, vol 29, No 2, 1986, pp 434-438.

[17] Climate Data of Indian locations: Ludhiana; Pune, Bangalore, Jaisalmer, Midnapur; Indian Meteorological Department, Pune.

[18] M. Steinbach, P-N Tan, V. Kumar, S. Klooster and C. Potter, “Discovery of Cluster indices using clustering”, In ACM SIGKDD Conf on Knowledge Discovery and Data mining, 2003, pp 446-455..

Proc. of the 2012 7th International Conference on System of Systems Engineering, Genoa, Italy - 16-19 July 2012

Basic Guidelines for Simulating SysML Models: An Experience Report

Mara Nikolaidou, George-Dimitrios Kapos, Vassilis Dalakas and Dimosthenis Anagnostopoulos
Department of Informatics and Telematics
Harokopio University of Athens
70 El. Venizelou St, Kallithea, 17671, Athens, GREECE
email: {mara, gdkapos, vdalakas, dimosthe}@hua.gr

***Abstract*—Though there are numerous efforts for simulating SysML models, the automated generation of executable simulation code for specific simulation environments without any interference by the system engineer is still an issue attracting the researchers' attention. To become efficient and easy to use, such an activity should be explored using standardized methods and tools, such as the utilization of MDA concepts for model transformation. In this paper, we identified some basic guidelines for the generation of executable simulation code based on existing SysML system models and the selection of related methods and tools for simulation and model transformation purposes. The proposed guidelines are incorporated in a three-step methodology that can be applied independently of the simulation framework selected. In the paper, we will discuss our experience applying it, based on examples from different system domains, where DEVS framework was chosen for simulation purposes. The reasons for its selection and the potential drawbacks and difficulties drawn from its adoption will also be discussed, to comment on the characteristics a simulation framework and language should obtain to be effectively applied for SysML model simulation.**

I. Introduction

Systems Modeling Language (SysML) is the emerging standard for model-based system engineering [1]. The Systems Modeling Language (SysML) is such It is a general-purpose graphical modeling language, for systems engineering applications and supports the specification, analysis, design, verification and validation of a broad range of systems and systems-of-systems. These systems may include hardware, software, information, processes, personnel, and facilities. It is defined as a profile of Unified Modeling Language (UML), the standard for modeling software intensive systems.

Using SysML the system engineer should perform all engineering activities based on a common model, according to Model Driven Architecture (MDA) concepts [2]. The common system model should be general enough to cover all engineering activities, while it should also be specialized to serve specific engineering activities, such as system validation.

Model-based system design is served by numerous methodologies adopting SysML as a modeling language. To validate the proposed system architectures, quantitative methods are usually applied, focusing on system performance. Thus, system validation is often performed using simulation. Since most simulation methodologies are model-based, such an approach is suitable for model-based system engineering. This justifies the increased interest in integrating SysML modeling environments and simulation tools. To this end, there are numerous efforts to simulate SysML models (for example [3], [4]). Most of them aim at transforming SysML system models to simulation models, executed in a specific simulation environment, as for example PetriNets [4] or ModelicaML [5].

Prominent efforts in this area include the definition of a SysML4Modelica profile endorsed by OMG [6] and the corresponding transformations to convert SysML system models, using the profile, to executable Modelica simulation code with standard MDA methods, as QVT language. The authors are working on a similar approach [7] targeting at transforming SysML models to executable DEVS simulation models [8]. A corresponding SysML DEVS profile and SysML-to-DEVS metamodel transformation in QVT have been implemented and tested using different system examples.

Though there are numerous efforts for simulating SysML models, automated generation of executable simulation code for specific simulation environments without any interference by the system engineer is still an open issue. To become efficient and easy to use, such an activity should be explored using standardized methods and tools, utilizing MDA concepts for model transformation. Furthermore, a variety of simulation methods and tools should be supported, depending on the systems under study. Either continuous or discrete simulation may be applied, while model libraries could be used instead of writing simulation code for all system entities. Though the simulation tools and system requirements may differ, model transformation tools and the process of integrating simulation-specific characteristics into SysML models may be standardized. Towards this direction, basic guidelines for the generation of executable simulation code based on existing SysML system models and the selection of related methods and tools for simulation and model transformation are identified. These guidelines, based on the authors experience, are analytically explained in the paper, based on specific examples. The proposed guidelines are incorporated in a three-step methodology that can be applied independently of the simulation framework selected. In the paper, we will discuss our experience applying it, based on examples from different system domains, where DEVS framework was chosen for simulation purposes. The reasons for its selection and the potential drawbacks and difficulties drawn from its adoption will also be discussed, to comment on the characteristics, a simulation framework and language should have to be effectively used for SysML model simulation.

The paper is structured as follows: In the section II a short review of SysML model simulation is given. The proposed guidelines and the related benefits and potential considerations are discussed in section III. In section IV the three-step methodology for SysML model simulation is presented. An example applying the proposed methodology when simulating SysML models using DEVS is described in section V. Conclusions and Future work reside in section VI.

II. Related Work

There are a lot of efforts from both research and industrial communities to simulate SysML models [9], [10]. Apparently SysML supports a variety of diagrams describing system structure and states, necessary to perform simulation, which are utilized by different approaches. In most cases, SysML models defined within a modeling tool are exported in XML format and, consequently, transformed into simulator specific models and forwarded to the simulation environment.

Depending on the nature and specific characteristics of systems under study, there is a diversity of ways proposed to simulate SysML models, utilizing different diagrams. In [11], a method for simulating the behavior of continuous systems using mathematical simulation is presented, utilizing SysML parametric diagrams, which allow the description of complex mathematical equations. System models are simulated using composable objects (COBs) [12]. It should be noted that in any case SysML models should include simulation-specific information, which facilitates simulating them [13]. These approaches are better suited for systems with continuous behavior.

Simulation of discrete event systems is utilized, based on system behavior described in SysML activity, sequence or state diagrams. In [3], system models defined in SysML are translated to be simulated using Arena simulation software. SysML models are not enriched with simulation-specific properties, while emphasis is given to system structure rather than system behavior. Model Driven Architecture (MDA) concepts are applied to export SysML models from a SysML modeling tool and, consequently, transformed into Arena simulation models, which should be enriched with behavioral characteristics before becoming executable. In [4], the utilization of Colored Petri Nets is proposed to simulate SysML models. If the system behavior is described using activity and sequence diagrams in SysML, it may be consequently simulated using discrete event simulation via Petri Nets.

To enable the construction of executable simulation models, simulation capabilities should be embedded within SysML models utilizing profile mechanism. The formal method, proposed by OMG to extent or to restrict UML and consequently SysML models, is the definition of a profile, emphasizing the properties of a specific world or domain, a simulation methodology or tool in this case. SysML profiles contain stereotype definitions, which facilitate the formal extension/restriction of UML entity semantics. In [14], simulation is performed using Modelica [15]. To ensure that a complete and accurate Modelica model is constructed using SysML, a corresponding profile is proposed to enrich SysML models with Modelica-specific properties, utilizing ModelicaML.

Ideally, the simulation models extracted from SysML models should become executable without any additional programming effort from the system engineer, while SysML models should be easily transformed to executable simulation models. Both, [3] and [10] utilize MDA concepts for model transformation.

Furthermore, the simulation methodology adopted should be popular and facilitate the execution of simulation models on a variety of simulators, while the existence of model libraries may also enhance the capabilities of the system engineer to produce simulation code.

The SysML4Modelica profile endorsed by OMG [6] enable the transformation of SysML models to executable Modelica simulation code. The relevant standard also proposes the transformation of SysML models, defined using the profile, to executable Modelica models using standard MDA methods, as QVT language. The authors are working on a similar approach [7] targeting at transforming SysML models to executable DEVS simulation models [8]. A corresponding SysML DEVS profile and SysML-to-DEVS metamodel transformation in QVT have been implemented and tested using different system examples.

Since SysML profiles are based on formal UML extension mechanisms, they can be implemented in any standard UML modeling tool, such as Rational Modeler [16] or MagicDraw [17], enabling the integration of simulation tools with any of them.

In the following, we try to identify the requirements for effectively simulating SysML models, independently of the simulation method applied, and suggest some basic guidelines for selecting the methods and tools to use.

III. Guidelines

The proposed guidelines are suggested to enable engineers facing the challenge of evaluating existing SysML system models via simulation to choose the proper simulation and model transformation methods and tools. Issues concerning SysML model enrichment with simulation specific information and the generation of simulation models are addressed. Model enrichment must be based on appropriate SysML profiles, while, in order to generate simulation models, a reference meta-model for the corresponding simulation framework must be defined or selected, if one already exists. Each guideline is analyzed and justified, while open issues that may affect applicability of the guidelines are also discussed.

A. SysML profiles with simulation-specific properties

As already identified in the literature, to simulate a SysML system model the system engineer should incorporate simulation-specific properties in it. This task corresponds to the specialization of the common model to serve a specific engineering activity, in this case system validation using simulation. The appropriate way to achieve this, is by defining and using simulation-related SysML profiles, related to a specific simulation methodology, as for example SysML4Modelica [6] or DEVS-SysML [18]. Such profiles should be applied in modeling tools (e.g, MagicDraw) and enable the validation of simulation-enriched SysML models prior to their transformation to simulation code.

Benefits:

- Definition and use of appropriate profiles allow the addition of simulation-specific properties in the common model, instead of redefining the simulation model.
- The adoption of such formal methods, offers credibility and the possibility to learn, extend, restrict and employ modeling tools, without specialized knowledge about specific simulation environments.
- Model validation is supported.
- Focus remains on analysis/design of the system.

Considerations:

- Definition of a SysML simulation-related profile is not a trivial procedure. Required simulation-specific attributes must be identified and positioned in the appropriate elements of the system model.
- Structural correlation between SysML and simulation models does not suffice for the definition of a profile. Semantics of the simulation formalism and those of SysML should also be taken into account.
- There are no SysML profiles available, for the majority of the simulation environments.
- The profiles should be applied and tested in the majority of well known modeling tools.

B. Complete enrichment of SysML models

It is more efficient to include all simulation-related information within the SysML model and automatically produce fully executable simulation code, since in this case the simulation method becomes transparent to the system engineer. There is no need to write simulation code or learn the specifics of the simulation environment. The engineer only uses SysML notation to describe the system model and its simulation-specific behavior. This feature is not supported by most of the approaches recorded in the literature, that focus on transforming system structure to the simulation model and not simulation-related system behavior.

Benefits:

- Successful incorporation of simulation behavioral (besides structural) aspects in the central model enables automated generation of executable simulation models.
- Such an approach enhances credibility and offers transparency and usability on the whole process.

Considerations:

- Although models may be validated against a profile, conceptual completeness of the profile itself in regard to the simulation framework cannot be easily proved.
- Defining simulation model behavior in a precise, yet generic manner is a challenging task.

C. Generating simulation models from system models

Transformation of SysML to simulation-specific models should be accomplished in a standardized fashion based on MDA concepts, using existing languages and tools, as QVT [2]. Such languages offer high-level mapping constructs that create simple and maintainable transformations that can be executed in several environments. To facilitate such a transformation, the existence and acceptance of a MOF meta-model for the applied simulation methodology is essential. Such a standardized meta-model for a specific simulation domain is of greater value, when the same simulation methodology can be applied using different tools, e.g., DEVS methodology.

Benefits:

- Simpler transformation definition, in comparison to custom code writing.
- Enhanced credibility, transparency and (re-)usability on the whole process.

Considerations:

- Completeness of such transformations is not always easy to check.
- Such an approach introduces overhead that is, however, overridden by the benefits, as the complexity of system models rises.

D. Reference meta-models

In the case where there is a variety of different simulators supporting the same simulation methodology, the simulation-specific MOF meta-model may serve as the basis for platform-independent models (PIMs) towards two directions: a) enable SysML to simulation-specific model transformation, and b) enhance interoperability between different simulators independently of the language they use (e.g. C++, Java) and the way they are executed (centrally or distributed). The progressive development of such MOF meta-models for different simulation environments may further contribute to interoperability and exploitation of a common SysML profile, serving all of them.

Benefits:

- The existence of reference meta-models enables QVT-based transformations.
- Interoperability between different simulation tools within a specific simulation framework is promoted.

Considerations:

- Currently, there are not many simulation-specific reference meta-models available.
- Reference meta-models are fully utilized once they have become widely accepted standards. However, this standardization may be a long-lasting procedure.

IV. A Methodology for SysML System Model Simulation

Having these issues in mind, a three-step methodology is proposed for SysML model simulation independently of the simulation framework or language adopted, as depicted in Fig. 1. The system engineer should specify the system model using SysML via a modeling tool and receive valid simulation results from the execution of the corresponding system model in an appropriate simulation environment. As shown in the figure, the three discrete steps are:

Step 1: *Constructing a simulation-specific profile to enrich SysML system models with simulation capabilities.* The system engineer should enrich system models with simulation information. This should be performed during system modeling, within the SysML modeling tool, according to the specific profile. The simulation-specific profile must consist of both a set of stereotypes and constraints. Stereotypes are used to characterize specific SysML model elements, while constraints specify how valid simulation models should be created.

Step 2: *Transforming SysML system models (usually represented in XMI) to simulation-specific system models.* All standard modeling tools facilitate exporting SysML models in XMI format. It is considered as a standard platform-independent representation of UML/SysML models. To enable

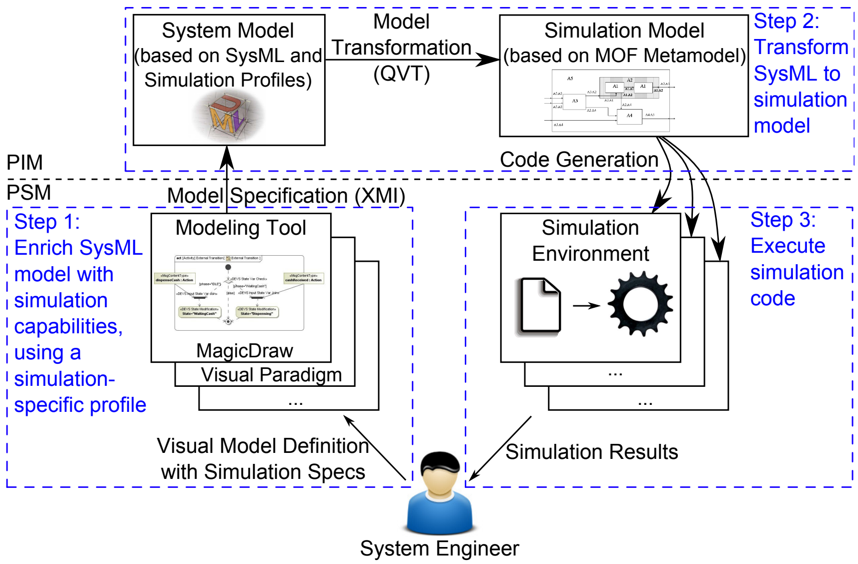


Fig. 1. A methodology for simulating SysML models

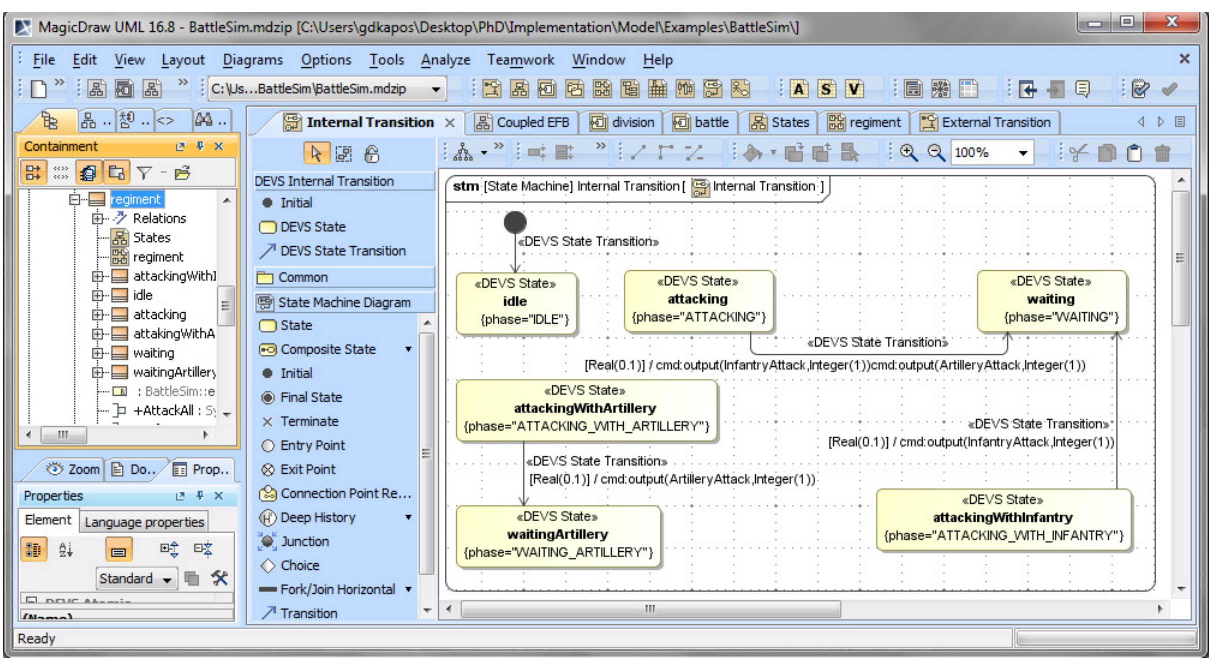


Fig. 2. Enriching SysML models using the DEVS-SysML profile in MagicDraw

model transformation in a standardized fashion, a platform-independent MOF 2.0 meta-model for the corresponding simulation environment should be available. Provided that such a meta-model is available, OMG's Query/View/Transform (QVT) language can be used for model transformation using existing transformation tools [7].

Step 3: *Creating executable simulation code.* The transformation of simulation-specific MOF models to executable simulation code constitutes the last step of the proposed approach. This transformation heavily depends on the target simulation environment. We argue that the maturity of a simulation community, in providing required tools that enable such transformations and defining a MOF meta model, plays a significant role in selecting it. After all, using SysML is closely related to the utilization of existing standards and tools and simulation frameworks should move towards their integration in model-driven engineering frameworks.

V. Simulating SysML Models with DEVS

Based on the presented methodology, the automated DEVS executable simulation code generation from enriched SysML models has been realized [7]. DEVS was selected as the target simulation framework for a series of reasons. DEVS is a formalism allowing a hierarchical and modular description of the models, which lacks a standardized and easy-to-use interface. The authors have identified similarities between DEVS and SysML [18]. Combining alternative SysML diagrams to define all aspects of coupled and atomic DEVS models facilitated the definition of a SysML profile. Moreover, employment of proper stereotypes and constraints have made enrichment of SysML models feasible.

There are several simulation environments for DEVS, as well as attempts to provide generic representations of DEVS models, mainly targeting DEVS simulators interoperability. Many of them, as the one presented in [19], are based on XML representation of DEVS entities. These efforts provided the foundations for the definition of a reference DEVS meta-model in XMI [7].

A SysML profile and corresponding constraints for DEVS simulation framework have been defined and implemented in MagicDraw [17] modelling tool, enabling proper enrichment of SysML system models (step 1). This required to locate the exact parts of the SysML models, where DEVS-specific attributes should be added. Also, specification of behavioral details in a generic, yet precise manner has been a difficult task, in contrary to structural aspects. Fig. 2 illustrates a screen shot of a *state machine diagram* defining an aspect of the DEVS simulation model for a *regiment* component in a *battle* simulation example.

A MOF 2.0 meta-model for DEVS models has been defined, based on previous attempts to standardize DEVS model representations in XML format and mainly [20]. The meta-model is a central element in the overall approach and needs to be finalized first. Changes on the meta-model would affect both the transformation from system models to DEVS models and from DEVS models to executable format. Additionally, a QVT transformation that generates DEVS models from enriched SysML models has been defined (step 2). The definition of the whole set of QVT relations, constituting the transformation, required an effort investment that is not returned for few, simple models. However, applying the transformation in larger, complex models justifies such an approach. Enriched system models have been extracted from MagicDraw in XMI format and transformed to the respective DEVS models. Fig. 3 presents a part of the DEVS model that is generated from the SysML model shown in Fig. 2.

```
<DEVS_ATOMIC>
 <MODEL_NAME text="regiment"/>
 <INPUTS>...</INPUTS>
 <OUTPUTS>...</OUTPUTS>
 <STATES>
  <STATE_SET>
   <STATE_SET_NAME text="STATE SET"/>
   <STATE_SET_VALUES>
    <STATE_SET_VALUE text="idle" initial="true"/>
    <STATE_SET_VALUE text="attackingWithInfantry"/>
    <STATE_SET_VALUE text="attacking"/>
    <STATE_SET_VALUE text="attakingWithArtillery"/>
    <STATE_SET_VALUE text="waiting"/>
    <STATE_SET_VALUE text="waitingArtillery"/>
   </STATE_SET_VALUES>
  </STATE_SET>
  <STATE_VARIABLES>...</STATE_VARIABLES>
 </STATES>
 <INTERNAL_TRANSITION_FUNCTION>
  <CONDITIONAL_FUNCTION>
   <STATE_CONDITION text="attacking"/>
   <TRANSITION_FUNCTION>
    <NEW_STATE text="waiting"/>
    <STATE_VARIABLE_UPDATES/>
   </TRANSITION_FUNCTION>
  </CONDITIONAL_FUNCTION>
  ...
 </INTERNAL_TRANSITION_FUNCTION>
 <OUTPUT_FUNCTION>
  <CONDITIONAL_OUTPUT_FUNCTION state="attacking">
   <PORT_OUTPUTS>
    <SEND port="InfantryAttack">
     <VALUE type="Integer" value="1"/>
    </SEND>
    <SEND port="ArtilleryAttack">
     <VALUE type="Integer" value="1"/>
    </SEND>
   </PORT_OUTPUTS>
  </CONDITIONAL_OUTPUT_FUNCTION>
  ...
 </OUTPUT_FUNCTION>
 <TIME_ADVANCE_FUNCTION>
  <CONDITIONAL_TIME_ADVANCE>
   <STATE_CONDITION text="attacking"/>
   <TIME_ADVANCE>
    <VALUE type="Real" value="0.1"/>
   </TIME_ADVANCE>
  </CONDITIONAL_TIME_ADVANCE>
  ...
 </TIME_ADVANCE_FUNCTION>
 <EXTERNAL_TRANSITION_FUNCTION>
  ...
 </EXTERNAL_TRANSITION_FUNCTION>
</DEVS_ATOMIC>
```

Fig. 3. DEVS model, derived from the enriched SysML model

In order to implement the third step of the methodology, the approach presented in [21] has been utilized. This approach was selected, since it supports an XML-based language, named XLSC, for describing DEVS models that can be executed into

```
<atomicModel name="regiment">
 <statePart>
  ...
 </statePart>
 <propertiesPart/>
 <portsPart>
  ...
 </portsPart>
 <functionsPart>
  ...
  <internalTransitionFunction>
   <action>
    <if>
     <condition>
      <equal>
       <retrieve state="phase"/>
       <string>attacking</string>
      </equal>
     </condition>
     <then>
      <update state="phase">
       <string>waiting</string>
      </update>
     </then>
    </if>
    ...
  </internalTransitionFunction>
 </functionsPart>
</atomicModel>
```

Fig. 4. DEVS executable model in XLSC format

a DEVSJava simulator, promoting model transformation. An XSL transformation from DEVS models in XMI format to executable XLSC simulation models in XML format has also been defined and tested. Fig. 4 presents a part of the *regiment* model in XLSC (executable) format.

VI. Conclusions

Following the proposed guidelines and the three-step methodology, tools of simulating SysML models with DEVS have been successfully applied and implemented. The DEVS-SysML profile and DEVS MOF meta-model were proven to the fundamental elements of the approach, enabling in practice automated code generation. The choice of DEVS as a simulation methodology also derived from the proposed guidelines. Standard model transformation languages, such as QVT, also proved to be efficient, provided that corresponding XML-based tools are available for simulation. Furthermore, they promote interoperability between different simulation methodologies, provided corresponding MOF meta-models are defined.

Future work include the application of tools developed in different domains, such as military and information system simulation, and the support of simulation model libraries. The integration of the proposed DEVS-SysML profile with system design profiles is also under investigation. The definition MOF metamodels for different simulation environments and the exploitation of a common SysML profile, serving all of them, will also be addressed.

VII. Acknowledgements

The authors would like to thank Nicolas Meseth, Patrick Kirchhof, and Thomas Witte for their valuable help. Not only they provided us with their XLSC prototype interpreter, but also they eagerly answered every question we posed.

References

[1] OMG, "Systems Modeling Language (SysML) Specification. Version 1.0," September 2007.

[2] ——, "Model Driven Architecture. Version 1.0.1," *Available online via http://www.omg.org/cgi-bin/doc?omg/03-06-01.pdf*, June 2003.

[3] L. McGinnis and V. Ustun, "A simple example of SysML-driven simulation," in *Winter Simulation Conference (WSC), Proceedings of the 2009*. IEEE, 2009, pp. 1703–1710.

[4] R. Wang and C. Dagli, "An executable system architecture approach to discrete events system modeling using SysML in conjunction with colored petri nets," in *IEEE Systems Conference 2008*. Montreal: IEEE Computer Press, April 2008, pp. 1–8.

[5] W. Schamai, "Modelica Modeling Language (ModelicaML): A UML Profile for Modelica," Tech. Rep., 2009. [Online]. Available: http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-20553

[6] OMG, *SysML and Modelica Integration*, Dec. 2008. [Online]. Available: http://www.omgwiki.org/OMGSysML/doku.php?id=sysml-modelica:sysml_and_modelica_integration

[7] G. D. Kapos, V. Dalakas, M. Nikolaidou, and D. Anagnostopoulos, "An integrated framework for automated simulation of SysML models using DEVS," 2012, under Submission.

[8] B. P. Zeigler, H. Praehofer, and T. Kim, *Theory of Modeling and Simulation*, 2nd ed. Academic Press, 2000.

[9] E. Huang, R. Ramamurthy, and L. F. McGinnis, "System and simulation modeling using SysML," in *WSC '07: Proceedings of the 39th conference on Winter simulation*. Piscataway, NJ, USA: IEEE Press, 2007, pp. 796–803.

[10] O. Schonherr and O. Rose, "First steps towards a general SysML model for discrete processes in production systems," in *Proceedings of the 2009 Winter Simulation Conference*, Austin, TE, USA, December 2009, pp. 1711–1718.

[11] R. Peak, R. Burkhart, S. Friedenthal, M. Wilson, M. Bajaj, and I. Kim, "Simulation-based design using SysML part 1: A parametrics primer," in *INCOSE Intl. Symposium*, San Diego, CA, USA, 2007, pp. 1–20.

[12] R. Peak, C. J. Paredis, and D. R. Tamburini, "The composable object (COB) knowledge representation: Enabling advanced collaborative engineering environments (CEEs), COB requirements & objectives (v1.0)," Georgia Institute of Technology, Atlanta, GA, Technical Report, Oct. 2005.

[13] D. R. Tamburini, "Defining executable design & simulation models using SysML," *Available online via http://www.pslm.gatech.edu/topics/sysml/*, March 2006.

[14] A. A. Kerzhner, J. M. Jobe, and C. J. J. Paredis, "A formal framework for capturing knowledge to transform structural models into analysis models," *Journal of Simulation*, vol. 5, no. 3, pp. 202–216, 2011.

[15] Modelica, *The Modelica Language Specification Version 3.2*, revision 1 ed., February 2012.

[16] IBM, "Rational Software Modeler," *Available online via http://www.ibm.com/developerworks/rational*, 2010.

[17] MG, *SysML Plugin for Magic Draw*, 2007.

[18] M. Nikolaidou, V. Dalakas, L. Mitsi, G.-D. Kapos, and D. Anagnostopoulos, "A SysML profile for classical DEVS simulators," in *Proceedings of the Third International Conference on Software Engineering Advances (ICSEA 2008)*. Malta: IEEE Computer Society, October 2008, pp. 445–450.

[19] S. Mittal, J. L. Risco-Martín, and B. P. Zeigler, "DEVSML: Automating DEVS execution over SOA towards transparent simulators," in *DEVS Symposium, Spring Simulation Multiconference*. ACIMS Publications, March 2007, pp. 287–295.

[20] J. Martín, S. Mittal, M. López-Peña, and J. De la Cruz, "A W3C XML schema for DEVS scenarios," in *Proceedings of the 2007 spring simulation multiconference-Volume 2*. Society for Computer Simulation International, 2007, pp. 279–286.

[21] N. Meseth, P. Kirchhof, and T. Witte, "XML-based DEVS modeling and interpretation," in *SpringSim '09: Proceedings of the 2009 Spring Simulation Multiconference*. San Diego, CA, USA: Society for Computer Simulation International, 2009, pp. 1–9.

System of Systems Information Interoperability using a Linked Dataspace

Edward Curry
Digital Enterprise Research Institute,
National University of Ireland, Galway.
ed.curry@deri.org

Abstract - *System of Systems pose significant technical challenges in terms of information interoperability that require overcoming conceptual barriers (both syntax and semantic) and technological barriers. This paper presents an approach to System of Systems information interoperability based on the Dataspace data management abstraction and the Linked Data approach to sharing information on the web. The paper describes the fundamentals of the approach and demonstrates the concept with a System of Systems for enterprise energy management.*

Keywords: System of systems, interoperability, linked data, energy management.

1 Introduction

Emerging next generation smart environments such as Smart Grids, Smart Cities, and Smart Enterprises are complex systems that require a complete and holistic knowledge of their operations for effective decision-making. Multiple information systems currently operate within these environments and real-time decision support will require a System of Systems (SoS) approach to provide a functional view of the entire environment to understand, optimize, and reinvent processes. The required system of systems will need to connect systems that cross-organizational boundaries, come from multiple domains, (i.e. finance, manufacturing, facilities, IT, water, traffic, waste, etc.) and operate at different levels (i.e. region, district, neighborhood, building, business function, individual). These SoS pose many significant challenges, including the need for flexible mechanisms for information interoperability.

This paper presents an approach to SoS information interoperability based on two emerging trends, Dataspace and Linked Data. A Dataspace is a conceptual approach to information management that supports the co-existence of heterogeneous data with an incremental approach to interoperability. Linked data is a technology that leverages the web architecture to share data in a flexible and incremental manner to reduce technological barriers. When used together the resulting Linked Dataspace can provide a viable approach to SoS information interoperability. This paper discusses the approach with section 2 examining SoS interoperability requirements; section 3 discussing dataspaces and linked data. Section 4 details the role of a linked dataspace for SoS interoperability, with section 5 detailing a case study for an enterprise energy management SoS. Section 6 concludes and discusses future work.

2 SoS Interoperability

Within many modern information systems, interoperability is not seen as a strong requirement within their design. This results in limited interoperability between systems and a high cost associated with system alignment. The classic interoperability challenge can be see within existing enterprises systems. An enterprise can use multiple information systems to support its business processes including sales, marketing, management reporting, research and development, financial reporting, customer loyalty and satisfaction, and similar areas. Integrating these systems is a key challenge for any IT department as the number of systems involved can be significant, and it can be time-consuming and expensive to stitch together the legacy systems so they use common data and processes.

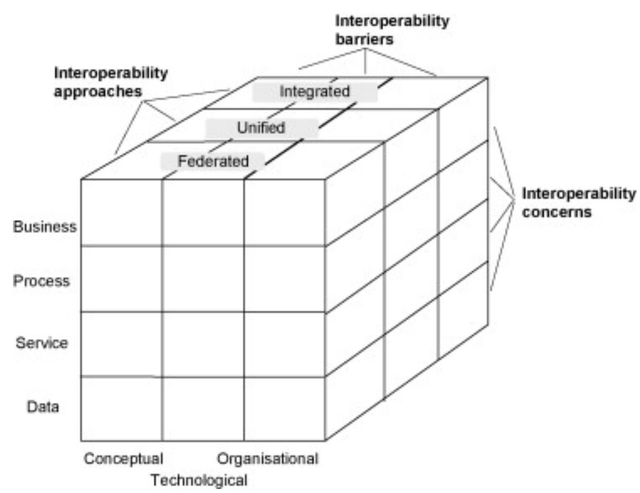


Figure 1. Enterprise interoperability framework [1]

Interoperability is a complex issue with many facets. The INTEROP framework [1], see Figure 1, identifies the barriers and approaches to interoperability. INTEROP identifies three dimensions of interoperability barriers:

1) *Conceptual barriers* represent the syntactic (format) and semantic differences (interpretation of meaning) of exchanged information.

2) *Technological barriers* refer to the incompatibility of information technologies such as protocols, encoding, platforms, or infrastructures.
3) *Organizational barriers* concern the organizational incompatibility of definitions of responsibility, authority, and organizational structures.

Interoperability approaches are the means and solutions to address interoperability barriers. INTEROP classifies interoperability approaches into:

- *Integrated approaches* where a common format for all models exists.
- *Unified approaches* where an upper model exists to facilitate mapping between models.
- *Federated approaches* where no common format exists but rather is done on the fly.

2.1 Types of System of Systems

Many definitions of a SoS exist, within this work we use the definition of a SoS as a collaborative set of systems in which its components are independent dedicated systems that are separately acquired and integrated to form a single system, yet maintain a continuous operational existence independent of the collaborative system [2]. A SoS can take many forms, SoS classifications [3] include *Virtual*, *Collaborative, Acknowledged,* and *Directed*.

- *Directed SoS* are built and centrally managed to fulfil specific purposes. Component systems maintain an ability to operate independently, but their normal operational mode is subordinated to the central managed purpose.
- *Acknowledged SoS* have defined objectives and dedicated resources. The constituent systems retain their independent ownership and objectives. Changes in the systems are based on collaboration between the SoS and the systems.
- *Collaborative SoS* component systems interact voluntarily to fulfil agreed upon central purposes. The central players collectively decide the means of enforcing and maintaining standards.
- *Virtual SoS* have no central management authority and a centrally agreed upon purpose. Virtual SoS rely upon relatively invisible mechanisms to maintain them.

2.2 SoS Interoperability Challenges

The differing SoS types present different interoperability requirements. On the one hand a Directed SoS with dedicated resources and central coordination can implement an Integrated or Unified approach to interoperability (i.e. Many-to-One). At the other extreme a Virtual SoS has no central authority or resources and may require a Federated approach to interoperability (i.e. Many-to-Many), in such scenarios a constitute system may not even be aware they are involved in a SoS.

The existing system design mind-set views interoperability as an external responsibility, and outsources interoperability to external systems. Within SoS, interoperability needs to be a fundamental requirement to their design and operation. The key challenge is to simplify interoperability without increasing complexity, hierarchy, control, or acquisition cost. This will require a change in mind-set of system design to embrace interoperability concerns. An effective interoperability approach for SoS will minimize complexity and ensure constituent systems do not need to be re-engineered as other constituent systems are added, removed, modified, or replaced. To this end, a common interoperability infrastructure is needed to support flexible SoS information interoperability. The infrastructure needs to be a lightweight technology platform supporting the minimum set of requirements to enable the maximum flexibility of the interoperability approach. In this initial work we focus on the foundation level of the INTEROP framework and investigate information interoperability for SoS. However, improving the conceptual and technical interoperability of systems is an important step to support organizational interoperability.

3 Background

3.1 Dataspace

A Dataspace is an emerging data management architecture that is very distinct from current approaches. The dataspace approach recognizes that in large-scale integration scenarios, involving thousands of data sources, it is difficult and expensive to obtain an upfront unifying schema across all sources [4]. Dataspace shifts the emphasis to providing support for the co-existence of heterogeneous data that does not require an upfront investment into a unifying schema. Data is integrated on an *"as needed"* basis with the labor-intensive aspects of data integration postponed until they are required. Dataspaces reduce the initial effort required to setup data integration by relying on existing matching and mapping generation techniques. This can result in a loosely integrated set of data sources, when tighter integration is required it can be achieved in an incremental *"pay-as-you-go"* fashion by more closely integrating the required data sources. Similarly a dataspace may only provide weak guarantees of consistency and durability. As stronger guarantees are desired, more effort can be put into making agreements among the various systems, see Table 1.

Table 1. Dataspace comparison

	Model	**Formats**	**Control**	**Query**	**Integration**
DBMS	Relational	Homog.	Complete	Precise	Explicit
Dataspace	All	Heterog.	Partial	Approx.	Implicit/ Incremental

3.2 Linked Data

Information integration projects typically focus on one-off point-to-point integration solutions between two or more systems in a customized but inflexible and ultimately non-reusable manner. The fundamental concept of Linked Data is that data is created with the mind-set of sharing and reuse. Emerging from research into the Semantic Web, Linked data leverages the existing open protocols and W3C standards of the Web architecture for sharing structured data on the web. Linked Data proposes an approach for information interoperability based on the creation of a global information space [5]. The main components are 1) *Universal Resource Identifiers* (URIs) to name things, 2) *Resource Description Framework* (RDF) for representing data, 3) *Linked Data principles* for publishing, linking, and integration, 4) *Vocabularies* to establish and share understanding, and 5) *Bottom-up incremental agreement*. Linked Data has the following advantages:

- Separate systems that are designed independently can be later joined/linked at the edges.
- Interoperability is added incrementally when needed and where it is cost-effective.
- Data is expressed in a mixture of vocabularies.

3.3 Linked Open Data

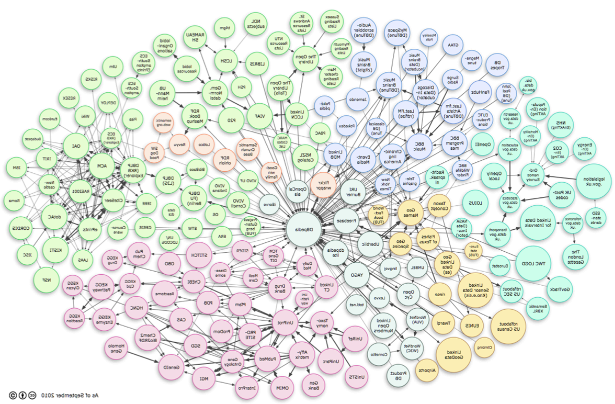


Figure 2. The linked open data cloud

Linked Data is facilitating the publishing of large amounts of structured data on the web. The resulting Web of Data can be considered as a web scale dataspace supported by Semantic Web technologies. The Linked Open Data Cloud, see Figure 2, represents a large number of interlinked datasets that are being actively used by industry, government and scientific communities [6]. The cloud has been growing considerably; as of September 2011 it has over 300 open data sets with more than 35 billion facts, interlinked by 500 million links.

3.4 Linked Data Principles

Linked data technology uses web standards in conjunction with four basic principles for exposing, sharing and connecting data. These principles are:

1. Use URIs as names for things: the use of Uniform Resource Identifier (URI) (similar to URLs) to identify things such as a person, a place, a product, an organization, an event or even concepts such as risk exposure or net profit, simplifies reuse and the integration of data.

2. Use HTTP URIs so that people can look up those names: URIs are used to retrieve data about objects using standard web protocols. For an employee this could be their organization and job classification, for an event this may be its location time and attendance, for a product this may be its specification, availability, price, etc.

3. When someone looks up a URI, provide useful information using the standards: when someone looks up (dereferences) a URI to retrieve data, they are provided with information using a standardized format. Ideally in Semantic Web standards such as RDF.

4. Including links to other URIs so that people can discover more things: retrieved data may link to other data sources, thus creating a data network e.g., data about a product may link to all the components it is made of, which may link to their supplier.

3.5 Resource Description Framework

The Resource Description Framework (RDF) is the basic machine-readable representational format used to represent information. RDF is a general method for encoding graph-based data that does not follow a predictable structure. RDF is schema-less and self-describing, meaning that the labels of the graph describe the data itself. Data and facts are specified as statements and are expressed as atomic constructs of a subject, predicate and object, also known as a triple. The statement "Edward Curry is the Occupant of Room 202e" is expressed in triple format as:

Subject - "Edward Curry"
Predicate - "is the Occupant of"
Object - "Room 202e"

RDF is designed for use in web-scale decentralized graph data models. For this reason the statement parts need to be identified so that they can be readily and easily reused. RDF uses URIs for identification, expressing the previous statement in RDF then becomes:

http://www.deri.ie/about/team/member/edward_curry#me
http://vocab.deri.ie/rooms#occupant
http://lab.linkeddata.deri.ie/2010/deri-rooms#r202e

4 Linked Dataspace for SoS

This paper proposes that Dataspaces together with Linked Data can be the basis for a SoS interoperability approach that would help to overcome technical and conceptual barriers to information interoperability.

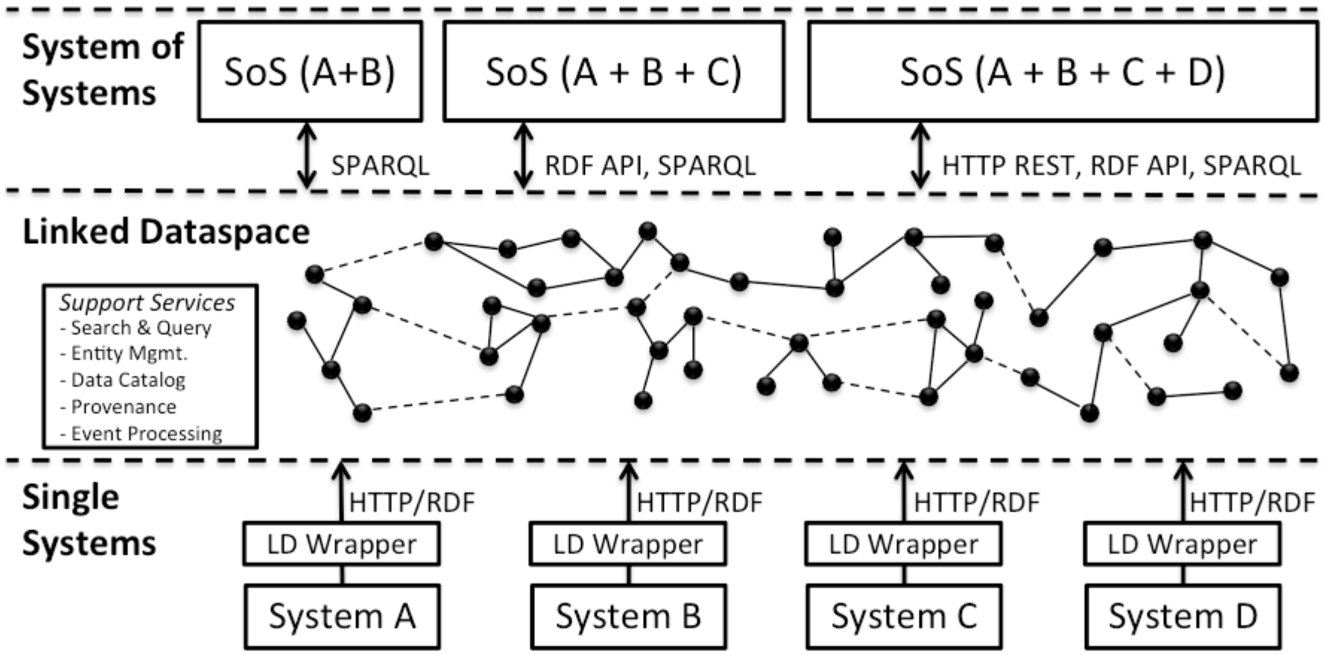


Figure 3. Linked dataspace for system of systems

A linked dataspace can be accommodated with minimal disruption to existing information infrastructure, as a complimentary technology for data sharing, it should not be seen as a replacement for current infrastructure (relational databases, data warehouses, etc.). The objective is to expose the data within existing systems, but only link the data when its needs to be shared. Once a system exposes its data within the linked dataspace it can participate in many SoS, knowingly or unknowingly

4.1 Architecture

Figure 3 shows the proposed placement of a linked dataspace layer for interoperability of SoS. The linked dataspace serves as an independent layer placed above the existing systems layer, but below the system of systems layers. The approach can support the three approaches to interoperability (Integrated, Unified, Federated) with linked data providing a common syntactic and access protocols. The main components of the architecture are the wrapper on existing systems, the Linked dataspace consisting of a linked data cloud & support services, and the resulting system of systems.

4.2 Linked Data Wrappers

At the bottom of the architecture are the existing operational legacy information systems. Wrappers perform the "RDFization" process, which transforms multiple formats and legacy data and lifts it to the dataspace.

4.3 Linked Data Cloud

The Linked Dataspace links at the information-level (data) not the infrastructure-level (system) by focusing more on the conceptual similarities (shared understanding) between information. This is achieved by following an entity-centric approach that focuses on the concepts that exist within the systems, for example, business entities like employees, products, customers, intellectual property, assets, etc. Entities within the dataspace are enriched with data from multiple systems. This results in a cloud of interlinked resources that reflect virtual or actual entities with links to relevant knowledge and contextual information from across all the information systems that have exposed linked data. Sources may be added in an incremental manner to the cloud where they can be reused. Each entity within the cloud has a dereferenceable URI that returns data in a machine-readable format describing the resource identified.

Agreement on data schemas can be achieved in a number of ways, from minimal agreement on a small number of simple vocabularies, to the use of agreed upon domain ontologies. The most appropriate approach will depend on the type of SoS involved and the interoperability approach undertaken. In keeping with the minimal agreement ethos of a dataspace, and the triple structure of RDF, the Entity- Attribute-Value (EAV) model is appropriate for implementing an entity-centric dataspace [7]. EAV can facilitate a *pay-as-you-go* methodology for data integration allowing the incremental enrichment of entity profiles. Multiple triples can be joined together to build up an entity-centric graph of information. Figure 4 provides an example of a entity-centric linked data graph segment, the example shows how relevant energy data can be exposed and linked to build up the energy profile of a user from across corporate, building and office IT systems.

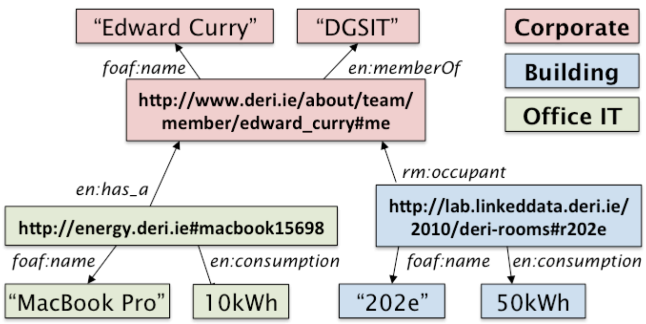


Figure 4. Entity-centric linked data graph

4.4 Support Services

Dataspace support services are designed to simplify the consumption of the linked data cloud by encapsulating common services for reuse. Some example support services used in this work are:

- *Entity Management Service* to improve data quality and inter-linkage between entity data scattered among legacy systems. The EMS can leverage automatic entity consolidation algorithms that are supported by humans for collaborative data management [8].
- *Complex Event Processing* engine [9] is used to assess situations of interest that are encoded as event\action rules. Real-time information from sensors networks are also supported via the Semantic Sensor Network Ontology [10].
- *Data Catalogue and Provenance service* [11] to query the catalogue about data sources with specific attributes such as freshness and publisher, and track data back to its origin.
- *Search and Query services* [12] allow users to interact with the dataspace using structured or natural language interfaces.

4.5 System of Systems

At the top of the architecture is the SoS that consume the resulting data and events from linked data cloud. A SoS developed using the Linked Dataspace is now discussed.

5 Enterprise Energy Management

An enterprise can use energy in many ways, from manufacturing and logistics, to building and IT. Within each of these areas an enterprise may have information systems deployed to help manage energy usage. However, in order to manage energy holistically, a real-time energy view of the entire enterprise is needed to provide an ability to understand, change, and reinvent business processes to better support energy efficient practices. This requires a SoS with support for a cross-domain view that helps users to interpret the information and understand how it can be used to optimize energy usage.

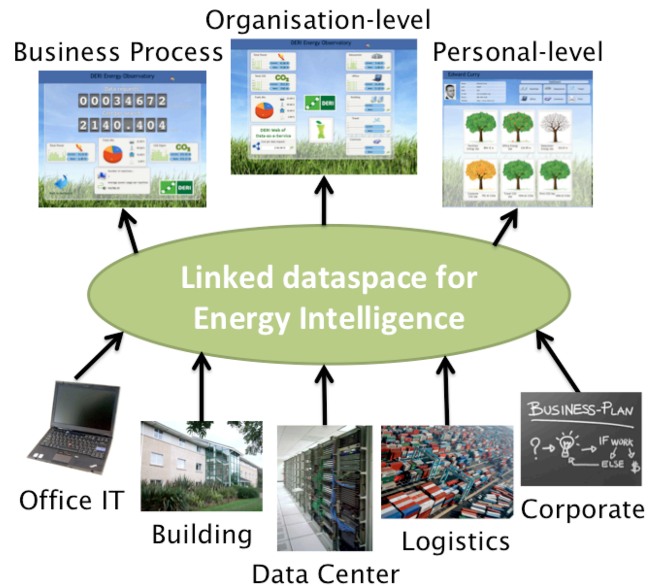


Figure 5. Enterprise energy system of systems

To demonstrate a linked dataspace for SoS interoperability we have implemented an enterprise energy management SoS within the Digital Enterprise Research Institute (DERI). The resulting Linked dataspace for Energy Intelligence (LEI), see Figure 5, enables a holistic SoS view of the energy management within DERI (facilities, office IT, Data Centre, business travel, commute, etc.), which can be used to manage energy at different-levels (organization, business process, individual) and from different domains perspectives (energy, finance, sustainability, human resources, etc). The LEI dataspace connected over 10 existing systems and was used to deliver a number of other energy management tools including buildings, office IT, data centre, and cloud services. In all cases the Linked Dataspace provided the linkages required to integrate operational legacy information system, and real-time data streams in a straightforward and flexible manner.

6 Related Work

The Global Information Grid (GIG) represents a globally interconnected, end-to-end set of information capabilities and processes for collecting, processing, and managing information on demand. The GIG follows a network-centric approach to data sharing, replacing the "one-to-one," interoperability approach (protocols, standards, etc.) with a "many-to-one" interoperability approach in which each system will interface with the GIG. The Net-Centric Adapter for Legacy Systems (NCALS) [13] is designed to enable net-centric operations, it focuses on enabling legacy systems to interoperate with the GIG. NCALS provides a service-oriented architecture connection for a legacy system to the GIG. However, it does not modify the legacy components to comply internally with the net-centric technical standards. It allows the legacy system architecture to remain largely undisturbed.

The Global Earth Observation System of Systems (GEOSS) [14] seeks to be a global public infrastructure that generates comprehensive, environmental data, information, and analyses for a wide range of users. The GEOPortal disseminates information and analyses directly to users, making it easier to integrate diverse data sets, identify relevant data and portals of contributing systems, and access models and other decision-support tools. Access to data is through service interfaces with interoperability specifications agreed among contributing systems.

Both the GIG and GEOSS are large-scale SoS taking different implementation approaches. We believe a linked dataspace can have a role for data interoperability within such systems and is particularly suited where the web architecture is already a part of the SoS. Within GEOSS, EuroGEOSS is currently investigating RDF as a basis for semantic interoperability.

7 Summary

Within system of systems significant technical challenges exist in terms of information interoperability that require overcoming conceptual (syntax and semantics) and technological barriers. This paper presented an approach to System of Systems information interoperability based on the dataspace approach to information management and linked data approach to sharing information on the web. The resulting linked dataspace supports the co-existence of heterogeneous data with an incremental approach to interoperability. The paper describes the fundamentals of the approach and demonstrates the concept with a system of systems deployment for Enterprise Energy Management.

Future work will focus on the development of appropriate support services for linked dataspaces using approximation techniques [15], [16], and the interoperability of system services as RESTful services.

Acknowledgements: This work has been funded by Science Foundation Ireland under Grant No. SFI/08/CE/I1380 (Lion-2).

References

[1] D. Chen, G. Doumeingts, and F. Vernadat, “Architectures for enterprise integration and interoperability: Past, present and future,” Computers in Industry, vol. 59, no. 7, pp. 647-659, 2008.

[2] E. Rechtin, “The art of systems architecting,” IEEE Spectrum, vol. 29, no. 10, pp. 66-69, 1992.

[3] M. W. Maier, “Architecting Principles for Systems-of-Systems,” Systems Engineering, vol. 1, no. 4, pp. 267-284, 1998.

[4] M. Franklin, A. Halevy, and D. Maier, “From databases to dataspaces: a new abstraction for information management,” Sigmod Record, vol. 34, no. 4, pp. 27-33, 2005.

[5] T. Heath and C. Bizer, “Linked Data: Evolving the Web into a Global Data Space,” Synthesis Lectures on the Semantic Web: Theory and Technology, vol. 1, no. 1, pp. 1-136, Feb. 2011.

[6] E. Curry, A. Freitas, and S. O’Riáin, “The Role of Community-Driven Data Curation for Enterprises,” in Linking Enterprise Data, D. Wood, Ed. Boston, MA: Springer US, 2010, pp. 25-47.

[7] P. M. Nadkarni, L. Marenco, R. Chen, E. Skoufos, G. Shepherd, and P. Miller, “Organization of heterogeneous scientific data using the EAV/CR representation.,” Journal of the American Medical Informatics Association, vol. 6, no. 6, pp. 478-93, 1999.

[8] U. ul Hassan, S. O’Riain, and E. Curry, “Leveraging Matching Dependencies for Guided User Feedback in Linked Data Applications,” in 9th International Workshop on Information Integration on the Web (IIWeb2012), 2012.

[9] S. Hasan, E. Curry, M. Banduk, and S. O’Riain, “Toward Situation Awareness for the Semantic Sensor Web: Complex Event Processing with Dynamic Linked Data Enrichment,” in 4th International Workshop on Semantic Sensor Networks 2011 (SSN11), 2011, pp. 69-81.

[10] L. Lefort, C. Henson, K. Taylor, and P. Barnaghi, “Semantic Sensor Network XG Final Report, W3C Incubator Group Report (28 June 2011),” 2011.

[11] A. Freitas, T. Knap, S. O’Riain, and E. Curry, “W3P: Building an OPM based provenance model for the Web,” Future Generation Computer Systems, vol. 27, no. 6, pp. 766-774, Jun. 2011.

[12] A. Freitas, J. Oliveira, S. O’Riain, E. Curry, and J. Pereira da Silva, “Querying Linked Data using Semantic Relatedness: A Vocabulary Independent Approach,” in Proceedings of the 16th International Conference on Applications of Natural Language to Information Systems, NLDB 2011, 2011, vol. 6716, pp. 40-51.

[13] A. Thomas, T. Turner, and S. Soderlund, “Net-Centric Adapter for Legacy Systems,” IEEE Systems Journal, vol. 3, no. 3, pp. 1-7, 2009.

[14] M. L. Butterfield, J. S. Pearlman, and S. C. Vickroy, “A System-of-Systems Engineering GEOSS: Architectural Approach,” IEEE Systems Journal, vol. 2, no. 3, pp. 321-332, 2008.

[15] S. Hasan, S. O’Riain, and E. Curry, “Approximate Semantic Matching of Heterogeneous Events,” in 6th ACM International Conference on Distributed Event-Based Systems (DEBS 2012), 2012.

[16] A. Freitas, E. Curry, J. G. Oliveira, and S. O’Riain, “Querying Heterogeneous Datasets on the Linked Data Web: Challenges, Approaches, and Trends,” IEEE Internet Computing, vol. 16, no. 1, pp. 24-33, 2012.

Business Interactions Modeling for Systems of Systems Engineering: Smart Grid Example

Edin Arnautovic, Davor Svetinovic
Computing and Information Science
Masdar Institute of Science and Technology
Abu Dhabi, United Arab Emirates
{earnautovic, dsvetinovic}@masdar.ac.ae

Ali Diabat
Engineering Systems and Management
Masdar Institute of Science and Technology
Abu Dhabi, United Arab Emirates
adiabat@masdar.ac.ae

Abstract - *Systems of systems typically contain many technical systems but also span across and affect different legal and economic entities (e.g., companies, public bodies, etc.). The economic entities are involved in exchanges of goods, services and value. To develop an envisioned or to understand an existing system of systems, systems engineers should also be able to model and analyze these value exchanges and the business interactions related to them. Business interactions are strongly related not only to the requirements on the systems of systems, but also on the systems within them. In this paper we propose to use business models, business process models and value network models to analyze business interactions in the context of systems engineering, and to elaborate how they can be related to requirements. We examine our proposal on an example system of systems: a smart grid.*

Keywords: System of Systems, Business Process Modeling, Value Network Models, Requirements Engineering.

1 Introduction

Business interactions and the transfer of economic value between actors in complex systems have not received enough attention in systems engineering in the past. Besides technical issues, systems engineers focused typically only on few economic aspects such as costs for the systems' development, or return of investment.

Systems of systems (SoS) typically include many technical systems but also span across different economic and legal entities, which are involved in exchanges of goods, services and value. To develop an envisioned, or to understand an existing SoS, systems engineers should also be able to model and analyze these exchanges and the business interactions related to them.

In this paper we argue that the business interactions including flows of goods, services and money in an envisioned SoS should be analyzed during the systems engineering process. This flow usually takes place between different enterprises participating in a SoS. Examples of technical systems within SoS are, e.g., energy systems, information technology systems, transportation systems, buildings, etc., and the economic and legal entities are companies, infrastructure providers, municipalities, etc. Economic and legal entities are themselves socio-technical systems and contain people, IT systems and organizational structures. An economic or legal entity is often related to different regulative policies and infrastructures and it can be a part of a larger organization of customers and suppliers.

Just as the requirements engineering is essential and difficult for the single-system development, it is even more challenging for the SoS. For example, some of the systems that are contained in a SoS may be legacy systems that are developed based on particular sets of requirements and functionality in mind and might be hard to integrate in a SoS.

According to their *managerial control* SoS can be divided into [1]:

- **Directed,** which are centrally managed during long term operation**;**
- **Collaborative,** where a central management organization exists but it is only responsible to develop standards and does not have the power to run the system; and
- **Virtual,** which do not have a central management authority.

We consider *collaborative* and *virtual* types of SoS, where there is no central control entity and the decision making is done in cooperation among many actors. Having business interactions among these actors represented in more or less formal models will improve the *cooperative decision making* needed for the development of such systems.

2 Background and Related Work

A basis for any consideration of business aspects of an organization and related technical systems is its *business model.* Business model consists of four elements that together create and deliver value: customer value proposition, profit formula, key resources, and key processes [2]. It also defines the key external relationships of the organizations, which are further on the basis for the concrete value exchanges between organizations and the inter-organizational business processes. If the rationale for the establishment of these different external relationships between organizations is well understood, the business interactions can be better analyzed and designed.

To integrate the concepts of *value* into requirements engineering, the *e^{3}value* methodology [3] is proposed, where in addition to "pure" requirements engineering, the analysts should perform an analysis of economic value creation, distribution and consumption in a *value network*. A value network represents a web of direct and indirect ties among various participants delivering value either to their immediate customer or to the end consumer. The e^{3}value methodology includes so-called value network models (VNM), which are a graphical, conceptual representations showing what is offered to whom and what is requested for that in return in the economic sense.

Business process models have attracted a lot of attention in the field of e-commerce and e-services. A *business process* is a set of *activities* performed by *actors* playing particular *roles*, consuming some *resources* and producing others [4]. Activities are related to *events* and can react to, or generate them. The actors operate within an *organization* and the organizations perform specific *business functions.* The most commonly used language for business process modeling is the Business Process Modeling Notation (BPMN) [5]. BPMN can represent control and data flow, activities, and decision points in a business process. Diagrams made using BPMN can be seen as a flowchart with constructs for business process modeling, e.g., AND-split, AND-join, XOR-split, XOR-join, and event-based choice. Another relevant methodology for business process representation that is based on business protocols is Amoeba [6]. Protocols capture the interactions among autonomous parties via *commitments* and deal with the evolution of requirements via an application of protocol composition. Process models can be used for describing, analyzing, or enacting a process with the former two being more relevant for SoS engineering.

The work on *Stakeholder Value Networks* [7] is related to our approach, in particular in analyzing the stakeholders that are involved in the value exchange during the system's runtime. Our work is also related to requirements engineering for SoS. For example Yang *et al.* [8] use semi-autonomous agents to model the elements in SoS and simulate the emergent behaviors, and use the results to examine the SoS requirements. Other related area is the *Organizational Systems Engineering,* presented by Hubbar *et al.* [9]. However, they focus on the general relations between organizations and systems rather than on the business interactions as we do in this paper. Rhodes *et al.* [10] present an interesting research that deals with an enterprise as a SoS (the "SoS enterprise"). We look at the interactions of enterprises within SoSs.

Integrating business and technical aspects of a complex system is often challenging. For the smart grid, a systematic analysis of business domains and their realization in technical domains is presented in [11].

3 Business Interactions in SoS

As mentioned above, most SoSs include not only technical systems and components, but also different economic actors (e.g., companies, customers, regulation agencies). To first understand a system from the commercial perspective, the modeling and analysis of business interactions should be performed together with (or even prior to) the analysis and specification of systems requirements. Focusing too early on technical properties can lead to suboptimal solutions that can seem reasonable from the technical perspective, but do not satisfy the economic (or, e.g., environmental) requirements of all stakeholders in an optimal way. For the analysis of business interactions we propose to use two forms of business interaction models (BIMs): value networks models (VNM), and business processes. BIMs can improve the requirements engineering and architecting of SoS.

For example, if we associate the value object in a VNM with the flow of money and goods or services per time unit, we can reason about the profitability of a corresponding actor and we can evaluate several scenarios, e.g., by associating different amounts of money to value objects. Iteratively, we can investigate different versions of a VNM and choose the most suitable one.

Using business process models, we can go deeper and model the exchange of concrete messages between economic entities and the corresponding systems in business process model, and see if, e.g., they are compliant with the business procedures within a single system. The experts can analyze such business process models for their feasibility, complexity and potential areas for improvement. Another way to analyze such processes is with business process simulation [12]. Such, business process models can be analyzed according to some metrics such as costs or time. In addition, simulation can be used to investigate the consequences of process changes by modeling and simulating different variants of one process model, e.g., using the "what-if" questions.

If business processes are supported by IT systems and the corresponding events generated during interactions are saved, e.g., in the *logs, business process mining* [13] techniques can be used to monitor and measure the performance of the running processes. The information acquired during measurement can be utilized to improve existing processes and/or introduce new processes to meet the to meet the requirements of the SoS.

Complex SoSs are typically not designed by one central organization, and they emerge through cooperation, so that one entity cannot create business models and resulting business interactions for the complete SoS and all its sub-systems. Nevertheless, the entities within a SoS can jointly develop rules, regulations, which, e.g., could drive standards and technical protocols. They should also work on potential conflicting requirements. In addition, BMIs can be of interest for government regulators and standard bodies.

Figure 1 shows the relation between BMIs and SoS requirements. The relations between the SoS requirements and requirements on each single sub-system have been recognized as the important, yet still not sufficiently investigated [14]. If value exchanges between actors would lead to high profits for some actors but would not acceptable for others, or, e.g., increase general environmental impact, the design of BMIs has to be reconsidered. Such redesign can then result in new or changed requirements on the SoS. This again drives the requirements on each single sub-system within the SoS. However, many of the current and future SoSs are not designed from scratch and the SoS requirements (or even a running SoS) would exist before the BIMs are created. Therefore, this relation is bidirectional. Some approaches have even proposed formal or even automatic *transformation* of requirements models into business processes (an overview of such approaches is given by Decreus *et al.* in [15]).

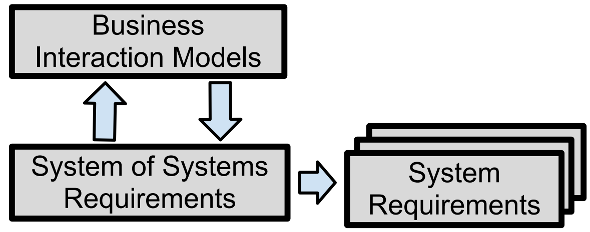


Figure 1. Business Interaction Models and Requirements.

After identifying the requirements, they have to be carefully *engineered* and modeled, e.g., using goal driven approaches such as KAOS [16] and i* [17], use cases [18] or natural language. Using such models, the requirements are analyzed for their consistency, potential conflicts, etc. If a change of some requirements is needed, the affecting business process models have to be modified and investigated again. Special attention has to be given to the interoperability requirements, which fulfillment enable interactions in business interaction models. Another issue is that for a simple system, it is typically clear who is responsible for the requirements elicitation and management (e.g. a chief systems engineer or program manager). For a complex SoS however, it might not be clear who the *owner* of requirements is.

4 Example of Business Interactions in Smart Grid

Smart grid is a general term to describe the future complex SoS responsible for the delivery of electric energy. Smart grid includes many subsystems such as central and distributed generators, high-voltage network and distribution system, industrial and building automation systems, energy storage installations, smart meters, etc. It also includes many economic actors such as customers, energy producers, distributors, service providers, etc.

National Institute of Standards and Technology presents a conceptual model for the smart grid and defines seven different domains: bulk generation, transmission, distribution, customer, service provider, markets, and operations domains [19]. The domains are connected with communication networks and energy power lines. The relation between business actors and domains can be diverse. The economic actors can span over several domains, and each domain can contain several actors. For our example we focus on the operations, service provider, and customer domain.

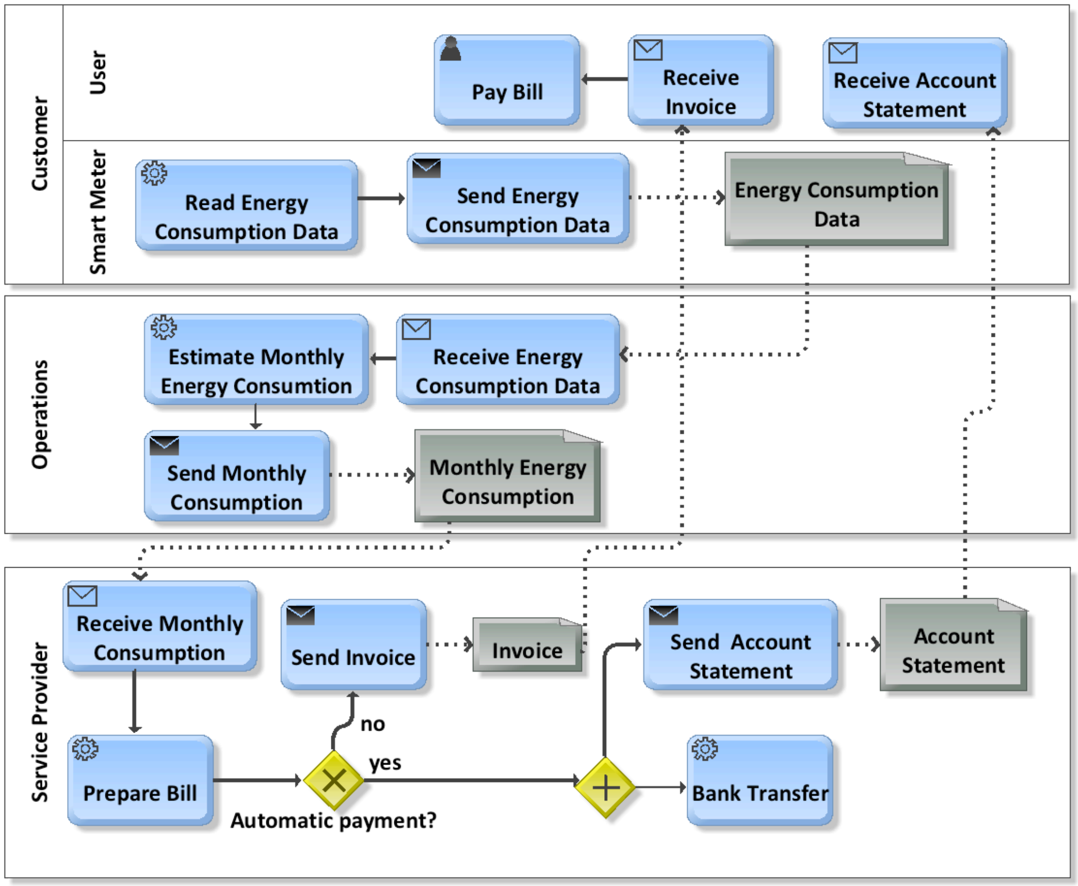


Figure 2. Energy Billing Business Process

Traditionally, an electric utility company was responsible for the generation, transmission, and distribution of electrical energy. This included also the customer relationship management tasks such as account management and the billing of electrical energy consumption. With the deregulation of the energy sector and the envisioned smart grid, this will be changed. If we apply the *unbundling* business model pattern [20], which advises separating the three different types of businesses: customer relationship management, infrastructure business and product innovation business, we will get separate business actors: Service Provider dealing with the dealing with the customer relationships and product innovation (latter not relevant for our example) and the Operations dealing with the infrastructure business (for the sake of the simplicity of our example, we omit product innovation business type). This is in line with the smart grid domains mentioned above.

The transfers of money and services between three actors have to be analyzed in *value network models*. Most importantly, a profitability analysis for each actor can be estimated using the expected net cash flow, based on the actor's need. To calculate profitability of each actor, a formula to each value object exchanged has to be assigned, and based on it, the total value of the outgoing objects from the total value of incoming objects has to be extracted. Kartseva *et al.* [21] have made such investigation for the liberalized energy market and the distributed power generation (two scenarios, which are some of the most important drivers for the smart grid). We have analyzed the utilization of value network models for sustainable systems on the example of Masdar City [22].

Resulting from the separation of business models and the organizations, the business processes have now to span across organizational boundaries and the systems related to them. From the systems engineering perspective, now we have interactions between systems within a SoS. Out of this separation, several new or changed business processes emerge. In Figure 2, we show an example of a business process of energy *billing* and include corresponding business interactions. The business process interactions now take place between three entities: Customer, Service Provider and Operations. Further, the Customer contains two actors: the actual human user and the smart meter. The smart meter is responsible for the automatic readings of the energy consumption and communicates with the corresponding systems of the Operations entity. Such business process representation has a global perspective (also known as a *choreography*), where the actors act as peers without central control and each peer is responsible for the execution of its activities.

So, if the business process and such proposed separation of actors prove to be satisfactory after analysis and simulation, they can be used as a basis to drive the systems requirements. For example, this might lead to new requirements on regulations concerning the legal status and practices of newly identified actors. From the technical viewpoint, this leads to new requirements on the systems associated with the actors. For example, IT systems of the *Operations* actor will have to able to perform activities *Send Invoice* and *Send Account Statement* and be able to communicate with the actor *Customer*, and also be able to communicate with the *Bank* (this actor is external to the SoS and not represented in the figure). Thus, there are new requirements on the interfaces of these systems. Moreover, there are new requirements on the systems' functionality to perform the actual billing activities (*Prepare Bill*, etc.). In addition, there is a requirement to introduce new sub-system (smart meter).

5 Conclusion

For complex SoSs that can include several economic organizations dealing only with technical aspects is not sufficient. To develop and maintain successful and long-lasting SoS, it is important to integrate technical and organizational aspects. As an important task within SoS engineering, we propose to analyze business interactions using business interaction models such as value network models and business process models. We presented a simple example of interactions between the smart grid domains.

In our future work we will work on more complex smart grid scenarios such as distributed energy generation and energy markets. Moreover, we will investigate how to relate business models, business processes and value networks with their environmental impact (e.g., carbon emissions) and how we can use these models to analyze and reduce the environmental impact of a SoS.

References

[1] Mark W. Maier, "Architecting principles for systems-of-systems," *Systems Engineering* Vol. 1, No. 4, 1998, pp. 267-284.

[2] Mark W. Johnson, Clayton M. Christensen, and Henning Kagermann. "Reinventing Your Business Model,"*Harvard Business Review* 86.12, 2008, pp. 50-59.

[3] J. Gordijn, H. Akkermans, "Value-based requirements engineering: exploring innovative e-commerce ideas," *Requirements Engineering Journal* 8(2), 2003, pp. 114-134.

[4] Hafedh Mili, Guy Tremblay, Guitta Bou Jaoude, Éric Lefebvre, Lamia Elabed, and Ghizlane El Boussaidi, "Business process modeling languages: Sorting through the alphabet soup," *ACM Comput. Surv.* 43, 1, 2010, 56 pages.

[5] OMG. Business Process Modeling Notation (BPMN) Version 1.0. OMG Final Adopted Specification. OMG, February 2006. Available via http://www.bpmn.org/

[6] Nirmit Desai, Amit K. Chopra, and Munindar P. Singh, "Amoeba: A methodology for modeling and evolving cross-organizational business processes," *ACM Trans. Softw. Eng. Methodol.* 19, 2, Article 6, October 2009, 45 pages.

[7] Wen Feng, Edward F. Crawley, Olivier de Weck, Rene Keller and Bob Robinson, "Dependency Structure Matrix Modelling for Stakeholder Value Networks - Managing Complexity by Modelling Dependencies," - *Proceedings of the 12th International DSM Conference* 2010, Cambridge, UK, 2010, pp. 3-16

[8] Ke-wei Yang, Ying-wu Chen,Yan-jing Lu, Qing-song Zhao, "The study of guided emergent behavior in system of systems requirement analysis," *5th International Conference on System of Systems Engineering (SoSE), 2010*.

[9] E-M. Hubbard, C.E. Siemieniuch, M.A Sinclair, A. Hodgson, "Working towards a holistic organisational systems model," *5th International Conference on System of Systems Engineering (SoSE),* 2010.

[10] D.H. Rhodes, A.M. Ross, D.J. Nightingale, "Architecting the system of systems enterprise: Enabling constructs and methods from the field of engineering systems," *3rd Annual IEEE Systems Conference, 2009*, pp.190-195.

[11] Husam Suleiman, Ahmed Khaja, Nauman Zafar, Emine Phillips, Davor Svetinovic, and Olivier L. de Weck, “Inter-Domain Analysis of Smart Grid Domain Dependencies and Mappings Using Domain-Link Matrices,” *IEEE Transactions on Smart Grid,* vol.3, no.2, June 2012, pp.692-709.

[12] W.M.P. van der Aalst. “Business Process Simulation Revisited,” In J. Barjis, editor, *Enterprise and Organizational Modeling and Simulation*, volume 63 of *Lecture Notes in Business Information Processing*, Springer-Verlag, Berlin, 2010, pp. 1-14.

[13] W.M.P. van der Aalst, “Intra- and Inter-Organizational Process Mining: Discovering Processes within and between Organizations,” In P. Johannesson and J. Krogstie, editors, *IFIP Conference on the Practice of Enterprise Modelling (PoEM 2011)*, volume 92 of *Lecture Notes in Business Information Processing*, Springer-Verlag, Berlin, 2011, pp. 1-11.

[14] Judith Dahmann and Kristen Baldwin. “Implications of systems of systems on system design and engineering.” *2011 6th International Conference on System of Systems Engineering*, 2011, pp.131–136.

[15] Decreus, K., Snoeck, M. and Poels, G, “Practical Challenges for Methods Transforming i* Goal Models into Business Process Models,” *17th IEEE International Requirements Engineering Conference*, 2009, pp. 15-23.

[16] A. Dardenne, A. van Lamsweerde and S. Fickas, “Goal-Directed Requirements Acquisition.,” *Science of Computer Programming*, 20(1-2), April 1993.

[17] E. Yu, “Towards Modelling and Reasoning Support for Early-Phase Requirements Engineering,” *Proceedings of the 3rd IEEE Int. Symp. on Requirements Engineering (RE'97)* Jan. 6-8, 1997, Washington D.C., USA. pp. 226-235.

[18] Ivar Jacobson, M. Christerson, P. Jonsson, G. Övergaard, *Object-Oriented Software Engineering - A Use Case Driven Approach*, Addison-Wesley, 1992.

[19] “NIST Framework and Roadmap for Smart Grid Interoperability Standards , Release 1.0. United States National Institute of Standards and Technology,” 2010.

[20] Osterwalder and Y. Pigneur, Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers, 1st ed, Wiley, Jul. 2010.

[21] V. Kartseva, J. Gordijn, & Y.-H. Tan, “Value-based business modelling for network organizations: lessons learned from the electricity sector,” *Proceedings of the Twelfth European Conference on Information Systems,* pp. 906-918.

[22] Edin Arnautovic, Davor Svetinovic, Value models for engineering of complex sustainable systems, *Conference on Systems Engineering Research 2012*, *Procedia Computer Science*, Volume 8, 2012, pp. 53-58, ISSN 1877-0509, 10.1016/j.procs.2012.01.013.

Proc. of the 2012 7th International Conference on System of Systems Engineering, Genoa, Italy - 16-19 July 2012

On Dynamic Models for Wind Farms as Systems of Systems

Kalev Rannat
Research Lab for Proactive Technologies
Tallinn University of Technology
Tallinn, Estonia
kalev.rannat@dcc.ttu.ee

Merik Meriste
Research Lab for Proactive Technologies
Tallinn University of Technology
Tartu, Estonia
merik.meriste@ut.ee

Leo Motus
Research Lab for Proactive Technologies
Tallinn University of Technology,
Estonian Academy of Sciences
Tallinn, Estonia
leo.motus@akadeemia.ee

Jürgo-Sören Preden
Research Lab for Proactive Technologies
Tallinn University of Technology
Tallinn, Estonia
jurgo@preden.ee

Abstract – *The authors consider a wind-farm as a system of systems consisting of a number of wind turbines as sub-systems. Each wind turbine comprises its own sensor network and automation units for control and monitoring of turbine's work, structural integrity and environmental situation. Sensing of environment is hierarchical – a single wind turbine senses local conditions and informs both neighbors and central control. The wind-farm's central control owes its situational awareness to single turbines at the sites and areal environmental sensors. The wind turbines, embodied into the same environment, get influenced by and influence both the environment and neighbors. Interaction with the environment, influencing the neighbors and being subjected to central control lead to agent-based modeling and simulation as a convenient method for better monitoring and understanding the work of wind-farms.*

Keywords: Dynamic models, system of systems, wind-farms, wind turbine control, mediated interactions, agent-based modeling, simulation.

1 Introduction

One can pose a question, what has the wind-farm and its modeling to do with a concept of System of Systems? The answer depends on a viewpoint how deeply we go into the details. Do we consider a wind farm as a fixed system functioning in some environment or as a coupled collection of autonomous (possibly adaptive) agents interacting with the environment through some physical devices within that environment? The second alternative suggests as one of approaches, to consider a wind-farm as a system of systems with participating autonomous systems embodied in their environment and influenced by activities of systems in their local neighborhood.

Wind energy systems are moving away from centralized and hierarchical systems towards networks of heterogeneous energy producers and consumers. Naturally it leads to new challenges in design and modeling the energy systems for contemporary needs. The models and monitoring systems must take into account wind speed fluctuations, must be able to monitor and predict output power of the wind-farm influenced by several random factors and the status of the power distribution grid. Decentralization leads to more independence in decision making at the sites and more effort to monitoring the surrounding environment and the behavior of the neighbors. These facts support an idea to consider a wind-farm as a system of systems. The single sub-systems (wind-turbines with local sensor network and control) must be adapted to both local environment and should get information about the working parameters of the neighbors and react on commands over central control. Central control needs an overview about the prevailing meteorological conditions but should act also on signals about excesses at single turbines and the needs of distribution network (e.g. temporary need for power production). From the modeling point of view, we need to work simultaneously with different sub-models, describing always changing meteorological conditions, power grid actual needs, the turbines in the system with their local control and communication facilities. To design better to learn the behavior of wind-farm, we may need to simulate collaboration of different adaptive and dynamic models.

In this work, we consider dynamic modeling of wind-farms. One of the central issues, how to handle aerodynamic interactions, is supposed to be resolved by a concept of mediated interaction, applied to modeling, simulation and monitoring of wind farms as systems of systems.

Modeling a wind-farm as a system of systems needs appropriate tools and methods for describing different parts of the system, interactions and linking the subsystems. Designing, as well as monitoring of the wind-farm and its subsystems gets often challenging due to a huge number of independent parts, interactions and communications between the nodes. The subsystems interact continuously with the surrounding environment and possibly have a certain influence to each other (either caused by the environment or initiated by control processes). The task is rather complicated because it is needs to reckon with changes in the environment. The environment may influence the subsystems' interaction capabilities - it may put some of the subsystems (nodes) to work in totally different conditions compared to all others. It is hard to specify all the effects coming from the environment, but while designing a system (an application), attention must be paid on a full spectrum of nuances coming from it. It is obvious that situated in some physical environment the turbine depends first of all on particular environmental conditions (influenced, in turn, by turbines nearby).

The wind-farm represents a certain category of system of systems – the subsystems (turbines) belong to the same category of autonomous devices. More general case is that of modeling systems of systems with heterogeneous subsystems (e.g. using different types of wind-turbines and accessories).

2 Wind-farm as a System of Systems

In the following we analyze in detail a question, why the wind-farm should be considered system of systems.

If looking at a modern wind-turbine as a set of control units, sensors and actuators we can see quite a sophisticated system with integrated data communication facilities. Additionally, each wind-turbine can act as an independent situation-aware system with its given degrees of freedom in decision making but still working in favor of the wind-farm (i.e. they work under central control and optimization of power production). For a wind-farm as a large system, consisting of a number of wind-turbines, the control and monitoring automation of each wind turbine (including the blade pitch control, generated electric power control, etc) work all together to optimize energy production with acceptable quality. The subsystems of a wind-turbine perform specific predesigned subtasks. To keep the technical parameters (at a single turbine and a wind-farm level) within technical limits, we need a detailed overview about the environmental and wind-turbines' state. In addition, the control and monitoring of a single wind-turbine gets technically more complex for so-called "smart" turbines with a lot of additional sensors for individual fine-tuning and monitoring of structural integrity. All it leads to the needs for appropriate control algorithms and environmental models for both optimal monitoring/control and simulation/design.

2.1 Control and monitoring of a wind-farm

Wind turbine control algorithms design relies usually on the basis of a physical (single) turbine dynamics modeling. This is transparent, intuitive and leads to practically simple control structures.

However, the uncertainties associated with physical parameters (e.g. local environmental conditions, adaptive abilities of wind-turbines, etc.) involved as part of such simplified models may considerably decrease optimal performance of a single turbine and a wind-farm. A noticeable performance improvement is awaited from using more accurate (i.e. less uncertain) models and real-time *in situ* measurements from the operational wind turbine. It is hard, if not impossible, optimally to drive a wind-farm by statistical wind parameters obtained from wind models due to the heterogeneity of actual environmental conditions and turbulence – dynamic feedback has a significant role in real adjustment of operating turbines. For example, each turbine works in its own "local environment", influenced by the prevailing wind conditions and neighboring turbines.

Different counterparts (often with conflicting interests) operate the wind-farm – wind power producers concern about minimizing of forecast errors about power production over a long time interval for better scheduling of power generation (maximizing income). On the other hand, the system operators concern with guaranteeing the security of the system, minimizing operational costs while maintaining a high level of reliability [1]. Modeling a wind-farm should coincide with the needs of both groups of users. This leads to issues of scaling and situation-aware interaction of different models, driven at different levels (e.g. from marketing to system's operator and local situation at a certain wind-farm and a single wind-turbine). If looking more ahead, the models may need to interact at wind-parks and wind-parks clusters' level (it would be in interest of producers/operators having several wind-parks at different geographical locations. Here, the scaling is not related to the number of wind-turbines or wind-parks involved in the model only, but also to the timescales. It is quite a different task to model a wind-turbine with its control subsystems reacting on the real-time situation at the site and the productivity of a turbine or wind-farm for power generation.

The nature of the tasks (for wind-farm modeling and simulation) points to a need to apply modeling of different individual components in spatiotemporal context. Due to the abovementioned reasons the modeling in software agents could be considered a convenient and most flexible method [2].

2.2 Control algorithms and dynamic feedback

As a common praxis, most control algorithms depend on measurements from turbine structure and drive-train for use in the control feedback. However, these turbine measurements can be unreliable or exhibit delayed response to disturbances acting on the turbine. This puts the control units to react to complex atmospheric disturbances after their effects have been "sensed" by the turbine. There is always a delay between the time that a disturbance arrives, and the time that the control actuators begin to mitigate the resulting loads.

Practical need to reduce the "reaction time" for load mitigation has led to an idea to use environmental parameters (i.e. wind characteristics) upwind of the turbine before they impact the turbine rotor. It offers a possibility to prepare the needed control actuation signals in advance. Lidar technology is capable of measuring velocity upwind of the turbine with sample rates in the 10's of Hz [3]. With these measurements, it is possible to design preview controllers that can adjust pitch (and/or torque) as necessary before wind disturbances arrive at the turbine (i.e. feed-forward techniques [4]).

Assessing the wind resource for any wind farm project is a key factor for both technical performance and financial profitability. Remote sensors have shown their potential by reducing vertical uncertainty through measurement of the wind up to and above hub height, and by reducing horizontal uncertainty through measurement of the wind at multiple locations across the wind farm. The better the measurements quality, the better will be the analysis and accordingly higher predictability and control facilities for wind-farms. The measured real-time data can be also assimilated to the general wind-model of the wind-farm, resulting in better forecast and control.

One of the main goals for modeling and simulation of wind-farms is to develop models that allow real-time predictions of flows and incorporate measurements from a set of spatially distributed sensor devices. For these purposes, the flow information can be a basis for new control paradigms that acknowledge the uncertainty in the modeling and dynamically manage the flow resource in order to optimize specific control objectives.

Let's describe typical designing a wind-farm (Figure 1) from wind-turbines with some known technical characteristics. First of all, the design of wind-farms starts from detailed studies about the energetic potential of winds at the site (geographic area) [5]. It gives a statistical description about wind conditions (as well about the sea state for off-shore installations). The statistical means (and extremes) in the environment may serve as guidelines for the engineers, how to determine the maximum loads for installations and both mechanical and electric devices. For energetic efficiency of the wind-farm, (integrated system of single wind-turbines) we need to know each turbine's mechanical characteristics and actual wind conditions for them. The wind over the wind-farm area could obviously be not homogeneous. Each turbine has a certain impact to the local wind characteristics (initiating additional turbulence for the next down-wind turbine, decreasing the mean wind speed) [5]. For example, for mean wind speed 10-15 m/s, the turbulence pattern from a neighbor 500 meters afar can arrive in less than a minute. During intensive and sudden wind gusts during nearby thunderstorms, the potentially mechanical overload initiating wind-event can reach remarkably earlier. The reaction-speed using dynamic feedback depends highly on the technical details of wind-turbines. Depending on the technical characteristics of the turbines the detection-decision-action cycle (estimated) should not exceed 10-30s. It also depends on the type of situation detected and the type of action. For example, the wind-orientation of the rotor cannot be influenced quickly due to gyroscopic forces.

2.3 Interactions in wind-farm and modeling of interactions

To provide effective control and functionality of a wind-park embodied in the environment, an appropriate environmental model (air flow model), models of subsystems (wind turbines), a model and method for advising the control and related situational information in space and time domain are needed. It comprises

a) Information about the relations between a subsystem and its local environment;

b) Information about each subsystem's behavior depending on behavior of its neighbors;

c) Information about the state of the subsystem and its control methods.

The challenge here relates to always changing wind (and/or sea-state) conditions. For lower wind speeds we still want to keep the energy production in technically acceptable limits and for the higher extremes we need to avoid damages. The optimum can be granted only if we can adapt each turbine with actual changes in the air flow. In practice, it means real-time monitoring of wind characteristics not only over the wind-farm area, but at each turbine. For efficient (and safe) energy production, the control system of the turbine can use wind information (also mechanical load characteristics) from its neighbors.

Connecting wind-farms into distribution network may modify power flows. In certain conditions, it could result in under or over-voltage on specific points of the network. Additionally, it could increase the cases of power quality problems and produce any type of alterations regarding voltage stability [6]. Impact of wind energy on power systems must focus on issues like security, stability, power quality and operation of power systems.

Moreover, we need to have methods to describe the interactions (environment – turbine, turbine – turbine) and communication facilities. The interactions may be immediate physical interactions (environment with a turbine) or mediated. Changes in settings may be initiated due to control parameters measured and transferred over communication channels to a certain turbine for decision-making and local control. Additionally, the interaction can be mediated by the environment - the turbulence initiated by turbine_11 reaches turbine_12 by the mean air flow. In practice, mediated interactions [7] can be proposed as a key concept for applying such a method for monitoring, modeling and simulation of system of systems like wind-farms.

2.4 Modeling mediated interactions

A promising approach in modeling wind-farms as systems of systems like this is to apply networked autonomous agents embodied into a certain environment. In this case, each subsystem could be represented by its situation-aware agent. The agent is aware about the current state of the subsystem and supported with some means for partial control of subsystem's functions if necessary (obviously depending on a particular situation). The agent-subsystem pairs (Figure 1) embodied in an environment (and, situated into the dynamic spatiotemporal model of the system of systems give an operator a possibility for continuous real-time monitoring of each subsystem's working characteristics and simultaneously the state of the surrounding environment. The representative agent of each turbine has given a property to monitor and interpret the environmental conditions to offer maximum output with granted quality. The input data for decision-making and control actions may come from agent-turbine's own sensors' and/or obtained/mediated from the neighbors or central control.

The wind-turbines, covering a large area of the wind-farm (Fig. 1), being affected by the same always changing environment and each-other, have supported (due to their technical nature) the choice in favor of agent-based modeling and simulation (ABMS). The concept of intelligent agents [8] coincides well with all the basic needs and functions assigned to a single wind-turbine (e.g. the ability to interact, to sense the environment and decision-making).

The agent-based approach to the modeling and simulation of complex coupled systems offers several advantages. *First*, the facility to integrate heterogeneity among the agents (in general, there can be different types of wind-turbines in the wind-farms). *Second*, easy to create a modular structure, interoperable with other platforms (using JAVA). *Third*, the ability to represent different time scales with the same model. *Fourth*, a possibility to use several approaches combined in the same model. *Fifth*, the easy scalability of the model (allowing to add and remove agents dynamically, e.g. failures, scenarios of enlargement of the farm, etc.) [2]. Additionally, the emphasis on modeling the heterogeneity of agents (representing different wind-turbines), environmental conditions and the emergence of self-organization are the distinguishing features of agent-based simulation as compared to other simulation techniques such as discrete-event simulation and system dynamics. Agent-based modeling offers a way to model systems composed of agents who interact with and influence each other, learn from their experiences, and adapt their behaviors, getting better suited to their environment [9]. This is what the modern wind-turbines have to do – to sense the environment and adapt itself to the aerodynamic loads for optimal and safe power production.

Modeling and simulation of a wind-farm is a task that arises a following question – how and where to integrate the sensors (and information from the sensors) and what is the best way to describe the aerodynamic interactions (turbulence initiated by the neighbors and mediated by the environment).

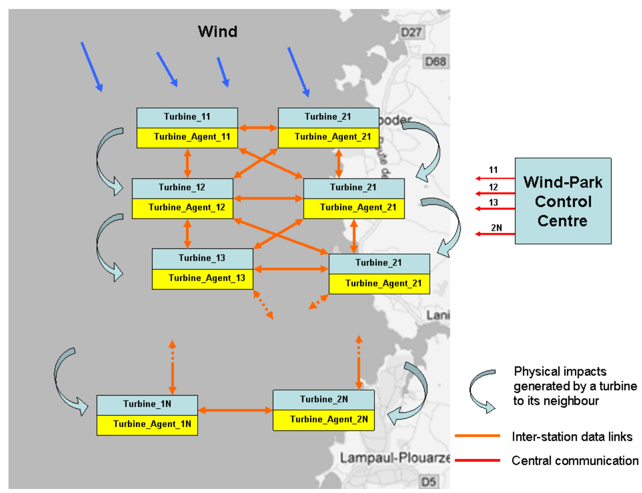


Figure 1. Model of the Wind-Farm with turbine-agent pairs, intercommunications and a control centre.

Agents interact with each other offering (mediating!) to subsystem-agent pair adequate situational information as well as information about the current sate of neighbors to support subsystem's situation awareness. The information about the environmental situation around and at the site of one subsystem and the state of the subsystem will be made available to the neighbors (if necessary). The information can be used to tune/control the environmental model (wind model) itself. Similar agent based approach can be enhanced for managing the clusters [2] of wind-farms.

3 Dynamic models

The control facility (or model) of the wind-farm is convenient to realize as a dynamic map with active agents embodied. Such a map is much more than just a GUI of a wind farm – it is an agent-based model of the wind-park presented by some set of proactively interacting autonomous agents. It supports (selective) simulation, real-time monitoring and tuning of single turbines (at the site or for a subset of turbines) according to the prevailing hydro meteorological conditions. The workflow of the dynamic map itself is based on the concept of mediated interaction - functionalities of the dynamic map are implemented by agents comprising the proactive middleware for mediated interactions. The proactive middleware serves for turbine agents (includes them) and is responsible on data acquisition from the neighbors, collaborative processing of several environmental parameters, signaling about excesses, alerting the neighbors, analyzing the wind situation, etc.

In order to model the situation in the wind farm correctly, both at the local level of the individual turbines but also globally in the wind-farm as a whole, the information required for generating an adequate situational picture must be collected and fused appropriately. It appears to be feasible to organize the information fusion hierarchically using the concept of situation parameters, the parameters characterizing a specific property of the conditions in the wind farm in a specific location (e.g. the local wind characteristics and structural state for a single turbine). The concept of distributed creation of situation awareness can be well applied in the context of wind farms where the information available (and required) for achieving situation awareness stems from a variety of sources (or independent models and sub-models). The information can be collected and preprocessed by autonomous nodes (single turbine's data processor for local control) and/or centrally for a whole wind-farm.

The nodes (represented by appropriate autonomous agents) equipped with sensors, communication and computing facilities of the system can be considered as a distributed system. For pragmatic and technical reasons those autonomous nodes are not really synchronized, and communication links between those nodes are not quite stationary – i.e. we have to deal with ad hoc networks that may comprise mobile nodes (e.g. for service).

The difficulty is in imposing a suitable hierarchical structure to situations that enable to define the (automatically) measurable factors for basic situations from which more sophisticated situations can be constructed, measured and reasoned about, and then used to build up situational awareness for the end users (which may be natural or artificial agents).

The notion of the situation and the hierarchy of situations form a meta-structure that is built from the information collected about basic (or elementary) situations by the information acquisition system (sensors at each turbine and measuring points at the wind-farm).

The concept of hierarchical composition of situations has been proposed in [9]. If looking at the hierarchy of situation parameters from the data abstraction perspective – parameters at the higher level characterize more abstract concepts (e.g. wind conditions at wind-farm level, the actual situation and state of the energy distribution grid) than the lower level parameters. The lower-level situations depict local wind conditions and structural load parameters for a single turbine. Applying the concept of situation parameters at different levels enables incremental situation analysis and evaluation. At the lower level we acquire data from sensors located at a single wind-turbine, where the data is initially processed and decisions taken for control actions and decided, if and what information should be sent to the neighbors or central control. The central control senses the wind-farm environment using single sites (with additional checkpoints and sensors if those exist), puts collected information into a certain context, checks the power distribution grid requirements and tries to guarantee safe and high quality energy production.

In order to exchange the situation parameter values a specific middleware is required that takes care of situation parameter source discovery, parameter value delivery and propagation of the validity information associated with the situation parameters.

The middleware has a service-oriented architecture, enabling easy, self-adjusting communication between dynamic collections of interacting autonomous agents – those interactions are required to deliver the situation parameter values to individual agents for fusion of situational information.

Some examples of the application-oriented services that must be provided by the middleware are:

- validation of information acquired from different sources, assigning tags to data items if necessary
- transformation (and compression) of validated (and fused, or otherwise processed) information into the interim format defined by the middleware
- keeping track of the access rights of all the agents linked to the middleware and checking the rights during any transaction
- remembering the preferred formats of messages, specific subscriptions for information from the agents, and processing capabilities of each involved agent
- delivering the subscribed information and satisfying all the constraints and requirements imposed by the agents.

Another set of services in the middleware is for intrinsic use (for handling the interim data format). These services are required for creating, maintaining, updating, and partitioning according to the subscriptions from, and position of the agents.

The peculiarity of this middleware is in its autonomous and smart operation that pays attention to individual properties and requirements of the clients, and in (situation sensitive) on-line validation of the outcome of its services. This becomes possible due to application of the "mediated interaction" concept. This concept is built on a situation-aware interactive model of computation [10] and on a well-established message exchange paradigm where consumer has to subscribe to a message. More details about data exchange for shared situation awareness, subscription based data exchange model and hierarchical buildup of situation parameters can be found from [11].

The self-organizing dynamic digital map with active agents of turbines can serve for both testing and visualizing of simulated or real-time situations. It also can be applied as an interface for long-term data monitoring and collecting to the databases – later the time-series can be post-processed and analyzed. The analysis helps to optimize the working characteristics of the wind-farm in general or in a certain subarea (up to the level of a single wind-turbine) [2].

4 Conclusions

The paper discuses the issues and solutions involved in creating networked situation aware agents for the described above approach. Means for mediated interactions for both modeling and simulation of wind-farms, situational information representation, computation and exchange on the basis of dynamic maps are described. Methods for information exchange are also considered especially in the context of the interface with computational simulations applied for evaluating physical aspects of the environment.

The application area is chosen just to illustrate some technical issues and a possible way for a solution. Different applications pose different technical constraints, but we stay in limited scope of problems in the article.

The ongoing and future work is targeted to realization of a virtual wind-farm with modeling of aerodynamic interactions (mediated by the environment) between the neighboring wind turbines. Environmental conditions (the wind-field) are generated by a wind-model. Agent-based modeling and ongoing simulations are planned to finalize with Java-based REPAST [12], having good possibilities for work on different platforms and numerical models.

References

[1] Use of Wind Power Forecasting in Operational decisions, Argonne National Laboratory, Report ANL/DIS-11-8, September 30, 2011.

[2] E. Kremers, et al., "Agent-Based Simulation of Wind Farm Generation at Multiple Time Scales," in Wind Farm – Impact in Power System and Alternatives to Improve the Integration, Edited by Gastón Orlando Suvire, pp. 313-330, InTech 2011.

[3] M. Harris, M. Hand, and A. Wright, *Lidar for turbine control, NREL Technical Report*, NREL/TP-500-39154, 2006.37

[4] J.H. Laks, L.Y. Pao, and A.D. Wright, "Control of Wind Turbines: Past, Present, and Future," Proceedings of 2009 American Control Conference, Hyatt Regency Riverfront, St. Louis, MO, USA, June 10-12, 2009.

[5] From Turbine to Wind Farms - Technical Requirements and Spin-Off Products, *Edited by Gesche Krause,* InTech 2011.

[6] Wind Farm – Impact in Power System and Alternatives to Improve the Integration, *Edited by Gastón Orlando Suvire*, InTech, 2011.

[7] J.L. Blackhurst, J.S. Gresham, M.O. Stone, "The autonomy paradox", *The Armed Force Journal*, October 2011.

[8] D. Poole, A. Mackworth, *Artificial Intelligence: Foundations of Computational Agents*, Cambridge University Press, 2010.

[9] L. Motus, M. Meriste, J. Preden, "Towards Middleware Based Situation Awareness," Proc. Military Communications Conference - 2009. MILCOM 2009, 18-21 Oct. 2009: 5th IEEE Workshop on Situation Management (SIMA 2009), Boston, 19-21 October. Boston, USA: IEEE Operations Center, pp. 1–7, 2009.

[10] L. Motus, M. Meriste, W. Dosch "Time-awareness and Proactivity in Models of Interactive Computation" Electronic Notes in Theoretical Computer Science, vol. 141, pp. 69-95, 2005, www.elsevier.com/locate/entcs

[11] J.-S. Preden et al, "Data Exchange for Shared Situation Awareness," Proc. Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support, New Orleans, LA, pp. 228-231, 2012.

[12] M.J. North, T.R. Howe, N.T. Collier, and J.R. Vos, "A Declarative Model Assembly Infrastructure for Verification and Validation," in S. Takahashi, D.L. Sallach and J. Rouchier, eds., Advancing Social Simulation: The First World Congress, Springer, Heidelberg, FRG (2007).

Ground properties evaluation for the design of geothermal heat pump systems and uncertainty measurement during the Thermal Response Test

Marco Fossa
Dime
University of Genova
marco.fossa@unige.it

Davide Rolando
Dist
University of Genova
davide.rolando@unige.it

Abstract – *The exploitation of low enthalpy geothermal resources for building heating and cooling purposes represents an important opportunity for saving energy in the domestic and service sectors, that cover about 30% of the demand of primary energy in the European Union. Vertical Borehole heat exchangers (BHE) are the most frequently adopted solution for ground coupled heat pump (GCHP) applications. The heat transfer between the borefield and the surrounding ground is driven by the ground thermophysical properties, the borefield geometry and the temporal distribution of heating and cooling loads. For these reasons the estimation of the ground properties, thermal conductivity above all, is crucial in the borefield design analysis. The thermal response test (TRT) is well assessed technique for evaluating the ground conductivity together with the equivalent borehole resistance. In this paper the ground modeling and measurement analysis are discussed with respect to TRT experiments. Starting from a set of real TRT investigations carried out in different parts of Italy, an uncertainty analysis is presented with special attention devoted to the measurement errors and disturbances, including the important effects of the typical non steady condition of the applied heat transfer rate.*

Nomenclature

H	Borehole depth, $[m]$
R_b	Borehole thermal resistance, $[(m \cdot K)/W]$
T	Temperature, $[K]$
T_f	Mean fluid temperature, $[K]$
$T_{g,\infty}$	Undisturbed ground temperature, $[K]$
α	Ground thermal diffusivity, $[m^2/s]$
$\dot{Q}$	Heat transfer rate, $[W]$
$\dot{Q}'$	Heat transfer rate per unit length, $[W/m]$
$\dot{m}$	Mass flow rate, $[kg/s]$
ρ	Ground density, $[kg/m^3]$
c	Ground specific heat, $[J/(kg \cdot K)]$
c_l	Fluid specific heat, $[J/(kg \cdot K)]$
k	Ground thermal conductivity, $[W/(m \cdot K)]$
r	Radius, $[m]$
r_b	Borehole radius, $[m]$
t	Time, $[s]$
t_c	Characterist time for ILS validity, $[s]$
E_1	Exponential integral
Fo	Fourier Number, $Fo = (\alpha\, t)/r^2$
S	Angular Coefficient
δ	Measurement uncertainty
γ	Euler Constant, $\gamma \simeq 0.5772$

1 Introduction

Ground coupled heat pumps (GCHP) are systems combining a heat pump with a ground heat exchanger for building heating and cooling purposes. Due to the favorable ground temperatures all along the year, GCHP are able to perform coefficient of performance, as a seasonal average, up to 4, so dramatically reducing the energy consumption and related emissions with respect to traditional fuel burning systems or air coupled air conditioning units.

Vertical Borehole heat exchangers (BHE) are the most frequently adopted solution for ground coupled heat pump applications. These heat exchangers are buried in the ground with depth variable typically between 80 and 150 meters, and they are constituted by single or double polyethylene pipes, disposed in a drilled well (0.1 m diameter, typically) and filled with heat enhancing grouting material. The systems of boreholes, where a suitable mixture of water and antifreezing is circulating, are usually disposed in regular matrixes at reciprocal distances in the range 5 to 10 m. The heat transfer process to the surrounding ground is driven by the ground thermophysical properties, the borefield geometry and the temporal distribution of heating and cooling loads. In particular the ground thermophysical properties play a fundamental role in this thermal process, especially with reference to the ground thermal conductivity: the higher is the thermal conductivity, the higher (winter mode) are the fluid returning temperatures (to the heat pump), so enhancing the efficiency of the system and reducing the equivalent greenhouse gas emissions.

The knowledge of the ground properties, thermal conductivity above all, is also crucial in the borefield design procedure. In fact, since highly conductive soils allow the same amount of heat to be transferred (at same temperature differences) with a reduced number of BHEs and since the borefield strongly affects the overall plant cost, the correct sizing of the ground heat exchangers, trough a preliminary investigation on the ground properties, is a compulsory task in order to obtain the shorter payback periods.

The thermal response test is well assessed technique for evaluating the ground conductivity and the so called borehole thermal resistance. This measurement procedure was first proposed by Mogensen (1983) [1] and it is based on the Infinite Line Source model (ILS, Ingersoll et al. 1954) [2].

The ILS model main assumptions are to consider pure conduction and constant ground properties in all directions. Under those hypotheses, ILS is able to describe the thermal response of an infinite ground medium to a step heat pulse coming from a single heat source or heat sink, which extends infinitely in the depth direction. According to the Mogensen approach (here after referred as ILS2R, i.e. ILS 2 resistance model), the inner part of the borehole is modelled as an invariant thermal resistance (the borehole resistance), while the surrounding ground has a time varying behavior, that can be described by the ILS solution. Finally the ILS solution provides the radial temperature variation as a function of a dimensionless time and according to the ground properties (Gehlin 1998, Pahud and Matthey 2001) [3] [4]. By applying the ILS solution to a set of measurements (temperatures of the heat carrier fluid at the BHE top surface) taken while a constant heat transfer rate is applied, a quite simple inverse technique can be applied for ground parameter recognition. Even if the application of the simple ILS2R model could be questionable, mainly because of the different thermophysical properties of the inner BHE content with respect to the ground ones (Signorelli et al. 2007, Bauer et al 2011) [6] [7] the Mogensen approach is able to provide reliable estimations (at least at engineering level) of the depth average thermal conductivity of the ground.

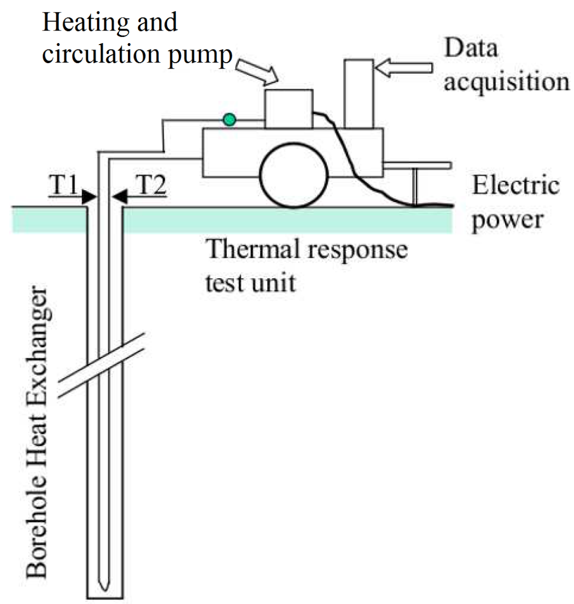


Figure 1: Thermal Response Test set up. [5]

The uncertainty on the ILS2R estimates depends on a number of causes, related to the non fulfillment of the initial hypotheses: measurement errors (temperatures, flow rate, heat transfer rate), wrong estimation of the undisturbed ground temperature, non uniform ground properties, non constant heat transfer to the fluid during the experiment, not suitable selection of the time window where to apply the inverse method (Beier and Smith, 2003, Signorelli et al, 2007, Bauer et al., 2011, Acuña and Palm, 2009) [8] [6] [7] [9]. The above uncertainty conditions always exist, at different extents depending on the particular test carried out and related to the particular features of the TRT machine employed for the test.

This paper presents the analysis of a number of TRT runs, taken in different part of Italy and Switzerland by third part research groups. The aim is to highlight from the analysis of these data the typical uncertainty sources that arise in real TRT test and to demonstrate, through the uncertainty analysis, the relative importance of each error contribution. Some strategies for uncertainty estimation are discussed and a number of simulations are performed in order to assess, trough the temporal superposition of basic temperature response factors, the influence of a non constant heat load to the ground during the TRT experiment.

2 Thermal Response Test Theory

The thermal interaction between the ground and a vertical heat exchanger, when underground water circulation can be neglected, is governed by the three-dimensional time-dependent conduction equation. Due to its complexity this equation is often solved numerically. However, a number of one-dimensional (radial direction) and two-dimensional (radial and axial directions) analytical solutions have been proposed, able to simulate the ground response to a single constant heat pulse.

In this paper the Infinite Line Source model (ILS2R) theory is presented. ILS2R model is based on Kelvin's line source theory and has been applied to simulate the behaviour of BHEs. Mogensen proposed this solution to perform a Thermal Response Test experiment.

The thermal test, in a brief description, consists in injecting heat into a fluid circulating inside a BHE and to register the fluid temperatures (from and to the ground) in time.

The analysis of the Thermal Response Test data is based on a description of the heat as being transfered from a line source. According to this model the temperature field in the radial direction r after a time duration t of a constant heat injection (or extraction) rate per length of borehole $\dot{Q}'$ is given by [10]:

$$T(r,t) - T_{g,\infty} = \frac{\dot{Q}'}{4\pi k}\int_{\frac{r^2}{4\alpha t}}^{\infty}\frac{e^{-u}}{u}\,du = \frac{\dot{Q}'}{4\pi k}E_1\left(\frac{r^2}{4\alpha t}\right) \quad (1)$$

where $T_{g,\infty}$ is the undisturbed ground temperature, $\alpha = k/(\rho\, c)$ is the thermal diffusivity and E_1 is the so-called exponential integral. The thermal diffusivity is hence a function of the unknown ground conductivity and of the ground volumetric heat capacity, which can usaually be deduced with adequate precision from the geological data of the site. For large values of the parameter $\alpha\, t/r_b^2$, E_1 can be approximated with the following expression valid for the time criterion $t_c \geq 5r^2/\alpha$ [11]:

$$\begin{aligned} E_1(X) &= -\gamma - ln(X) - \sum_{n=1}^{\infty}(-1)^n\frac{(X)^n}{n\cdot n!} \simeq \\ &\simeq -ln\left(\frac{r^2}{4\alpha t}\right) - \gamma \simeq ln(4Fo) - \gamma \end{aligned} \quad (2)$$

where $X = r^2/(4\alpha t)$, γ is the Euler constant [$\gamma \simeq 0.5772$], Fo is the Fourier number defined by:

$$Fo = \frac{\alpha\, t}{r^2} \quad (3)$$

The temperature at the borehole wall ($r = r_b$) can be calculated as:

$$T(r_b,t) = \frac{\dot{Q}'}{4\pi k}\left(ln\left(\frac{4\alpha t}{r_b^2}\right) - \gamma\right) + T_{g,\infty} \quad (4)$$

The thermal caracteristics of a borehole heat exchanger are determined by its effective borehole thermal resistance R_b which defines the proportional relationship between the temperature difference of the fluid (T_f) and the borehole wall (T_b), and the heat transfer rate per unit length.

The effective borehole thermal resistance takes into account both the geometrical parameters of the borehole heat exchanger (pipe spacing, diameter, number of pipes, depth) and various physical parameters (thermal conductivity of grout materials, flow rate in the borehole, fluid properties) [4].

Adding the effect of the effective borehole thermal resistance R_b between the fluid and the borehole wall the fluid temperature as a function of time can be written:

$$T_f(t) = \frac{\dot{Q}'}{4\pi k}\left(ln\left(\frac{4\alpha t}{r_b^2}\right) - \gamma\right) + \dot{Q}'\cdot R_b + T_{g,\infty} \quad (5)$$

If $\dot{Q}$ is constant, the Eq. (5) becomes a simple linear relation:

$$T_f(t) = S\cdot ln(t) + m \quad (6)$$

where $S = \dot{Q}'/(4\pi k)$ and m being a constant related to the R_b quantity as shown in Eq. (7).

$$m = \left(\frac{\dot{Q}}{H}\left(\frac{1}{4\,\pi\,k}\left(\ln\left(\frac{4\alpha}{r_b^2}\right) - \gamma\right) + R_b\right) + T_g\right) \quad (7)$$

where H is the borehole length.

The R_b model assumes that T_f correspondes to the average between the inlet and outlet temperature of the carrier fluid.

Taking into account the logarithmic correlation in Eq. (6), the plot of the fluid temperature versus the time hints the possibility of estimate the slope parameter (S) through a (log)linear regression.

Thus, the effective ground thermal conductivity is given by:

$$k = \frac{\dot{Q}'}{4\pi}\cdot\frac{1}{S} \quad (8)$$

Once the ground thermal conductivity is known, the effective borehole thermal resistance R_b can be estimated as:

$$R_b = \frac{H}{\dot{Q}}(T_f - T_{g,\infty}) - \frac{1}{4\pi k}\left(ln(t) + ln\left(\frac{4\alpha}{r_b^2}\right) - \gamma\right) \quad (9)$$

3 Case Studies and Uncertainty Analysis

Thermal Response Test analysis provides effective ground thermal conductivity and effective borehole thermal resistance quite easily but no level of confidence in these estimates is given.

The ILS2R model does not take into account the thermal heat capacity of the circulating fluid neither the differences in thermal properties between the grout and the soil. Thus, the model is not able to match the early-time data, which are affected by the properties of the inner borehole. [12] [7]

According to some authors, the ILS2R model is associated to a maximum error less than 10% for $t > t_c$, which means $\alpha t/r_b^2 \geq 5$, and a maximum error of 2.5% when $\alpha t/r_b^2 \geq 20$ [5] [11]. Neverthless no check is usually performed.

TRT name	Duration [h]	t_{samp} [min]	$T_{g,\infty}$ [°C]	H [m]	r_b [m]	Bhe type
HR1	142	1	16	100	0.07	U
HR2	141	1	16	100	0.07	U
DPH	122	10	12.4	160	0.07	UU
GNT	48	0.25	15.6	185	0.07	U
GR1	123	4	9.4	100	0.07	U
GR2	142	4	13.9	100	0.07	U
GR3	117	4	13.2	300	0.07	U

Table 1: Experimental TRT data set employed in this study.

The results of seven Thermal Response Test data set concerning different sites in northern Italy and Switzerland have been analysed according to the standard TRT theory and to a statistical analysis devoted to overall uncertainty estimation. The TRT experiment and related BHE type are summarized in Tab.1.

3.1 Sensitivity Analysis

The evaluation of the overall uncertainty of a measurement can be assessed with a *sequential perturbations method* [13] which consists in the alternate introductions of a perturbation on each parameters involved in a measurement. The quadratic mean of the individual uncertainties gives the overall uncertainty.
The following uncertainties are considered:

$$\begin{aligned} \delta\dot{Q} &= 1.96 \cdot \sigma(\dot{Q})\, [W] \\ \delta H &= 0.5\, [m] \\ \delta T_{g,\infty} &= 0.5\, [°\mathrm{C}] \\ \delta\rho c &= (0.2 \cdot \rho c)\, [J/(m^3\, K)] \end{aligned} \tag{10}$$

where $\dot{Q}$ is the electrical heat power measured as a function of time during the TRT and σ is the standard deviation of measurements.
As shown in Fig.2 the heat power variations and the undisturbed ground temperature uncertainty mostly contributes to the final borehole thermal resistance overall uncentainty. While the constant heat flux is one of the main hypothesis of ILS2R model and therefore often discussed, further studies would be necessary for the correct evaluation of undisturbed ground temperature [14] [15].

4 Variable Ground Heat Flux Analysis

The ILS2R model, as stressed before, does not take into account changes in the heat input rate, even if those changes may occur in field tests due to unexpected or controlled events. The capability of analyzing these anomalies could allow to achieve important information about the soil thermal properties and to further develop the TRT analysis method.

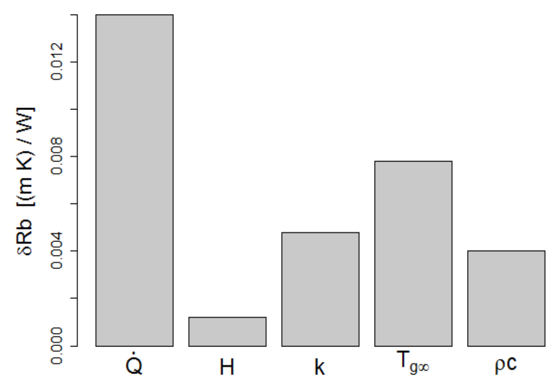


Figure 2: Sensitivity analysis related to borehole resistance estimation.

Fig.3 shows a typical electric power curve during a real TRT experiment. The signal presents the typical “noise” due to the electrical power instability. According to the guidelines suggested by Ashrae this variation should be less than 1.5% of the average power for the power input to be considered constant.
Applying the temporal superposition principle [10] the temperature evolution in time can be calculated according to the ILS2R model solution taking into account different non-constant heat input situations as found in real field test or according new experiment strategies [16].
The first case here considered is a situation where the heat transfer to the carrier fluid is stopped after a given amount of hours, while the fluid is still circulated in the BHE (Fig.5). The superposition technique allows the fluid temperature profile to be simulated and decribed also in the ”decay” period, during which further estimates of the ground conductivity can be obtained.
The second case here presented is related to power cut off during the test (Fig.4), according to the power profile (continuous line) shown in the same figure. It can be observed that the superposition technique is able to describe the fluid temperature evolution (dashed line) even in this case. Worth noticing, since the superposed ILS method is adopted as the model for ground behaviour description, a proper iterative technique can be applied to solve the inverse problem of the ground conductivity estimation.
The last case considered in this preliminary investigation (Fig.6) is the continuous variation of the heat transfer rate as measured in a real TRT experiment among those described in Table 1. It can notice from figure inspection that the superposed ILS is able to generate realistic fluid temperature profiles, hence suitable for proper optimization techniques to be applied for bhe parameter estimation.
Further investigations and technique refinements are necessary and will be tackled in future papers.

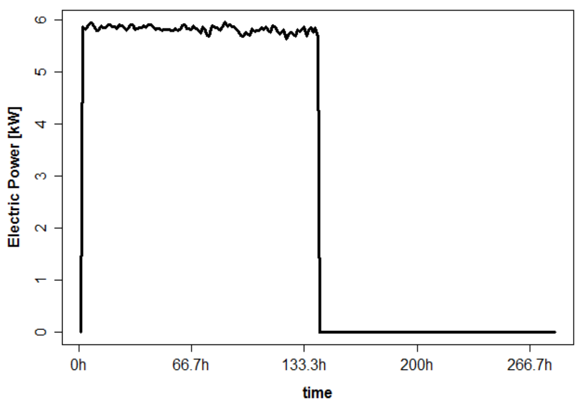


Figure 3: Electric power measured during a real TRT experiment among those of Tab.1.

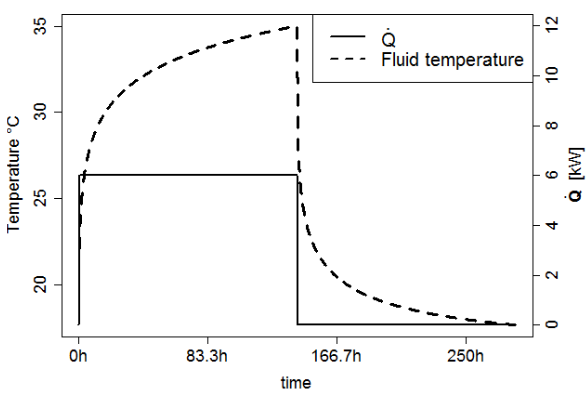


Figure 4: Simulated fluid temperature profile (dashed line) for stepwise non linear heat input.

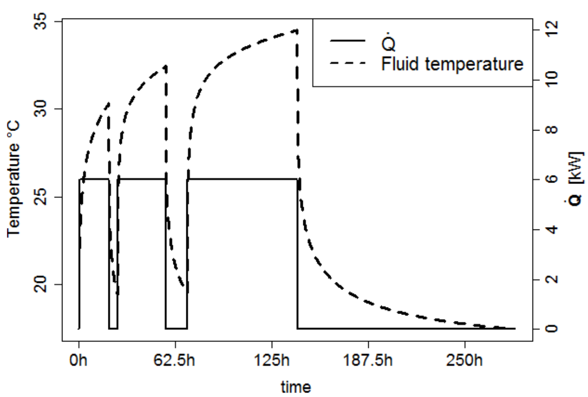


Figure 5: Calculated fluid temperature profile (dashed line) according to a complex, stepwise like, heat input function.

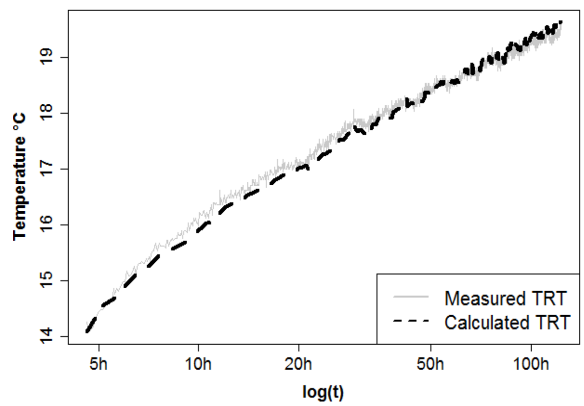


Figure 6: Superposed (calculated) solution compared to measured values (continuous line) as a result of an optimum research on ground conductivity value.

5 Conclusions

In this paper, the thermal response test theory has been discussed with reference to the common error sources of measurement uncertainty. To this aim a series of real TRT data have been analysed in order to assess the importance of each measured quantity on the uncertainty related to the estimated parameters, namely the ground thermal conductivity and borehole resistance. It is demonstrated that the errors due to the fluctuations of the thermal power delivered to the geothermal fluid are the most relevant ones. The final part of the paper has been thus addressed to defining a possible strategy for taking into account those variations in the heat transfer rate to the fluid, by applying a temporal superposition of the base solution of the heat conduction problem. Even if this is a preliminary study and further technique refinement is request, the results show that a proper superposition technique can be successfully applied to reduce the overall uncertainty in the estimated parameters and even to manage unfavourable situations (e.g. power failure) or to force power step input to the ground in order to enhance the standard measuring technique based on the assumption of strictly constant heat transfer rate to the borehole.

Acknowledgments

The Authors acknowledges the financial support of the Italian Erde company and the its collaboration in developing an innovative TRT machine.

References

[1] P. Mogensen, “Fluid to duct wall heat transfer in duct system heat storage,” pp. 652–657, International Conference on Surface Heat Storage in theory and Practice, Stockholm, Sweden, 1983.

[2] L. R. Ingersoll, *Heat conduction*. McGraw Hill, 1948.

[3] S. Gehlin, *Thermal Response Test, in-situ measurements of thermal properties in hard rock*. Licentiate thesis, Lulea University, 1998.

[4] D. Pahud and B. Matthey, “Comparison of the thermal performance of double u-pipe borehole heat exchangers measured in situ,” *Energy and Buildings*, vol. 33, no. 5, pp. 503 – 507, 2001. Proceedings of International Conference.

[5] S. Gehlin, *Thermal Response Test - Method Development and Evaluation*. Doctoral thesis, Lulea University, 2002.

[6] S. Signorelli, S. Bassetti, D. matthey, and T. Kohl, “Numerical evaluation of thermal response tests,” *Geothermics*, vol. 36, pp. 141–166, 2007.

[7] D. Bauer, W. Heidemann, and H.-J. Diersch, “Transient 3d analysis of borehole heat exchanger modeling,” *Geothermics*, vol. 40, no. 4, pp. 250 – 260, 2011.

[8] R. A. Beier and M. D. Smith, “Minimum duration of in-situ tests on vertical boreholes,” *ASHRAE Transactions*, vol. 109, pp. 475–486, 2003.

[9] J. Acuña and B. Palm, “Distributed thermal response test on a u-pipe borehole heat exchanger,” Effstock Conference, October 2009.

[10] H. S. Carslaw and J. C. Jaeger, *Conduction of Heat in Solids*. Oxford Science, 1947.

[11] P. Eskilson, *Thermal Analysis of Heat Extraction Boreholes*. Doctoral thesis, University of Lund, 1987.

[12] R. A. Beier, M. D. Smith, and J. D. Spitler, “Reference data sets for vertical borehole ground heat exchanger models and thermal response test analysis,” *Geothermics*, vol. 40, no. 1, pp. 79 – 85, 2011.

[13] R. J. Moffat, “Describing the uncertainties in experimental results,” *Experimental Thermal and Fluid Science*, vol. 1, no. 1, pp. 3 – 17, 1988.

[14] S. E. Gehlin and B. Nordell, “Determining undisturbed ground temperature for thermal response test,” *ASHRAE Transactions*, vol. 109, pp. 151–156, December 2003.

[15] D. Rolando, *Stima delle proprietá del terreno nella realizzazione di scambiatori verticali per applicazioni geotermiche a bassa entalpia*. Thesis, University of Genova, Italy, 2011.

[16] P. M. Monzó, *Comparison of different Line Source Model approaches for analysis of Thermal Response Test in a U-pipe Borehole Heat Exchanger*. Thesis, Royal Institute of Technology, KTH, Sweden, 2011.

The automatic identification system of maritime accident risk using rule-based reasoning

IDIRI B.
MINES ParisTech
Centre for Crisis and Risk Research
Sophia Antipolis, France
bilal.idiri@mines-paristech.fr

NAPOLI A.
MINES ParisTech
Centre for Crisis and Risk Research
Sophia Antipolis, France
aldo.napoli@mines-paristech.fr

Abstract - *Current maritime traffic monitoring systems are not sufficiently adapted to the identification of maritime accident risk. It is very difficult for operators responsible for monitoring traffic to identify which vessels are at risk among all the shipping traffic displayed on their screen. They are overwhelmed by huge amount of kinematic ship data to be decoded. To improve this situation, this paper proposes a system for the automatic identification of maritime accident risk. The system consists of two modules. The first automates expert knowledge acquisition through the computerized exploration of historical maritime data, and the second provides a rule-based reasoning mechanism.*

Keywords: Maritime monitoring systems, automatic rule-based reasoning, data mining, maritime accident risk identification.

1 Introduction

Maritime transport is an essential element in economic development and maintaining links with remote territories. Shipping operates in a context that is both spatial-temporal (i.e. the physical environment) and strategic (i.e. deployment modes). The maritime zone is a complex environment that consists of many coastal states with different regulations and involves many different actors. Ships are exposed to the risk of maritime accidents such as sinking or grounding in coastal waters.

With the advent of localisation technologies, telecommunications, technological advances in embedded systems and digital maps, new maritime tracking systems such as the French *Surveillance des espaces sous juridiction nationale et des approches maritimes* (SPATIONAV), the American Integrated Control and Safety System (ICSS) and the Spanish *Sistema Integrado de vigilancia Exterior* (SIVE) have been developed. These systems make it possible to track vessel movements (position, heading, speed, home port, etc.) in almost real-time. They consist of a data acquisition infrastructure that can capture and transmit shipping data, and an information system for the storage, processing and display of ship movements on a control interface. Despite the widespread use of these systems (for example, all European Union countries have implemented the technology) the number of maritime accidents remains significant. Worldwide, there are still many thousands of maritime accidents each year and in 2008-2009 the amount of oil and hazardous substances accidentally spilled at sea increased by 400% [1]. The most recent accidents include the cruise ship Costa Concordia, which ran aground off the coast of Italy on 13th January, 2012 causing 32 deaths. The vast area to be monitored, the number of ships at sea (41 million ship positions/day for 62,000 ships according to Lloyds) and the multiplicity of scenarios make it very difficult to identify maritime accident risk. Although the timely identification of these risks is crucial, the current context makes this task challenging and complicated.

This paper proposes an approach to the identification of maritime accident risk that is based on automated reasoning, which is in turn founded on knowledge rules. Knowledge rules are obtained through data mining analysis methods applied to maritime databases. The goal is to provide a tool that offers a comprehensive picture of risk events that are scattered and fragmented in time and space, in order to help maritime shipping controllers to understand the latent meaning of the facts displayed on their monitors. Risk events indicate an imminent risk of a maritime accident and the tool can also display alarms that alert operators to events of interest.

Definitions

- Automatic reasoning is a subdomain of Artificial Intelligence[1]. It makes it possible to simulate human reasoning with a computer in order to deduce new knowledge from a sequence of input events and existing knowledge.

- Analogous to the search for nuggets in a goldfield, data mining aims to extract hidden information through a global analysis of large datasets. It seeks to discover models (patterns) that are difficult to identify directly due to the large volume of data, the number of variables to be taken into account and the fact that there are unknown assumptions [2].

2 Background

Several earlier studies have addressed the issue of improving maritime tracking systems – either at the level of the data acquisition infrastructure (the integration of new sensors, sonar networks, aerial drones, etc.) or at the information processing level. Among the work that has been carried out to improve information processing systems, there is the initial Defense Advanced Research Projects Agency (DARPA) project [3], French led DCNS projects [4][5], the work of Vandecasteele and Napoli [6], the work of Etienne, Devogele and Bouju [7] and the research carried out by Roy [8]. The literature also offers several other, more specific studies on the analysis of maritime accident risk. Among these studies, several different approaches can be distinguished: probabilistic [9], statistical [10] and those based on numerical simulation [11]. However, most of these approaches to the modelling and analysis of maritime accidents either do not take into account lessons that can be drawn from historical data or are restricted to confirmatory analyses[2]. This fact makes the discovery, formalization and use of expert knowledge difficult and complicated. Moreover, these approaches make it impossible to discover models that are not already known by experts.

Our research is inspired by the work of Roy [8] who applied rule-based reasoning to the automatic identification of abnormal ship behaviour. However, it should be noted that abnormal ship behaviour does not necessarily help in the identification of maritime accident risk, while normal behaviour may indicate or pose a risk. For example, the entry or exit of a Ro-Ro (Roll-on/Roll-off) type vessel into or out of a port can be considered normal behaviour, while 76% of accidents involving this type of vessel occur in, or near ports.

[1] Artificial Intelligence is a research domain that aims to create intelligent machines. The central problems include the modelling of traits such as reasoning, perception, recognition, etc.

[2] This is an analysis that aims to confirm or negate an initial hypothesis.

3 Methodology

In order to automate the identification of maritime risks we chose a computerised reasoning approach. Our choice was influenced by the fact that this type of approach is much easier to implement than mathematical models. The modularity of the knowledge discovered simplifies system maintenance and makes it easy for users to add or remove knowledge.

Our work is distinct from automation studies of maritime monitoring [3][8] in that it focuses on maritime safety (e.g. grounding, sinking) and not security (e.g. terrorist attacks, illegal immigration).

We chose to use automatic data exploration methods for expert knowledge acquisition. This approach simplifies the task of experts, who only have to validate the knowledge generated. The literature often describes brainstorming methods that bring together domain specialists in order to acquire and define expert knowledge. While these methods are clearly interesting, they are also complicated and expensive to implement and the output scenarios are highly dependent on the personal experiences of the experts. The expert knowledge acquired using our method is not limited to ship behaviour but also includes risk situations and areas. A risk situation is defined as one that meets the conditions (sea conditions, vessel characteristics) conducive to risk, while risk areas are areas where there is a high accident rate.

3.1 Data mining

The extraction of knowledge for the identification of maritime accident risk uses data mining methods that transform historical maritime data into knowledge. This approach solves the problem of knowledge acquisition that is known to be a bottleneck in the process of building expert automated reasoning systems.

In order to maximize the potential for risk identification, the selected approach aims to extract three types of knowledge, namely: risk situations, risk areas and risk behaviours. Risk behaviours in this context describe ship behaviour that indicates a risk (unusual change of course, slowing down, etc.). If conditions do not exist to indicate a risk situation (for example if parameters are missing), the system can draw upon knowledge of risk behaviours (e.g. entry into a high risk area, abrupt change of course) to identify potential risks.

There are several data mining methods available to extract data related to these different types of knowledge. They include: association rule mining, sequence mining, data clustering and trajectory clustering. Whichever method is used, the knowledge discovered through data mining must be validated by domain experts.

3.2 The data

A consequence of the selected approach to automatic knowledge extraction is that the system implementation must include a data acquisition phase. Broadly speaking, the data acquired can be divided into two categories:

- **Static geo-spatial data**: This data is the result of maritime accident investigations. The data mining analysis of this category of data generates knowledge of risk situations and risk areas. Knowledge related to risk situations links information about navigation conditions (wind, currents), vessel characteristics (type, age, etc.) and maritime risks (type, category, etc.) for the forecasting and targeting of maritime accidents. Knowledge of risk areas can enable closer monitoring of shipping in these areas and planning of intervention and rescue measures.

- **Dynamic geo-spatial data**: This category of data tracks ship movements (position, speed, heading, etc.) and supplementary data describing navigation conditions (wind, current, etc.). The data mining analysis of this category of data discovers knowledge describing the risk behaviour of ships.

3.3 Rule-based reasoning

In an automated reasoning system knowledge is often encoded either as rules (generalized examples) or cases (examples). A rule takes the form ‘if *antecedent* then *consequent*’ where ‘antecedent’ and ‘consequent’ express conjunction or disjunction of instances of database objects. Knowledge provides the link between known information (the antecedent) and the information to be deduced (the consequent) or actions to be executed (e.g. display alarm). A case describes a problem and its associated solution. The case paradigm records details of resolved source cases in a database, which are used to solve new problems (target cases).

Once knowledge has been formulated, several types of reasoning can be applied. Among these, the most frequently used are deductive reasoning and analogical reasoning. In the case of deductive reasoning, output values are deduced from input values and in the case of analogical reasoning, the new problem is reduced to a problem that has been resolved. The known solution is then adapted to the new problem. Table I shows a simple comparison between case-based reasoning (CBR) and rule-based reasoning (RBR).

Table I: Comparison of RBR and CBR

	CBR	RBR
Knowledge	Case	Case generation
Modularity	Problem	Rule
Problem resolution	Adapted case	Rule application (fast)
Reasoning	Non-deductive	Deductive
Acquisition	Easy (episodic problem solving)	Difficult (how to go about resolving a problem)

The adopted approach used RBR rather than CBR. This decision was based on several factors. First, it is easy to understand as humans often reason in the form of rules (i.e. if condition then action). Second, knowledge rules facilitate modularity as a complex problem can be decomposed into simple rules. Finally, reasoning is deductive and not analogical (analogical reasoning can sometimes lead to erroneous conclusions). The weakness of RBR is linked to the knowledge acquisition problem, which the data mining method described earlier is intended to overcome. Moreover, the data mining approach means that the majority of the rules generated using this method can easily be formalized as knowledge rules.

In a rule-based reasoning system, three essential components are defined: a knowledge base, the facts and an inference engine. To make the analogy with human reasoning, the knowledge base is everything that the human being knows. Facts correspond to their perception of the environment (sight, taste, touch, etc.) and the inference engine is the human’s capacity to use logic and make judgments related to knowledge.

As Figure 1 shows, the inference engine continuously monitors incoming events (Facts) to see if there are applicable rules in the rule base (Knowledge). Before executing these rules, the engine must resolve any conflicts (run priority, avoid loops, etc.) that may arise between them. After execution of the rules selected by the engine, new rules and/or new facts will enrich, respectively, the rule base and the fact base.

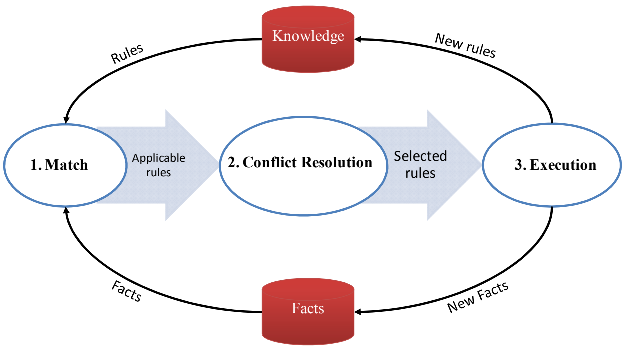


Figure 1. The rule-based reasoning process

4 System Architecture

The system consists of two main modules that are shown in Figure 2. The first module involves expert knowledge acquisition from historical data (accident investigations, ship movements, environment, etc.). The second module puts this knowledge to use in the task of automatic risk identification. The objective is to provide a system that can identify and link, from end-to-end risk events that are scattered and fragmented in time and space, and that can raise alerts to warn operators or experts of potential risks.

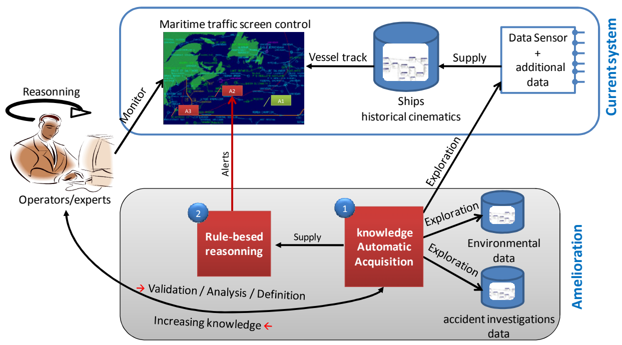


Figure 2. System architecture

The system improves upon existing maritime traffic surveillance systems used by maritime safety organizations (such as the French *Centres Régionaux Opérationnels de Surveillance et de Sauvetage*) in that it helps in risk identification. In addition to the paths taken by shipping that are typically displayed to maritime controllers, the operator is alerted to risk situations and is able to detect shipping that is in risk areas or that displays risk behaviours (slowing down, drifting off course, etc.).

5 Experimentation

In order to test our approach we used a data set from which we were able to extract both knowledge rules and facts. This data was cleaned[3] and pre-processed before being analysed using data mining methods.

5.1 Databases

- **MAIB**: The Marine Accident Investigation Branch (MAIB) database holds details of accidents and incidents that have occurred on ships between 1991 and 2009. This data relates to British vessels located anywhere in the world and all ships in United Kingdom territorial waters at the time of the accident. The database contains records of 14,900 accidents and incidents involving 16,230 ships.

- **MERRA**: Modern-Era Retrospective analysis for Research and Applications (MERRA) is a NASA project that provides meteorological data. Historical data from the period 1991-2009 helped to complete data missing from the MAIB database (winds, surface currents, etc.) and a daily download of weather forecasts made it possible to feed weather data into the fact base.

- **AIS**: The Automatic Identification System (AIS) database provides information about ship kinematics (position, course, speed, etc.) which is transmitted in almost real-time by AIS sensors installed on-board vessels. AIS is mandatory worldwide on all merchant ships over 300 tons and, in the European Union, on all fishing vessels over 15 meters.

5.2 Knowledge acquisition

In an earlier study, we outline how knowledge describing risk situations was acquired [12]. In that study we applied the association rule learning method to the MAIB database. Association rule learning is an unsupervised data mining method that makes it possible to extract knowledge rules of the type 'if *condition* then *result*' from *itemsets* that frequently appear together in a database (an *itemset* is a set of *items* and an *item* is an instance of a database object).

[3] Data cleaning involves the identification and correction of anomalies in the data (incompleteness, incoherencies, inaccuracies, etc.).

We applied the Apriori algorithm (developed by Christian Borgelt) implemented in the Rattle 2.6.4 software package (Rattle provides a graphical user interface for data mining) to the pre-processed MAIB data. The result of the analysis was rules that linked known parameters such as the type of vessel, its position, the geographical area and a specific context (e.g. oceanographic and meteorological conditions) with predictive parameters (e.g. type of accident). An example of the rules acquired in this way is shown below (Tableau II).

Tableau II: An example of the rules

Rule	{Location = Coastal waters, Vessel_Category = Fishing/processing, Age_Slice_Of_Vessel = 11 to 18 years} → {Incident_Type = Machinery Failure}
Measures	support[4] = 0.086 confidence[5] = 0.725 lift[6] = 1.47
Interpretation	If there is a fishing vessel, aged 11-18 years, sailing in coastal waters then there is a risk that it will break down.

This rule has been carefully chosen to show the value of combining the various aspects of knowledge (as described in Section 3.1). However, the simple application of this rule in the inference engine will return a huge number of results as many events will meet these conditions. Therefore, before creating an alert, the inference engine must also examine the behaviour of selected vessels (stopping at sea, slowing down, drifting off course, etc.).

5.3 Implementation of rule-based reasoning

Automatic reasoning was implemented using the Drools 5.4.0 (JBoss Rules) software package. Drools is a free, open source and powerful business rule management system. The components of the Drools platform that were implemented were the Business Rules Engine (Drools Expert) and the Business Rules Repository (Drools Guvnor). In our current work rules are added using the Guvnor user interface, however we plan in the near future to integrate all knowledge rules directly through the Java application programming interface (API).

[4] The *support* of an itemset is a reliability indicator; it is defined as the proportion of occurrences of the itemset in the database.

[5] *Confidence* is an indicator of the precision of the rule; it is equal to the probability that the consequent will occur together with a particular antecedent.

[6] The *lift* of a rule is an indicator that makes it possible to verify that the results obtained are not due to chance.

5.4 Example of rule deployment in Drools

Population of the rule and fact base in Drools can be done in several ways: through the Java API, by extending the Drools codebase, or by using the Business Rules Management System (BRMS) interface. The example given below describes the deployment of the rule described in Section 5.2 and shows how to add a fact. This example demonstrates how easy it is to deploy and understand rule-based reasoning.

- **Add rules:**

```
Rules "Risk of mechanical failure"
when
   $s: ship (location= = "Coastal waters", ship class
  =="fishing", age ≥ 11, age ≤ 18)
then
   check_behaviour ($s.id_ship);
end
Rules "Behaviour: drifting of course"
when
  $r: risk (Type=="Mechanical failure")
   course (behaviour=="Drifting")
then
   channel["alerts"].send(new Alert());
 end
```

- **Add a fact:**

```
Rule "Add Ship Course"
when
then
     insert (new Course ());
end
```

6 Conclusion

This paper proposes a system for the automatic identification of maritime accident risk based on rule-based reasoning with the aim of improving maritime tracking systems. The system continuously applies acquired knowledge rules to the flow of events describing the movement of ships and changing sea conditions, which makes it possible to identify risks in real-time. The system enables maritime traffic controllers to focus on important events and reduces the complexity of maritime risk monitoring.

The knowledge rules that feed the automated reasoning system are generated using data mining methods. This resolves the problem of knowledge acquisition that is often found to be a bottleneck in automated expert reasoning systems. While brainstorming methods are clearly useful, they are complicated to implement, expensive and the output scenarios rely heavily on the expert knowledge of individuals.

To maximize the potential for risk identification our knowledge discovery method divides knowledge into three aspects that describe: risk situations, risk areas and risk behaviours. The combination of these three aspects improves risk identification.

In the longer term further improvements could be added to the system. These include an alert scheduling system that would prioritise alarms and provide recommendations for intervention measures. We can use Bayesian networks[7] to manage the recommendations for intervention measures as proposed in [13].

References

[1] Centre de documentation de recherche et d'expérimentations sur les pollutions accidentelles des eaux, "Analyse annuelle des déversements de polluants en mer et sur le littoral" 2009. [Online]. Available: http://www.cedre.fr//fr/accident/pollution-accidentelle-mer-littoral.php#combien.

[2] G. Gardarin, *Internet/intranet et bases de données*, Eyrolles edition, pp. 264, 1999.

[3] Darpa, "Predictive Analysis For Naval Deployment Activities (PANDA)", 2005. [Online]. Available: https://www.fbo.gov/index?s=opportunity&mode=form&tab=core&id=7c1de14f8cdfd3d14f112f99a2bbd782&_cview=0.

[4] M. Morel, V. Flori, O. Poirel, A. Napoli, P. Salom, and G. Proutiere Maulion, "Trairtement et Authentification des MenAces et RISques en mer", in Workshop Interdisciplinaire sur la Sécurité Globale (WISG08), 2011.

[5] M. Morel, "SisMaris: Système d'Information et de Surveillance MARitime pour l'Identification des comportements Suspects de navire", in première Conférence Méditerranéenne Côtière et Maritime CM 2, pp. 261-264, 2009.

[6] A. Vandecasteele and A. NAPOLI, "Using Spatial Ontologies for Detecting Abnormal Maritime Behaviour", in Proceedings OCEANS 2012, 2012.

[7] L. Etienne, T. Devogele, and A. Bouju, "Spatio-Temporal Trajectory Analysis of Mobile Objects following", in Proceedings of the International Symposium on Spatial Data Handling (SDH), 2010.

[8] J. Roy, "Rule-based expert system for maritime anomaly detection", in Proceedings of Sensors, and Command, Control, Communication, and Intelligence (C3I) Technologies for Homeland Defense VIX, SPIE, Paper 7666-97, Orlando, F.L, USA, April 2010.

[9] M. D. Amrozowicz, A. Brown, and M. Golay, "A Probabilistic Analysis Of Tanker Groundings", in 7th International Offshore and Polar Engineering Conference, pp. 1-19, 1997.

[10] L. A. LeBlanc and C. T. Rucks, "A multiple discriminant analysis of vessel accidents", *Accident Analysis & Prevention journal*, vol. 28, no. 04, pp. 501-510, 1996.

[11] C. Nishizaki, H. Itoh, K. Yoshimura, K. Hikida, and N. Mitomo, "Development of a method for marine accident analysis with bridge simulator," in 6th International Conference on System of Systems Engineering (SoSE), pp. 31-36, 2011.

[12] B. IDIRI and A. NAPOLI, "Découverte de règles d'association pour l'aide à la prévision des accidents maritimes," in 12ème Conférence Internationale Francophone sur l'Extraction et la Gestion de Connaissance (EGC 2012), Revue des Nouvelles Technologies de l'Information RNTI-E-23 Éditions Hermann, pp. 243-248, 2012.

[13] X. Chaze, A. Bouejla, A. Napoli, F. Guarnieri, T. Eude, and B. Alhadef, "The Contribution of Bayesian Networks to Manage Risks of Maritime Piracy against Oil Offshore Fields," in International Workshops Information Technologies for the Maritime Sector (ITEMS 2012), Database Systems for Advanced Applications, Lecture Notes in Computer Science Volume 7240, pp. 81-91, 2012.

[7] A Bayesian network is a system for the representation of knowledge and the calculation of conditional probabilities.

Real-time risk definition in the transport of dangerous goods by road

Chiara Bersani, Claudio Roncoli
Department of Communication, Computer and System Sciences
University of Genova, 16145 Genova, Italy
chiara.bersani@unige.it

Abstract - *The definition of risk in the transport of dangerous goods is an open issue. No international standard is currently defined. In addition, the definition of risk is directly related to the possibility to its control at decisional level, for example, by rerouting the traffic. In this work, a proposal to define risk at strategic, tactical, operational and realtime level is proposed. A system of systems vision of the definition at operational/realtime level is particularly promising of research aspects both from a SoSE and a technological viewpoint.*

Keywords: Tracking, filtering, estimation, information fusion, resource management.

1 Introduction

To define a risk able to manage aid decision makers in the transport of dangerous goods, it is useful to refer to a classification hierarchy of the decisional levels that may be associated with the management of that type of transport. This classification consists of four levels:

– The strategic level

– The tactical level

– The operative level

– The level of control in real-time

The four levels are ordered according to the time horizon (in decreasing order) considered and the level of detail of the model used (in increasing order, see table 1).

	Time Horizon	*Level of Detail*
Strategic level	years (>2)	national scale
Tactical level	Months, years (≤2)	multi-regional
Operational level	days	regional scale
Level of control in real-time	seconds, minutes, hours	local scale

Table 1. The four decisional levels

In the strategic level decisions are taken on models on a national scale with a time horizon of a few years (generally 5-10 years). In general, the strategic level decisions involve the highest levels of governments as Decision Makers (DMs) and requires considerable capital repayable only in the long terms. Those decisions may also define the country's transport policies, decide on the investments in major infrastructures, the type of transport services to be provided and pricing policies involving shippers and carriers. Strategic decisions in the context of DG transport by road may be related to prevent or reduce this transport on certain road infrastructures in certain time windows; to prevent or authorize the establishment of a production entity (eg. subject to Seveso legislation) that requires DG transport; definition of a risk index associated to each stretch of the road infrastructures in order to classify them according to the risk exposure.

In the tactical level a time horizon of a few months or a maximum of one or two years is considered. Decisions are made on a model with a level of detail corresponding to a multi-regional geographic area. The tactical planning is based on aggregated and forecasted data that make the decisions sensitive to significant changes in the values considered, eg. due to seasonal changes in demand for transport. At this level typically the DM decides on: the routes to be served, the frequency and times of services to be provided, the general rules for terminal management and allocation of the workforce, the repositioning of the vehicle "empty". Decisions related to DG transport concern the definition of strategies and policies for the DG vehicle travel schedules to minimize the maximum risk exposure in a region; impose reductions or completely inhibition of DG flows on specific stretch of the road, changing hours for transits on some strings at certain times of the day to minimize the risk for people.

The operational level considers a time horizon of a few days and the activities of this level aim at scheduling of the transport in the current week/month. The decisional models are at a regional level or, at maximum, a multi-regional level. The operational decisions concern the planning daily procedures and management of resources according to the daily demand and/or anomalous events that occur. The DM has a detailed knowledge of the status of personnel, vehicles and infrastructure and it has to decide on the scheduling of services, training and maintenance activities;

on routing and dispatching of drivers and vehicles and the allocation of resources.

The level of control in real-time aims at the continuous monitoring of the transport resources through suitable hardware instrumentation installed on the vehicles and on the infrastructure. Potential decisions in the control of this traffic are taken on a local scale and regard a time horizon of a few seconds, minutes or at most some hours. The strategies present at this decisional level are necessary to contrast any hitchs in the transport network such as the temporary unavailability of infrastructure or excessive density of vehicles that transport DGs in a certain stretch of the network. This phase is initially run by the transport companies able to interact with the vehicles in a fast and efficient way, but with the view of preventing risk the local Administration and managers of road infrastructure would also need to participate in this phase.

Three different definitions are proposed for these levels. The paper addressed the real-time definition for which technologies and a system of systems engineering vision could help to enhance its proper application.

2 Methodology

In order to describe the proposed methodology to compute the DG risk associated to road infrastructure, it is necessary introduce some specific definitions.

There is most literature which define risk *R* as a function of set of triplets:

$$R=f(s, p, c) \quad (1)$$

where *s* is a scenario, *p* its probability and *c* its consequences. Risk analysis can be viewed as the process of enumerating all triplets of interest within a spatial and temporal envelope [1]. Territorial vulnerability denotes susceptibility to losses of all above units and structures contained in a territorial entity as well as of their interconnections and linkages. Some researchers emphasize the “exposure” dimension of territorial vulnerability, others consider equally the “exposure” and “coping capacity” dimensions and there is a third group [2] advocating a three dimensional essence of vulnerability (i.e. one comprising “exposure”, “sensitivity” and “adaptive capacity” or “exposure”, “resistance” and “resilience”).
According to the above various conceptual interpretations, different procedures of assessment of risk and territorial vulnerability exist.

In this paper, vulnerability is related to the concept of risk [3]. According to a generally accepted perception, risk includes two components: accident probability and magnitude of the event. The magnitude concerns the event consequences and usually it is related to people, environment and proprieties. If these issues are assigned economical values, consequences for persons and environmental can be quantified as inhabitants and environmental elements involved in the accident. The risk associated with a DG accident X, R(X), then is a combination of the probability of the accident, Pr(X), and the related cost, C(X). A common way of computing the risk is to calculate the product of the two components, R(X) = Pr(X) C(X).

In view of the above, the vulnerability can be treated as the component associated to the exposure of the risk.

So, in the proposed approach, the main components to compute risk associated to DG transport on road are:

- The definition of territorial vulnerability indexes. The vulnerability assessment shall be calculated according to the three types of exposures:

    – social vulnerability (in numbers of inhabitants and the number of road user in the section of the infrastructure);

    – environment vulnerability (in numbers of specific sensible elements within the impact area);

    – economical vulnerability (in numbers of specific elements or propriety within the impact area).

- Data on traffic flows on the infrastructures associated to common vehicles, heavy vehicles and DG vehicles. Obviously, these data should be obtained as a function of time horizon related to the decision levels. In particular, as regards the definition of risk at the strategic level, it will refer to annual average flow data, for the tactical level to monthly average data, for the operating level at daily data and for the real-time level the data will come in real time from the monitored DG vehicles and traffic flows from traffic detector (eg. inductive-loop detectors embedded in the pavement of the roadway).

- The calculation of the accident probability which is the result of a procedure which receives as input the data flow and accident historical and statistical data per km.

The figure 1 explains the methodology to be applied in the four decisional levels.

2.1 Impact area

There are different methods to create the impact zones. In this application, the method proposes by the Italian Civil Protection has been used [4]. This method evaluates, for different kinds of transported DG product and for different accident scenario (explosion, Bleve, fire ball, spills), the distances from the release point in which individuals might be exposed. In that case, the impact distance serves as the radius that defines the impact zone. It is possible to consider the DG shipment over a road segment as the movement of a danger circle along that road segment. These movement carvers out a band on both sides of the road segment thereby defining the region of possible impact.

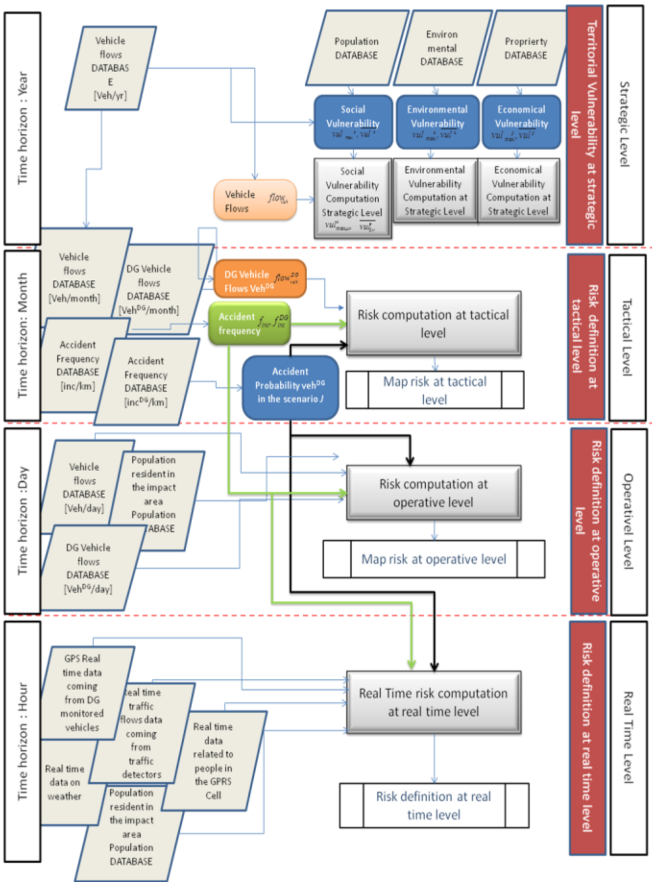


Figure 1. Methodology description

3 Risk definition for DG transport on road at different decisional levels

DG routing analyses should adequately integrate the road network with its surroundings, since risk levels strongly depend on the characteristics of the region traversed by the shipments. In the proposed approach a significant analyses of the impact area is proposed in order to classified each segment of the road infrastructure according to the risk for DG transportation.

3.1 Vulnerability at strategic level

At strategic level, the proposed approach aims to define six different indices related to each section of the road infrastructure. In details, for each limited segment of the road (typically for a DG accident the impact zone has a radius of 100 mt. and the damage zone of 200 mt) the following values should be identified:

1a. Maximum social vulnerability $vul^{s}_{\max}$ [inhab];

1.b. Avarage social vulnerability $\overline{vul^{s}}$ [inhab];

2.a . Maximum environmental vulnerability $vul^{e}_{\max}$ [km^2];

2.b Average environmental vulnerability, $\overline{vul^{e}}$ [km^2];

3.a. Maximum economical vulnerability, $vul^{p}_{\max}$ [€];

3.b Avarege economical vulnerability $\overline{vul^{p}}$ [€].

The social vulnerability considers the sum of two type of information: persons who live in the impact area obtained by Statistics Census and also common users of the specific road infrastructure. Those value results from the classical relationship of macroscopic traffic models among traffic density, speed and flow (such as, for example, in [5]).

$$flow[veh/h] = \delta[veh/km] v[km/h] \tag{1}$$

From the average yearly flows on the specific section of the road, the average value of flows in [veh/h] can be computed. Given an average speed value, it is possible to define the number of vehicle which transit on the specific road section. Hypotizing an average value of two persons for vehicle, it possible to quantified the number of potential users exposed during an DG accident on that road section.

3.2 Risk definition at tactical level

The definition of risk for DG transportation at the tactical level refers to an analysis of data on a monthly horizon.
The risk at tactical level can be formulated as follows:

- Social risk at the tactical level:

$$risk^{s}_{tactical} = \sum_{j} \left(Pinc_{veh} vul^{s}_{\max} flow^{HAZ}_{veh} P_j \right) \tag{2}$$

- Environmental risk at the tactical level:

$$risk^{e}_{tactical} = \sum_{j} \left(Pinc_{veh} vul^{e}_{\max} flow^{HAZ}_{veh} P_j \right) \tag{3}$$

- Economic risk at the tactical level:

$$risk^{p}_{tactical} = \sum_{j} \left(Pinc_{veh} vul^{p}_{\max} flow^{HAZ}_{veh} P_j \right) \tag{4}$$

where

- $Pinc_{veh}$ is the accidents probability per kilometer [accident km-1];
- $flow^{HAZ}_{veh}$ monthly data traffic flows for DG vehicles per km;
- $vul_{\max}$ maximum vulnerability computed on montly data traffic flows;
- P_j is the probability of occurrence for the scenario *j*.

The scenario is related to the DG product involved in the accident and its possible evolution (explosion, Bleve, fire ball, spills).

3.3 Risk definition at operational level

The definition of risk at operational level should be reported at a daily horizon. The analysis of social vulnerability should not be limited to the person resident in the surrounding of the occurrence point of the incidental event but should also take account of all those people that really can be found in the area potentially exposed. The main object is to quantify correctly people, environmental sensible elements and enumerated buildings and other industrial or residential structures located in the impact area. This is possible by the analysis of Geografical Information System (GIS) or simply, when possible, by the visualization of aerial maps or orthophotos.

- Social risk at the operational level:

$$risk^{s}_{operative} = \sum_{j} \left[Pinc_{veh} vul^{s}_{operative} flow^{HAZ}_{veh} P_j \right] \quad (5)$$

- Environmental risk at the operational level:

$$risk^{e}_{operative} = \sum_{j} \left(Pinc_{veh} vul^{e}_{operative} flow^{HAZ}_{veh} P_j \right) \quad (6)$$

- Economic risk at the operational level:

$$risk^{p}_{toperative} = \sum_{j} \left(Pinc_{veh} vul^{p}_{operative} flow^{HAZ}_{veh} P_j \right) \quad (7)$$

3.4 Risk definition at real time level

The risk definition at real time is the most complex from the computational viewpoint because it requires the acquisition of different types of data monitored over a hourly period time at maximum.

As at operational level, in the real time case, the social risk is calculated observing people exposed not only in the centers of vulnerability but also on the road infrastructures.

A possible method to quantify the expected number of people involved is making a census, at various times of the day, of mobile phones in each GPRS cell contained in the impact area. The goal is to get a correct evaluations of people exposed in the area to be enable rapid decision making and response in case of a DG accident.

The quantification of this issue is important and essential especially in those situations where the number of residents at risk is modest, and in which the contribution of people present on the infrastructure is comparable or greater than the resident population.

Social risk at real time decision level:

$$risk^{s}_{realtime} = \sum_{j} \left[Pinc_{veh} * g\left(\alpha vul^{s}_{\max} + \beta vul^{GPRS}_{\max}\right) * flow^{HAZ}_{veh} * P_j \right] \quad (8)$$

Environmental risk at real time decision level:

$$risk^{e}_{realtime} = \sum_{j} \left(Pinc_{veh} * vul^{e}_{\max} * flow^{HAZ}_{veh} * P_j \right) \quad (9)$$

Economic risk at real time decision level:

$$risk^{p}_{realtime} = \sum_{j} \left(Pinc_{veh} * vul^{p}_{\max} * flow^{HAZ}_{veh} * P_j \right) \quad (10)$$

where

$g\left(\alpha vul^{e}_{\max} + \beta vul^{s}_{GPRS}\right)$ is a linear combination of the values associated to persons who live in the impact area and the expected number of persons computed by the census of mobile users in GPRS cells. The parameters g, α, β have to be calibrated according to time windows.

4 Case study to define social risk for DG transport on road at different decisional levels

The case study focuses on a typical Italian highway long on 105 km. The highway is divided on 14 stretches. The following table shows the average daily traffic flows computed on annual data.

		2010 Traffic Flows data				
Highway		Direction 1		Direction 2		
Stretches	KM	Total vehicles	Freight Vehicle (%)	Total vehicles	Freight Vehicle (%)	Total on the two directions
1	0,0	22367,15	15%	16802,6	16%	39169,75
2	4,1	8677,53	2%	10069,0	21%	18746,55
3	12,0	15689,08	17%	14925,4	17%	30614,48
4	18,8	14557,24	16%	13721,4	16%	28278,69
5	22,5	14299,15	18%	13428,8	18%	27727,97
6	35,2	11972,76	20%	11538,4	19%	23511,19
7	43,4	11758,90	19%	10996,2	20%	22755,15
8	49,1	11253,42	20%	10732,1	20%	21985,60
9	56,1	10698,32	21%	9831,00	21%	20529,32
10	68,4	10440,00	21%	9413,53	21%	19853,53
11	76,8	9010,31	22%	8585,21	21%	17595,52
12	88,7	9655,19	20%	9008,21	20%	18663,40
13	97,0	8647,77	22%	8135,60	22%	16783,37
14	105	9656,85	18%	9168,77	18%	18825,63

Table 2. Average daily traffic flows computed on annual data (2010).

Highway strecthes	1	2	3	4	5	6	7
Population density [inhab/km2]	914	329	339	887	951	613	213
Highway strecthes	8	9	10	11	12	13	14
Population density [inhab/km2]	272	861	761	408	923	974	455

Table 3. Population density [inhab/km2]

4.1 Social vulnerability computation at strategic level

This highway has two lanes for carriageway which run parallels in the two direction so, considering an impact area with a radius of 100 mt, the persons who transit in both two directions can be exposed.

Taking into account 60 km /h as average speed vehicle and two persons for vehicles, the index associated to the road users potentially exposed on each stretch of the highway, in case of DG accident, can be computed. The maximum social vulnerability at strategic level can be calculated as in the table 4.

4.2 Social risk computation at tactical level

At the strategic decisional level, stretches 1 resulted the main vulnerable so the tactical social risk will be computed on this stretch of highway for each month. For the selected road segment, the DG traffic represents the 3,45% of the total freight traffic. Like at strategical level, the social vulnerability at tactical level is computed as the sum of people who transit on the highway (based on monthly traffic flows data) and population density in the impact area. Besides, the accident probability for the selected highway is 8,63E-07 [acc km-1] and the probability that scenario associated to the GPL explosion has 10^{-2} order of magnitude. Table 5 shows value to compute social risk at tactical level according to eq. (2).

stretch	Traffic flows Two direc tions	Vehicular Density (average speed at 60 km/h)	Persons on the stretch (2 inhab per veh)	Persons on the strech (2 inhab per veh)	Resident persons (pop density)	vul^{s}_{max}
	[Veh/h]	[veh/km]	[inhab/km]	[inhab/hm]	[inhab/hm2]	Value per hm
1	1632,07	27,20	54,40	5,44	9,136	14,58
2	781,11	13,02	26,04	2,60	3,294	5,90
3	1275,60	21,26	42,52	4,25	3,388	7,64
4	1178,28	19,64	39,28	3,93	8,874	12,80
5	1155,33	19,26	38,51	3,85	9,506	13,36
6	979,63	16,33	32,65	3,27	6,132	9,40
7	948,13	15,80	31,60	3,16	2,128	5,29
8	916,07	15,27	30,54	3,05	2,724	5,78
9	855,39	14,26	28,51	2,85	8,607	11,46
10	827,23	13,79	27,57	2,76	7,606	10,36
11	733,15	12,22	24,44	2,44	4,075	6,52
12	777,64	12,96	25,92	2,59	9,232	11,82
13	699,31	11,66	23,31	2,33	9,737	12,07
14	784,40	13,07	26,15	2,61	4,549	7,16

Table 4. Social vulnerability at strategic level

	DG Traffic flows Two directions	Persons on the stretch (2 persons for veh)	Resident Persons	Accident probability	Probability for GPL explosion	Social Vulnerability per month
	[Veh/h]	[inhab/hm]	[inhab/hm2]	[Acc/hm]	[Acc/hm]	value per hm
January	184,67	4,87	9,14	8,63E-08	0,001	2,23E-07
February	208,07	4,74	9,14	8,63E-08	0,001	2,49E-07
March	235,18	5,18	9,14	8,63E-08	0,001	2,90E-07
April	237,42	6,66	9,14	8,63E-08	0,001	3,24E-07
May	233,17	6,70	9,14	8,63E-08	0,001	3,19E-07
June	234,54	7,53	9,14	8,63E-08	0,001	3,37E-07
July	253,22	9,24	9,14	8,63E-08	0,001	4,01E-07
August	234,64	9,44	9,14	8,63E-08	0,001	3,76E-07
September	261,92	6,75	9,14	8,63E-08	0,001	3,59E-07
October	250,67	5,53	9,14	8,63E-08	0,001	3,17E-07
November	237,36	4,69	9,14	8,63E-08	0,001	2,83E-07
December	212,55	4,99	9,14	8,63E-08	0,001	2,59E-07

Table 5. Social risk at tactical level for GPL explosion on a specific stretch of the highway

From the table 5, it is evident that the social risk definition for the same sector of the infrastructure is subject to variation due to the seasons and, for economical reasons, also to the different periods of the year.

4.3 Social risk computation at operational level

At operational level, the time horizon is daily and a deep analysis of the exposed sensible targets in the neighbouring of the accident point has to be done. In the selected impact area (see figure 2), two training schools and a industry appear. At operational levels, the definition of residents can be refined according to the specific working day and location.

Figure 2. Arial visualisation of the selected impact area.

From table 5, the social risk at operational level increases during the working days Tuesday and Friday. Special attention, for this stretch of the highway, should be given forcing a reduction of DG flows or allowing transits during the night to minimize social risk.

DG Traffic flows Two directions	**Persons on the stretch (2 persons for veh)**	**Resident Persons from arial visulization**	**Accident probability**	**Probability for GPL explosion**	Social risk at operational level
[Veh/h]	[inhab/hm]	[inhab/hm2]	[Acc/hm]	[Acc/hm]	Value per hm

Monday 01/07/2010	276,3	7,9	750,0	8,63E-08	0,001	1,413E-04
Tuesday 02/07/2010	262,0	9,3	750,0	8,63E-08	0,001	1,573E-04
Wednesday 03/07/2010	114,0	10,5	750,0	8,63E-08	0,001	7,777E-05
Thursday 04/07/2010	79,7	9,6	750,0	8,63E-08	0,001	4,964E-05
Friday 05/07/2010	261,5	9,2	750,0	8,63E-08	0,001	1,563E-04
Saturday 06/07/2010	275,0	6,5	500,0	8,63E-08	0,001	7,701E-05
Sunday 07/07/2010	285,9	6,7	500,0	8,63E-08	0,001	8,270E-05
Monday 08/07/2010	292,7	7,7	750,0	8,63E-08	0,001	1,453E-04
Tuesday 09/07/2010	263,6	10,1	750,0	8,63E-08	0,001	1,724E-04
Wednesday 10/07/2010	107,2	10,7	750,0	8,63E-08	0,001	7,426E-05
Thursday 11/07/2010	79,3	9,8	750,0	8,63E-08	0,001	5,029E-05
Friday 12/07/2010	263,9	9,6	750,0	8,63E-08	0,001	1,643E-04
Saturday 13/07/2010	275,1	6,8	500,0	8,63E-08	0,001	8,059E-05
Sunday 14/07/2010	243,9	7,0	500,0	8,63E-08	0,001	7,360E-05

Table 6. Social risk at operational level in a selected impact area for 15 day in July.

4.4 Social risk computation at real time level

There are many new Information Technology Systems (ITS) that promise to reduce the effects of transportation hazards. The suite of geospatial technologies including the global positioning system (GPS), geographic information systems (GIS), and remote sensing also hold much promise to improve the amount of information available to transportation users, planners, and emergency responders.

At real time level, in fact, the definition of risk for DG transportation implies to receive timely data about DG vehicle positions, traffic flows, and the expected value of people really present in the impact area, e.g. by quantification of GPRS mobile users in the specific cell. Knowing the DG vehicle position by GPS on the stretch, it possible to outline the danger circle along that road segment. Besides, receiving real time traffic flows data by inductive-loop detectors, it is possible to estimate highway users in that section. Given the current impact area, the proposed methodology suggests to take advantages by the new technologies which, in real time, can estimated the number of GPRS mobile users in a selected cell. The average value for mobile traffic in a cell is 10 erl/km^2, so it can be assumed 600 min of mobile traffic available. Assuming an average value of 3 min of calls for user, 200 users for km^2 is estimated. Taking into account the same values for accidents probability per kilometre and probability of DG accident scenario, the social risk definition at real time level can be computed as eq. 8.

Due to the low probability values, the risk indices have low order of magnitude. The main goal of this approach is define a series of thresholds values to classify each sectors of the road infrastructures.

Traffic flows Two directions	DG Traffic flows Two directions	Mean speed	Persons on the stretch (2 persons for veh)	Mobile Users	$Pinc_{veh}$	P_j	DG risk at real time level
[veh/min]	[veh/h]	[km/min]	[inhab/hm]	[inhab/hm2]	[Acc/hm]	[Acc/hm]	value per km
14,9	9,65600	1,7	1,7	2,0	8,6E-08	1,0E-03	3,1E-09

Table 7. Social risk at real time level.

5 Conclusion and Future results.

The proposed methodology to estimate the risk index associated to each road section in the different decision level aims at proposing an objective method to categorize each infrastructure versus DG transportation risk. The presented approached promise to be useful to support governments and the different decision makers involved in the DG transport in allocating resources to the phases of management: mitigation, preparedness, emergency response and recovery.

As a future development of this work, the current analysis will be implemented from the DG fleet manager viewpoint. Each DG transportation company could be able to certificate its transport providing objective informations about the planned routing of its vehicles. For each planned routing vehicle, the DG company could compute the impact of its DG transports on the social, environmental and economic exposure. By the daily transmission of those parameters to the competent public authorities, the DG company should certificate an effective effort to minimize risk in its vehicle routing planning, obtaining, when possible, economical or operative facilitations.

References

[1] Kaplan, S., and Garrick, B.J. (1981) On the quantitative definition of risk, Risk Analysis, 1: 11-27.
[2] Berdica, K. (2002) An introduction to road vulnerability: what has been done, is done and should be done, Transport Policy, 9: 117-127.
[3] Cova, T.J. and Conger, S. (2004) Transportation hazards, in *Transportation Engineers' Handbook*, M. Kutz (ed.), *in press*.
[4] Pianificazione dell'emergenza esterna degli stabilimenti industriali a rischio d'incidente rilevante. Presidenza del Consiglio dei Ministri, Dipartimento Protezione Civile. Linee guida 2004, Italy.
[5] Sanwal K.K., Petty K., Walrand J., An extended macroscopic model for traffic flow, Transportation Research Part B, (1996), 30(1), l-9.

Integration of a Bayesian network for response planning in a maritime piracy risk management system

Chaze X.
MINES ParisTech - CRC
BP 207. 1 Rue Claude Daunesse
06904 Sophia Antipolis Cedex, France
xavier.chaze@mines-paristech.fr

Bouejla A., Napoli A., Guarnieri F.
MINES ParisTech - CRC
BP 207. 1 Rue Claude Daunesse
06904 Sophia Antipolis Cedex, France.
amal.bouejla@mines-paristech.fr
aldo.napoli@mines-paristech.fr
franck.guarnieri@mines-paristech.fr

Abstract - *This article describes an innovative system to protect offshore oil infrastructure against maritime piracy. To detect and respond efficiently to this threat, many factors must be taken into account, including the potential target, the protection methods already in place and operational and environmental constraints, etc. To improve the handling of this complex issue, we have designed a system to manage the entire processing chain; from threat identification to implementation of the response. The system implements Bayesian networks in order to capture the multitude of parameters and their inherent uncertainties, and to identify and manage potential responses. This article describes the system architecture, the integrated Bayesian network and its contribution to response planning.*

Keywords: Maritime piracy, Oil platforms, SARGOS, Bayesian networks, International Maritime Organisation, Expert knowledge.

1 Presentation of the SARGOS system

1.1 Context

Offshore oil extraction currently accounts for about one-third of global oil production. Despite its scarcity, this source of energy is under active exploration in many parts of the world, notably in hazardous territorial waters such as the Gulf of Guinea, and particularly off the Nigerian coast.

Since 2005, the number of acts of piracy against oil fields and especially ships has grown steadily (in 2011, 552 attacks on ships and platforms were registered with the International Maritime Bureau[1]). Attacks on infrastructure generate significant additional costs arising from the payment of ransoms, the installation of security equipment, and increased insurance premiums, etc. These additional costs directly affect the international price of oil [1] and [2]. Although attacks on oil platforms are less frequent and certainly less publicised, they are extremely disturbing because of the severe impact on the crew (personnel may be taken hostage, injured or even killed), infrastructure (facilities may be damaged or destroyed), the economy (price spikes) and the environment (oil spills). The lack of effective tools for infrastructure protection means that actors involved in the offshore oil and gas industry find themselves helpless. One example is the attack on the Exxon Mobil platform [3] off the coast of Nigeria, which led to the kidnapping of nineteen employees and extensive damage to the facility caused by the explosive devices used by the pirates. Such incidents are prime examples of the weaknesses in current anti-piracy systems. At the present time, oil installation security is provided by so-called classical tools (radio identification, radar, Automatic Identification Systems, etc.), which, despite their usefulness in detection, cannot provide a response tailored to different types of threats (fishing boat, jet ski, etc.). Moreover, their effectiveness depends on many parameters related to both the environment and technical and operational constraints.

1.2 SARGOS objectives

To meet this new need for the protection of civilian infrastructure, the French National Research Agency (ANR) has funded the SARGOS[2] system. The project is approved by French regional bodies and brings together a multi-disciplinary consortium[3] of partners with complementary skills. The aim is to design a system to improve infrastructure protection and offer a new method that is able to both detect threats and plan a response because at the present time, there is no comprehensive system capable of managing the entire threat processing chain.

[1] International Chamber of Commerce International Maritime Bureau's Piracy Reporting Centre (http://www.icc-ccs.org)

[2] Graduated Offshore Response and Alert System (*Système d'Alerte et de Réponse Graduée OffShore*).

[3] The SARGOS project includes participants from private sector organisations such as DCNS (a French naval shipbuilder) and SOFRESUD (a supplier of high-tech equipment to the defence industry), and public research centres including ARMINES (a French contract research organisation) and TéSA (Telecommunications for Space and Aeronautics).

To achieve this, the system must be capable, in the case of a confirmed intrusion, of generating an alarm and initiating an internal and external response appropriated to the danger level of the situation. This response has to be implemented through a graduated series of non-lethal counter-measures (sonic cannons, barring infrastructure access, etc.).

1.3 System architecture

The SARGOS system architecture consists of two major sub-systems. First, a module for the detection, tracking and classification of threats in the marine environment: using powerful instrumentation (FMCW[4] radar, infrared cameras, etc.) the SARGOS system can identify a potential intrusion and generate an alert report that provides an inventory of all the relevant parameters necessary to characterise the threat. And a module to formalise and model graduated responses: taking into account the evolution of the situation, regulatory constraints and the operational infrastructure, the data contained in the alert report just described is used to define an appropriate response to deter or repel attackers.

The functional diagram of the SARGOS system demonstrates this threat processing cycle (Figure 1).

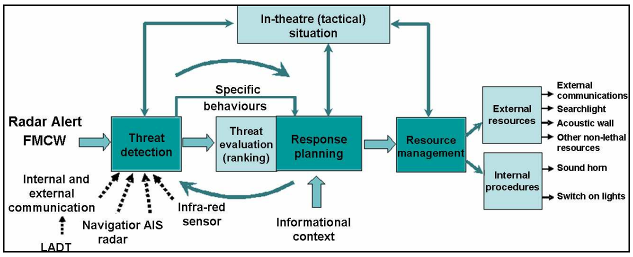


Figure 1. Functional diagram of the SARGOS system

The safety and security of the infrastructure is assured by the application of a response plan generated by the system, which triggers a series of progressive and reversible actions and reactions.

2 Contribution of a Bayesian network

Actually, the great weakness in this process lies in the preparation of the diagnosis used for planning the response. To overcome this shortcoming, we propose a new approach that is able to generate automated response plans, tailored to the nature of the detected intrusion.

2.1 Why a Bayesian network?

A detailed investigation of the issues highlights significant constraints that the SARGOS system must take into account in order to fully reflect the complexity of a situation [4]. On the one hand, the large number of variables to be included (representing the threat, the target, the environment, etc.) and the dependencies that may exist between them suggest the development of a decision support system based on graph theory. On the other hand, the uncertainty inherent in certain variables (threat identification, intervention options, etc.) emphasises the need for a system based on probability theory and probabilistic calculations [5]. With these two approaches in mind, a process for the automatic preparation of response plans tailored to the nature of the detected intrusion, based on Bayesian networks was explored [6] and [7].

We focus particularly on the contribution of Bayesian inference techniques that are applied to, on the one hand, a maritime database and on the other to expert knowledge in the domains of offshore oil and maritime safety. Data from the database and expert knowledge are modelled using Bayesian networks, tools based on Thomas Bayes' theorem (1).

$$P(A/B) = \frac{P(B/A)\,P(A)}{P(B)} \qquad (1)$$

The theorem is used in statistical inference to update probability estimates from observations and the probability distributions applicable to these observations. A Bayesian network represents knowledge in a way that makes it possible to calculate conditional probabilities [8]. Widely used for diagnosis (medical or industrial), Bayesian networks capitalise and exploit knowledge, and are particularly suitable for capturing and reasoning with uncertainty inherent in many complex problems [9], [10] and [11].

2.2 Software used

Among the existing softwares specialised in Bayesian networks, it was decided to choose the Bayesia software series which proposes on one hand an intuitive desktop solution that experts have easily learned to use in a very short time and on the other hand an Application Programming Interface (API) which can include a previously created Bayesian network into a standalone module.

Indeed, BayesiaLab[5] software was first used to automatically generate an initial network (from existing piracy data) by suggesting dependencies between the principal variables [12]. The software was used again in the second stage (by experts) to determine the complete architecture of the final SARGOS network (cf. §2.3).

This desktop version was then used to test the results of the model developed using simulated scenarios. Thanks to the graphical interface, experts can create realistic attack scenarios by determining the modalities of their choice. The

[4] Frequency Modulated Continuous Wave

[5] BayesiaLab software is developed by the French company Bayesia (http://www.bayesia.com/)

Bayesian network then calculates the resulting probabilities which are analysed by the experts to improve iteratively the initial modalities and probabilities (cf. §3.2).

Finally, the API provided by the Bayesia software series provides an efficient tool to operate an existing Bayesian network automatically and integrate it in a standalone system, making so possible a real-time use (cf. §3.3).

2.3 Implementation method

The approach used to construct the SARGOS Bayesian network consists of two complementary steps. First, an initial Bayesian network was constructed using data from the 'Piracy and Armed Robbery database' of the International Maritime Organisation (IMO[6]). This is the only database in existence that holds historic records of pirate attacks in the maritime environment. On 15th July, 2011 the database contained records of 5,502 attacks (dating back to 1994) and the data noted for each attack included: the name of the asset under attack, the number of attackers, the weapons used, the measures taken by the crew to protect themselves, the impact on the crew and the pirates, etc.

This approach served two purposes: first, it made it possible to determine the principal tools and measures used by the crew to protect themselves, to evaluate their effectiveness and to define the probability of certain types of attack; and secondly it helped to define an initial framework for the formalisation of knowledge related to acts of maritime piracy.

The second step leveraged expert knowledge in the oil and safety domains. As the information contained in the IMO database related primarily to attacks on shipping, the contribution of knowledge from domain experts made it possible to extend the system to include oil fields [13]. Using the Bayesian network created from the IMO data, experts were able to share and transfer knowledge that was then used to build the final Bayesian network and complete the architecture in order to make it as versatile as possible (nodes and arcs were added, modalities and probabilities were modified) [14]. In this way, the data extracted from the IMO database was combined with the experience of experts, through the course of multiple brainstorming sessions, in order to address the a priori lack of knowledge and experiential feedback.

Furthermore, in future, we could also imagine improve our specific knowledge. Once SARGOS systems will be operational on several platforms or offshore infrastructures, all events which will be treated will come enrich a database of existing cases. From this historical data and continuous real-time flow of data, it would be possible to define datamining rules in order to discover new knowledge [15]. These rules will supply an automated reasoning rule-based knowledge to allow the automatic identification of abnormal behavior of vessels typical of a risk of maritime piracy attack.

[6] http://www.imo.org

3 Results and integration in the SARGOS system

3.1 Model developed

The basic architecture of the SARGOS response planning network consists of four modules and five sub-modules (Figure 2).

The modules are: Basic parameters; Aggravating factors and constraints; the Overall danger level of the situation; the Countermeasures. Basic parameters are static or dynamic physical data that characterise the threat and the target. They are either obtained directly from the alert report or are derived from it. Aggravating factors make it possible to take into account the potential deterioration of the situation, while constraints are parameters that must be taken into account to ensure the effectiveness of the response both technically and operationally. The overall danger level of the situation is derived from the basic parameters. Its assessment takes into account the potential consequences of the problem created by the threat and the vulnerability of the target. Countermeasures are all the defences implemented by the target in order to return as quickly as possible, and in the best condition possible, to a safe situation. Countermeasures are classified into five sub-modules according to the danger level of the situation and the operational availability of on-board equipment.

These sub-modules are: Communication and distress calls; Deterrence and low-impact repulsion measures; Repulsion, anti-boarding and neutralisation measures; Procedure management; Ensuring the safety and security of the facility. From these sub-modules, the Bayesian network proposes a set of countermeasures that may be activated according to the estimated danger level (for example: activate the safety system and silent alert, etc.).

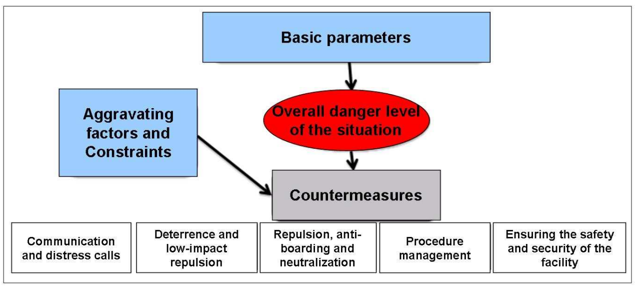


Figure 2. Functional diagram of the SARGOS system

The planning report generated by the network makes it possible to make a rational assessment of probabilities and formalises expert knowledge. This report constitutes the interface between the processing of the alert report and exploitation of the results of the Bayesian network. It is a summary of the essential information needed to actually trigger response procedures. Consequently the probability of activating a particular countermeasure will obviously vary according to the situation.

3.2 Simulation of attack scenarios

Once the probability distribution of the various modalities has been established, an interesting exercise is to test the Bayesian network by using it to simulate different attack scenarios through the selection of certain criteria. An examination of these scenarios made it possible to finalise the network before integrating it into the SARGOS system.

The example below (Figure 3) shows how response planning is tailored to the danger level of the situation and can adapt to changes in parameters representing the threat and the target. Specifically, it shows the results of setting parameters to simulate an attack on a Floating Production, Storage and Offloading (FPSO) unit by an unknown vessel. This example shows that the danger level of the situation, at time T1, was 2 with a 64.68% probability of occurrence. In this case the counter-measures to be applied were: inform the crew master, request the intervention of the security vessel, broadcast a strong message by loudspeaker, turn on the searchlight and activate the security post.

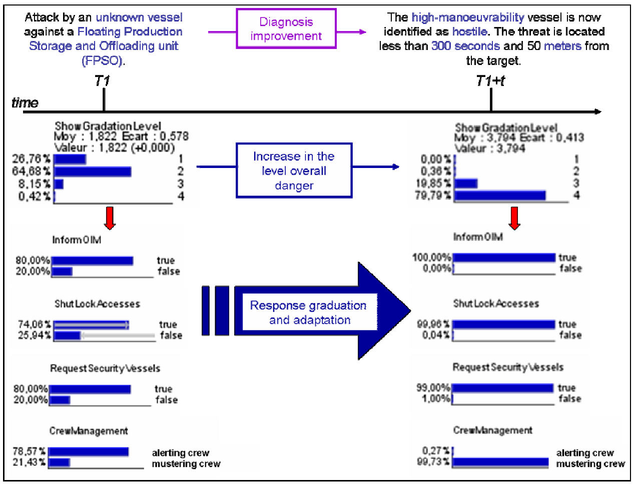


Figure 3. Evolution of response planning as more information about the situation becomes available

At time T1+t the attacker has been identified as hostile and equipped with a highly maneuverable boat. The parameters that impacted response planning were: the ranking between the threat and the target (i.e. the time required for the threat to cover the remaining distance to the target), the distance between the threat and the target and the response time of the security vessel. The danger level is now 4 with a 79.79% probability of occurrence. This higher level requires a more forceful response, reflected in the recommended measures: assemble the crew, secure the installation and block access to sensitive areas.

The creation of attack scenarios makes it possible to refine the probability of an attack and test the response of the Bayesian network by changing the parameters that represent the threat, the target, the environment, etc.

3.3 Integration of the Bayesian network into the SARGOS system

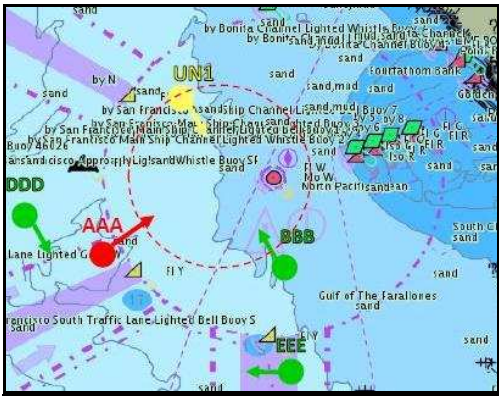
In order to integrate the Bayesian network into the SARGOS system, a prototype was developed that included an alert report as input and a planning report (which listed all the counter-measures to be applied either by the crew or automatically by the system) as output. The BayesiaEngine software provides a module that makes it possible to select and set attack parameters. This module consists of an application programming interface (API) and a Java library. Intermediate calculations are carried out on the basis of these parameters and the results are fed into the enhanced Bayesian network created from expert knowledge.

The resulting list of counter-measures varies according to the attack scenario. Consequently, a threshold must be set in order to only activate those measures that provide the most relevant response at a particular time, and in a particular situation. This threshold was set at 70%. In other words, only those counter-measures where one of the modalities had a probability greater than 70% were selected for further processing. This threshold was arrived at by domain experts as it reflects actual events in more than two-thirds of real-life cases. Following an extensive period of testing, the selected counter-measures were found to correspond to realistic and reliable responses.

The SARGOS system can handle multiple threats contained in a single alert report. Consequently, priorities must be established. In the system, the first threat to be treated is always the one where time available to react is the shortest for the target that is most exposed. Figure 4 shows the user interface of the SARGOS system, and demonstrates how multiple threats can be processed simultaneously.

In this example, the system has detected several potential threats heading towards the oil field and has classed them into ‘Enemy’, ‘Unknown’ or ‘Friend’. An alert is only generated following a classification of Enemy or Unknown.

Once a threat has been detected and analysed, the counter-measures are selected and added to the response planning report prepared in a specific order. The main factors determining this order of priority were: the action mode of the counter-measure, its ease of implementation, the degree of automation or the need for a large number of crew members to activate it, the time required for it to become effective and its potential additional functions.



Class	ID	CPA	TCPA	Type
Enemy	AAA	2. 3 m	3. 30 mn	Small Boat
Unknown	UN1	1.5 m	2.20 mn	
Friend	BBB	2.8 m	4.15mn	Tanker
Friend	CCC	0		Tanker
Friend	DDD	0		Small boat

Figure 4. The user interface of the SARGOS system showing threat prioritisation

The planning report is divided into two parts: the first concerns communication and a general request for assistance directed at the entire oil field; the second concerns the specific asset at risk. The response planning report also displays the counter-measures to be activated in chronological order.

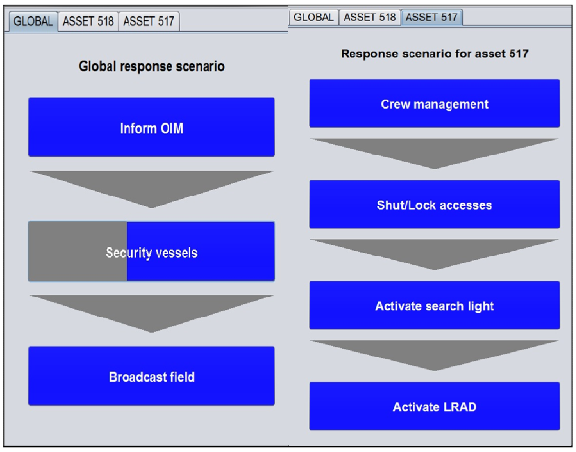


Figure 5. The SARGOS user interface showing global (left-hand side) and specific (right-hand side) countermeasures to be deployed

In the above example (Figure 5) the global counter-measures are, in order: inform the crew master, request the intervention of the security vessel and broadcast information about the attack to other installations in the field. The specific measures are: assemble the crew, block access to the infrastructure, activate searchlights and activate the noise cannon (Long Range Acoustic Device; LRAD). The representation of the probability that a particular measure will be implemented can be seen in the counter-measure ‘Security Vessels’, where the proportion of the blue segment suggests a 60-70% probability that this method will be called upon.

4 Conclusions

Response planning in the SARGOS system results in the preparation of a response planning report based on an intelligent assessment of the alert report. The response planning report includes all the information necessary for the physical implementation of measures to protect against a threat.

Using a Bayesian network for response planning is a major benefit of the SARGOS system as the network is able to manage all possible interactions between threat characteristics, the target, the environment, the crew and the facilities. It can adapt to real-time changes in the danger level of the situation.

Network scalability is also made possible through integration of feedback related to the processing of attacks previously managed by the system. The planning module can therefore be updated and improved iteratively.

Finally, in order to improve the modeling of knowledge embedded in the Bayesian network, an interesting approach would be to draw upon an appropriate ontology [16]. The use of a suitable ontology would make it possible to formalise knowledge upstream of the Bayesian network in order to consolidate the steps of threat detection and identification.

References

[1] One Earth Future, (2011) “The Economic Cost of Piracy”. [online]. Available from http://oneearthfuture.org/images/imagefiles/11%2001%20OBP-Brochure-A4.pdf (Accessed 15 May 2012)

[2] L’Institut Supérieur d’Economie Maritime, (2010) “Piraterie: perturbation de l’économie maritime ? [Piracy: Disruption to the Maritime Economy?]” [online]. Mer et Marine, October 2010. Available from http://www.meretmarine.com/article.cfm?id=114482 (Accessed 15 May 2012)

[3] M.A. Giraud, B. Alhadef, F. Guarnieri, A. Napoli, M. Bottala Gambetta, D.Chaumartin, M. Philips, M. Morel, C. Imbert, E. Itcia, D. Bonacci, P. Michel, (2011) “SARGOS: Securing Offshore Infrastructures Through a Global Alert and Graded Response”, System Workshop MAST Europe, June 27-29 2011

[4] B.S. Ware, A.F. Beverina, L. Gong, B. Colder, (2002) “A Risk-Based Decision Support System for Antiterrorism”, *Digital Sandbox*, 8 pages, August 14 2002

[5] S. Scott, “A Bayesian paradigm for designing intrusion detection systems”, (2004) *Computational Statistics & Data Analysis*, vol. 45, no. 1, p. 69–83, 2004.

[6] R. Dantu and P. Kolan, (2005) “Risk management using behavior based bayesian networks”. *Intelligence and Security Informatics*, Vol. 3495, pp. 115-126.

[7] L.D. Hudson, B.S. Ware, S.M. Mahoney and K.B. Laskey, (2002) “An Application of Bayesian Networks to Antiterrorism Risk Management for Military Planners”. [online] Available at http://digilib.gmu.edu:8080/jspui/bitstream/1920/268/1/Antiterrorism.pdf (Accessed 15 May 2012)

[8] P. Naïm, P.H. Wuillemin, P. Leray, O. Pourret and A. Becker, (2007), *Réseaux bayésiens*. [*Bayesian Networks*], 3rd ed., Eyrolles.

[9] P.R. Kannan, (2007) “Bayesian networks: Application in safety instrumentation and risk reduction”. *ISA Transactions*, Vol. 46 No 2, April 2007, pp. 255-259.

[10] C.J. Lee and K.L. Lee, (2006) “Application of Bayesian network to the probabilistic risk assessment of nuclear waste disposal”. *Reliability Engineering & System Safety*, Vol. 91 No 5, May 2006, pp. 515-532.

[11] J.E. Martín, T. Rivas, J.M. Matías, J. Taboada and A. Argüelles, (2009) “A Bayesian network analysis of workplace accidents caused by falls from a height”. *Safety Science*, Vol. 47 No. 2 February 2009, pp. 206-214

[12] SûretéGlobale.org, (2008) “Apport des réseaux bayésiens dans la prévention de la délinquance. [The Contribution of Bayesian Networks to Crime Prevention]”. [online]. Available from http://www.sureteglobale.org/pdf/reseaux%20bayesiens.pdf (Accessed 15 May 2012)

[13] X. Chaze, A. Bouejla, A. Napoli, F. Guarnieri, T. Eude et B. Alhadef, (2012) “The Contribution of Bayesian Networks to Manage Risks of Maritime Piracy against Oil Offshore Fields”, in ITEMS 2012 (Information Technologies for the Maritime Sector), Busan, South Korea, *Database Systems for Advanced Applications*, Lecture Notes in Computer Science Volume 7240, 2012, pp 81-91

[14] L. Torti and P.H. Wuillemin, (2009) “Modélisation de réseaux bayésiens de très grandes tailles [Large Scale Bayseian Network Modelling]” in MajecSTIC, 16-18 November 2009, Avignon, France.

[15] B. Idiri, A. Napoli, (2012) « Découverte de règles d'association pour l'aide à la prévision des accidents maritimes », in Conférence Internationale Francophone sur l'Extraction et la Gestion de Connaissance (EGC 2012), Bordeaux, France, *Revue des Nouvelles Technologies de l'Information* RNTI-E-23[2], Editions Hermann, pp 243-248.

[16] A. Vandecasteele and A. Napoli, (2012) “Spatial ontologies for detecting abnormal maritime behaviour” in OCEANS 2012, 28 May 2012, Yeosu, South Korea.

Access and Monitor Vulnerability of Urban Metro Network System in China

Zhiru Wang
Construction Management and Real Estate Department
Southeast University
Nanjing, Jiangsu, China
zhiru.wang@seu.edu.cn

Qiming Li
Construction Management and Real Estate Department
Southeast University
Nanjing, Jiangsu, China
njlqming@163.com

Jingfeng Yuan
Construction Management and Real Estate Department
Southeast University
Nanjing, Jiangsu, China
101011337@seu.edu.cn

Zhipeng Zhou
Construction Management and Real Estate Department
Southeast University
Nanjing, Jiangsu, China
zhou_zhipeng@seu.edu.cn

Ruoyu Jia
Construction Management and Real Estate Department
Southeast University
Nanjing, Jiangsu, China
220111050@seu.edu.cn

Ying Lu
Construction Management and Real Estate Department
Southeast University
Nanjing, Jiangsu, China
luying_happy@126.com

Abstract – *It is a transition period for UMNSs in China, as operation mode is changed from single line operation to network operation. With the booming construction of UMNS in many big cities, interdependence and cascading relationship is more and more clear. Impropriate handled of failures in one line would induce malfunction in related lines or even cascade the whole operation network. This research takes UMNS as object and focuses on vulnerability analysis of it. The objectives of this research contain: exploring the connotation and formation mechanism of vulnerability; identifying the relationship between characters of the network and vulnerability of UMNS; assessing physical, structural, and social functional vulnerability of UMNS; establishing visualized dynamic monitoring system and case study.*

Keywords: Urban Metro Network System (UMNS), vulnerability, near-miss, complex network, dynamic monitor.

1 Introduction

Increasing in density of people in cities and expansion of city range encouraged building of high capacity public transportation infrastructures. With the growth of the population and private cars in large cities, ground public transportation system cannot avoid congestions which would reduce the travel time reliability of passengers. Therefore, right-of-way and track based public transportation was encouraged as most of them run underground which could solve the land use and provide reliable service. Additionally, large competitive sports events and international meetings is another reason for the booming construction of UMNS.

Track based public transportation can be divided into two classes from whether it is running in one city. It is considered as commuter, intercity, and high-speed rail if it is running in two or more city, if not, it is considered as trams, light rail and metro, underground, subway, and rapid transit. UMNS in this research is defined as railways operated in an urban area with high capacity, high frequency, right-of-way, long operation time, and short distance between two stations. Both metro (subway, underground, tube, and rapid transit) and light rail are included in the UMNS in this research.

UMNS is one of the fundaments of modern society. It is important to the welfare of people and the economic efficiency of businesses. As a result, once it fails function by the disturbance, it would lead to increases in travel time that impair the ability of individuals to take part in their daily activities, including commuting, dropping off and picking up children from day-care and school, doing the shopping, etc. for businesses, the impacts can include delayed deliveries and supplies, loss of customers and manpower, increased freight cost.

UMNS is a system of system which is composed of railway system and other four subsystems containing: traction power system, telecommunication system, auxiliary power system, and electrical in-feed system [1]. The function of a subsystem influences the function of other systems. Disturbances in one system would transverse to other dependent system and degraded the normal operation of railway system finally. The impacts of disturbance is variable and can either be big or small which had been proved that components in the telecommunication system are generally found to be the most critical ones for the railway system when considering single, two, and three simultaneous failures; components in traction power system is also found to be highly critical; no consequences arise for the railway system for any disturbance to auxiliary power system [1].

Therefore, this research tries to answer the following questions: how to properly and accurately identify the most vulnerable components in the UMNS? How to dynamically simulate and monitor the vulnerability of UMNS?

2 Literature Review

Researches on disruption related transportation network is abundant. They mainly focus on risk analysis [2], safety analysis [3], reliability assessment [4], and emergency management [5]. However, vulnerability in transportation network was firstly proposed by [6] in the proceedings of the 1st international symposium on transportation network reliability (2001) defined as:

"A node is vulnerable if loss (or substantial degradation) of a small number of links significantly diminishes the accessibility of the node, as measured by a standard index of accessibility; a network is considered to be vulnerable if degradation of a small number of network links, perhaps a single link, causes significant reduction in network performance"….

They argued that the vulnerability of the network was more important than "reliability" because of potentially severe adverse consequences of network degradation in sparse networks and low probability of occurrence of freak events does not offsetting the consequences of a network failure. This definition is related to the consequences of node or link failure, irrespective of the probability of failure which was supported by researchers like Berdica (2002)[7].

The related terms of vulnerability of transportation network have not been detailed discussed here. But the relationship among them is shortly described in Figure 1. **Robustness** always used in computer technology. It is the ability of a system to withstand strain [8]. Resilience is often an issue in ecology and expresses the capability of an ecosystem to "return to normal" after having been disturbed [7].

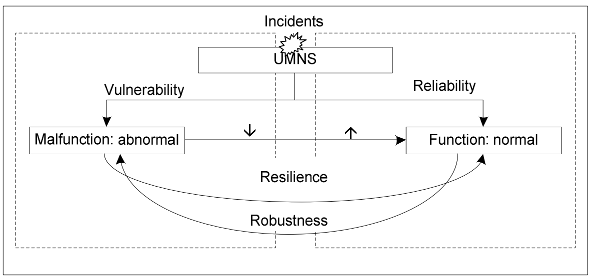


Figure 1. Relationship among the related terms

Vulnerability assessment is widely applied in networked infrastructures such as power supply system [9], Internet [10], and transportation networks. Researches on vulnerability of transportation networks mainly focus on road network [6-8, 11-25]. Complex network theory, Game theory, and other simulation and numerical computation based approaches were used as approaches to fulfill their research aim.

However, research on vulnerability of UMNS is seldom. Interdependence among the systems were modeled by employing graph theory and evaluated by the number of travelers not able to reach their destination in an electronic railway system [1]. The telecommunication system are generally found to be the most critical ones for the railway system when considering single, two, and three simultaneous failures; components in traction power system is also found to be highly critical; no consequences arise for the railway system for any disturbance to auxiliary power system. The methodology is static data statistic based without considering the impact of travel behavior to the statistic result. Methodology for identifying critical lines and stations was proposed by Shimamoto et al. (2008). User equilibrium was used to model the mathematical function. Passengers fail to board and impacts of overcrowding were taken as indicators of vulnerability. In the complex network theory based researches, topological connectivity such as degree of the node, shortest path, betweenness, and cyclomatic number were used as vulnerability indicators [26-28]. A general conclusion that UMNS is robust to random failure but fragile to targeted attacks could be found in these papers.

Although many efforts have been made to provide a more effective approach for vulnerability analysis of UMNS, there are some limitations on the researches, which include the following:

- Lack of dynamic research on multi-layer transportation system

Most research was conducted from one layer system except Johansson and Hassel (2010). However it was conducted from static viewpoint based on data statistical analysis. It is no longer sufficient to analyze the multi-layer TN only from only one layer and static viewpoints. A dynamic based approach is needed in vulnerability evaluation in multi-layer transportation system.

- Lack of complex network theory based research from service function perspective

Topological features are often used as indictor of vulnerability in these complex network theory based researches. The mission of transportation system is a tool for travel. So, vulnerability analysis from topological viewpoint should also consider the service function of transportation system by considering travel supply and travel demand together. A complex network theory based dynamic research from service function perspective is needed.

3 Research framework

3.1 Characters of UMNS operation

UMNS is a system of system, which is composed of railway system and other four subsystems containing: traction power system, telecommunication system, auxiliary power system, and electrical in-feed system. The function of a subsystem influences the function of other systems. Disturbances in one system would transverse to other dependent system and degraded the normal operation of railway system finally. The impact of disturbance is variable and can be either big or small. Components in the telecommunication system are found to be the most critical ones for the railway system when considering single, two, and three simultaneous failures; components in traction power system is also found to be highly critical; no consequences arise for the railway system for any disturbance to auxiliary power system [1].

Operation mode changes from single line operation to networked operation. There is essential difference between single line operation and networked operation. Interaction effect cannot be identified, as there is no transfer stations in single line operation UMNS. Interaction effects will be stronger with the increasing of transfer stations. Disturbances on one station would transverse to other dependent stations and lines and degraded the normal operation of UMNS finally. Impropriate handled of failures in one line would induce malfunction in related lines or even cascade the whole network operation.

There are weak points in UMNS. Weak points are the fundamental reason for malfunction of UMNS, which can easily be destroyed by disruptions from technique, social, or nature. Stations or links failure in one line would transverse to other dependent lines or even cascade the whole system. Therefore, impacts of failures of these weak points can be either small or big. Those stations and links whose failure would cause big impacts to the whole system are considered as weak points.

3.2 Research aim and objective

This research focuses on vulnerability analysis. Identification of physical, structural, and social vulnerability of UMNS is the aim of this research. The result can be used by government planer and operation company manager to provide a safety, reliable, and efficient UMNS to society. The specific objectives of this research study include:

●To explore the connotation and formation mechanism of vulnerability of UMNS

●To provide an approach for gathering and monitoring near-miss

●To provide an approach for assessing the vulnerability of UMNS from physical, structure, and social functional perspectives

●To identify vulnerable components and locations

●To simulate and dynamically monitor the vulnerability of UMNS

3.3 Research content

According to the above description of the research objectives, the detailed content analysis is conducted as following:

Objective 1: Explore the connotation and formation mechanism of vulnerability of UMNS. Analysis the interdependent and connectivity characters of UMNS based on complex network theory from both topological and weighted structure viewpoints. Explore the connotation of vulnerability of UMNS based on Set theory and dissect the relationship with other related terms like reliability, safety, robustness, resilience, and risk. Analyze the formation mechanism and expression of vulnerability of UMNS based on Grounded theory. Analysis of key influencing factors of vulnerability of UMNS based on Fuzzy Fault Tree. Analyze the mechanism of action among the influence factors based on System Dynamics. Finally, conduct quantitative and qualitative relationship analysis between the characters of UMNS and the vulnerability of UMNS.

Objective 2: Build System of gathering and monitoring near-miss of vulnerability of UMNS. Near-miss incidents happen daily in UMNS leaded by unsafe behavior of workers and passengers and unsafe condition of objects. The occurrence of near-miss is the source of disturbance; therefore, gathering and monitoring near-miss is the first step for disruption simulation. Build the near-miss database by Microsoft SQL Server 2008 and conduct demand analysis of information gathering and monitoring based on case analysis. Build wireless senor network under Zigbee agreement based on a combined information gathering technique.

Objective 3: Assess physical vulnerability of UMNS based on improved Accident Sequence Precursor (ASP). Establish improved Accident Sequence Precursor model using the information collected by wireless senor network. Improve the Accident Sequence Precursor by conditional probability to meet the requirement for quantitative assessing of physical vulnerability. Identify the vulnerable components in the physical systems based on sensitive analysis. Conduct empirical analysis on UMNSs in Nanjing, Shanghai, and HK.

Objective 4: Assess structural vulnerability of UMNS based on complex network theory. Dissect the topological characters of UMNS using software for complex network analysis (like Pajek and Ucinet). Compare the topological character of UMNS and weighted UMNS. Establish evaluation index system based on information collected by wireless senor network and Delphi method. Analyze the performance of UMNS under random failure and targeted attack. Identify the vulnerable locations from topological perspective. Conduct empirical analysis on UMNSs in Nanjing, Shanghai, and HK.

Objective 5: Assess social functional vulnerability of UMNS based on the theory of fluid dynamics. Establish speed, density, and flow function using the information collected by wireless senor network based on the theory of fluid dynamics. Disruptions to the trains are considered as the vibration source and propagate in the same way with vibration wave. Establish social functional vulnerability evaluation model based on percolation model. Identify the vulnerable locations from social functional perspective. Conduct empirical analysis on UMNSs in Nanjing, Shanghai, and HK.

Objective 6: Simulate and dynamically monitor the vulnerability of UMNS based on GIS. The research combines vulnerability assessment with GIS in order to fulfill visualized vulnerability management. Identify the distribution character of vulnerable components and locations. Conduct dynamic simulation on WebGIS using the information gathered by wireless senor network and identify the vulnerable components and locations in dynamic operation simulation system. Propose engineering suggestion to reduce the vulnerability of UMNS by the results and fulfill visualized dynamically monitoring.

The above research content is an organic integrity and the relationship among them is described in Figure 2. The beginning of the arrows provides data, results, or models to the end of the arrows. For example, UMNSs of Nanjing, Shanghai, and HK provide topological structures to the second research object for structural analysis; Near-miss data base provides delay cases for three types of vulnerability evaluation.

4 Conclusion

With the development of economies, size of cities in China have become larger and larger, and more and more people are living in cities, especially in developing countries, so traffic congestion has become a very serious problem in many large cities in China. UMNS has become the most effective solution for relieving traffic pressure. It is important to the welfare of people and the economic efficiency of businesses. As a result, once it fails function by the disturbance, it would induce serious impacts. This research study aims to evaluate vulnerability of UMNS and identify the vulnerable components and locations in terms of its physical, structural, and social functional perspectives. If the scheme is implemented successfully, the topological of UMNS can be optimized and limited resources can be assigned efficiently. Additionally, economy and wealth can be saved.

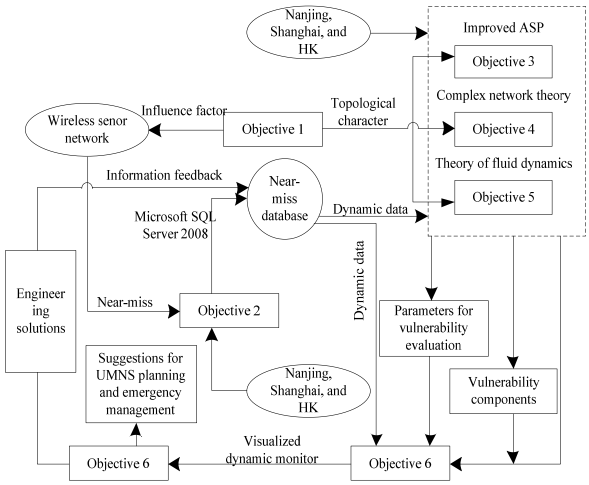


Figure 2. Relationship among the research objectives

5 Acknowledgment

The research work for this paper is funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD). It is also funded by the National Natural Science Foundation of China (Grant No. 51178116 and No.71001027).

References

[1].Johansson, J. and H. Hassel, An approach for modelling interdependent infrastructures in the context of vulnerability analysis. Reliability Engineering \& System Safety, 2010. 95(12): p. 1335--1344.

[2].Soons, C.J., et al., Framework of a quantitative risk analysis for the fire safety in metro systems. Tunnelling and Underground Space Technology, 2006. 21(3-4): p. 281.

[3].Can O S, J.H. and F. De Zulueta, Using hypermedia to improve safety in underground metropolitan transportation. Multimedia Tools and Applications, 2004. 22(1): p. 75--87.

[4].Wang, Z. and R. Xu, Reliability Simulation Analysis of Urban Rail Transit Networks Based on Complex Network. Journal of System Simulation, 2009(20): p. 6670-6674.

[5].Xu, R., M. Zhang and Z. Wang, Design and Implementation of Emergency Management System of Urban Rail Transit Based on Workflow Modeling. Journal of Tongji University(Natural Science), 2008(6): p. 754-759.

[6].D'Este, G.M. and M. Taylor, Network vulnerability: An approach to reliability analysis at the level of national strategic transport networks, in NETWORK RELIABILITY OF TRANSPORT, PROCEEDINGS. 2003, ELSEVIER SCIENCE BV: AMSTERDAM. p. 23-44.
[7].Berdica, K., An introduction to road vulnerability: what has been done, is done and should be done. Transport Policy, 2002. 9(2): p. 117--127.
[8].Sullivan, J.L., et al., Identifying critical road segments and measuring system-wide robustness in transportation networks with isolating links: A link-based capacity-reduction approach. Transportation Research Part A: Policy and Practice, 2010. 44(5): p. 323--336.
[9].Adachi, T. and B.R. Ellingwood, Serviceability of earthquake-damaged water systems: Effects of electrical power availability and power backup systems on system vulnerability. Reliability Engineering \& System Safety, 2008. 93(1): p. 78--88.
[10].Shahriari, H.R. and R. Jalili, Vulnerability Take Grant (VTG): An efficient approach to analyze network vulnerabilities. Computers \& Security, 2007. 26(5): p. 349--360.
[11].Abdel-Rahim, A., et al., Survivability Analysis of Large-Scale Intelligent Transportation System Networks. Transportation Research Record: Journal of the Transportation Research Board, 2007. 2022(-1): p. 9--20.
[12].Bell, M., et al., Attacker--defender models and road network vulnerability. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2008. 366(1872): p. 1893--1906.
[13].Erath, A., et al., Vulnerability assessment methodology for Swiss road network. Transportation Research Record: Journal of the Transportation Research Board, 2009. 2137(-1): p. 118--126.
[14].Jenelius, E., Network structure and travel patterns: explaining the geographical disparities of road network vulnerability. JOURNAL OF TRANSPORT GEOGRAPHY, 2009. 17(3): p. 234-244.
[15].Jenelius, E. and L.G. Mattsson, Road network vulnerability analysis of areacobering disruptions. A grid-based approach with case study. 2009.
[16].Jenelius, E., T. Petersen and L.G. Mattsson, Importance and exposure in road network vulnerability analysis. Transportation Research Part A: Policy and Practice, 2006. 40(7): p. 537--560.
[17].Luathep, P., et al., Large-scale road network vulnerability analysis: a sensitivity analysis based approach. Transportation, 2011: p. 1--19.
[18].Matisziw, T.C. and A.T. Murray, Modeling s-t path availability to support disaster vulnerability assessment of network infrastructure B-2993-2012. COMPUTERS & OPERATIONS RESEARCH, 2009. 36(1): p. 16-26.
[19].Schintler, L.A., et al., Using raster-based GIS and graph theory to analyze complex networks. Networks and Spatial Economics, 2007. 7(4): p. 301--313.
[20].Scott, D.M., et al., Network robustness index: A new method for identifying critical links and evaluating the performance of transportation networks. Journal of Transport Geography, 2006. 14(3): p. 215--227.
[21].Sohn, J., Evaluating the significance of highway network links under the flood damage: An accessibility approach. Transportation Research Part A: Policy and Practice, 2006. 40(6): p. 491--506.
[22].Tampere, C.M.J., et al., Methodology for identifying vulnerable sections in a national road network. Transportation Research Record: Journal of the Transportation Research Board, 2007. 2012(-1): p. 1--10.
[23].Taylor, M.A.P., Critical Transport Infrastructure in Urban Areas: Impacts of Traffic Incidents Assessed Using Accessibility-Based Network Vulnerability Analysis. Growth and Change, 2008. 39(4): p. 593--616.
[24].Taylor, M.A.P. and G.M. D Este, Transport network vulnerability: a method for diagnosis of critical locations in transport infrastructure systems. Critical Infrastructure, 2007: p. 9--30.
[25].Taylor, M.A.P., S.V.C. Sekhar and G.M. D'Este, Application of accessibility based methods for vulnerability analysis of strategic road networks. Networks and Spatial Economics, 2006. 6(3): p. 267--291.
[26].Derrible, S. and C. Kennedy, The complexity and robustness of metro networks. Physica A: Statistical Mechanics and its Applications, 2010. 389(17): p. 3678--3691.
[27].Angeloudis, P. and D. Fisk, Large subway systems as complex networks. Physica A: Statistical Mechanics and its Applications, 2006. 367: p. 553--558.
[28].Zhang, J.H., et al., Networked analysis of the Shanghai subway network, in China. PHYSICA A-STATISTICAL MECHANICS AND ITS APPLICATIONS, 2011. 390(23-24): p. 4562-4570.

DRIHM: Distributed Research Infrastructure for Hydro-Meteorology

Andrea Clematis
Daniele D'Agostino
Emanuele Danovaro
Antonella Galizia
Alfonso Quarati
IMATI - CNR
Genova - Italy
andrea.clematis@ge.imati.cnr.it

Antonio Parodi
Nicola Rebora
Tatiana Bedrina
Cima Foundation
Savona - Italy
antonio.parodi@cimafoundation.org

Dieter Kranzlmueller
Michael Schiffers
LMU
Munich - Germany
kranzlmueller@ifi.lmu.de

Bert Jagers
DELTARES
Bert.Jagers@deltares.nl

Quillon Harpham
HR Wallingford
Q.Harpham@hrwallingford.com

Pierre-Henri Cros
CERFACS
cros@cerfacs.fr

Abstract – *One of the main challenges of the 21st century is represented by accurate weather predictions together with the estimate of extreme phenomena and their impacts on the environment and on the society. The key point of this challenge is to enable the acceleration of advances in hydrometeorological research, and to integrate these advances in the everyday forecasts thus improving the protection of civilians and of the environment.*

The DRIHMS (Distributed Research Infrastructure for Hydro-Meteorology Study) project suggests that a step forward in this direction lies on the ability to easily access hydrometeorological data, to share predictive models, and to facilitate the collaboration among different experts in this area. At this aim it is necessary the support of an e-infrastructure permitting to deal with the massive amount of information needed and providing the adequate level of systems interoperability. These are the goals of the DRIMH, Distributed Research Infrastructure for Hydro-Meteorology, project presented hereafter.

Keywords: Hydrology, Meteorology, e-infrastructure

1 Introduction

Predicting weather and climate and its impacts on the environment, including hazards such as floods, droughts and landslides, continues to be one of the main challenges of the 21st century. Extreme precipitation and flooding events are among the greatest risks to human life and property, with significant societal and economic implications [1]. United Nation (UN) agencies, national governments responsible for large regions, and local governments for their local needs, ask for global certified management tools based on scientific collaborations to deal with processes leading to extreme hydrometeorological events [2].

At the heart of this challenge lies the ability to accelerate scientific advances in Hydro Meteorological Research (HMR) and the use of these advances in protecting civilians and the environment. From a research point of view dealing with the forecast of flood events need to address manifold issues that involve not only HM scientists but requires a strong collaboration with the Information and Communication Technology (ICT) community to provide new technological solutions, [3][4]. The DRIHMS (Distributed Research Infrastructure for Hydro-Meteorology Study) project [5] was initiated to provide a detailed analysis of the needs for such a step forward in the cooperation between HMR and ICT scientists, aiming at the creation of a HMR Virtual Research Community, and to provide a foundation for the deployment of an e-Science environment for HMR on extreme events, as well as on climate change impacts and adaptation. The main aims of DRIHMS were: to understand how to overcome the current limitations in the sharing of data, tools and knowledge at European level; to create a common understanding of what is available; to map out an e-Science based path to extract new knowledge from the latest generation of HM observational and modeling systems. These goals are based on the needs expressed by the HMR and ICT scientific communities of reference, and results have been presented in [6].

With these ideas in mind, the DRIHM (Distributed Research Infrastructure for Hydro-Meteorology) project [7] started. DRIHM would act as a 'system of systems' linking together existing and planned observing and predictive systems around Europe, and supporting the development of new systems able to fill the gap among different steps of the modeling chain from observation to impact. To prove the full extent of the DRIHM environment capability, the completion of the flood forecasting chain will be enabled. It considers the forecast of severe weather events over complex orography areas and the assessment of their impact, with the formulation of flooding risk scenarios relevant from early-warning and Civil Protection perspectives.

In the rest of the paper we present the DRIHM project. We firstly present the DRIHM vision to achieve the mentioned goals; in Section 3 we shortly consider related projects. We introduce the Experiment Suites adopted in DRIHM to implement the full modeling chain in Section 4; a view on the collection of the requirements of the target DRIHM scenarios is provided in Section 5. Section 6 defines the architectural view of the DRIHM e-Science environment; while the last Section concludes the paper.

2 Shaping the DRIHM vision

The DRIHM project intends to develop a prototype e-Science environment to facilitate the collaboration among scientists involved in HM research activities. The DRIHM aim is to provide end-to-end HMR services (models, datasets and post-processing tools) at the European level, with the ability to expand to global scale. This e-Science environment would consist of resources, simulation models, methods and tools to be supplied by various institutions over European e-Infrastructures like EGI (European Grid Infrastructure) [8], and the NGIs (National Grid Initiatives) and exploiting available high performance computing resources such as those provided by PRACE (Partnership for Advanced Computing in Europe) [9]. A simplified architectural view of the DRIHM e-Science environment is provided in Figure 1.

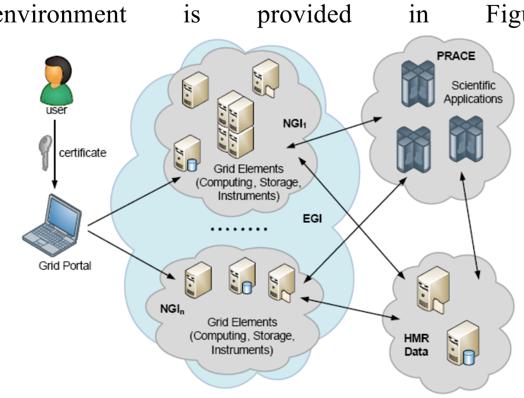


Figure 1. DRIHM e-Science architectural view.

However, the HMR e-Science environment will also have specific HMR requirements not yet or only partly addressed within the actual existing e-Infrastructures. To make a step forward in this direction, DRIHM takes into account the information collected in the DRIHMS project [6], and in particular the necessity of four application service families:

- Simulation services: enabling users to run simulation tools, as meteorological models, hydrological and hydraulic models, stochastic downscaling and data fusion codes;

- Model Gallery services: allowing visualization of the results provided by HMR modeling and post-processing tools. This enables the understanding of how to use and setup a particular model to solve a specific problem;

- Teaching and Learning services: giving access to a library of multimedia educational information, as tutorials, manuals and demonstrations, along with video and audio media describing modeling techniques and principles;

- Data Access services: including dedicated services, for accessing dynamic grid data corresponding to severe HM events into an HMR e-Science environment data repository and needed for calibration, validation, or evaluation of the suitability of different modeling tools.

With this vision in mind, the general objectives of DRIHM are to lead the definition of a common long-term strategy, to foster the development of new HMR models and observational archives for the study of severe HM events, to promote the execution and analysis of high-end simulations and analysis, and to support the dissemination of predictive models as decision analysis tools. To shape this vision, the key points of the DRIHM e-Science environment are:

- The provisioning of integrated HMR services (such as meteorological models, hydrological models, stochastic downscaling tools, decision support systems, observational data) enabled by unified access to and seamless integration of underlying e-Infrastructures;

- The design, development and deployment of user-friendly interfaces to abstract HMR service provision from the underlying e-Infrastructure complexities and specific implementations;

- The support of an HMR e-Science environment enabling the user-driven "composition" of virtual facilities in the form of HM forecasting chains, composed by different HMR resources;

- The establishment of HMR e-Science support centers and corresponding training activities to attract a broad end-user audience comprising of scientists and non-specialists including relevant ESFRI (European Strategy Forum on Research Infrastructures [10]) communities.

The project considers three suites of experiments acting as laboratories for the integration of a new HM modeling chain and the development of new working practices in the HMR e-Science environment based on innovative ICT services. By the end of the project, the added value will be the integration of all three, i.e. the modeling of a probabilistic forecasting chain, on the DRIHM e-Science environment. This should be operational for selected European watersheds and tested on severe rainfall events. Such scenarios (geographical area and severe events) have been selected by the HMR partners, and from now on we will call them Golden Cases.

3 Related Projects

The proposed DRIHM project builds on the findings of DRIHMS and promises to establish a rigorous and sustained framework towards the adoption and exploitation of e-Infrastructure by the HMR and related Earth sciences communities. DRIHM builds on the results of the DRIHMS project, but it will take the collaboration between the ICT and HMR communities several steps further.

Leading HMR initiatives, such as HyMeX (HYdrological cycle in the Mediterranean Experiment [11]), CASA project [12], COST 731 “Propagation of uncertainty in advanced meteo-hydrological forecast systems” [13], EUMETNET (Network of European Meteorological Services) Opera (Operational Programme for the Exchange of Weather Radar Information) project [14], EUROGEOSS [15], HEPEX (Hydrologic Ensemble Prediction Experiment) project [16], MAP D-PHASE (MAP Forecast Demonstration Project), MEDEX (MEDiterranean Experiment) project [18], OPERA (OPerational Eo-based Rainfall) - ASI (Italy Space Agency) project [19] have already expressed interest in contributing to DRIHM models and data, as part of the DRIHM virtual community and in using the services and tools deployed by DRIHM.

DRIHM will also look for collaboration with ICT projects as EGI, PRACE, MAPPER (Multiscale APPlications on European e-infRastructures [20]), SCI-BUS (SCIentific gateway Based User Support [21]) and others to assure long-term sustainability of project efforts and results.

4 DRIHM Experiment Suites

DRIHM combines the European expertise in HMR modeling, in HMR remote sensing, in Grid and HPC, and in studies of severe weather impacts. The final objective of the DRIHM project is to enable a step beyond the state of the art in the modeling of a probabilistic forecasting chain on the DRIHM e-Science environment. For this reason, the interaction between HMR and ICT scientists focus on three layers composing the forecasting chain, designed to prove the full extent of the DRIHM e-Science environment capability. Actually, many attempts have been made in different countries to build up a sound HM probabilistic chain, notably in Southern Europe and in US [22],[23]; there is an increasingly pressing need, driven by societal and economical expectations, to predict possible impending floods, with a quantitative measure of uncertainty.

In Figure 2, a schematic diagram of the forecast chain is presented; it considers the **Rainfall layer**, **Discharge layer** and **Water level, Flow & Impact Layer**. These address the interdisciplinary and international challenges of HMR in forecasting severe HM events over complex orography areas and assessing their impact, with the formulation of flooding risk scenarios relevant from early-warning and Civil Protection perspectives. They correspond to the mentioned Experiment suites, and will consider the Golden Cases selected by the HMR partners.

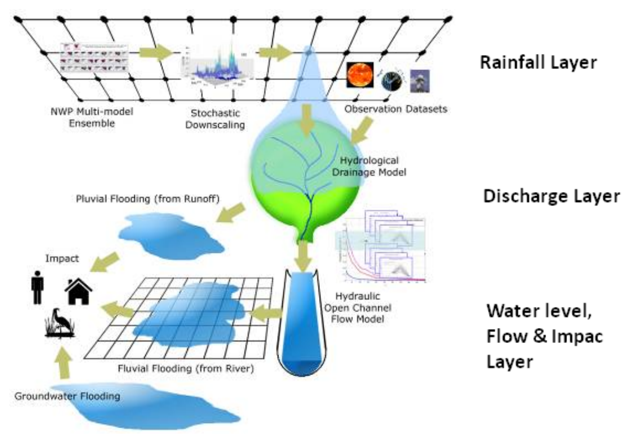


Figure 2. DRIHM Experiment Suites.

Experiment Suite 1 - Rainfall: Combination of different Numerical Weather Prediction (NWP) models to form a high-resolution multi-model ensemble together with stochastic downscaling algorithms to enable the production of more effective quantitative rainfall predictions for severe meteorological events.
Experiment Suite 2 - Discharge: Fusion of rainfall predictions (potentially from experiment suite 1) with corresponding observations, which are input into multiple hydrological models to enable of the production of more accurate river discharge predictions.
Experiment Suite 3 - Water Level, Flow and Impact: Execution of hydraulic model compositions in different modes to assess the water levels, flow and impact created by the flood events. Indeed, this process can be driven from data produced by experiment suite 2.

For the first two suites, we distinguish between the baseline versions of experiment and the advanced versions of experiment. As for the baseline experiment, we mimic standard HMR activities without DRIHM e-Infrastructure, i.e. the HMR services (e.g. meteorological and hydrological models, stochastic downscaling and data fusion tools, and observational data) that partner are currently used to employ, over the Golden Cases data repository. As for the advanced versions of experiment suite, each partner will be able to use all HMR services provided by DRIHM virtual community without the constraints of distance, access, usability and scientific barriers. The third suite represents the completion of the flood forecasting chain, initiated in Experiment Suites 1 and 2, from rainfall through discharge to water level, flow and impact.

5 DRIHM User Requirements Collection Process

One of the goals of the present first phase of the project is the collection of the requirements of the baseline and targeted DRIHM scenarios. The main steps of this phase are:

- collect information about the existing practices and methods, collect shortcomings of such methods;
- describe high level user stories to derive a set of new Use Cases, related to the execution of baseline version of the experiment suites 1 and 2;
- derive requirements from shortcomings of the existing methods and from the new Use cases.

Each HMR partner provided a synthesis of baseline experiments and their classification based on the usage description; they also defined a "Golden Case" derived by the events that interested Catalonia, France, and Italy between the mid October and mid November 2011.
Let us focus on the collection of the requirements of the baseline experiment suites. This step has twofold impacts: 1) drive the integration/combination of modeling techniques used in different disciplines (Meteorology, Hydrology and Hydraulics) on a distributed infrastructure; 2) provide evidence of the step forward that the DRIHM e-science infrastructure will enable in doing HMR when we will collect the requirements of the advanced experiments.

The process to collect requirements follows the standard scenario based methodology, [24]. The starting point is the current situation "as-it-is"; we derive the requirements according to the following items:

- The scenarios of the DRIHM baseline experiments constitute the target situation DRIHM will be aiming at. We therefore consider the models relevant for that, as grouped in two parts: those relevant to Suite 1, and those relevant to Suite 2.
- We also take into account the HMR Hot Topics outcome of the DRIHMS polls [6];

The process utilized for requirements elicitation and selection is the following:

- we collect description of the baseline: hydro and meteo models and their "Model Use Cases", enriched with a list of their problems/limitations;
- we describe scenarios of the DRIHM baseline experiments in form of user stories and then as use cases.

The following table provides a synthesis of the DRIHM baseline experiments; for each HMR partner we outline the Golden Case to consider, the employed models, the corresponding experiment suites. We also mapped the related problems on the HMR hot topics suggested by DRIHMS. As for the table, CIMA and RHMSS feature a full hydro-meteorological forecasting chain covering both baseline suite 1 and 2, while UPM is contributing to baseline suite 2, DLR and CNRS to suite 1.

We organized collected requirements in four classes: Data Management, Model Management, Workflow Management and Post processing. Data management mainly refers to services for facilitating interaction between existing data and baseline methods. Model management mainly refers to services for interoperability of Meteorological and Hydrological models. Workflow management, i.e. the ability to chain two or more models to obtain complex simulations, heavily depends on Data and Model requirements. Post-processing focuses on basic

Partner	Case study	Baseline Suite 1	Baseline Suite 2	HMR Hot Topics
CIMA	Genoa 2011	COSMO (Deterministic) Continuum (Deterministic) RainFARM (Probabilistic)	DRIFT (semi-distributed model)	Probabilistic forecasting (full chain) Precipitation downscaling Model verification metrics
RHMSS	Serbia 2010	WRF (Deterministic)	HBV (distributed model)	Probabilistic forecasting (full chain) Precipitation downscaling
RHMSS	Serbia 2010	WRF-NMM (Deterministic)	HYPROM (shallow water)	Probabilistic forecasting (full chain) Precipitation downscaling
UPM	Catalunia 2011	-	RIBS probabilistic-near real-time	Probabilistic forecasting (hydrology)
CRNS	France 2011	AROME (Probabilistic) MESO-NH (Probabilistic)	-	Probabilistic forecasting (meteorology)

services for checking model execution and to export/analyze the outcome of models execution.

6 Overview of DRIHM architecture

The DRIHM e-Science environment architecture is organized by the well-known layer pattern. Figure 4 depicts the complete layered architecture including the usage-relations between the layers.

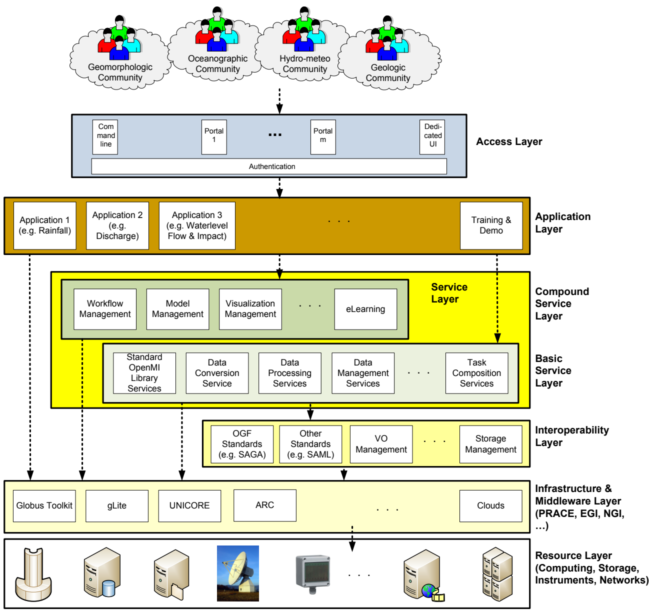


Figure 3. DRIHM architecture.

The **Resource Layer** contains all physical resources (computing elements, storage elements, instrument elements, sensor networks, communication networks) that are provided by the various partners in the HMR e-Science environment. Typically, these resources are provided over European e-infrastructures; for accessing them a Grid middleware system like Globus, www.globus.org, gLite, http://glite.cern.ch, UNICORE, www.unicore.eu, is required. Some of the resources may also be available in a Cloud (public or private). The middleware and the access services are part of the **Infrastructure and Middleware Layer**.

The **Interoperability Layer** assists in transparently accessing the resources via the underlying layers, thus hiding the heterogeneity of Grid resources/middleware. The interoperability layer is based on standards like those defined in the OGF, Open Grid Forum, www.ogf.org/, or in the OASIS, Organization for the Advancement of Structured Information Standards, www.oasis-open.org/.

The HMR related services are comprised in the **Service Layer** which itself is separated in the **Basic Service Layer** and the **Compound Service Layer**. The Service Layer typically interacts with the Interoperability Layer to transparently access the HMR resources. However, it may also interact with the Middleware Layer directly. The Basic Service Layer provides all services, which are fundamental for the HMR community. Examples are data conversion, model access or task composition. The Compound Service Layer provides more complex services assembled from basic services. Typical examples in the HMR context are the creation and the management of workflows, the management of model sets, or the visualization of simulation results.

The **Application Layer** contains the specific HMR applications for simulations, experiments and e-Learning. It is accessed via the **Access Layer** that provides the required capabilities to authorize the access to the HMR applications via various portals or clients and through an authentication interface.

DRIHM will optimize the performance, use and access to an eco-system of HMR services, thus serving cross-disciplinary needs for the development of new modeling, post-processing and analysis tools to be translated in the medium-term in operational facilities for extreme hydro-meteorological event risk prediction, prevention and mitigation.

7 Conclusions

In this paper we presented the DRIHM project, and we described the DRIHM e-Science environment planned to support the HM community to enable multidisciplinary and global collaboration between meteorologist, hydrologists and possibly other Earth science scientists. DRIHM will foster the development of new HMR models and observational archives for the study of severe hydrometeorological events, will promote the execution and analysis of high-end simulations, and will support the dissemination of predictive models as decision analysis tools. By the end of the project, the main achieved result will be the deployment, on a dedicated e-Science environment, of a probabilistic forecasting chain enabling the combination of different meteorological, hydrological and hydraulic models for a rich set of European watersheds and severe rainfall events. To drive this process, the project has defined the Experiment suites and a set of Golden Case. In this work we presented the preliminary results of the collection of the user requirements. The project would attract also non-specialist as the "citizen scientists" interested in HMR and related Earth science disciplines. This new research e-Infrastructure has the potential to enable a level of international research collaboration and fosters a strong and unprecedented collaboration and partnership between HM scientists and ICT scientists, thus serving cross-disciplinary needs.

Acknowledgments

DRIHM is an EU FP7-Infrastructure Project under grant, Contract No: RI-28356.

References

[1] United Nation International Strategy for Disaster Reduction, *Disasters in numbers*, http://www.unisdr.org/files/24692_2011disasterstats.pdf

[2] United Nations International Strategy for Disaster Reduction Secretariat (UNISDR), *Reducing Disaster Risks through Science: Issues and Actions, The Full Report of the ISDR Scientific and Technical Committee* 2009. http://www.unisdr.org/we/inform/publications/11543

[3] Shukla, J., Hagedorn, R., Hoskins, B., Kinter, J., Marotzke, J., Miller, M., Palmer, T. N., Slingo, J., "Revolution in climate prediction is both necessary and possible: A declaration at the World Modelling Summit for Climate Prediction." *Bulletin of the American Meteorological Society*, Vol. 90, pp. 175 - 178, 2009.

[4] Shukla, J., Palmer, T. N., Hagedorn, R., Hoskins, B., Kinter, J., Marotzke, J., Miller, M., Slingo, J., "Toward a new generation of world climate research and computing facilities". *Bulletin of the American Meteorological Society*, Vol. 91, pp. 1407 – 1412, 2010.

[5] The DRIHMS, Distributed Research Infrastructure for Hydro-Meteorology Study, project: www.drihms.eu

[6] Parodi A., Morando M., Rebora N., Trasforini E., Molini L., Siccardi F., Craig G., Tafferner A., Kranzlmüller D., Schiffers M., Clematis A., D'Agostino D., Galizia A., Quarati A., *The DRIHMS White Paper*, ISBN 978-88-906068-0-9, April 2011.

[7] The DRIHM, Distributed Research Infrastructure for Hydro-Meteorology, project: www.drihm.eu

[8] EGI, European Grid Infrastructure, www.egi.eu

[9] PRACE, Partnership for Advanced Computing in Europe, www.prace-project.eu

[10] ESFRI, European Strategy Forum on Research Infrastructures: http://ec.europa.eu/research/esfri

[11] HyMeX home page: http://www.hymex.org/

[12] CASA project: http://www.casa.umass.edu/

[13] COST 731, Propagation of uncertainty in advanced meteo-hydrological forecast systems: http://cost731.bafg.de/servlet/is/Entry.9691.Display/

[14] Opera, Operational Programme for the Exchange of Weather Radar Information, project: http://www.knmi.nl/opera/

[15] EUROGEOSS project: http://www.eurogeoss.eu/default.aspx

[16] HEPEX project: http://www.hepex.org/

[17] MAP D-PHASE project: http://www.isac.cnr.it/dinamica/projects/dphase/

[18] MEDEX project: http://medex.aemet.uib.es/

[19] OPERA, OPerational Eo-based Rainfall, project: http://www.operaproject.it

[20] MAPPER project: http://www.mapper-project.eu

[21] SCI-BUS project: http://www.sci-bus.eu/

[22] Cloke HL, Pappenberger F. 2009. Ensemble flood forecasting: a review. Journal of Hydrology 375(3–4): 613–626.

[23] Siccardi, F., Boni, G., Ferraris, L., Rudari, R., 2005. A hydrometeorological approach for probabilistic flood forecast. Journal of Geophysical Research – Atmospheres,110. D5.

[24] I. Jacobson, M. Christerson, and P. Jonsson, Object-Oriented Software Engineering. A Use Case Driven Approach; ACM Press, 1992; ISBN 978-0201544350

Modeling and Simulation for System Reliability Analysis: The RAMSAS Method

Alfredo Garro
Department of Electronics, Computer and System Sciences (DEIS)
University of Calabria
Via P. Bucci 41C, 87036, Rende (CS), Italy
garro@deis.unical.it

Andrea Tundis
Department of Electronics, Computer and System Sciences (DEIS)
University of Calabria
Via P. Bucci 41C, 87036, Rende (CS), Italy
atundis@deis.unical.it

Abstract - *The paper presents the up-to-date version of RAMSAS, a recently proposed model-based method for the reliability analysis of systems through simulation. RAMSAS can be easy plugged in various phases of a typical system development process ranging from the design to the testing phases so to complement other well-known and wide adopted techniques for system reliability analysis (e. g. FMECA, FTA, RBD) by providing additional analysis capabilities. The present version of RAMSAS is the result of an intensive experimentation phase in several application domains (avionics, automotive, satellite) which allowed improving the effectiveness of the method especially in the modeling of both the intended and dysfunctional system behavior. The paper concludes with a discussion about the specific aspects of the reliability analysis of System of Systems (SoS), and how RAMSAS can be further extended to effectively support it.*

Keywords: System Reliability Analysis, Model-Based Systems Engineering, Simulation, Failure Modeling.

1 Introduction

Large-scale systems are usually designed and implemented as a set of interconnected parts forming an integrated whole in order to obtain the desired functions with acceptable performance and costs. The increase in both the heterogeneity and numbers of system components raises several and challenging engineering issues and demands for suitable analysis, design, verification and validation methods and techniques to guarantee system integration and requirements satisfiability in a flexible way. In particular, non-functional requirements play an important role as they are decisive in the achievement of system objectives. More specifically, Reliability which represents the ability of a system to perform its required functions under stated conditions, for a specified period of time, is a key requirement to satisfy especially for mission critical systems where system failures could cause even human losses [2]. To perform system reliability analysis several quantitative and qualitative techniques are currently available (such as Series-Parallel system reliability analysis, Markov Chains, FMECA and FTA) [2]. Nevertheless, the increase in both system complexity and accuracy required in the reliability analysis often goes beyond the capabilities of the so far mentioned techniques which are mainly based on statistical and probabilistic tools and on the hierarchical system decomposition. As a consequence, new techniques are emerging which are centered on model-based approaches so to benefit from the available modeling practices and which incorporate the use of simulation to flexibly evaluate the system reliability indices and compare different design choices [1, 7]. In this context, RAMSAS, a model-based method for the reliability analysis of systems through simulation has been recently proposed [5]. RAMSAS has been already experimented in the avionics [4, 5] and automotive domain [3]; moreover, an ongoing experimentation concerns the reliability analysis of an Attitude Determination and Control System (ADCS) of a Satellite. These experiences allowed improving RAMSAS so to increase its effectiveness in system reliability analysis. In particular, in the present version of RAMSAS, which is for the first time presented in details in this paper, new approaches for modeling both the intended and dysfunctional behavior of the system are proposed. In particular, the modeling of the dysfunctional system behavior is based on the introduction of specific behavioral templates of dysfunctional tasks and related fault/failure patterns. A discussion about the specific aspects of the reliability analysis of System of Systems (SoS), and how RAMSAS can be further extended so to effectively support it, is also provided and concludes the paper.

2 The RAMSAS method

The RAMSAS method, which enables the reliability analysis of systems through simulation, is based on a classical iterative process consisting of four main phases which are reported in Table 1 along with their input and output work-products. In the first phase (*Reliability Requirements Analysis*), the objectives of the reliability analysis are specified and the reliability functions and indicators to evaluate during the simulation are defined. In the *System Modeling* phase, the structure and behavior of the system are modeled in SysML (OMG Systems Modeling Language) by using *zooming* in-out mechanisms

[9]; moreover, beside the *intended* system behaviors, specific *dysfunctional* behaviors and related tasks, which model the onset, propagation and management of faults and failures, are introduced. In the *System Simulation* phase, the previously obtained models of the system are represented in terms of the constructs offered by the target simulation platform. Finally, simulation results are analyzed with respect to the objectives of the reliability analysis; if necessary, new partial or complete process iterations are executed.

In the following sub-sections each of the above sketched phases will be described by focusing on the enhancement introduced in the present version of RAMSAS respect to the previous ones. Complete examples of the execution of the RAMSAS phases and related work-products can be found in [3, 4, 5].

Table 1. RAMSAS phases and related work-products

Phases	Input work-products	Output work-products
Reliability Requirement Analysis	System Design Model (SDM), System Requirements (SR)	Reliability Analysis Objectives (RAO)
System Modeling	System Design Model (SDM), Reliability Analysis Objectives (RAO)	System Model for Reliability Analysis (SMRA)
System Simulation	System Model for Reliability Analysis (SMRA), Reliability Analysis Objectives (RAO)	Simulation Results (SIRE)
Results Assessment	Simulation Results (SIRE), Reliability Analysis Objectives (RAO)	Design Suggestions Report (DSR), Reliability Analysis Report (RAR)

2.1 Reliability Requirements Analysis

In the *Reliability Requirements Analysis* phase the objectives of the system reliability analysis are specified. The inputs of this phase are the work-products typically resulting from previous *System Design* phases. Starting from this documentation, the scenarios to be analyzed, the functions that the system has to perform, the related operative conditions, and the reference time horizons should be clearly individuated along with the reliability functions and indicators to be derived from the analysis of the simulation results. In particular, the *Reliability Requirements Analysis* phase takes as input a description of the system under consideration in term of both *System Requirements* (*SR*) and *System Design Model* (*SDM*). *SR* includes functional (*FR*) and non-functional requirements (*NFR*), whereas *SDM* provides a system representation in terms of its architecture and behavior. Among the *NFR*, the Reliability Requirements (*RR*) specify the ability required for the system in performing the functions identified by the *FR* under specific stated conditions and for a given period of time. In addition, a Failure modes and effects analysis (FMEA) [2] can be also provided to highlight the potential failure modes of the system along with their severity and likelihood. Starting from the above mentioned *SR*, in the *Reliability Analysis Objectives* (*RAO*) work-product, the reliability indicators and the scenarios of interests are identified along with the main analysis techniques to be applied to the data gathered from simulation. In the *RAO*, a visual representation of the *SR* can be also provided through *SysML Requirements Diagrams* along with the allocation of these requirements (especially the reliability ones) to main system components.

2.2 System Modeling

In the *System Modeling* phase the structure and both the *intended* and *dysfunctional* behavior of the system under consideration are represented in SysML by executing four modeling activities (see Figure 1): *System Structure Modeling*, *Intended Behavior Modeling*, *Dysfunctional Behavior Modeling* and *Behavior Integration*.

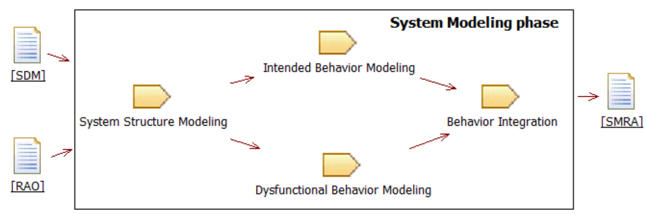


Figure 1. Activities of the System Modeling phase

The specifications concerning the *structure* and *intended behavior* of the system are derived from the System Design Model (*SDM*) resulting from previous design phases; these activities can be straightforward if during the system design similar structural and behavioral reference models have been adopted along with a UML based modeling notation. With reference to the *dysfunctional behavior* of the system, its modeling can be based on the results reported in the *RAO* document concerning the analysis of the failure modes of the system (see Section 2.1). Finally, the intended and dysfunctional system behaviors are suitable *integrated* so to provide a complete system specification for the subsequent simulation phase. In the following Sub-Sections, each of these modeling activities is discussed more in details.

2.2.1 System Structure Modeling

In the *System Structure Modeling* activity, the system structure is modeled by using SysML *Blocks* following a *top-down* approach. To this aim, several decomposition levels should be considered by applying *in-out zooming* mechanisms [9] such as *system, subsystems, equipment, and components;* however to allow system analysis at the desired level of details, further abstraction levels along with different and deeper hierarchies can be also introduced. Each system entity is defined by both a *Block Definition Diagram* (BDD) and an *Internal Block Diagram* (IBD). Specifically, for a given abstraction level, a BDD describes a block with its interfaces, attributes, operations, constraints, parts and relationships with other blocks; whereas, an IBD provides a description of the block internal structure in terms of the organization of its component blocks.

2.2.2 Intended Behavior Modeling

In the *Intended Behavior Modeling* activity the intended (or normal or expected) behavior of the system is represented following also a layered approach but combining the *top-down* with a *bottom-up* strategy. The reference model is service and task-oriented: the behavior of each entity is modeled in terms of the *services* (or *functions*) that the entity is able to provide and which are performed through *tasks*. In order to specify the behavior of the system and its component entities, two levels of decomposition are considered: *leaf level* (e.g. component level) and *non-leaf level* (e.g. equipment, subsystem or system level).

In particular, for each entity at the *leaf decomposition level* (the lowest decomposition level):

- the services (or functions) provided by the entity, in terms of their input and output work-products along with pre and post conditions, should be specified;
- each task (flow of activities/actions) performed by the entity for providing a specific service (or function) has to be specified through an *Activity Diagram*; each task can be *scheduled* or *triggered* by incoming events/messages;
- the exchange of messages between the entity and the external environment (which can be another entity at the same or at a higher decomposition level) should be represented through *Sequence Diagrams*;
- in case the behavior of the entity depends on its internal state, a state machine which models the entity life cycle can be specified through a *Statechart Diagram*.

Moving from the *leaf decomposition level* to higher decomposition levels (*not-leaf decomposition levels*), the representation of the entity behavior is similar; however, how the component entities (i.e. sub-entities) participate and determine the behavior of the considered enclosing entity should be taken into account; as a consequence, for each entity at a *not-leaf decomposition level*:

- beside the services (or function) provided by the entity, how these services can be obtained by composing the services provided by the sub-entities should be also specified;
- each task (flow of activities/actions) performed by the entity for providing a specific service (or function) has to be specified by an Activity Diagram possibly highlighting through *swimlanes* the responsibility of each sub-entity in carrying out the activities of the task;
- the exchange of messages between the entity and the external environment should be represented through Sequence Diagrams possibly highlighting the role played by its sub-entities in each interaction (e.g. by explicitly representing them as participants in the diagrams);
- in case the behavior of the entity depends on its internal state, a state machine which models the entity life cycle can be specified through a Statechart Diagram; the diagram can adopt advanced constructs, such as *composite states*, *AND/OR-decomposition* and *history pseudo-states*, for representing how the behavior of the entity is related to the behavior of its sub-entities.

2.2.3 Dysfunctional Behavior Modeling

In the *Dysfunctional Behavior Modeling* activity, the focus is on the modeling of *faults* and *failures*, which are key concepts of the system reliability analysis [2]. Specifically, the behavior, concerning faults and failures of each system entity (i.e. the *dysfunctional behavior*), is specified as a set of specific *tasks* (which can be modeled through *Activity Diagrams*). The reference model of a generic system entity, regardless on the considered abstraction level, is shown in Figure 2. An entity is represented by a SysML Block which provides a set of services/functions; the tasks performed by the entity for providing these services are modeled during the *Intended Behavior Modeling* phase (see Section 2.2.2). Beside the intended behavior specified through these tasks, a set of *dysfunctional tasks* are added so to model the *dysfunctional behavior* of the entity [1, 6]. In particular, each block could receive as input a set of *failure events* (e.g. due to the failures of other blocks) and could, in turn, produce in output other *failure events* due to its *failure*; moreover, internal *faults* (represented as *fault events*) can be generated and treated inside the block possibly producing block failures (and thus output *failure events*). With reference to the above described behavioral reference model (see Figure

2), six templates of *dysfunctional tasks* have been individuated (see Figure 3): *Fault Generation*, *Failure Generation, Failure Management, Fault Management Failure Propagation*, and *Failure Transformation.*

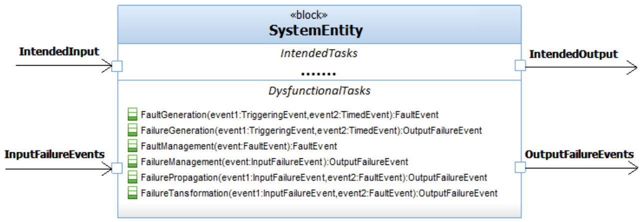


Figure 2. The reference Behavioral Model of a system entity

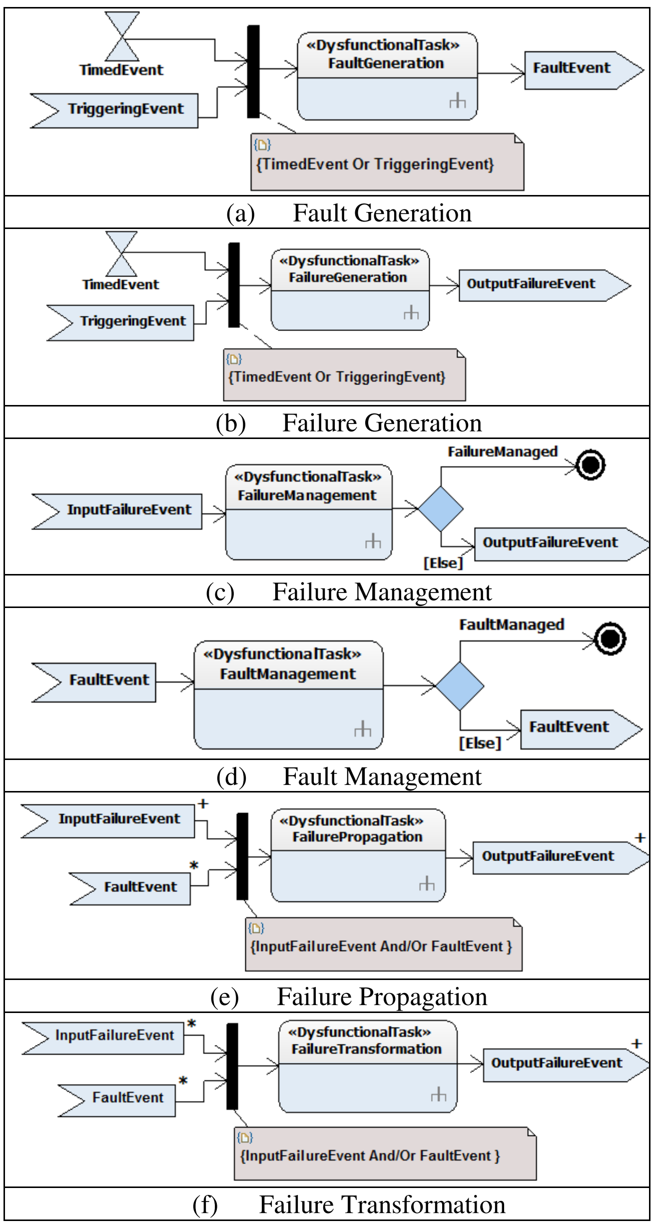


Figure 3. Templates of dysfunctional tasks

Tasks of the *Fault* and *Failure Generation* type (see Figure 3.a and 3.b respectively) can generate a fault/failure as a result of specific causes occurring according to a given probability function. In order to allow driving, during the simulation, the processes of fault/failure generation so as to study the reliability performances of the system, these tasks can be scheduled or triggered by specific events. However, whereas a *failure* is directly associated to an *output failure event* (see Figure 3.b), to produce a *failure event* starting from a *fault*, the fault has to be taken in input from either a *Failure Propagation* or a *Failure Transformation task*. Incoming *failures* as well as internally generated *faults* can be (successfully or not) handled by tasks of the *Failure* and *Fault Management* type (see Figure 3.c and 3.d respectively). Finally, tasks of the *Failure Propagation* and the *Failure Transformation* type take *dysfunctional events* in input and produce *dysfunctional events* in output. In particular, tasks of the *Failure Propagation* type (Figure 3.e) generate *output failure events* either through the propagation of incoming failure events or by combining such incoming failures with internal faults. Tasks of *Failure Transformation* type (Figure 3.f) produce *output failure events* derived from the transformation or combination either of incoming failure events or of internal faults.

To further support this crucial modeling activity, a set of *patterns* to associate to each of the above discussed six types of *dysfunctional tasks* should be defined. The definition of each of these patterns should take into account the type of the dysfunctional task (e.g. Failure Generation, Propagation or Transformation) as well as the specific nature of the fault/failure to which the pattern refers to; in other word, a basic pattern is associated to a couple *(dysfunctional task type; fault/failure type)*. As a consequence, for the definition of these patterns, beside the individuated six dysfunctional task types (see Figure 3), a (possibly hierarchical) classification of faults/failures needs to be introduced. A first solution could consider the following fault/failure types [6]: (i) *reaction too late*; (ii) *reaction too early*; (iii) *value failure*; (iv) *commission*; and (v) *omission*. By combining the individuated six dysfunctional task types with these five fault/failure types, thirty different basic fault/failure behavioral patterns can be defined (their complete description, which is the result of an ongoing research activity, is beyond the scope of this paper).

The modeling of the dysfunctional behavior of each system entity, in terms of a set of dysfunctional tasks of the above described types and possibly based on available fault/failure behavioral patterns, is essential to evaluate through simulation the dysfunctional behaviors of the system and analyze the possible consequences of failures as well as feasible solutions for their management in order to improve system reliability.

2.2.4 Behavior Integration

In the *Behavior Integration* activity, the normal/intend behaviors and the dysfunctional behaviors modeled in the previous modeling activities (see Section 2.2.2 and 2.2.3) are integrated to obtain an overall behavioral model of the system and its component entities. As an example, starting from the *Activity* and *Sequence diagrams* which have been used to model both the intended and dysfunctional behaviors of the system entities, a complete and integrated definition of the life cycle of each entity, regardless on the considered abstraction level, can be obtained and represented through a *Statechart diagram*. This activity closes the System Modeling phase by delivering the *System Model for Reliability Analysis* (*SMRA*) work-product.

2.3 System Simulation

The objective of the *System Simulation* phase is to evaluate through simulation the reliability performance of the system and, possibly, compare different design alternatives and parameters settings; in particular, the following three (see Figure 4) main activities are performed: *Model Transformation, Parameters Setting,* and *Simulation Execution*.

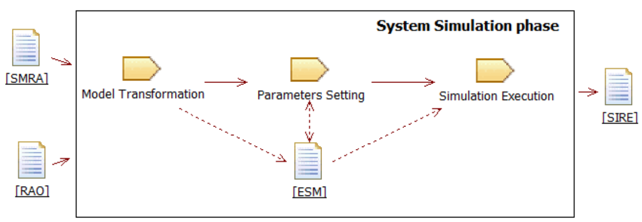


Figure 4. Activities of the System Simulation phase

In the *Model Transformation* activity the previously obtained Models of the System in the *SMRA* are represented in terms of the constructs offered by the target simulation platform which is, currently, MathWorks Simulink, so producing an *Executable System Model* (*ESM*). This transformation is based on a mapping between the basic SysML and Simulink constructs (see Table 2) [3, 4, 10]. It is worth noting that not only the intended behavior of the system but also the dysfunctional tasks (see Section 2.2.3) which are essential for analyzing during the simulation the reliability performance of the system are generated. Indeed, this allows to suitably *injecting* faults and failures during the simulation and setting the parameters of the related generation, management, propagation and transformation tasks (see Section 2.2.3).

In the *Parameters Setting* activity the *ESM* is refined so to allow the flexible setting of system configuration and simulation parameters which can be tuned according to both the characteristics of the operative scenario to simulate and the dysfunctional behaviors to analyze.

In the *Simulation Execution* activity the *ESM* is executed by varying the desired parameters according to the analysis objectives reported in the *RAO*. Specifically, the *ESM* is executed by Simulink according to a synchronous reactive model of computation: at each step, Simulink computes, for each block, the set of outputs as a function of the current inputs and the block state, then it updates the block state. During the simulation *faults* and *failures* can be injected and/or caused to stress and analyze the reliability performance of the *system*. At the end, the data generated from the simulations are reported in the *Simulation Results* (*SIRE)* work-product to be analyzed in the next phase.

Table 2. Mapping among SysML and Simulink constructs

Entity	SysML	Simulink
System/Subsystem/ Equipment/Component	Block, Part	Block, Subsystem Block
Behavior/Constraint	Activity Diagram, Sequence Diagram, Statechart Diagram, Parametric Diagram	S-Function, State Flow diagram
Input/Output Interface	Flow Port	Input/Output Simulink Block
Association/Binding	Connection	Line

2.4 Results Assessment

In the *Results Assessment* phase, the results reported in the *SIRE* are elaborated with reference to the objectives of the reliability analysis identified in the initial phase of the process so to obtain important information on the reliability properties of the system under consideration. A great part of these analyses can be directly performed in Simulink, whereas more advanced analyses can be performed by external tools by exporting the obtained results through the MATLAB Workspace. As a result, the following two work-products are produced in output:

- *Reliability Analysis Report* (*RAR*), which provides a detailed analysis about the reliability performance of the system under consideration;

- *Design Suggestions Report* (*DSR*), which provides a set of suggestion to improve the design of the system and/or choose among different design choices. It is worth to note that the *DSR* exploits typical FMECA and FTA notations and representation formats so to ease the use of RAMSAS in conjunction with classical RAMS techniques.

As for any iterative process, new (partial or complete) iterations of RAMSAS can be executed for achieving new or missed analysis objectives.

3 Discussion and Conclusions

The current version of RAMSAS, described in Section 2, has been originally conceived for large-scale systems, i.e. systems which are constituted by a multitude of components organized so to form a whole with clearly defined boundaries. Examples of this kind of systems are military and commercial aircraft, spacecraft, satellites, power plant automobiles, etc. The reliability analysis of these systems is a challenging task which, as proved by RAMSAS, could benefit both from model-based approaches and from simulation [3, 4, 5]. However, although the structure of these large-scale systems is rather complex (or better complicated) it remains quite the same during the system life cycle; moreover, a great part of the system components manifest a reactive behavior and pro-activeness is limited to a narrow subset of components. Moving from large-scale system to the System of Systems (SoS) context, these assumptions typically do not hold [8]. Indeed, a SoS (e.g. a Coast Guard Integrated Deep-water System or an Air Traffic Management System) is constituted by a set of interconnected, and often geographically distributed, systems which interact for achieving common goals and are capable of autonomous and independent behaviors; moreover, the set of involved systems typically changes during the SoS life as new systems join the SoS and other dynamically leave it. As a consequence, the reliability analysis of a SoS presents different and peculiar aspects respect to that of a large-scale system that need to be accurately taken into account and that require for specific approaches and techniques. Some features of RAMSAS are particularly suited for the reliability analysis of SoS such as the adoption of zooming in-out mechanisms for the structural and behavioral modeling of the system, and the exploitation of simulation both for the analysis of system properties and for the evaluation of alternative scenarios and design choices. However, new features need to be added; as an example, the possibility to define the potential changes in the system structure during the system life cycle. Moreover, specific concepts to explicitly model the pro-active (and thus autonomous) part of the behavior of the SoS entities (which are themselves systems) should be introduced (such as the *goals* that the entity will to achieve). These novelties could lead to evolve the current *vision* of system adopted in RAMSAS to a more *agent-oriented* one in which a SoS is treated, and thus modeled and simulated as a Multi-Agent System [9]. Ongoing and future research activities are devoted to investigate how to further extend RAMSAS in the above mentioned directions so as to effectively support the reliability analysis of SoS.

Acknowledgments

Andrea Tundis was supported by a grant funded in the framework of the "POR Calabria FSE 2007/2013".

References

[1] R. Cressent, V. Idasiak, F. Kratz, and P. David, "Mastering safety and reliability in a model based process," Proc. of the Reliability and Maintainability Symposium (RAMS), Lake Buena Vista, FL (USA), January 2011.

[2] B. Dodson and D. Nolan, *Practical Reliability Engineering*, John Wiley & Sons Ltd, England, 2001.

[3] A. Garro and A. Tundis, "Enhancing the RAMSAS method for System Reliability Analysis: an exploitation in the automotive domain," Proc. of the 2nd Int. Conf. on Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH), Rome (Italy), July 2012.

[4] A. Garro and A. Tundis, "A Model-Based method for System Reliability Analysis," Proc. of the Symposium On Theory of Modeling and Simulation (TMS), Orlando, FL (USA), March 2012.

[5] A. Garro, A. Tundis, and N. Chirillo, "System reliability analysis: a Model-Based approach and a case study in the avionics industry," Proc. of the 3rd Air and Space International Conference (CEAS), Venice (Italy), October 2011.

[6] L. Grunske and B. Kaiser, "Automatic Generation of Analyzable Failure Propagation Models from Component-Level Failure Annotations," Proc. of the 5th Int. Conf. on Quality Software (QSIC), Melbourne (Australia), September 2005.

[7] IEC 61508, *Functional safety of electrical/electronic/programmable electronic safety-related systems, Parts 1-7*, 2010.

[8] M. W. Maier, "Architecting Principles for Systems of Systems," *Systems Engineering*, Vol 1, No. 4, pp. 267-284, 1999.

[9] A. Molesini, A. Omicini, A. Ricci, and E. Denti, "Zooming multi-agent systems," Proc. of the 6th Int. Workshop on Agent-Oriented Software Engineering, Utrecht (The Netherlands), July 2005.

[10] A. Sindico, M. Di Natale, and G. Panci, "Integrating SysML with Simulink using Open-Source model transformations," Proc. of the 1st Int. Conf. on Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH), Noordwikerhout (The Netherlands), July 2011.

SoS in Disasters: Why Following the Manual Can Be a Mistake

Antonella Cavallo, Vernon Ireland
antonella.cavallo@adelaide.edu.au
vernon.ireland@adelaide.edu.au

Entrepreneurship, Commercialisation
and Innovation Centre (ECIC)
The University of Adelaide
Adelaide SA 5005, Australia

Abstract –*According to both the US Geological Survey and the World Bank, $280 billion dollars could have been saved if $40 billion dollars had been invested in disaster prevention. Natural and human-made disasters that have occurred over the last few years show that there is a gap in disaster prevention caused by the interconnected nature of risks, which cannot be foreseen with current risk management methods. In this paper we point out how disaster management could benefit from a SoS approach in emergency response and preparedness strategies. Using recent disasters as case studies, we identify some keys to success in managing a SoS in preparation, during and in the aftermath of a disaster. In particular, we discuss the idea of the interconnectedness of risks in independent and interdependent systems and the application of Boardman and Sauser's concept of "creative disobedience", which are fundamental for goal achievement of systems belonging to a SoS.*

Keywords: SoS dynamics, disaster management.

1 Introduction

In many countries, risks related to disasters are managed following a paradigm of traditional risk management. This is based on a cycle in which risk identification, risk evaluation, risk mitigation and policy adjustment [1, 2] are the four steps which are followed in a logic of continuous improvement towards the evolving system conditions. The corollary of this approach is that only identified risks can be managed and embedded into a policy. However, disasters are part of complex systems involving a network of risks, rather than linear causal relationships [3]. The difference between complicated and complex risks can be inferred [4] from Snowden and Boone's distinction between 'complicated and complex cognitive domains' [5]. In a risk network, there are some risks which can be identified and some which stay unforeseen. The former are 'complicated' as they can be ascertained a priori. The latter are 'complex' as they involve a higher degree of uncertainty which involves the cause-effect relationship to stay unknown until after the risk occurrence [6]. This approach is particularly effective in the case of linear causal relationships, i.e. when an identified cause is connected to one or more identified effects. However, there are at least two problems with this approach: first the way lessons learned are currently managed does not work as expected [7], which means that even after repeating the risk management cycle, some risks are left out and hence not managed. Second, a deterministic approach for risk assessment [8], does not consider the possibility of unforeseen risks due for instance to factors like climate change.

The fact that we need to come up with new approaches to face a complexity which includes risks that we cannot necessarily foresee has been recognised by many [3, 4, 9, 10]. Particularly, Boteler [11] highlights the need for holistic approaches in disaster management. Leveson [7] points out that safety is an attribute which applies to the whole system and not just to the individual components. Salmon [12] suggests that a disaster can be considered as a system which cannot be broken down easily into its components as it needs to be analysed as a whole.

From an organisational point of view, the system of a disaster involves several organisations, institutions and agencies such as the Red Cross, government, fire services and police. These are all systems in an evolving system of systems (SoS), which is the disaster context itself, immersed in ever-changing environmental and contextual conditions. The coordination of different agencies is often organised hierarchically and it can be challenging, particularly in the occurrence of an unexpected disaster. Problems can arise in areas such as communication, situation awareness and cultural issues [13]. Differences in the goal achievement criteria of each organisation can be an obstacle as well. For this reason, it is important that each agency is aware of the action scope of the others. Also it should be discussed previous to a disaster which agency is going to take the lead in the case of an emergency [13]. In the context of a disaster, agencies are expected to act on the basis of regulations and policies following command-control logics. In this paper, however, we argue that these formal rules correspond to an appropriate response to 'complicated risks', but that they are less valuable in the context of complex risks.

2 Complexity in recent disasters

The US Geological Survey and the World Bank estimated that $280 billion dollars could have been saved if $40 billion dollars had been invested in natural disaster prevention worldwide [Benson and Clay 2003 in 14]. Natural and human-made disasters have dominated global media reporting in recent times. The 9/11 terrorist attack, 2004 tsunami near Sumatra, Victorian bushfires in Australia, earthquake in New Zealand, flooding in New Orleans and earthquake, tsunami and nuclear disaster in Japan are just some of the catastrophic events which have occurred over the past few years. The number of natural disasters are believed to be increasing because of climate change [15, 16], but human mistakes are exacerbating their effects. For example, pre-existing management issues contributed worsening the nuclear disaster in Japan and flooding consequences in Queensland both in 2011 [17]. So what can entities such as government authorities or crisis monitoring agencies really do to prevent death, destruction, loss, mental illness and long-term effects in the world? Recent disasters have shown a lack of efficiency in managing situations before they turn into disasters. For example, in 2005 despite the adequate available information, the public officials of New Orleans failed to evacuate the population before flooding caused a landfall. By then the inhabitants did not have any way to escape and as a consequence hundreds of people lost their lives [18].

In a time where information flows very quickly from one side of the earth to the other, people are increasingly critical of how disasters are managed and concerned with what can be done to prevent critical situations from turning into crises. Traditional risk management has developed preparedness brochures, simulations and other types of awareness programs (e.g. Red Cross, Country Fire Services, US Geological Survey). Most of these programs focus on specific risks and do not take into consideration the multi-causality of risks [19]. Many risk programs are also organized around manuals, rules and laws which are assumed to be followed in the case of a disaster. In reality, however, the disasters cited above demonstrate that these measures are not followed by all parts of society, including the various institutional, corporate and demographic groupings. For instance, despite the existence of regulations in Japan, Prime Minister Kan delayed disclosing important information about the radioactivity of the Fukushima power plant, causing a part of the population to move into the radioactive wake [20]. Costa Crociere, the company which owns the cruise ship which crashed in Italian waters earlier this year, was aware of the illegal sail-past practice, but it did not do anything to stop captains from continuing doing it [21].

While the last examples seem to suggest simply that the law should have been applied to avoid potential disasters, there are other situations where the ideal behaviour or response is not as easy to identify. For example, after the cyclone in 1977, many Indian laborers moved from rural Southern Indian to the coast, which was more likely to be hit by disasters [22]. This raises a number of difficult questions for risk managers. It seems straight forward that these people should not move to the coast as it will be more dangerous when the next natural disaster occurs. However, the correlation between disasters and poverty [14] suggests that they might have moved to the coast to access the minimal resources to live. In that case, 'life preservation' has to form an additional goal of a SoS in the case of a disaster.

In this context, the efficiency of disaster management depends on awareness of the SoS risk interdependencies and prioritization. In other words, risks need to be contextualized in their network and managed, taking into consideration the different patterns which can result as a consequence of a crisis. When the Icelandic volcano Eyjafjallajökull erupted in 2010, the authorities used the precautionary principle to stop most of the air traffic after the eruption. Although some airlines had verified the sky conditions and ascertained the absence of significant danger, they were not allowed to transport passengers for days, causing heavy economic losses. Eventually, authorities decided to reopen airports without having any more data about ash in the sky than when they decided to close them [3]. Like in the case of the Eyjafjallajökull, many institutions refuse to manage complex risks as they do not think to be able to cope with the complexity and uncertainty involved [3, 19]. The reason for this is because contingent programs are based on the identification and analysis of specific known risks, whereas high uncertainty requires the identification and analysis of different alternative risk patterns [19]. For this reason, we use the lens of Complex Risk Management to focus on the management of foreseeable and unforeseeable risks (also called 'unknown unknowns' or 'unk unks' in aerospace engineering [3]) and their effects in a disaster.

3 A SoS during the Fukushima Daiichi disaster

Reports on the Fukushima Daiichi disaster, which have been published since [20, 23], reveal the structure of a System of Systems (SoS), where the interconnectedness and interactions between the systems and their environment made risks evolve and develop into a network of unanticipated causes and effects over a very short period of time. Despite the fact that earthquake and tsunamis are not uncommon in Japan (in fact they have much better anti-seismic buildings than in Italy), neither the government nor the Nuclear Safety Commission (NSC) nor the company at the centre of the nuclear disaster, Tokyo Electric Power Company (TEPCO) had put in place adequate measures of disaster prevention [20].

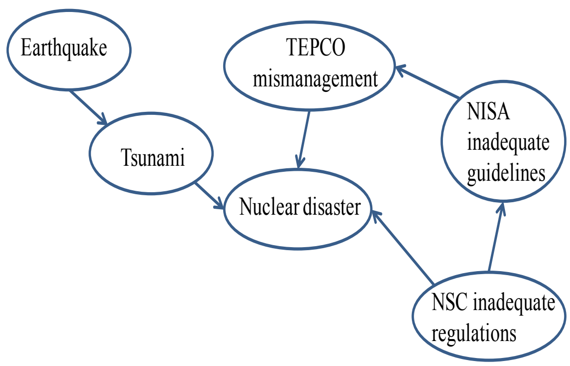


Figure 1. A partial risk network of the Fukushima Daiichi disaster inferred from [20, 23].

In Japan, the Nuclear Safety Commission (NSC) oversees the Nuclear and Industrial Safety Agency (NISA). NSC's inadequate regulations impacted on the guidelines that NISA imposed on TEPCO [20]. So, for example, NSC's regulations did not take into consideration the possibility of electricity outages as it was judged that temporary blackouts would have been quickly solved. As a result of poor regulations and controls, TEPCO managed the disaster inappropriately, contributing to worsening the nuclear disaster which had been caused by the tsunami following the initial quake [20, 23] (Figure 1). Here we could assume that if more regulations had been put in place, then the disaster would have had reduced dimensions. However, the fact that appropriate regulations were available at an international level and that TEPCO for different reasons did not happen to abide by them should open a question about the effectiveness of international control and monitoring institutions. In this case, the decision not to abide by the international law had been consciously taken at the corporate and at the institutional level.

Command-control logic failed at different levels during the Fukushima disaster. When the disaster happened, the two most important people in TEPCO's top management were away on holidays or on business travel. For different reasons, they were not able to go back immediately to Fukushima, causing a slowdown in decision-making and information sharing between the government and TEPCO's management [20]. The nuclear disaster involved a complexity that TEPCO was not prepared to face, to the point that the government had to call in external nuclear experts to help to manage the disaster [20]. However, there was a decisive point where the hierarchical logic of decision-making was broken by a subordinate who would not have normally had the authority to ignore the orders. In the middle of the disaster, the situation in the Daiichi plant was still extremely uncertain and the decisions taken seemed to be distant from their operational impacts. The lack of adequate information and the misuse of the available data put the decision-makers in a difficult position [20]. It was not clear which effect their decisions or the lack of decisions would have had on the whole system. So, at one of the most critical points of the crisis, a key decision about stopping or continuing the water injections to reduce the temperature of Unit 1 could not be taken.

Tensions between the government and TEPCO top management and then between the latter and the site management, made the decision even more complex. At some point the government authorized the injections, while the liaison of the company to the government said that they had to be stopped until a strategy to manage the crisis had been decided. However, the Fukushima Daiichi site management was convinced of the importance of continuing the water injections. In fact, during a teleconference, Fukushima Daiichi Director Yoshida decided to formally accept the orders to stop the water injections, but secretly told his staff in charge of this operation to continue the water injections. By doing this, he helped the company to manage the conflict with the government, whilst importantly preventing the worsening of the nuclear crisis [20].

This is an example of what Boardman and Sauser call 'creative disobedience' [24] which is, and must be, implicit in paradoxical situations such as crises. The idea is that in complex systems where available information is often not available at every level of the hierarchy, subordinates need to understand the main goal of their mission to the extent that they are able to act independently of orders for the highest of the system goals. In the case of Fukushima, Yoshida was probably in a more informed 'cognitive domain' [5] than the other people who were in a position of taking a decision. The complexity of this scenario raises some very interesting questions about the SoS dynamics during the disaster.

4 Why systems thinking in the crisis

In the midst of the uncertainty, Yoshida had a good intuition, but what if things had gone differently? The case of the Japanese crisis is just an example. There are others in which extreme uncertainty pushed people to make decisions against the manual. However, in the end the resulting effects of the disaster were interpreted by authorities as a result of the operators' misbehavior.

Much legislation is built around the responsibility of individuals managing the disaster. If on the one hand, this gives a decision-making framework to the responsible authorities, on the other it can push these to take decisions away from an informed common sense. This effect is caused by a hierarchical and command-control way of thinking which is based on the assumption that situations, risks, projects, disasters, systems can be broken down into manageable entities. By doing this, the sense of the effects on the overall system is lost.

As Gilpin and Murphy note [19], it is important to understand to what extent the whole can be reduced before we start losing essential information about the system. Current regulations are very likely to have unforeseen negative effects [25] if a global perspective is not taken. There is a big limitation, though. Thinking in terms of system of systems about a disaster can be overwhelming, especially when a generalization is attempted. In fact, some authors state that systems theory is not adequate for a crisis as it would involve the system being subject to "rationalism and control" [19] which, as our examples show, might not be enough to protect a system from a crisis. This is true when the SoS is considered with a command-control lens, i.e. a hierarchical one. Indeed, it is not realistic to have a central control for a SoS in a disaster because each organization and institution has a separate set of goals, i.e. it is independent and obeys the organisation's peculiar hierarchy.

There is an underlying risk network in every SoS in which risks have an impact beyond the individual systems which have triggered the risk occurrence [26]. For example, the nuclear disaster in Japan has had consequences on near countries (outgoing arrows in Figure 2). This raises a number of questions about the management of global risks as highlighted by Beck [26] who talks about 'world risk societies' to indicate those communities which share risks beyond political borders. Gaps and overlaps [27] emerge from the comparison between risk networks of a disaster and the organisational structure of different agencies dealing with disasters.

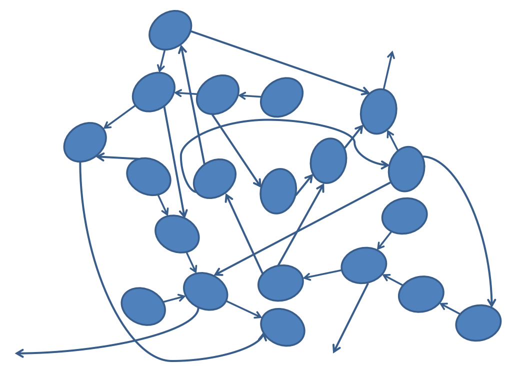


Figure 2. Stylized risk network for a disaster.

If we superimpose risk domains of a disaster on the organisational structure of the agencies managing disasters, we get two pictures which do not appear complementary (Figure 3). Big circles represent risk domains, whereas small circles with a stylized organisation chart in the middle correspond to organisations managing disasters. We assume that all organisations, institutions, communities, etc. belong to a system of systems.

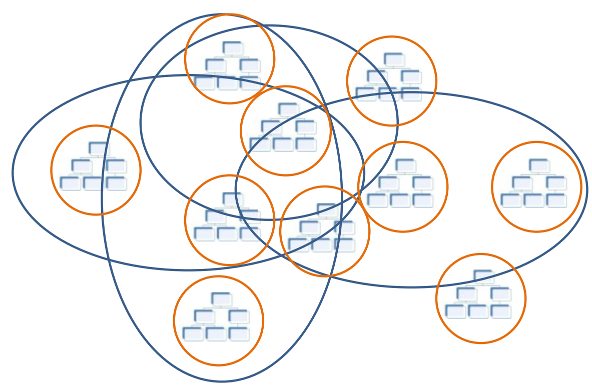


Figure 3. Example of risk domains and hierarchical systems in a SoS configuration.

That organisations need to work together to manage a disaster is clear to the most. However, what we are saying here is that they need to do it in a more integrated way, as has been acknowledged also by some humanitarian organisations [28]. Being aware of the ontology of complexity as explained at the beginning of this paper and acknowledging the fact that organisations operate in a system of system can support organisations in working towards a new way of integrated cooperation. For example, we know that in complex projects, the probability of success can be increased by getting team members to share goals and values [3]. Then, a SoS approach is a viable method when the goals of a SoS have been agreed in advance within different organisations and when it is possible to have an inductive approach to problems together with a deductive one, i.e. command-control. This involves individual organizations and individual members having the maturity and autonomy to allow them to take decisions away from predefined frameworks, but still appreciating and understanding the long-term goal of the underlying organization.

5 Conclusion

This paper suggests that current disaster management frameworks should be changed into paradigms which involve a SoS vision of situations potentially turning into disasters. Instead of using command-control management techniques, the perspective should be also inductive so that goals can be negotiated and organizations can achieve a higher degree of effectiveness in disaster SoSs. Such an approach would also allow a better integration of all functions planning and operating during a disaster.

Moreover, a global perspective supported by complexity science and systems theory shall focus on hazards taken in their risk network context. We need to be aware of global risks and have an approach which aims to mitigate some risks and to accept the uncertainty that others involve. As the number of disasters increases and technology development induces incalculable system evolutions [29] (e.g. nuclear disasters), we need to change our way to think about disasters. We need to collaborate with other 'risk societies', discuss priorities and goals and decide together which risks to manage. Finally, we need to get used to the fact that we cannot control everything and that disasters may happen. This will not necessarily mean that we failed.

References

[1] F. Romeike, *Modernes Risikomanagement: Die Markt-, Kredit- und operationellen Risiken zukunftsorientiert steuern*: Wiley-VCH, 2004.
[2] K. Hoffmann, *Risk-Management: neue Wege der betrieblichen Risikopolitik*. Karlsruhe: Verlag Versicherungswirtschaft, 1985.
[3] A. Cavallo, *Risk Management in Complex Projects. An exploratory study to managing unknown unknowns in uncertain environments*. Saarbruecken: Lambert Academic Publishing, 2010.
[4] C. H. Loch*, et al.*, *Managing the unknown. A new approach to Managing High Uncertanty and Risk in Projects*. Hoboken: Wiley, 2006.
[5] D. Snowden and M. J. Boone, "A Leader's Framework for Decision Making: Wise executives tailor their approach to fit the complexity of the circumstances they face " *Harvard Business Review,* 2007.
[6] J. K. De Rosa*, et al.*, "A Research Agenda for the Engineering of Complex Systems " in *SysCon 2008 - IEEE International Systems Conference*, Montreal, 2008.
[7] N. G. Leveson, "Applying systems thinking to analyze and learn from events," *Safety Science,* vol. 49, pp. 55-64, 2011.
[8] National Emergency Management Committee, "National Emergency Risk Assessment Guidelines," T. S. E. Service, Ed., ed. Hobart, 2010.
[9] S. M. Newell, "Changing the discourse of projects and project management " in *Asia-Pacific Research Conference on Project Management*, Melbourne, 2010.
[10] M. T. Pich*, et al.*, "On Uncertainty, Ambiguity, and Complexity in Project Management," *Management Science,* vol. 48, pp. 1008-1023, 2002.
[11] F. E. Boteler, "Building disaster-resilient families, communities, and businesses," *Journal of Extension,* vol. 45, 2007.
[12] P. Salmon*, et al.*, "New perspectives on disaster response: the role of systems theory and methods " in *Australian & New Zealand Disaster and Emergency Management Conference* Brisbane, 2012.
[13] P. Salmon*, et al.*, "Coordination during multi-agency emergency response: issues and solutions," *Disaster Prevention and Management,* vol. 20, pp. 140-158, 2011.
[14] M. E. Ibarraran*, et al.*, "Climate change and natural disasters: macroeconomic performance and distributional impacts," *Environment, Development and Sustainability,* vol. 11, pp. 549–569, 2009.
[15] M. L. Parry*, et al.*, "Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change," ed. Cambridge, UK: Cambridge University Press, 2007, p. 976.
[16] M. K. Van Aalst, "The impacts of climate change on the risk of natural disasters," *Disasters,* vol. 30, pp. 5-18, 2006.
[17] F. Norton. (2012, 21 March). Floods report opens way for legal action. *Lateline.* Available: http://www.abc.net.au/lateline/content/2012/s3455694.htm
[18] H. Donkin, "Match fit?," in *Project Manager*, ed. Sydney: The Australian Institute of Project Management (AIPM), 2012.
[19] D. R. Gilpin and P. J. Murphy, *Crisis Management in a Complex World*. New York: Oxford University Press, 2008.
[20] Y. Funabashi and K. Kitazawa, "Fukushima in review: A complex disaster, a disastrous response," *Bulletin of the Atomic Scientists,* 1 March 2012 2012.
[21] M. Imarisio and F. Sarzanini, *Concordia. The true story*. Milan: Corriere della Sera, 2012.

[22] P. Winchester, *Krakatoa. The day the world exploded: August 27, 1883*. New York: HarperCollins, 2003.
[23] Y. Hatamura*, et al.*, "Executive Summary of the Interim Report," Investigation Committee on the Accident at Fukushima Nuclear Power Stations of Tokyo Electric Power Company, 26 December 2011.
[24] J. Boardman and B. Sauser, "Paradox " in *Systems Thinking. Coping with 21st Century Problems*, ed Broken Sound Parkway, NW: CRC Press, 2008, pp. 169-185.
[25] Risk Response Network, "Global Risks 2012 - Seventh Edition," World Economic Forum, Cologny/Geneva2012.
[26] U. Beck, "Cosmopolitanism as imagined communities of global risk," *American Behavioral Scientist,* vol. 55, pp. 1346-1361, 2011.
[27] National Emergency Management Committee (NEMC), "National Strategy for Disaster Resilience. Building the resilience of our nation to disasters," Attorney-General's Department, Ed., ed. Barton: Commonwealth of Australia, 2011.
[28] UNICEF. (2012, 22 May). *Social Protection Strategic Framework*. Available: http://www.unicef.org/socialprotection/framework/
[29] D. Cayley, "Episode 5 - Ulrich Beck and Bruno Latour," in *How To Think About Science*, ed: Canadian Broadcasting Corporation Radio One, 2009.

The Holistic Military Capability Life Cycle Model

Jukka Anteroinen
National Defence University
Helsinki, Finland
jukka.anteroinen@mil.fi

Abstract - *This paper reviews existing military capability models and the capability life cycle. It proposes a holistic capability life-cycle model (HCLCM) that combines capability systems with related capability models. ISO 15288 standard is used as a framework to construct the HCLCM. The HCLCM also shows how capability models and systems relate to each other throughout the capability life cycle. The main contribution of this paper is conceptual in nature. The model complements the existing, but still evolving, understanding of the military capability life cycle in a holistic and systemic way. The model also increases understanding and facilitates communication among various military capability stakeholders.*

Keywords: Military capability, life cycle, system-of-interest, enabling system, capability model, holistic model

1 Introduction

'Capability' has become a prevailing paradigm in the military community. Since the beginning of the twenty-first century, armed forces have been developed and operated based on the various capability initiatives of nations such as the United States [1] and Australia [2]. The capabilities-based approach is considered an evolutionary advancement from the Cold War paradigm of the threat-based approach to planning [1],[3] by placing more emphasis on how an opponent may fight rather than on who the opponent may be or where a war may occur [1].

However, the term 'military capability' means different things in different contexts [4],[5],[6],[7]; therefore, it is only meaningful when it is defined within a context [8]. Capability is also used interchangeably to refer to objectives, tasks that need to be accomplished in support of these objectives and the means of conducting these tasks [9]. Consequently, a number of military capability models, such as joint capability areas in the United States [10] and the TEPIDOIL systems model in the United Kingdom [11], have been developed to help manage and understand capabilities. Nonetheless, only a few studies [4],[12],[13],[8],[14] provide a comprehensive view of military capability models (i.e., capability frameworks).

Moreover, military capability life cycle models, such as the CADMID in the UK [11] and the life cycle stage model in Finland [15] focus heavily on the acquisition stage of the capability life cycle and systems capability model, which is just one of the many capability models that are used to facilitate the evolution of military capability from cradle to grave. In addition, only two [8],[14] of the above mentioned capability framework studies incorporate life cycle considerations into the discussion on military capabilities.

Consequently, the objectives of this paper are to review existing capability models and the capability life cycle. This paper also proposes a holistic capability life-cycle model (HCLCM) that combines capability systems with related capability models. The HCLCM also shows how capability models and systems relate to each other throughout the capability life cycle. ISO/IEC 15288 standard [16] and its guide, ISO/IEC 24748 [17] report, are used as frameworks to construct the HCLCM.

This paper is organised as follows: Section 2 presents the theoretical background, Section 3 discusses the methodology and Section 4 introduces the holistic military capability life cycle model. Finally, Section 5 discusses the findings of the study and possibilities for future research.

2 Theoretical background

2.1 Existing military capability models

2.1.1 Capability as an effect or a function to execute tasks

Since the end of the Cold War, functional capability models have become a popular framework for developing transformational concepts [12]. One of the key concepts of the functional capability approach is that required capabilities are first considered effects or functions of armed forces rather than specific solutions in order to avoid potential bias towards particular developmental solutions [18]. Another key feature of this approach is that it attempts to provide capabilities that are suitable for a wide range of challenges while working within an economic framework that necessitates choice. Therefore, this approach contrasts other developing forces that are based on specific threats and scenarios [19].

Table 1 presents the functional capability models of various countries. In the United Kingdom, high-level operational concepts provide an effects-based framework for future military operations and a depiction of seven interdependent capability components [20], also known as functions of the Defence Capability Framework. In the

United States, joint capability areas are defined as functionally-grouped capabilities that support future force development and operational planning. In this context, capabilities constitute the ability to achieve the desired effects to perform a set of tasks [10]. The Finnish functional capability areas [21] are, in practice, similar to those of the United States. Australian capability areas usually include effects or functions, although they can also include technology or some other capability input [2].

Table 1. The functional capability models in various countries

USA Joint capability areas	UK Functions of defence capability framework	Australia Capability areas	Finland Capability areas
Command and control	Command	Command and control (EC)	Command and control
Battlespace awareness	Infrom	Information superiority (CP)	Situational awareness
Net centric		Networked capability (CDP)	Net centric
Logistics	Sustain	Logistics (EC)	Logistics
Building partnerships			Partenrships
Force support	Protect	Repair and maintenance elements (EC),	Protection
Force application	Project Operate	Transport and movement (EC) Precise Force Application (CDP)	Generation of force Force application
Corporate management and support			Corporate management and support
Legend: EC = enabling capability, CDP = capability development principle, CP = capability priority			

2.1.2 Capability as systems

Table 2. The systems capability models in various countries

USA (DOTMPLFI)	UK DLoD (TEPID OIL)	Australia FIC	Finland
D = doctrine	D = concepts and doctrine I = information	Command and management	Doctrine Will for defence
O = organisation	O = organisation	Organisation	
T = training	T = training	Collective training	
M = materiel	E = equipment L = logistiscs	Major systems Supplies	Materiel
L = leadership and education P = personnel	P = personnel	Personnel	Personnel
F = facilities	I = infrastrcuture	Facilities Support	Infrastrcture Support
Legend: DLoD = Defence lines of development, FIC = Fundamental inputs to capability			

Capability systems models such as US DOTMPLFI [10] and Australian FIC view capability as a system of interlocking and interdependent components [2]. Therefore, these systems capability models shift attention away from traditional platforms and technical-systems-focused approaches to the non-material aspects of capabilities [9]. The UK systems capability model [11] essentially has the same components as the US DOTMPLFI. In Finland, the capability systems model is defined in a more basic manner than the three other systems models shown in Table 2 [15].

2.1.3 Capability as military units

The physical realisations of capabilities are manifested in military units led by military commanders. Military units are the building blocks of the force structure [8]. Military units have standard operating procedures and tactics that are used in a designed operating environment, such as urban terrain in crisis response operations. Correspondingly, the operating environment is a wider concept than a physical environment. At the highest level of the military chain of command, operating procedures are known as doctrines. These operating procedures differ from military concepts because they describe how existing capabilities—i.e., units—are employed, whereas military concepts refer to the distant future [22].

2.1.4 Other military capability models

In addition to the three capability models mentioned in Section 2.1.1 – 2.1.3, at least two other capability models — the military power model and the weapons model—have been identified [14]. The military capability is one element of the concept of power in international relations [23]. It refers to the hard power of capabilities [5], [24], and the term ‘military capability’, in this context, is used interchangeably with the term ‘military power’ [5]. Military capability is also thought of in terms of materiel [25] i.e., as a function of a specific weapons system and its performance. This paper does not discuss these models further because the weapons model is not used directly to manage the capability life cycle and the military power model is used primarily by political decision-makers.

2.2 Capability life cycle

ISO/IEC 15288 standard introduces a systems life cycle concept and is applicable to any man-made system that provides services in defined environments for the benefit of users and other stakeholders [16]. A life cycle model is a conceptual segmentation of the definition of a system’s purpose, its realisation as a product or service and its utilisation, evolution and disposal. The life cycle model is typically segmented into stages to facilitate planning, provisioning, operation and support of the system. These segments ensure that a system progresses through established checkpoints to ensure satisfactory progress. A life cycle model can help an organisation to think of its work and its processes within a larger framework [17].

In addition to the CADMID life cycle model and the Finnish military capability life cycle model mentioned earlier, the United Kingdom’s through life capability management (TLCM) concept is similarly focused on specific stages— the acquisition and in-service — of the life cycle and leans primarily on the systems capability model [26]. The UK Ministry of Defence defines TLCM as “*an approach to the acquisition and in-service management of military capability in which every aspect of new and existing military capability is planned and* *managed coherently across all Defence Lines of Development (DLoD) from cradle to grave* [27].”

2.3 Capability frameworks

Studies, which provide a comprehensive and systematic view of military capability models, are often referred to as capability frameworks. Russell et al. [13] present a capability model in which systems capability model components are seen as resources in a unit that fulfils responsibilities. These resources, in turn, are transformed into capabilities when they are integrated with each other to fulfil operational responsibilities. The upper-level ontology in Suzic and Svenson’s [28] capability-based plan resembles the model by Russell et al., but ignores the responsibility component that exists between the capability and resource components. Fasana [12] presents an alternative capability model based on the interaction between effects, time and space. These three models do not build on existing capability models, but represent analytical constructs created by the authors.

Kerr et al. [4] propose a framework for military capability that consists of building blocks, a function layer, an effects layer and an influencers layer. Their paper provides a multi-layered capability model, but approaches the issue primarily from the bottom-up, i.e., how military systems (material) create functions and effects, rather than viewing issues from the top-down as the actual capability development process does. Yue and Henshaw [8] propose a holistic view of the UK’s military capability development by articulating what TLCM means for defence capability planning, development and delivery. They focus on systems and military unit capability models and issues related to capability development in a specific country. Touchin and Dickerson [7] discuss capability architecting, which links the customer’s needs with the structure of the capability solution to be implemented. They compare various architecting approaches and propose a definition of the term ‘architecting for capability’, but do not propose a capability architecting or life cycle model, as done in this paper.

Anteroinen [14] presents the comprehensive capability meta-model (CCMM), which compiles existing military capability models into an integrated whole using the well-known Zachman Framework of Enterprise Architecture as the structure of the model. The focus of the CCMM is the hierarchical order of capability models. Therefore, this paper builds on the CCMM, but expands the comprehensive capability model by connecting the system life cycle model with existing capability models.

3 Methodology

The ISO 24748 technical report approaches life cycles from the systems perspective. It identifies two systems terms: system-of interest and enabling system. System-of-interest refers to a particular system that is of interest to an observer. Each enabling system enables a part, e.g., a stage of the life cycle of a system-of-interest. Therefore, enabling systems facilitate the progression of the system-of-interest through its life cycle. As with any system, each enabling system also has its own life cycle. Each life cycle could be linked to and synchronised with that of the system-of-interest, as illustrated in Figure 1. During any particular stage in the system’s life cycle, the relevant enabling systems and the system-of-interest should be considered together. Since they are interdependent, they can also be viewed as a system.

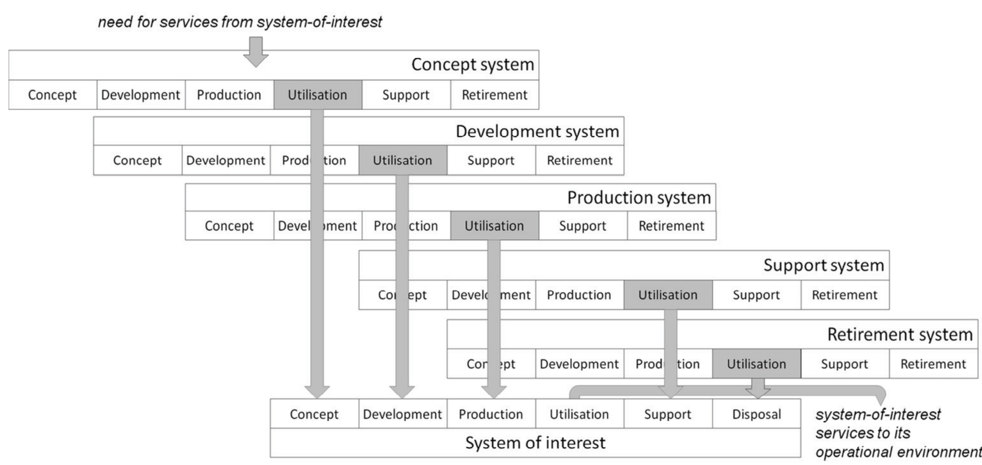


Figure 1. Interaction between system-of-interest and enabling systems (redrawn from [17])

Based on this description, the systems concept of ISO 24748 is considered a viable framework for incorporating existing military capability models, enabling capability systems and capability systems-of-interest into a holistic capability life cycle model. This holistic model will be presented below.

4 The holistic capability life cycle model

4.1 System of interest: military units

Despite the context-dependent definition of the term ‘capability’, the generally recognised objective of military capability is the ability or power to achieve a desired operational effect in a selected environment and to sustain this effect for a designated period [10], [29]. This objective is only achieved when real-life military units are engaged in an operation [7]. In other words, military units, i.e., force elements, represent the elements of military power that bring military power directly to bear [30] and manifest real-life fighting power, which is defined as the ability to fight and operate. Fighting power comprises a conceptual component (how to fight), a morale component (the ability to get people to fight) and a physical component (the means to fight) [22].

4.2 Enabling capability systems

4.2.1 Concept system: concept development and experimentation

The purpose of the concept stage in accordance with ISO 24748 report is to identify stakeholders' needs, explore concepts and propose viable solutions. In the military organisation, this stage corresponds to the concept development and experimentation (CD&E) system. This system plays an important role in driving strategic transformation in the military community by enabling the structured development of creative and innovative ideas into viable solutions for capability development [31]. Concepts connect strategic guidance with the development and employment of future military force capabilities and can, therefore, be seen as an "engine for transformation" [32]. Transformation can eventually lead to changes in doctrine, organisation, training, materiel, leadership, education, personnel, facilities and policy [26].

Military concepts can be seen as visualisations of future operations. They describe how a military commander, using military art and science, might employ the necessary capabilities to meet future military challenges. Therefore, functional capabilities and their requirements are defined through military concepts [32].

Concept development is a process [31] or life-cycle phase [33] that is aimed at identifying conceptual solutions for capability. Experimentation is seen as a controlled investigation to discover information, confirm or disprove a hypothesis or formally validate a concept based on a conceptual rationale [32].

4.2.2 Development system: capability planning

The second life cycle stage in the ISO 24748 is called development. Its activities include refining system requirements and creating solution descriptions. The corresponding enabling capability system in military organisation is the capability planning system and its related capability model is the functional model. The capability developers ultimately refine the functional capability requirements initially identified in the CD&E stage into viable functional capability goals [2] and systems solutions.

4.2.3 Production system: procurement

The capability procurement system enables the production of the life cycle stage, with the purpose of producing the system of interest, i.e., military units [17]. The capability solutions are procured as capability systems by taking into account the defence lines of development such as TEPIDOIL (capability systems model). The procurement of the capability systems are usually carried out in a dispersed manner. The procurement organisation takes care of procuring the material components supplied by the industry, while the required tactics and procedures are developed by the doctrine centres of the armed forces.

4.2.4 Support and retirement system: in-service support

The enabling support system provides sustained system capability, while the retirement system disposes of the system [17]. In order to sustain in-use military units, it is necessary provide continuous support to maintain the suitability and effectiveness of the units to meet contemporary and emerging requirements [2]. In the disposal stage, either the whole unit is disbanded or some of its elements are retired as unfit for purpose. The material part of the in-service support system is usually managed by a single in-service support organisation, such as DE&S in the United Kingdom or the Defence Materiel organisation in Australia. In contrast, the military command of the unit is usually responsible for the overall in-service support of the units. For instance, the Navy (command), with the help of its subordinate supporting units, manages the sustainment of the naval surface squadrons and their organic elements (e.g., training of personnel or up-keeping the operating procedures).

4.3 The entire model: HCLCM

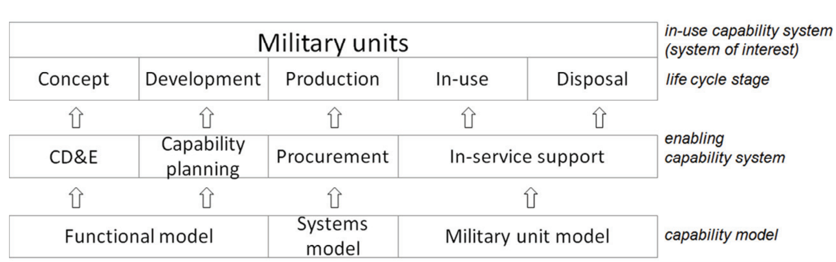


Figure 2. The holistic capability life cycle model, HCLCM

The entire HCLCM (Figure 2) shows how the in–use military capability system, i.e., the military unit and its evolution is facilitated from the concept to the disposal stage by enabling capability systems. It also illustrates how existing military capability models support each enabling capability system as explained in Sections 4.1 and 4.2. However, the life cycle stages of the enabling systems have not been included in Figure 2 for the sake of clarity.

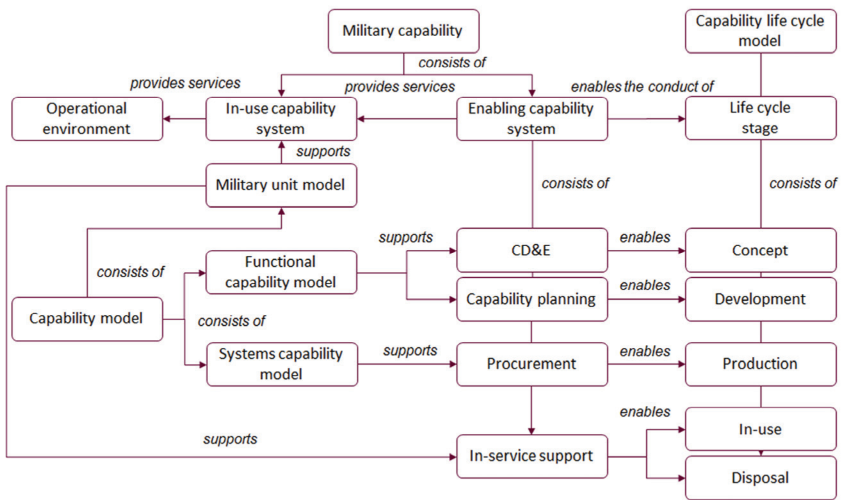


Figure 3. The information model of the HCLCM

Furthermore, to complement Figure 2, the HCLCM is also presented as an information model in Figure 3, which depicts the elements of the capability life cycle model in detail and shows how these elements relate to each other. The data and related data-flows of this information model have been presented in the earlier sections of the paper. The military capability models were reviewed in Section 2.1. The foundation for the capability life cycle model was introduced in Section 2.2. In-use capability system and enabling capability systems were presented in Sections 4.1 to 4.2. The relation between the in-use and enabling capability systems was in turn explained in Section 3.

5 Conclusions

This paper has proposed the holistic capability life cycle model (HCLCM). The model combines existing capability models and capability systems into a single model using ISO 15288 life cycle concept as the structure of the model. The HCLCM provides a holistic view of military capability throughout its life cycle and shows how existing capability models fit into this life cycle.

The main contribution of this paper is conceptual in nature. The model complements the existing, but still evolving understanding of military capability and its life cycle in a holistic and systemic way. First, the HCLCM is generic, i.e. it is applicable anywhere. Second, it is based on the existing capability models, and third, it is holistic. It connects a standard-based system life cycle model to the widely used military capability models. The model also shows that functional capability model relate primarily to the concept and the development life cycle stages, the systems model to the production stage and the military unit model in turn to the in-use and the disposal life cycle stages.

In practice, the model increases understanding and facilitates communication among various military capability stakeholders, such as concept developers, capability planners, acquirers and end-users, all of whom are focused on specific capability concerns and consider these concerns at particular stages in the capability life cycle. The HCLCM also helps to align each capability model in a larger context within the life cycle of military capability.

Future research on military capability should address the systems engineering (SE) issues related to the capability life cycle, such as requirements management or the systems engineering life cycle. These elements would further enhance the HCLCM by showing how stakeholders' needs are satisfied throughout the capability life cycle with the help of related capability models.

References

[1] Department of Defense, USA, 'Quadrennial Defense Review', Washington D.C, U.S.A, 2001.

[2] *Defence Capability Development Manual*, Director, Capability Operations and Plans, Canberra, Australia: Defence Publishing Service, 2006.

[3] *Collaborative Capability Definition, Engineering and Management (CapDEM)*, Fact Sheet CARDS 234 ed., Defence Research and Development Canada, Ottawa, Canada, 2005.

[4] C. Kerr, R. Phaal, and D. Probert, 'A Framework for Strategic Military Capabilities in Defense Transformation', 2006.

[5] S. Biddle, Military power: Explaining Victory and Defeat in Modern Battle: Princeton University Press, 2006.

[6] E. I. Neaga, M. Henshaw, and Y. Yue, 'The influence of the concept of capability-based management on the development of the systems engineering discipline', *in Proceedings of the 7th Annual Conference on Systems Engineering Research*, Loughborough University, UK, 2009.

[7] M. Touchin and C. Dickerson, 'Architecting for Capability', *in Proceedings of the Conference RNEC'08*, Leeds, UK, 2008.

[8] Y. Yue and M. Henshaw, 'An Holistic View of UK Military Capability Development', *Defense & Security Analysis,* Vol. 25, pp. 53–67, 2009.

[9] M. Fitzsimmons, 'Whither Capabilities-based Planning?', DTIC Document, 2007.

[10] Department of Defense, USA, 'Capability Portfolio Management', in *Directive 7045.20*, Ed. Washington D.C. U.S.A, 2008.

[11] *The Acquisition Handbook* Ed. 6, Ministry of Defence, UK, 2005.

[12] K. G. Fasana, 'Using Capabilities to Drive Military Transformation: An Alternative Framework', *Armed Forces & Society,* Vol. 37, pp. 141–162, 2011.

[13] D. Russell, N. Looker, L. Lu, and X. Jie, 'Service-Oriented Integration of Systems for Military Capability', in *11th IEEE International Symposium on Object Oriented Real-Time Distributed Computing (ISORC)*, pp. 33–41, 2008.

[14] J. Anteroinen, 'Integration of existing military capability models into the comprehensive capability meta-model', *in Systems Conference (SysCon), 2012 IEEE International,* pp. 156 – 16,2012.

[15] Defence Command, Finland, 'General Defence Doctrine - The Principles of the Defence System [Kenttäohjesääntö, Yleinen osa - Puolustusjärjestelmän toiminnan perusteet]', Helsinki, Finland, 2007.

[16] International Standardization Organization, 'ISO / IEC 15288:2008 - Systems and software engineering — System life cycle processes', Second Edition ed. Geneva, Switzerland, 2008.

[17] International Standardization Organization, 'ISO/IEC TR 24748: Guide to the application of ISO/IEC 15288 (System life cycle processes), part 1 ', ed. Geneva, Switzerland, 2010.

[18] M. F. Cochrane, 'Capability Disillusionment', *Defense AT&L,* Vol. 40, pp. 22–26, 2011.

[19] P. K. Davis, Analytic Architecture for Capabilities-based Planning, Mission-system Analysis, and Transformation: Rand Corp, 2002.

[20] Joint Doctrine and Concepts Centre, 'The UK Joint High Level Operational Concept', Ed. United Kingdom: Ministry of Defence, United Kingdom, 2004.

[21] H. Välivehmas. (2010, 10.11.2011). *Development programme of Finnish Defence Forces [Pv:n kehittämisohjelma ja näkymiä suunnittelusta]* [Online]. Available: http://www.finse.org/2010Spring/01_Finse_Valivehmas _PVn%20suorituskyvyn%20kehittamine.pdf

[22] Development, Concepts and Doctrine Centre, UK, 'Army Doctrine Publication Army Doctrine Primer', Ed. Swindon, United Kingdom: Development, Concepts and Doctrine Centre, UK, 2011.

[23] K. J. Holsti, 'The Concept of Power in the Study of International Relations', *Background,* Vol. 7, pp. 179–194, 1964.

[24] J. S. Nye, Jr., 'The Changing Nature of World Power', *Political Science Quarterly,* Vol. 105, pp. 177–192, 1990.

[25] S. De Spiegeleire, 'Ten Trends in Capability Planning for Defence and Security', *The RUSI Journal,* vol. 156, pp. 20–28, 2011/10/01 2011.

[26] D. Oxenham, 'Agile approaches to meet complex system of system engineering challenges: A defence perspective', in *System of Systems Engineering (SoSE), 2010 5th International Conference on*, 2010.

[27] A. Tetlay, 'Through Life Capability Management perspective for framework development for assessing and measuring System Maturity, System Readiness and Capability Readiness using Architecture Frameworks', in *System of Systems Engineering (SoSE), 2010 5th International Conference on*, 2010.

[28] R. Suzic and P. Svenson, 'Capabilities-based plan recognition', *in 9th International Conference on 9th International Conference on Information Fusion*, Florence, pp. 1–7, 2006.

[29] New Zealand. Defence Force, 'Statement of Intent 2008–2011', Ed. Wellington {N.Z.}: N.Z. Defence Force, p.v., 2008.

[30] 'Capabilities: What the UK Brings to Coalitions. Categories of Capabilities', *Whitehall Papers,* Vol. 50, pp. 44–64, 2000/01/01 2000.

[31] *MC Policy for NATO Concept Development and Experimentation,* NATO Military Committee, 2009.

[32] Chairman of the Joint Chiefs of Staff, USA, 'Joint Operations Concepts Development Process', Washington, D.C.: Joint Chiefs of Staff of the United States, 2006.

[33] International Standardization Organization, 'ISO Standard 19439:2006(E) - Enterprise integration - Framework for enterprise modelling', Geneva, Switzerland, 2006.

Proc. of the 2012 7th International Conference on System of Systems Engineering, Genoa, Italy - 16-19 July 2012

Aligning Analysis and Engineering Decision-Making within a Military Distributed System of Systems

Dr. Tim Rudolph Co-Author
ESC Chief Technical Officer & Chief Architect
Electronic Systems Center
Hanscom AFB, MA, USA
Tim.rudolph@hanscom.af.mil

Mr. Jon Salwen Co-Author
Enterprise Architect
The MITRE Corporation
Bedford, MA., USA
jsalwen@mitre.org

Mr. Murray Daniels Co-Author
Associate Technical Director
The MITRE Corporation
Bedford, MA., USA
mdaniels@mitre.org

Mr. Jeff Higginson Co-Author
Director of Engineering
The MITRE Corporation
Bedford, MA., USA
jhigginso@mitre.org

Abstract - *This paper describes and expands upon the proposed System of Systems Engineering (SoSE) Process Framework for the Air Force Materiel Command first outlined in IEEE SoSE 2011. Through repeated uses, it became quickly understood that systems-of-systems processes were foundationally similar to but different from basic systems engineering processes. The SoSE Process Framework, supporting processes, and products provide a context to integrate across modern complex systems-of-systems. In defining the SoSE Process Framework we tailor existing systems engineering processes and add a minimal amount of SoS-level engineering to it. The four core SoSE processes, Architecture and Analysis, Test and Evaluation, Management and Governance together can help to successfully deliver integrated capabilities at the SoS and Enterprise level within the U.S. Air Force. The key concept introduced is the alignment of analysis and engineering efforts across the Enterprise System Engineering and Architecting realm, and the synchronization of decision making within the SoSE components. The paper concludes by discussing implications of the SoSE Process Framework, as well as recommendations for actions, products, key stakeholder decisions and future challenges.*

Index Terms**:** Architecture, Capability, Governance, Integration, Management, System of Systems Engineering, and Testing.

1. Introduction

Military defense systems are increasingly complex and interdependent. Addressing capability gaps and future requirements through acquisition requires the ability to look beyond individual system capabilities to a system of systems (SoS) view. [1]

For example, The U.S. Air Force [2] defines a capability as:

> "*A capability is the combined capacity of personnel, materiel, equipment, and information in measured quantities, under specified conditions, that, acting together in a prescribed set of activities can be used to achieve a desired output.*"

A capability has a full range of Doctrine, Organization, Training, Materiel, Leadership and Education, Personnel, Facilities, and Policy (DOTMLPF-P) aspects.[1] More specifically, a capability typically includes a process flow that cuts across many systems. In the general case, the materiel aspect of the capability forms a SoS.

[1] Note that this use of the term "capability" is distinguished from the general English language use of the term referring to some aspect (usually functional) of some piece of materiel. A capability is an expression of operational need and desired functionality, not all of which is provided by materiel solutions. A capability is typically a superset of multiple system subsets in which the functionality of the capability can be found.

Note that capabilities can be specified at different levels of granularity. They can be very broad (e.g., "defend the area of responsibility from attack") or fairly narrow (e.g., "identify moving targets in an area of interest"). The SoS Engineering (SoSE) approach described here is oriented to the middle of this spectrum. At the most constrained end, a relatively small/constrained capability could be wholly supported by a single system that may be amenable to traditional systems engineering approaches. Broader problems may require other approaches (e.g., Enterprise Systems Engineering or Complex Adaptive Systems Engineering).

The military can no longer afford to build complete new systems that encompass the materiel portion of a capability. The cost of new development is untenable when existing solutions might be leveraged instead. A systems-of-systems approach with a capability focus is often the most effective approach to leveraging existing functionality from existing and new systems to meet a capability.

This paper uses the following definition of SoS Engineering (SoSE) [3]: (emphasis added)

> ***"System-of-Systems Engineering (SoSE)*** *– The process of* ***planning, analyzing, organizing, and integrating*** *the capabilities of a* ***mix of existing and new*** *systems into an SoS capability that* ***is greater than the sum*** *of the capabilities of the constituent parts. SoSE* ***emphasizes interoperability*** *among systems* ***developed under different*** *sponsorship, management, and primary acquisition processes."*

Resources for capabilities are not, in the general case, programmed at the capability-level. Instead, individual systems are programmed with specific requirements and budgets in a mostly stove-piped manner. SoSE can help deliver materiel (through these programs) that supports cross-cutting capabilities. The investment in SoSE is weighed off against investments in the individual systems. Both need to be addressed.

2. The SoSE Process Framework Approach

The SoSE approach described here is that of a SoSE Process Framework (hereafter termed "the Framework") that outlines broad process responsibilities while allowing for tailoring within specific problem domains. The Framework emphasizes the coordination of engineering efforts across the constituent systems. This includes elements such as architecture, risk management, and test & evaluation (among others). While the definition of SoSE above emphasizes interoperability, the approach described here also considers other significant SoS attributes such as end-to-end (E2E) security and performance. While it focuses on "intra-capability" integration, it does not focus on integrating across capabilities or across "the enterprise" – although the larger environment of the capability (to include its materiel aspects, which may include enterprise provided services) is a consideration when doing SoSE.

There are many different types of SoSs. The approach described here focuses primarily on 'Directed' and 'Acknowledged' SoSs[2]. For these cases, a critical part of the SoSE approach within the Framework is the governance of the engineering efforts across the constituent systems. The approach stresses the use of existing systems engineering processes and the addition of a minimal amount of coordination across those processes. It focuses on the practical execution of SoSE within the constraints of current high-level processes and builds off of existing insights into SoSE. While every attempt is made to leverage on-going program-level systems engineering processes, the SoSE processes described herein require some amount of SoS-specific resources. SoS sponsorship is required to support successful execution of these processes.

2.1 Approach to SoS Life-cycle

The Framework addresses the entire lifecycle of individual component system development "from cradle to grave." Specifically, it addresses SoSE-related actions to take as early as system concept development and as late as system sustainment and disposal.[3] A SoS exists at any point in time as the collection of systems that support a given capability. SoSE efforts can start at any point in the lifecycle of any individual constituent system and continue for the lifetime of the capability. They feed into the lifecycles of

[2] "Directed SoSs are centrally managed. In Acknowledged SoSs, the evolution of the SoS is the result of a collaboration among the constituent systems' managers.

[3] SoSE efforts generally have more influence on constituents in their earlier phases.

the constituent systems in a continuous, iterative manner.

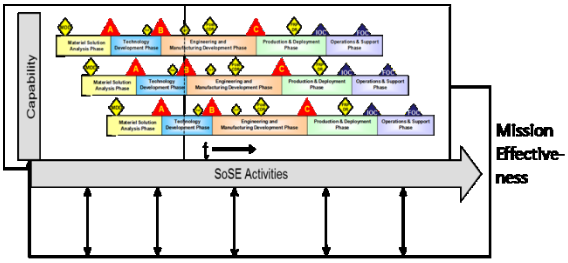


Figure 1- SoSE Life-cycle

As depicted notionally in Figure 1, the constituent systems in any given SoS generally have independent lifecycles. They can be at different stages of maturity and can evolve at different rates. Some can be in the early stages of concept development while others are in sustainment (or even get disposed). As a result, it is difficult, with the possible exception of a directed SoS, to describe the lifecycle of a SoS (the aggregate). The SoS has variable maturity over time. Significant collaboration among relevant stakeholders is needed to manage the evolution of the SoS. In addition, some SoSs have a transient, relatively short-lived existence independent from that of any of its constituents. An arrangement of systems can be constructed on relatively short notice to support a transient capability need. Once the need goes away, the SoS is effectively disbanded.

Some constituents evolve slowly (due to multiple inter-dependencies, high cost, technology limitations, etc.) The SoS Systems Engineer needs to consider the maturity of (and the rate of change of) the constituents. This produces many challenges (e.g., when to make governance decisions, keeping track of everything, testing SoSs with immature constituents) as well as injecting inherent risks (due to rapid change in an inter-dependent system). This can be ameliorated if the constituent systems apply certain architectural approaches that enhance system flexibility (e.g., open technology, loose coupling). Later in the constituent system's lifecycle (for example during sustainment) it is especially important that relevant programs participate in pertinent governance activities to ensure they maintain cognizance of the effects of their own and other programs' changes on the mission effectiveness of the SoS.

2.2 Goals

The goals of doing SoSE are to:

- effectively and efficiently engineer a SoS (and its constituents) that supports a mission capability and that (in the general case) transcends the boundaries of a single program
- proactively engage with appropriate stakeholders so as to eliminate unnecessary issues and avoid undue risk
- provide an input into investment decision-making processes

3 Four SoSE Processes

The Framework includes the following four SoSE processes:

- **Governance** – describes appropriate SoSE-related policy and direction, identifies roles and responsibilities
- **Management** – provides configuration, requirements, and risk management within a SoS
- **Architecture and Analysis** – models the capability and documents and analyzes constituent systems relationships
- **Testing and Evaluation** – tests SoS-level requirements at appropriate points in the lifecycle

3.1 Implementing SoSE Processes

The SoSE Implementation Model (Figure 2) shows the relationship among leadership, strategy, policy, process, and execution factors related to SoSE. The four SoSE processes support the strategic goal of acquiring SoSs that support cross-cutting capabilities. The execution of SoSE is guided by relevant policy and other guidance. SoSE is executed by engineers working at and above the program-level in conjunction with system-level SE processes.

The SoSE Implementation Model describes a continuous SoSE execution chain from high-level leadership down to system-level implementers. Leadership sets goals and objectives and approves the strategy. The strategy is embodied in the SoSE processes in accordance with relevant policy and guidance.

The processes are executed by program engineers as coordinated by engineers working above the program-level. The implementers of a SoS can be involved in various inter-relationships. Managing these relationships is a significant SoSE activity. The overarching management of SoS implementers varies by the type of a SoS. Only in the case of "directed" SoSs are SoS implementers primarily on a single project team. In other cases, all or most of the constituent systems may be under the control of a single organization that can exert some degree of management over the SoSE effort. In cases where the constituent systems aren't under the control of the same organization, the SoSE effort requires special constructs, such as working groups, to coordinate the implementation.

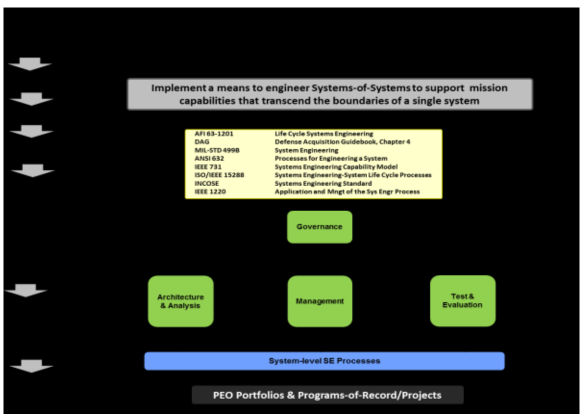


Figure 2 - SoSE Implementation

3.2 Implementation Critical Points

Some critical points about implementing SoSE via the Framework include:

- Governance of the SoSE effort is critical to establish authority and accountability of the SoSE stakeholders.
- The SoS Architecture represents the desired capability to include operational processes, desired effects, threats, and how the SoS functionality is allocated to its constituent systems.
- The SoS Architecture co-evolves with the architectures of the constituent systems. Changes to the constituent systems or in how they are interconnected correspond to changes to the SoS Architecture. Changes in the SoS Architecture may influence changes to one or more constituents.
- The SoS Architecture evolves in an additive manner (more or less gracefully) until some significantly disruptive event (e.g., new capability need, new technology) occurs requiring a significant change to the architecture.
- SoS requirements, risks, and relationships need to be managed.
- Test & Eval of the SoS needs to happen in conjunction with the Verification & Validation of the constituent systems.

3.3 SoSE Process Flow

The basic SoSE approach is to coordinate the system-level system engineering (SE) activities of constituent systems with a small number of SoS-level SE activities that focus on identifying stakeholders, developing the SoS Architecture, and evaluating the SoS. The four SoSE processes (Governance, Management, Architecture and Analysis, and Testing and Evaluation) overlay and integrate related SE processes to help realize the success of the SoS. Most existing SE processes already take into account some amount of consideration for systems as constituent members of SoSs (for example, by requiring configuration management of external interfaces). Depending on the specific SE processes used, the execution of SoSE, as described here, may require minor changes to how those system-level SE processes are executed.[4] The SoSE Process Descriptions identify the specific context in which they operate relative to system-level SE processes.

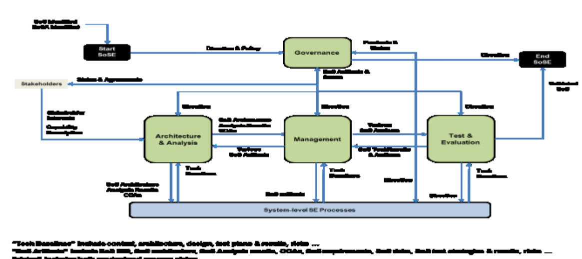


Figure 3 - SoSE Process Flow

[4] This may require direct intervention of engineering or program management leadership looking broader than a single system.

The SoSE Process Flow (shown abstractly in Figure 3) describes the SoSE processes in context relative to each other and to various system-level SE processes. The arrows in the diagram do not imply a strict time-sequenced flow of information and multiple iterations through the various SoS- and system-level processes will happen in parallel and at different rates depending on specific circumstances. One cause of multiple iterations is that constituent systems may be at different points in their lifecycles. The SoS is planned in the SoSE Architecture and Analysis process. SoSE is executed (in conjunction with various SE processes) as coordinated by the SoSE Management process. The SoS is checked in the SoSE Test and Evaluation process. The SoS is realized through various system-level processes executed on its constituent systems. The overall SoSE process is governed by and starts with the SoSE Governance process and includes the identification of a SoS, an appropriate SoS Authority (SoSA), and the need for associated SoSE. Generally this occurs as a result of capability planning through the identification of a capability need and the realization that the satisfaction of that capability need requires more than one system.

Once started, the SoSE Governance process governs the execution of the other three SoSE processes as well as the system-level Project Planning of constituent systems. The primary inputs to the SoSE Governance process are relevant policy/guidelines, the products from the other SoSE processes (most notably, the SoS Integrated Master Schedule (IMS) from the SoSE Management Process and the SoS Architecture from the SoSE Architecture & Analysis Process), and process status from the other SoSE processes. Interaction with all stakeholders is important at this point. The primary outputs are direction to the other SoSE processes and agreements to various stakeholders. Some of these stakeholders may be external to the execution of the SoSE Processes.

In the SoSE Architecture and Analysis process, SoS engineers develop the SoS Architecture and analyze it to identify issues, risks, and opportunities. The primary inputs to the SoSE Architecture and Analysis process include direction from the SoSE Governance Process, the description of the associated capability, the technical baselines of the constituent systems (to include their architectures), the interests of the relevant stakeholders (such as relevant CONOPS and higher-level architectures), and various SoS artifacts (e.g., SoS-level requirements, risks, test results). The primary outputs of this process are the SoS Architecture, which feeds both the SoSE Management and the SoSE Test and Evaluation processes, the SoS Analysis Results which support both investment and engineering decision-making, and various Courses of Action for how to evolve the SoS. The SoSE Architecture and Analysis process interacts with the system-level Design, Verification and Validation, and Decision Analysis processes.

In the SoSE Test and Evaluation process, engineers assess that the SoS satisfactorily supports the related capability. The primary inputs to the SoSE Test and Evaluation process include the SoS IMS and SoS-level Requirements (from the SoSE Management Process), the SoS Architecture (from the SOSE Architecture and Analysis Process), and direction from the SoSE Governance process. The SoS Architecture includes various capability performance measures that drive the SoS T&E process. The primary outputs of this process are SoS-related Problem Reports/Deficiency Reports, SoS-level risks discovered during the SoSE T&E process, and the validated SoS itself (to include a completed test report with a description of satisfied requirements).

In the SoSE Management process, engineers manage various SoS-level artifacts to include requirements, risks, test results, the SoS Architecture, analysis findings, and COAs. The SoSE Management process works primarily through interactions with the system-level Requirements Management, Configuration Management, Risk Management, and Technical Management and Control processes. The primary inputs to the SoSE Management process include outputs from those system-level processes, direction from the SoSE Governance process, and outputs from the SoSE Architecture & Analysis and SoSE Test & Eval processes. The primary outputs of this process are the SoS Integrated Master Schedule (IMS) which feeds the SoSE Governance process and the various managed SoS-level artifacts which feed various processes.

3.4 Use of the Framework

Existing systems engineering practice may need to be modified to varying degrees to support the use of the Framework. More specifically, the engineering resources to execute SoSE need to be identified. Some of these resources may be program-level systems engineers executing traditional systems engineering processes with an enhanced awareness of SoS considerations. There is also a need for some resources to execute SoSE at the SoS-level.

The SoSE processes are applied by engineers working at both the SoS-level and the program-level. At the SoS-level, engineers execute the SoSE processes for specified SoSs. At the program-level, systems engineers execute (perhaps slightly modified) traditional system-level SE processes with full awareness of relevant SoSs and in concert with engineers working at the SoS-level.

The systems engineering of the constituent systems of a SoS, if done right, takes into account the fact that the system is, in fact, a constituent of the SoS (to varying degrees depending on the specific system-level SE processes used). For example, individual projects are planned with awareness of their involvement in the SoS. The SoS provides context for each constituent (i.e., the other parts of the SoS form part of the environment in which the constituent operates). The capability that is being supported by the SoS provide operational context to frame the system's requirements. The risks of the SoS affect, and are affected by, the risks of the system. The SoS provides context for trade-off analysis at the system-level.

4 Summary

To a large extent, the military cannot afford redundancy of systems development – nor can it operate effectively with gaps in existing functionality. SoSE provides a manner to optimize interoperability between existing and new system implementations.

In lieu of significant changes to higher-level processes (DoD Program Planning, Budgeting and Execution process, requirements and acquisition policies), the objective of delivering integrated materiel support to mission capabilities can be at least partially realized by taking deliberate steps to execute SoSE and produce/provide related inputs to associated system-level governance mechanisms (e.g., proposed system requirements that align constituent systems with the SoSs that they support).

REFERENCES

[1] Dahmann, Dr. Judith, "Applying Business Process Modeling to Develop Systems Engineering Guidance for New DoD Acquisition Regulations," Proc. NDIA Systems Engineering Conference, San Diego, pp. 2, 5, October 2008.

[2] AFI 10-604, Capability-Based Planning, 10 May 2006,

[3] DoD Defense Acquisition Guidebook

[4] AFI 10-601, Operational Capability Requirements Development, 12 July 2010,

Systems Approach to the Safety of Complex Technical Facilities

Mikhail V. Belov
Deputy CEO
IBS
Moscow, RUSSIA
mbelov@ibs.ru

Abstract - *The paper is devoted to the safety of complex technical systems of systems (SoS) through the improvement of engineering processes at all system lifecycle stages. The safety of complex technical facilities (such as nuclear and conventional power plants, offshore oil and gas platforms and chemical processing plants) is analyzed based on the systems theory and SoS engineering. The safety concept discussed is the development and implementation of engineering process constraints to identify and prevent engineering inconsistencies which endanger systems safety. The method proposed consists of inconsistency classification as well as practical recommendations on how to formulate and implement safety constraints. The paper is based on the project implementation experience of IBS in 2010-2012 in the capital industry in Russia.*

Keywords: SoS safety, SoS design, engineering constraints.

1 Complex capital facility safety, engineering processes

The safety of technical systems has become extremely critical in recent years. Accidents involving the Fukushima NPP, Costa Concordia cruise ship, Sayano-Shushenskaya HPP and others have highlighted the importance of the safe operation of complex systems.

It is important to note that in the accidents mentioned – and in others – technical implementation issues played a considerable role, in addition to human error and natural catastrophic factors. The practice demonstrates that systems design is not sufficiently safety oriented.

The engineering community recognizes that systems safety might not be provided by traditional approaches like reliability theory, which worked very well in the past. Components (and even subsystems) reliability does not ensure safety: no considerable element failure was detected in the mentioned accidents (nor in many others).

Contemporary technical systems are electronics and software intensive, and human operators execute complicated mental functions. The “intellectual” domain of the systems is coming to predominate more and more. Mechanical or electromechanical means of ensuring safety and the probabilistic reliability approaches used formerly have become insufficient; thus, new safety methods are being developed.

System behavior analysis is the main idea of the STAMP approach [1], and accidents are considered as a system control issue which is solved by means of:

- accident modeling,
- system control hierarchy analysis,
- system operation processes modeling,
- operation process constraints development.

The constraints developed are exactly what ensure system safety.

The safety of complex capital facilities (CCFs) is now being considered from different viewpoints. The influence of engineering processes and design on system safety is of utmost importance, being considered in many of the documents of the US Department of Defense and Department of Energy, the IAEA and other organizations. The US DoD MIL–STD–882 standard [2] defines the selection of appropriate design solutions as the accident prevention method, focusing the safety issue on the design stage of the lifecycle. Due to the complexity of NPPs and extremely dire consequences of nuclear accidents, NPP safety is the focal point of the whole nuclear engineering community – designers, constructors, regulators and environmentalists. NPP development and utilization processes considerably influence safety, and so the IAEA devotes much attention to the engineering processes and supporting IT tools [3], [4]. In addition, the IAEA believes that the configuration management process is one of the most important for NPP safety. The main target [3] is to establish and to maintain conformance among requirements, design documents and data and a physically existing facility and its components (Fig. 1). Conformance failure during design, erection or operation is a very dangerous threat to capital facility safety.

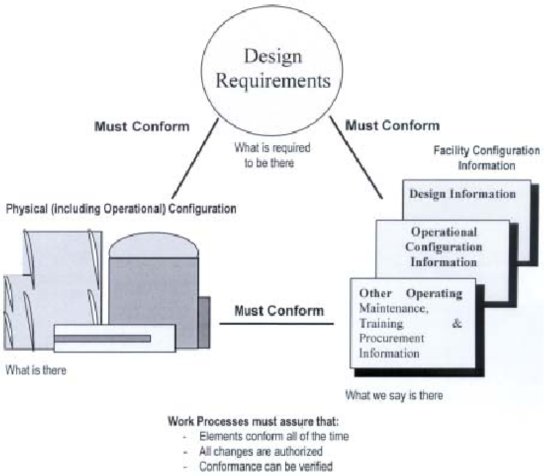


Figure 1. Configuration management equilibrium model.

This paper analyzes engineering processes and their influence on CCF safety at all lifecycle stages. The safety issue is considered through engineering inconsistencies, which are defined as any discrepancy between the engineering knowledge, data and documents (including their components) and the actual physical facility (its structure, systems and components), arising at the CCF design, construction or operation stages. Ensuring safer operations is regarded as the prevention and detection of engineering inconsistencies.

2 CCF engineering systems

Engineering processes take place in a very complex system environment at any stage of the CCF lifecycle, be it design, construction or operation. The CCF engineering system (CCF-ES) is defined as an organizational, technical, human machine system that executes various engineering functions at all lifecycle stages. The CCF-ES is a set of:

- organizational units involved in engineering activities,
- engineering processes (described in regulatory documents defining the processes),
- software which supports engineering activities,
- engineering data: the source and result of engineering activity.

The key characteristics of the CCF-ES are:

1. *The CCF-ES is software intensive.* The creation and operation of a contemporary CCF (fossil and nuclear power plants, petrochemical plants, oil & gas sea platforms and others) cannot be executed without intensive use of IT. Sophisticated IT systems are used to support CCF lifecycle processes, with the systems integrating tens and hundreds of different software platforms like CAD, CAM, CAE, PDM, MES, EAM and many others. Frequently, these systems are not considered systems at all, but a set of software platforms or even end user workplaces. This happens because the software platforms were implemented independently, or the architectural concepts were not initially developed. But even if an enterprise and its IT system are not considered an SoS, they do not cease being so.

2. *The CCF-ES is an "acknowledged" SoS.* A CCF-ES is a complex SoS, and at least two representations of the CCF-ES might be mentioned: CCF-ES as a set of organizational subsystems and CCF-ES as a set of software platforms-subsystems. The CCF-ES is a typical acknowledged SoS according to [5]: there are mutual targets and resources and a common management is assigned, but constituent systems have their own targets, resources and owners. The concept of an acknowledged SoS leads to a quite "soft" and intellectual control hierarchy corresponds to the current state of society and technology development and the considerable growth in such kinds of SoS in the US DoD is mentioned in [5]. It is important for this paper that all changes in an acknowledged SoS be implemented only based on cooperative principles between the SoS and constituent systems, but not directive command control.

3. *Engineering processes and data* are exactly the factors combining subsystems into a SoS. It is precisely processes that drive the system toward the main target of an enterprise and its main functions, so the processes require special attention and consideration.

4. *Engineering processes are "cross organizational"* and are performed by organizationally and geographically scattered units. The huge number of cross-organizational links is an important specific factor in design and engineering enterprises, and these links arise for a variety of reasons. First, by definition, there is a natural complexity in CCFs and it is impossible to concentrate all necessary expertizes in one firm. Second, there is a permanent shortage of highly qualified human resources, which is normal for a growing economy. All this leads to the necessity to establish cross-organizational engineering processes which are executed by different organizations that are geographically and organizationally scattered. It is significant that, due to the cross-organizational property of the CCF-ES, the system's scope might not be defined very clearly: for different reasons, some organizations or

software might be considered internal system components, and some might be seen as external entities.

5. *The CCF-ES is a human machine system*, with very complicated human heuristic activities, a wide variety and complexity of data models, and increasingly diversified operations executed with this data. The human-machine factor is another evident property of contemporary engineering activity. Engineering itself is a combination of sophisticated heuristic operations and labor intensive routines. IT systems play a supporting role, being a tool which facilitates routine operations. But the presence of a considerable share of intellectual heuristic activity in engineering makes engineering IT-systems much more sophisticated than ERP or other systems supporting financial and administrative activities. This complexity is represented by a greater variety of engineering data and operations with this data.

6. *The CCF-ES is a growing system*: rapid technology growth involves new tools and components, software platforms, data and more complex processes.

3 Systems approach to engineering inconsistencies

We have found above that the CCF-ES is a SoS with a considerable intellectual domain – sophisticated software platforms and high intellectual human activities. The CCF-ES also has a quite flexible control hierarchy and soft scope, comprises relatively independent constituent systems, and grows; in other words, it is a very complicated and sophisticated entity. Engineering inconsistencies are the undesirable result of CCF-ES operation, and their occurrence is considered an accident (as STAMP states [1]), the result of unsafe CCF-ES operations.

System behavior analysis and incident consideration from the system control point of view is the core idea of the STAMP.

The system behavior (or operation) of such a complicated system like the CCF-ES cannot be planned, predicted or prescribed due to the key role of the intellectual domain in system control mechanisms.

“Uncertainty” of the system’s behavior and its outcomes appears. Indeed, design solutions coming out of a CCF-ES cannot be described a priori: they are exactly “generated” by the system during design activities.

In these circumstances, appropriate “intellectual” tools should be developed and based on the specifics of the CCF-ES analyzed above to provide safe operation of the CCF-ES.

This problem is solved in [1] by the following:

- accident modeling,
- system control hierarchy analysis,
- system operation processes modeling,
- safety constraint development.

It is precisely constraints on the behavior or operation of the system that should provide system safety.

The task of improving CCF safety is carried out by developing and adhering to control constraints on the engineering processes. Engineering process constraints are imposed as additional parts of the process to be embedded in the main processes. The development of constraints involves process modeling, issuing instruction documents and putting information systems and other supporting tools into use.

4 Analysis of engineering inconsistencies

Let us define engineering inconsistencies within information-information, information-physical and physical-physical groups.

Information-information inconsistency is nonconformance of engineering information (knowledge, data and documents and their elements), for example, discrepancies between a design solution and the requirements that occur due to the introduction of new requirements.

Information-physical inconsistency is the nonconformance between engineering information and the actual facility (its structure, systems and components), for example, the lack of conformity between a pump as designed and the one actually purchased and supplied to the site.

Physical-physical inconsistency is the nonconformance among elements of the actual facility (its structure, systems and components), such as noncompliance between electrical and automatic control systems developed by an engineering company and those developed by subcontractors.

Information-information inconsistencies may appear from the very beginning of the CCF lifecycle, while information-physical and physical-physical inconsistencies arise at the construction/manufacturing/creation stage of the lifecycle.

Information-information inconsistencies appear among design data, knowledge and documents, which are generated in one or different organizational units using one or different software platforms.

Apparently, cross-organizational and cross-software inconsistencies appear more frequently, so more attention should be paid to developing safety constraints to control these processes.

For classification purposes, we define the following types of inconsistencies: solution-requirements (between design solution and requirements), schematics (between different P&ID designs), geometrical (between geometrical elements), and actuality (when some data becomes irrelevant).

Information-physical inconsistencies appear and are detected when purchasing, supplying, manufacturing, construction, assembling, erecting, etc. during processes which deal with physical components or subsystems of the CCF. The following types of inconsistencies might be defined: planning (between different working plans or schedules), purchasing and supply (between designed, ordered and supplied materials, components, equipment, etc.), construction and assembly (between design documents and physically existing construction elements, equipment, etc.).

Physical-physical inconsistencies are detected during different work, tests, checks, and so on. Physical-physical inconsistencies are not the focus of this article: they are often preceded by information and/or information-physical inconsistencies and must be identified and prevented as early as possible.

A list of typical engineering inconsistencies has been compiled based on IBS's project experience. Projects in 2010-2011 were focused on information systems implementation and improving engineering processes at facilities within the Russian oil & gas and energy sectors. Besides identifying the inconsistencies themselves, the laboriousness of fixing them and the frequency of their appearance might be estimated as well.

Such a list, of course, should not be considered to be exhaustive (any list of inconsistencies or mistakes is incomplete). However, it might be used as a starting point, to be updated by an organization implementing this approach.

5 Safety constraints development

For each type of engineering inconsistency, a special detection and prevention approach should be developed in the form of sub-processes or parts of the processes which are embedded in the main engineering processes. As stated above, these parts play the role of constraints and provide for the safety of CCF operation.

The concrete content of each constraint is very specific to each concrete CCF; an organization does not have the practical means to apply the content in other projects, so it is not described here. In this paper, we develop and present common approaches to be used in practice.

Constraints represent inspection rules which determine:

1) "What and how": the entity being inspected; the procedure, tools and data to be used; the results expected; the inspection criteria.

2) "When": at which stage of the main engineering process and between which tasks of the process are the inspections performed.

3) "Who": which organizational unit does the inspections.

The first part specifies the engineering information elements, comparing them to each other and to the actual system elements, and establishes the comparison rules. The other two parts define the constraint integration into the CCF-ES. A table of inconsistencies, classification of inconsistencies and other results from the previous section should be used to develop the first part of the constraints.

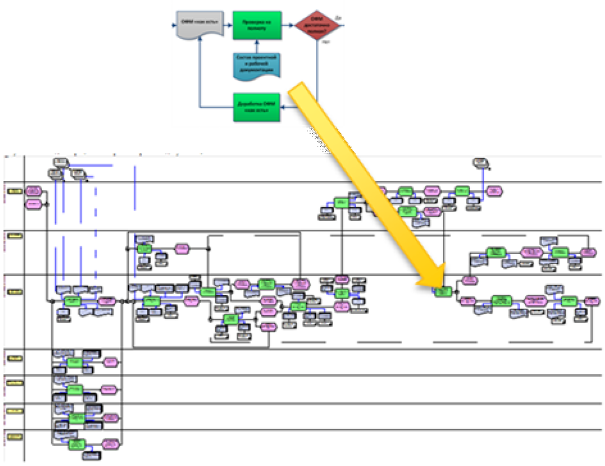


Figure 2. Safety constraints integration.

Engineering inconsistencies cause undesirable changes in the CCF lifecycle, and the negative consequences grow substantially from earlier to later lifecycle stages. Changes are processed according to the regular lifecycle from their identification to verification. To avoid this, it is necessary to execute detection operations as early as possible and to inform all parties to the engineering process about the design solutions in progress as early and as completely as possible. It is preferable to prevent inconsistencies than to fix them; in the former case, the procedure is stopped before beginning and undesirable changes – which use up resources – are minimal or even non-existent. This might be achieved based on parallel and collaborative engineering.

- Parallel Engineering is a special approach to the joint work of geographically and organizationally scattered parties with relevant but incomplete (unapproved) knowledge, data and documents.

- Collaborative Engineering is a special approach to the use of information technology for prompt and full interaction of geographically and organizationally scattered parties: online access to data and documents, video and audio conferences, text messaging, virtual blackboards, desktop sharing at meetings and online forums.

The following techniques might also be applied to identify and prevent inconsistencies:

- configuration management ideas: baseline control, configuration hierarchy and identifying and controlling items;

- requirements management, including requirements tracing;

- direct scrutiny of nonconformance among drawings, diagrams, geometry and work schedules, etc.

All this is done in a very complicated systems environment – the CCF-ES is an acknowledged SoS; thus, we should consider the core elements of an engineering SoS [5] to develop safety constraints:

1) Translating capability objectives (safer CCF operation) into "top-level" requirements, defining specific types of inconsistencies to be detected and prevented over time.

2) Understanding constituent systems and relationships (organizational and software) and their interactions and connections in terms of functions, data sharing, control and resources.

3) Assessing the extent to which SoS performance meets capability objectives: safer operation through the use of constraints to be developed.

4) Developing and evolving an SoS architecture in terms of functions, relationships, information flows and control: developing constraints and incorporating them into the processes and different organizational and software systems/subsystems.

5) Monitoring and assessing the impact of changes (in CCF systems/subsystems and outside the CCF) on CCF safety and identification of appropriate changes to be made in constraints. This element is particularly important for a CCF, as the CCF-ES is an acknowledged SoS, where changes in subsystems are not strictly controlled at the whole system level.

6) Selection of engineering inconsistency types and choosing possible solutions for constraints at both the subsystems level and the whole system level.

7) Orchestrating upgrades to SoS considering the independence of systems/subsystems (planning, resources, etc.), which is especially important for an acknowledged SoS of the CCF.

6 Examples

Let us consider two examples to illustrate the approach proposed to detect and prevent the typical engineering inconsistencies described in sections 4-5.

1. Inconsistency of the material and the geometry of the prefabricated pipe segment with the designed ones. This inconsistency happens at the pipe manufacturing plant but is usually detected only at the construction site during assembly work or during several acceptance tests (which are worth performing). For the detection of such an inconsistency, a direct check should be made to match pipe segments with the design data and documents. The constraint is established to prohibit the continuation of the process before ascertaining conformance. To make this solution more efficient, it is quite reasonable to execute such checks as early as possible. Thus, we recommend integrating this safety constraint into the appropriate process of the manufacturing plant as a quality control element.

This example demonstrates both the organizational aspect and the SoS specifics of constraints development and implementation. To develop adequate constraints, it is necessary to study the processes of the manufacturing plant (core elements 1 & 2 from the list presented in section V) and to formulate recommendations for the plant management (core element 6). It is extremely important to notice that this constraint is not mandatory for the plant management (acknowledged SoS), so regular monitoring of the plant processes should be undertaken to check for possible deviations (core element 5).

To prevent such inconsistencies, we recommend establishing joint workgroups of employees from the manufacturing plant with design entity employees (using collaboration engineering and parallel engineering). Safety constraints are established as requirements to regularly conduct virtual (or real) meetings of designers with technology developers, manufacturers and quality controllers from the manufacturing plant. The manufacturing and shipping processes of the prefabricated pipe segments should be discussed and approved at these meetings.

2. Changes which were made in construction (or assembling) work but not fixed in the "as built" documents. Such an inconsistency happens during construction, erection and assembling work, but might appear at the commissioning stage, or even later at the operation stage. Direct checking should be used to match the physically built or assembled component or elements to the appropriate design and requirements data.

The constraint is formulated as the mandatory checking of appropriate conformance which is embedded in the process. The implementation of this inspection as an element of the construction quality control system is a good approach to making the check as early as possible and to fixing the inconsistency as cheaply as possible. Furthermore, instrumental automated measurements are an appropriate extension of the constraint.

The example demonstrates the organizational aspect as well as the SoS specifics of the problem. SoSE core elements 1 & 2 and 5 & 6 are used to develop and to implement adequate constraint. To prevent the inconsistency, we also recommend establishing workgroups which should join designers and construction engineers (parallel and collaborative engineering) to discuss and to agree on the construction work technique. Different IT tools are appropriate for this joint activity, for example, online access to data and documents, video and audio conferences, text messaging, virtual blackboards and desktop sharing.

7 Conclusions

The systems safety approach for the oil & gas, chemical, utilities and other capital industries is the topic of this paper. The approach is based on engineering processes analysis and improvement during the whole lifecycle of capital facilities.

Systems theory, SoS engineering [5], systems safety [2] and the STAMP [1] method are highlighted in the paper.

The essence of the approach consists in the development of constraints on the engineering processes and implementation of those constraints to detect and prevent engineering inconsistencies which endanger systems safety.

The safety method includes:

- inconsistency definition and classification,
- a list of typical inconsistencies generated based on the project experience of IBS, where the author works,
- engineering process constraints definition,
- methods developed for identifying and preventing inconsistencies,
- the SoS engineering approach appropriate for developing and adhering to constraints.

Developing an organizational approach to ensure the efficient implementation of the proposed method can be considered as a future avenue of research.

References

[1] N. Leveson "Engineering a Safer World" http://sunnyday.mit.edu/safer-world

[2] DEPARTMENT OF DEFENSE STANDARD PRACTICE FOR SYSTEM SAFETY MIL-STD-882D 10 February 2000 http://www.everyspec.com on 2012-01-31 T10:17:03

[3] Application of configuration management in nuclear power plants. — Vienna : International Atomic Energy Agency, 2010.

[4] INFORMATION TECHNOLOGY FOR NUCLEAR POWER PLANT CONFIGURATION MANAGEMENT IAEA, VIENNA, 2010.

[5] Office of the Deputy Under Secretary of Defense for Acquisition and Technology, Systems and Software Engineering. Systems Engineering Guide for Systems of Systems, Version 1.0. Washington, DC: ODUSD(A&T)SSE, 2008.

Research on Capability Requirements Generation of Weapon System-of-systems Based on CRTAM Model

Yajie Dou
Department of Management
National University of Defense Technology
Changsha City, Hunan Province, China
wdxhxcbdyj@163.com

Long Li
Department of Management
National University of Defense Technology
Changsha City, Hunan Province, China
eric.Longlee@gmail.com

Qingsong Zhao
Department of Management
National University of Defense Technology
Changsha City, Hunan Province, China
zqszqr@163.com

Yingwu Chen
Department of Management
National University of Defense Technology
Changsha City, Hunan Province, China
ywchen@nudt.edu.cn

Abstract - *The thoughts and methods of Capability-Based requirement analysis have gone increasingly deep into the development of Weapon System-of-Systems (WSoS). Effective analysis and programming of capability are fundamental tasks for requirement analysis of WSoS. Under the background of operation, there are repeating capabilities among different operational tasks. Acquiring Capability Index Requirements (CIRs) from various Operational Task Requirements (OTRs) is difficult. To resolve this problem, the value of each capability index is measured as an interval and Meta-activity model is proposed, the operational activity is decomposed to Meta-activity collections and the capability requirement of a single operational task is captured using Meta-activity-Capability mapping. Based on this, the WSoS Capability indices are divided into five kinds and four types of Capability Requirements Transverse Aggregation Model (CRTAM) are built to deal with the CIRs acquired from the tasks decomposition. A Capabilities Indices Requirements List (CIRL) under the multi-tasks condition is gained. Finally, the method is validated by an example of anti-stealth air defense systems of systems (ASADSoS).*

Keywords: Weapons system-of-systems (WSoS), Meta-activity Decomposition, Capability Requirements Transverse Aggregation Model (CRTAM), Capabilities Requirement Generation

1 Introduction

Capability is an inherent yet static property regardless of the quality traits (performance indices/tactical and technical indices), quantities and operation process of System-of-Systems (SoS) [1]. However, Capability requirements are gained from the decomposition of missions. The relationship between capability requirement and mission is n to n. The capability requirement of Weapon System-of-Systems (WSoS) is demonstrated by an index system containing many indices and a hierarchical structure [2-4], which can be called as "Capability Index Requirements (CIRs)". Capability index requirement generation of WSoS is a key part in WSoS capability analysis [5]. In this paper, interval is used to represent the value of WSoS capability indices. Capability is an inherent yet static property regardless of the quality traits (performance indices/tactical and technical indices), quantities and operation process of SoS. Capabilities depict multiple aspects of SoS characteristics and involve component systems from different levels. Therefore, SoS capability is demonstrated by an index system containing many indices and a hierarchical structure.

Inside of the index system, top-level indices depict top-level capabilities, indices from lower level are the further decomposition and refinement of these from upper level and they embody a father-son relationship, indices from the lowermost level are directly related with the performance indices of equipment. In a word, capability indices are decomposed through different levels to the levels of performance, which is greatly related to specific equipment and of instability.

A sketch map of a SoS capability indices decomposition is shown in Fig 1. A capability index system can be divided into at least three levels: top capability level, sub capability level and performance capability level. S is the number of levels of a SoS capability indices system, $s \geq 3$ and level 1 is top capability level, level 2 to s-1 is sub capability level and level s is performance capability level.

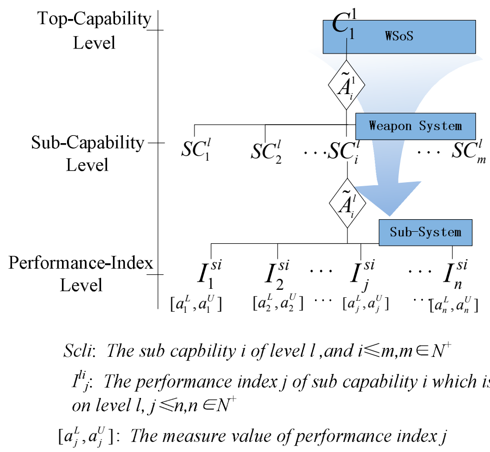


Figure 1.SoS capability indices decomposition

Capability index requirements (CIRs) represent the structure of capability index system and measure of index value in SoS requirements. CIRs are gathered from different stakeholders. They are uncertain and subjective. And moreover, Numbers and kinds of equipment with the same performance indices are actually not identical in SoS. Equipments with the same kind of indices differ with each other in value. Interval number [6-7] is an efficient means to depict complex and uncertain information, which is widely used in multi-attribute decision [6], etc. In this paper, interval is used to represent the value of SoS capability indices. The structure of the value of a capability index is shown as a pair of real numbers:

$$\tilde{q}_j^{li} = [a_j^L, a_j^U]^{li} \tag{1}$$

where, $\tilde{q}_j^{li}$ is the jth index value of the ith sub-capability in the lth level, l, $m \in N+, i \le m$,m is the number of the capabilities is the lth level, $n \in N+, j \le n$,n is the number of indices of the ith sub-capability in the lth level. aLj and aUj is the upper limit and lower limit of the interval respectively.

2 Decomposition Process of Operational Missions

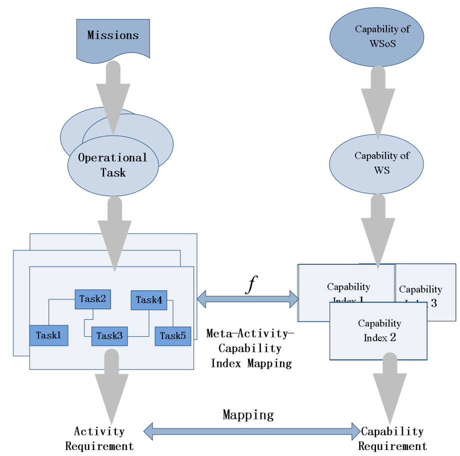


Figure 2.Meta-activity-Capability Index Mapping sketch map

This section illustrates a Capability Encapsulation method to define capability as the driven force of Meta-activity, which combines the capability and operational task. The new method also puts Meta-activity into operational task list in the process of decomposition, which accomplishes the distribution of capabilities at the same time. The new method makes the mapping process between tasks and capabilities more clear, differs from existing capability requirements analysis method which usually are based on qualitative ways.

The mapping rule is illustrated in Fig 2. capability requirement is defined in formula 2

$$Capability = f(ma_{i1}, ma_{j2}, \cdots\cdots, ma_{rn}) \tag{2}$$

i, j, r, n are natural numbers. *i, j, r* represent activity type identifiers.1,2,n represent identifiers of Meta-activities, f represents the index mapping rule from Meta-activity to capability performance.

3 Capability Requirements Transverse Aggregation Model (CRTAM) of sub-task

3.1 The classification and standardization of capability index

In this study, the capability indices are divided into two kinds-qualitative index and quantitative index. They both show incommensurate and contradictory features to each other. Therefore, each index cannot be directly for the calculation with the same integration method. The qualitative index and quantitative index should be normalized and standardized with the different treatment respectively. For the qualitative index, because of its possible various values, we can build one to one mapping or qualitative level quantization table for its standardization. For quantitative index, it can be divided into benefit and cost type according to the influence degree of the index value up on combat task. The quantitative index can be obtained by experiment, field measurement, statistical report analysis and other methods [8].For the two types of index, their standardized treatments require the establishment of corresponding normalized function.

In order to deal with qualitative index in a simple and practical way, a quantitative scale method was proposed in this study. The qualitative index results were directly mapped into a value between 0 and 1 through Table 1. It is recommended that the qualitative evaluation of level value often is set between 5 and 9, and the 9 level is used in the most frequency [8-9]. Experiments [10] show that in different objects identification, ordinary people can distinguish them when grades at level 5-9.Three kinds of quantification scales of qualitative index are shown in Table 1.

Table 1.Three kinds of quantification scales of qualitative index

Value \ level	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
9 scales	worst	worse	bad	A little bad	average	A little good	good	better	Excellenct
7 scales	worst	worse	bad		average		good	better	Excellenct
5 scales	worst		bad		average		good		Excellenct

In addition, the fuzzy number scaling quantization methods are also used in the language value quantification, such as triangular fuzzy number and trapezoid fuzzy number method [10], these quantitative methods can avoid the loss of information, but the computational process is more complex, especially the final sort process. In order to make this method simple and practical, we choose 7 scales as the quantification method in this study.

In the study, the cost index is divided into consumptive and reusable type index. In order to solve the problem conveniently, the ability index value was standardized into a common intervals, such as [0, 1], [0,100] and other quantitative indexes standardized function, which is suitable for the comparison between different weapon system capability development plans. However, this section no longer discusses the index standard function in detail.

3.2 Capability Requirements Transverse Aggregation Model (CRTAM)

Capability Requirements Transverse Aggregation Model (CRTAM) of sub-task in this study is not the aggregation progress from bottom capability index to upper capability index. The aggregation is the capability requirements aggregation of different tasks, which is not lengthways Aggregation (LA) but Transverse Aggregation (TA) [9]. There are some repetitive capabilities among deferent tasks. Though meta-tasks are the same, operational tasks are different. Implement level of meta-tasks is different. Transverse Capability Aggregation integrates capability requirement of different tasks, but not changes mapping rules between meta-task and capability, just as demonstrated in Fig 3:

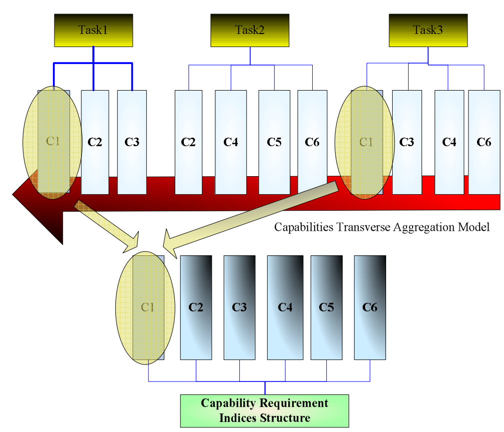


Figure 3. Capability Requirement Indices Structure sketch map(C is an abbreviation of Capability)

Capability Aggregation methods of capability requirement index are different according to different types of capability index. Hence, four types of CRTAM model are built in this study.

I. Maximization Aggregation Model(MAM)

Max value of this capability index involved in all tasks would be got as the value of capability index in Maximization Aggregation. Let $Value(C_i)$ indicate the value of aggregated capability index Ci($m \in N+$, $i \le m$)and m indicate total number of capability index in WSoS.

Values of C_i from Task1 to Task j are $Value(C_{i1})$, $Value(C_{i2})$, …, respectively, and $Value(C_i)$ can be calculated by the formula:

$$Value(C_i) = Max\{Value(C_{i1}), Value(C_{i2}), \cdots\cdots, Value(C_{ij})\} \quad (3)$$

Generally, this type of aggregation model would be used to get qualitative capability index. Qualitative capability index would be calculated with quantification measurement method. In quantification measurement method, Qualitative capability index can be got by directly mapping Table 1to the value between 0 and 1. Higher score indicates better performance of this index. Max value of Capability requirement index among all correlative tasks should be got as the Capability requirement index during capability index aggregation of sub-tasks.

II. Interval Union Aggregation Model(IUAM)

In Interval Union Aggregation Model, we get the union of capability index interval in all tasks as the value of aggregated capability index. Let $[x_{Li}, x_{Ui}]$ indicate the value of aggregated capability, and m indicate total number of capability index in WSoS. Values of C_i from Task1 to Task j are $[x_{Li1}, x_{Ui1}]$, $[x_{Li2}, x_{Ui2}]$…, $[x_{Lij}, x_{Uij}]$, …, respectively, and $[x_{Li}, x_{Ui}]$ can be calculated by the formula:

$$[x_{Li}, x_{Ui}] = [x_{Li1}, x_{Ui1}] \cup [x_{Li2}, x_{Ui2}] \cup \cdots\cdots \cup [x_{Lij}, x_{Uij}] \quad (4)$$

This formula means:

$$[x_{Li}, x_{Ui}] = [Max\{x_{Li1}, x_{Li2}, \cdots\cdots, x_{Lij}\}, Max\{x_{Ui1}, x_{Ui2}, \cdots\cdots, x_{Uij}\}] \quad (5)$$

In general, this kind of aggregation model is used to get efficient capability index. That is to say, if higher index value means better performance, index value of total task should be got by this kind of model. Take motor speed for an example, task4 and task6 both comprise a meta-task named main battle assault. Capability index, motor speed, valued [60, 65] (km/h) in Task4 and valued [50, 85] (km/h) in Task6, so the aggregated value of motor speed for total task is [60, 85] (km/h).

III. Interval Join Aggregation Model(IJAM)

In Interval Join Aggregation Model, we get the Join of capability index interval in all tasks as the value of aggregated capability index. Let $[x_{Li}, x_{Ui}]$ indicate the value of aggregated capability, and m indicate total number of capability index in WSoS. Values of Ci from Task1 to Task j are $[x_{Li1}, x_{Ui1}]$, $[x_{Li,} x_{Ui2}]$…, $[x_{Lij}, x_{Uij}]$, …,respectively, and $[x_{Li}, x_{Ui}]$ can be calculated by the formula:

$$[x_{Li}, x_{Ui}] = [x_{Li1}, x_{Ui1}] \cap [x_{Li2}, x_{Ui2}] \cap \cdots\cdots \cap [x_{Lij}, x_{Uij}] \quad (6)$$

This formula means:

$$[x_{Li}, x_{Ui}] = [Min\{x_{Li1}, x_{Li2}, \cdots\cdots, x_{Lij}\}, Min\{x_{Ui1}, x_{Ui2}, \cdots\cdots, x_{Uij}\}] \quad (7)$$

In general, this kind of aggregation model is used to get cost capability index. That is to say, if lower index value means better performance, index value of total task should be got by this kind of model. Take information upgrade time for an example, Task2 and Task4 both comprise a meta-task named information upgrade time. Capability index, information upgrade time, valued [0, 5] (min) in Task2 and valued [1, 3] (min) in Task4, so the aggregated value of motor speed for total task is [0, 3] (min).

IV. Interval Plus Aggregation Model(IPAM)

In Interval Join Aggregation Model, we get the sum of upper and lower limit of capability index interval in all tasks as the value of aggregated capability index. Let $[x_{Li}, x_{Ui}]$ indicate the value of aggregated capability, and m indicate total number of capability index in WSoS. Values of C_i from Task1 to Task j are $[x_{Li1}, x_{Ui1}]$, $[x_{Li2}, x_{Ui2}]$,……, $[x_{Lij}, x_{Uij}]$,……, respectively, and $[x_{Li}, x_{Ui}]$ can be calculated by the formula:

$$[x_{Li}, x_{Ui}] = [x_{Li1} + x_{Li2} + \cdots\cdots + x_{Lij}, x_{Ui1} + x_{Ui2} + \cdots\cdots + x_{Uij}] \quad (8)$$

In general, this kind of aggregation model is used to get consumption capability index, such as ammunition supplement. Take ammunition supplement amount for an example, task5 and task6 both comprise a meta-task named material supplement. Capability index, ammunition supplement, valued [30, 60] (t) in Task5 and valued [70, 150] (t) in Task6, so the aggregated value of ammunition supplement for total task is [100, 210] (t).

4 Case Study

Anti-stealth air defense systems of systems (ASADSoS) is a large-scale integrated system composed of heterogeneous and independent weapon systems which are worked together against the enemy air attacking. In ASADSoS, weapon systems are geographic distributed, effectively working together by interconnection, intercommunication and interoperation to achieve the goal of ASADSoS, and each weapon system can be operated and managed independently. Hence, ASADSoS is a typical SoS. Classified by defense scope, ASADSoS at least have two classes: National/Homeland Air Defense [12-13] and Area Air Defense [14-15]. In general, ASADSoS is consisting of three sub-systems: information system, command and control system, interception and strike system.

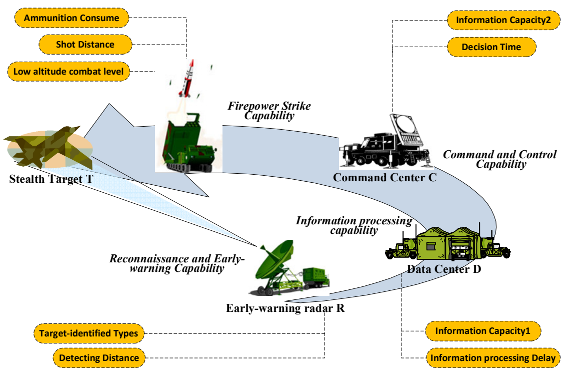


Figure 4. The anti-stealth air-defense system-of-systems (ASADSoS) and its performance index

This case is about anti-stealth air-defense system-of-systems (ASADSoS) with a specific goal that prevent enemy stealth aircraft destroy target T. From the view of capability analysis, ASADSoS can be divided into four capabilities which are shown in Fig 4: reconnaissance and early-warning capability, command and control capability, forces mechanization capability, firepower strike capability.

It is assumed that the anti-stealth air-defense system-of-systems (ASADSoS) receives the superior orders, it needs to complete three combat missions: Task1,Task2,Task3.Through the combat task decomposition, we can get the three kinds of ASADSoS performance requirements under different combat task conditions[14-15],which are shown in Table 2.

Table 2. The ASADSoS performance index selections list under three different task conditions

Top-capability	Sub-capability	Performance Index	Task1	Task2	Task3
ASADSoS Capability	Reconnaissance and Early-warning Capability	Target-identified Types(kind)	[15,30]	[20,35]	[20,30]
		Detecting Distance(km)	[10,20]	[10,45]	[10,55]
	Information processing capability	Information Capacity1 (KB/sec)	[224,400]	[224,500]	[336,400]
		Information processing Delay (sec)	[60,180]	[60,120]	[60,180]
	Command and Control Capability	Information Capacity2 (KB/sec)	[224,336]	[224,336]	[224,336]
		Decision Time (sec)	[180,600]	[120,320]	[180,540]
	Firepower Strike Capability	Ammunition Consume (ton)	[100,300]	[100,200]	[100,150]
		Shot Distance(km)	[10,50]	[10,45]	[10,60]
		Low altitude combat level	Strong	average	Very strong

Different tasks have the repetitive performance index, which is due to the same combat events appeared in different tasks. The combat activities name are the same, but different task requirements. So the combat activities are carried out in different levels. Different tasks have the different performance index requirements, thus, we need to deal with the performance index requirements with different aggregation model for different types of index, to achieve a capability index plan which meets the requirements from Task1,Task2 and Task3. The final capability index plan is shown in the following table:

Table 3. A final capability index plan which meets the requirements from Task1,Task2 and Task3

Top-capability	Sub-capability	Performance Index	Index type	Aggregation Model	Value
ASADSoS Capability	Reconnaissance and Early-warning Capability	Target-identified Types(kind)	Benifit	IUAM	[20,35]
		Detecting Distance(km)	Benifit	IUAM	[10,55]
	Information processing capability	Information Capacity1 (KB/sec)	Benifit	IUAM	[336,400]
		Information processing Delay(sec)	Cost	IJAM	[60,120]
	Command and Control Capability	Information Capacity2 (KB/sec)	Benifit	IUAM	[336,400]
		Decision Time (sec)	Cost	IJAM	[120,320]
	Firepower Strike Capability	Ammunition Consume (ton)	consumptive	IPAM	[300,650]
		Shot Distance(km)	Benifit	IUAM	[10,60]
		Low altitude combat level	Qualitative	MAM	[0.8,1]

5 Conclusions

The problem studied above is an important part during in the weapon system of systems (WSoS) capability requirement demonstration and analysis. In the study, a Capability Encapsulation method was introduced to define capability as the driven force of Meta-activity, which combines the capability and operational task. The new method makes the mapping process between tasks and capabilities more clear, differs from existing capability requirements analysis method which usually are based on qualitative ways. Based on this, Capability Aggregation methods of capability requirement index are proposed according to different types of capability index. Hence, four types of CRTAM model are built in this study. They are Maximization Aggregation Model (MAM), Interval Union Aggregation Model (IUAM), Interval Join Aggregation Model (IJAM), Interval plus Aggregation Model (IPAM).

The method is validated by a case study of ASADS system. The case proved that the method is available for design WSoS CIRs. The evolution and complexity of WSoS lead to the uncertainty of CR in WSoS. But, the clearer the value of CR is, the better effects the built system has while a WSoS designed. A Capabilities Indices Requirements List (CIRL) under the multi-tasks condition is gained to give helpful guidance on development of WSoS for the decision-makers.

References

[1] Maier, M.W., "Architecting Principles for Systems-of-Systems", Systems Engineering, vol 1, No 4, pp 267-84, 1998.

[2] Sage, A. P., and Cuppan, C. D., “On the Systems Engineering and Management of Systems of Systems and Federations of Systems,” Information, Knowledge, Systems Management, Vol. 2, No. 4, 2001, pp. 325-45.

[3] DeLaurentis, D.A., "A taxonomy-based perspective for systems of systems design methods," Systems, Man and Cybernetics, 2005 IEEE International Conference on, vol.1, No., pp. 86- 91 Vol. 1, 10-12 Oct. 2005.

[4] Boardman, J., Pallas, S, Sauser, B. J. and Verma, D.,2006, Report on System of Systems Engineering, Final Report for the Office of Secretary of Defense, Stevens Institute of Technology, Hoboken, NJ. http://www.boardmansauser.com/words/provocations.html [accessed Jun 30, 2011].

[5] Biltgen, P.T., Mavris, D.N., "Capability-Based Quantitative Technology Evaluation for Systems-of-Systems," System of Systems Engineering, 2007. SoSE '07. IEEE International Conference on, vol., no., pp.1-6, 16-18 April 2007.

[6] Ishibuchi H., Tanaka H. Multiobjective programming in optimization of the interval objective function. European Journal of Operational Research, 48: pp.219-225, 1990.

[7] Ramon E. Moore, R. Baker Kearfott, Michael J. Cloud. Introduction to interval analysis. Published 2009 by Society for Industrial and Applied Mathematics in Philadelphia.

[8] Griendling, K., Mavris, D.N., "Development of a DoDAF-based executable architecting approach to analyze system-of-systems alternatives," Aerospace Conference, 2011 IEEE, vol., no., pp.1-15, 5-12 March 2011.

[9] CJCSI 3170.01G: Operation of the Joint Capabilities Integration and Development System, Chairman of the Joint Chiefs of Staff (CJCS), March 1 2009.

[10] Joint Systems and Analysis Group of TTCP, Guide to Capability-Based Planning, Technical Cooperation Program Technical Report, TR-JSATP2-2-2004, 2004, pp. 3-6.

[11] Lizotte M.,Bernier F., Mokhtari M., Couture M.,Dussault G., Lalancette C., Lemieux F.and Lam S.,” CapDEM - Toward a Capability Engineering Process: A Discussion Paper” ,Report No.(s): ADA440003,Dec. 2004, http://handle.dtic.mil/100.2/ADA440003 [accessed Jun 30, 2011].

[12] Zhang Chengbin, Luan Liqiu,Zhang Hongzhou, "Multi-agent technology based design of hierarchy model of groups/teams in Air Defense System," Audio Language and Image Processing (ICALIP), 2010 International Conference on , vol., no., pp.1403-1406, 23-25 Nov. 2010 .

[13] Evans, R.C. 2004. National Air Defense: Challenges, Solution Profiles, and Technology Needs. The MITRE Corporation. Available at http://www.mitre.org/work/tech_papers/tech_papers_04/04_1108/04_1108.pdf. Accessed Jun 30, 2011.

[14] Barcio, B.T., Ramaswamy, S., MacFadzean, R. Barber, K.S., "Object-oriented analysis, modeling, and simulation of a national air defense system," Systems, Man and Cybernetics, 1995. Intelligent Systems for the 21st Century, IEEE International Conference on, vol.5, no., pp.3983-3988 vol.5, 22-25 Oct 1995.

[15] Xudong Miao, Demin Xu, Yongchun Wang, "Intuitionistic Fuzzy Likelihood Methods and Models for Naval Air Defense Situation and Threat Assessment," Fuzzy Systems and Knowledge Discovery, 2009. FSKD '09. Sixth International Conference on, vol.6, no., pp.311-315, 14-16 Aug. 2009.

The port as a system of systems: a System Dynamics simulation approach

Claudia Caballini, Simona Sacone, Silvia Siri
DIBRIS - Dipartimento interscuola di informatica, bioingegneria, robotica e ingegneria dei sistemi
University of Genoa, Italy
claudia.caballini@unige.it, simona.sacone@unige.it, silvia.siri@unige.it

Abstract – *The particular complexity that characterizes a maritime port makes it belonging to the category of a system of systems. In fact, the functioning of a port is made possible thanks to the interaction of numerous subsystems, which in turn depend on other systems. The goal of the present paper is to provide a description of a port as a system of systems and to model and simulate the port rail process in some Italian container terminals, by using the software Powersim Studio, implementing the System Dynamics methodology. "What-If" scenarios have been tested with the objective of underlining the system bottlenecks and proposing improvement suggestions.*

Keywords: system of systems, ports simulation, rail transportation, system dynamics, scenario analysis.

1 Introduction

The concept of system of systems has been in use for about one decade but only recently it has been applied to the domain of logistics and transportation. In this regard, a maritime port – as well as all the logistic nodes – can be considered as a "system of systems", meant as a collection of different systems with their specific goals, and which totally or partially share some resources and capabilities in order to make the whole system working. Numerous definitions and notions of system of systems have been proposed in the literature [1]-[5] and most of them well represent the characteristics of a port.

Systems of systems typically exhibit the behaviors of complex systems; this is again the case of a port that, by its nature, is a system characterized by a high degree of inherent complexity and, as such, requires efficient and effective management methodologies.

The complexity factors of a port are numerous, especially for what concerns the most Mediterranean ports. Such factors are primarily related to:

- the regulatory fragmentation, due to the multiple legal entities and aspects involved in regulating the port and its activities;
- the number of subjects involved, both of public (e.g. Port Authority, Customs, Coast Guard) and private type (e.g. terminal operators, freight forwarders, shipping companies), which have different interests and are often in conflict one with each other;
- the lack of a correlation – in the tenders for port concessions – between the assignment and maintenance of the concession and the achievement of predefined performance objectives. The presence of such correlation would trigger a virtuous circle able to make the throughput increase for productive terminals and to revoke the concession to negligent or unproductive ones;
- the organizational and process complexity of the port cycles, due to the different degree of computerization among the various players, the frequent lack of standard procedures, the bureaucratization;
- the fact that operations are often executed in accordance with not written behavioral practices;
- the relationship between the port system and the local institutional bodies for what concerns the institutional and planning activities, with particular reference to the infrastructures on the port-city interface.

Hence, the managing of the port sector requires suitable methods and technologies to address it and to face properly its criticalities. The paper is organized as follows. In Section 2 the rail port cycle is described, from the modeling and simulation point of view. Section 3 is devoted to the simulation results referred to three Italian container terminals, whereas in Section 4 some concluding remarks are drawn.

2 Modeling and simulation of the rail port cycle

The port system can be seen as composed of two main subsystems (Figure 1): the "truck-ship" cycle and the "train-ship" one, according to the transportation modes with which the goods reach the port or leave it. For these two cycles, two flows must be taken into consideration: goods that arrive by sea and continue by land (import cycle) and goods that are delivered to the port by truck or train and are embarked on a ship (export cycle). A further breaking down of such sub-systems can be done; each of them is in fact given by the interaction of four sub-sub-systems, which are managed by different actors in the port logistics sector: customs procedures, the interface of the port with the quay side (related to the transportation of goods by sea), the interface of the port with the land side (related to the transportation of goods by land) and the handling and storage of goods in the port/terminal.

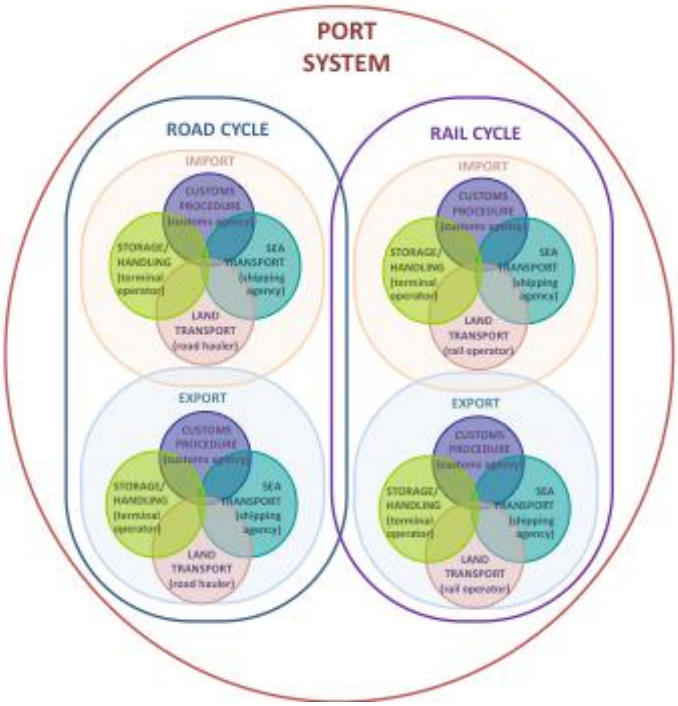


Figure 1. The “system of system” view of a port

Of course, the interaction and synergy of all these subsystems is not merely the sum of their performances and functionalities. Indeed, each of these subsystems constituting the entire system (the port) is itself a complex system. These individual systems are very different and can partially operate independently, however their properties and the kind of their interactions dramatically affect the whole system performance. From all these considerations, it becomes clear that the port is properly considered as a system of systems.

Among the different subsystems constituting the port system, the rail cycle is characterized by a higher complexity, in terms of procedures/processes, actors involved and management required, if compared with the road cycle. The choice of focusing on the port rail transportation is also motivated by the increasing importance that such transport mode has gained in the last decades, thanks to the advantages brought from the environment and congestion points of view. Besides, the increasing volumes passing through the ports impose the growth of traffic flows travelling by train: the possibility of transporting higher volumes of goods by rail allows freeing in a faster way the port areas, so allowing a higher competitiveness of the related port and, consequently, of its throughput.

Given the complexity of the context under study, the simulation methodology has proved to be a suitable tool to study and analyze this system. More specifically, the rail port system has been simulated by utilizing the System Dynamics (SD) methodology, which has been adopted by various authors, as in [6]-[8]. Born and developed in the late '50s thanks to Jay Forrester, professor at Massachusetts Institute of Technology of Boston, SD is a discipline devoted to the study and management of complex systems characterized by nonlinear dynamics, feedback loops and temporal delays. It allows to effectively model complex dynamic systems, in a continuous-time setting, and to analyze their behavior at an aggregate level.

In System Dynamics modeling, the dynamic behavior is due to the accumulation principle, which states that any dynamic behavior of the world occurs when flows accumulate in stocks. So, SD models each system in terms of levels and flows, behind which lie the concepts of integration and differentiation.

In the case presented here, the SD approach imposes to consider the rail port process as a continuous flow of containers/trains, with the final goals of analyzing different scenarios and comparing different management policies. Once having analyzed the processes characterizing the rail port cycle, they have been implemented according to the SD approach for three terminal containers[1]; the conceptual model of the import rail cycle is provided in Figure 2. As it can be seen, the process starts when containers, once unloaded from ships, are stored in the terminal yard (the ship loading process is neglected) and ends when trains depart under the electric line. More in detail, once containers are stacked in the yard, paperwork can be started: the Maritime Agency clearance, possible inspectorate checks and the customs clearance. The net operative times can be affected by some delays, which can vary according to different factors (i.e. the fact that freight forwarders do not start immediately to make the required procedures when containers are unloaded in the terminal or the documents availability). According to the particular customs result, one of the following four flows can be followed by each container: “No check”, “Document check”, “Scanner check” and “Physical check”; to each of them a different time is associated.

The import document operations are the same for the three container terminals analyzed. On the contrary, the physical process of picking up containers in the yard (including possible rehandlings) and loading them on the trains is peculiar of each terminal under investigation, in consideration of the particular layout and organizational modes adopted (in terms of distances, handling means utilized, etc.). For instance, one of the container terminals

[1] Each terminal mainly differs from the other ones in terms of typologies of cargo handled (full and empty, 20’, 40’, etc.), layout, available area, infrastructures, handling means number and internal procedures.

under study uses RMG-Rail Mounted Gantry cranes for loading/unloading containers on trains, while another one utilizes reach stackers. Once trains have been loaded with containers in the internal rail park of the terminal, they are moved through a diesel shunting to the external park where, after a document and physical check they can leave the port under the electric line on the basis of their scheduled rail slots.

The export cycle is specular to the import one. The particular complexity related to the rail cycle is also due to the resource sharing issue. In fact, in addition to the utilization of the same handling means between the import and export cycles, a problem of sharing the same rail infrastructure for both the cycles arises; if we consider that usually, in the majority of container terminals, only one rail track connects the internal rail park of the terminal with the external rail one, the complexity of synchronizing and managing the flow of trains in both directions is evident.

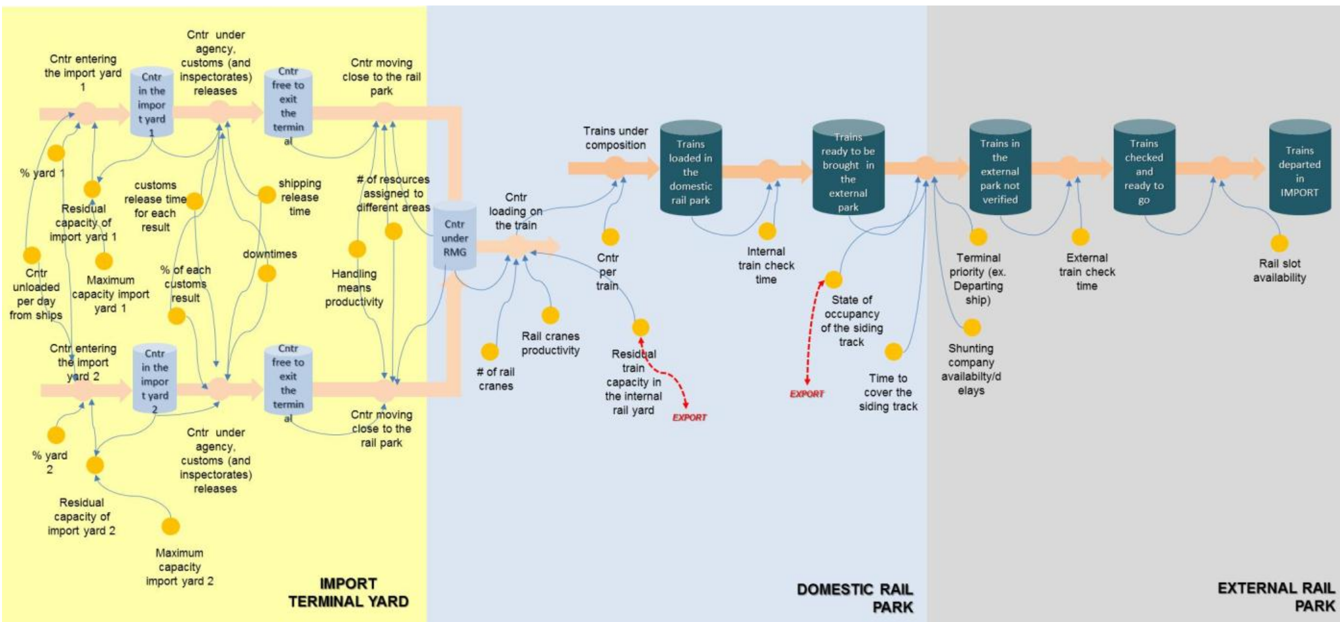


Figure 2. The SD conceptual model of the import rail port cycle

3 Scenarios evaluation and results

The rail port cycles of three container terminals located in Northern Italy have been modeled and simulated by using the Powersim Studio software for a time period of one year. Besides the models have been validated by using available real data.

With the goal of pointing out the bottlenecks of the cycle under study and, consequently, proposing some improvements, the three AS-IS container terminal models have been simulated as well as the following *"what-if"* scenarios:

- "Resources" scenario. This scenario considers the system performances by varying the number of resources available in the terminal, both in terms of moving equipment (rail cranes, trailers or reach stackers) and infrastructural resources (number of tracks in the internal and external rail parks, and number of crossing tracks connecting the two parks).
- "Technological" scenario. Considering the negative impact of the diesel shunting in terms of times and costs, this scenario assumes to electrify the root of the internal rail tracks, in order to completely eliminate the shunting activity.
- "Dry port" scenario. The possibility of disposing of a dry port to free the port areas is considered. This scenario has been investigated only for one of the terminal containers under analysis.

The Key Performance Indicators (KPIs) considered for all the above mentioned scenarios are the following:

1. dwell time (in import and export), that means the average time spent by containers in the terminal yard;
2. cycle time (in import and export), that means the average time spent by containers in the terminal, considering their stop both in the yard and on the rail wagons before leaving under the electric line;
3. trains left (import);

4. trains arrived (export).

The simulation outputs of the AS-IS configuration have shown that all the terminals are able to manage the trains scheduled (rail slots bought). Moreover, in order to understand the current potentiality of the system, the three terminals' models have been simulated without the constraint on the number of incoming and outgoing railway slots, so obtaining the maximum number of trains in import and export that the various terminals would be able to perform with their current resources (means and infrastructures), process times and delays. The simulation outputs have shown that all the terminals do not saturate their rail capacity: this means that a higher number of trains per day could be realized, both in import and export (Table 1).

Table 1 - Annual maximum potential vs. current rail throughtput for the various terminals

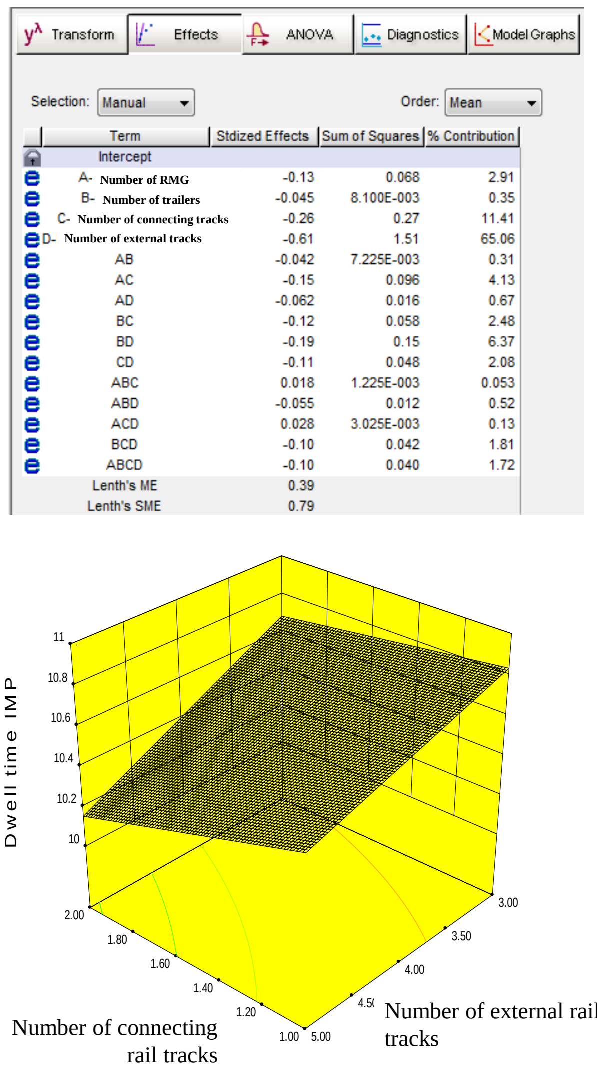
	Terminal 1	Terminal 2	Terminal 3
Annual maximum capacity (trains)	**4316**	**4809**	**2950**
Daily maximum capacity (trains)	11,82	13,18	8,08
Dwell time IMP (day)	3,12	1,99	3,57
Cycle time IMP (day)	3,98	2,2	3,8
Dwell time EXP (day)	2,56	2,15	2,90
Cycle time EXP (day)	2,65	2,76	3,2
Number of actual trains per year	**1374**	**1299**	**1560**
Number of actual trains per day	3,76	3,56	4,27
Dwell time IMP (day)	10,66	6,52	6,72
Cycle time IMP (day)	12,30	7,65	7,45
Dwell time EXP (day)	9,24	8,39	6,1
Cycle time EXP (day)	9,51	8,67	6,33
% of rail capacity utilization	**31,84%**	**27,01%**	**52,88%**

It should also be noted that, since the number of containers in the yards remain constant, for both the phases of import and export, the dwell time and cycle time will decrease. This is possible because the constraint of respecting a particular rail slot determines a slowing down of the flow of containers; in other words each train has to stay in the rail parks waiting for its slot to leave the port, so delaying all the cycle upstream operations.

3.1 "Resources" scenario

Let us now analyze the "resources" scenario for one of the terminal under study. By increasing the number of resources and keeping constant the number of import and export rail slots (that means binding the number of trains entering and leaving the port), the terminal is obviously able to handle the number of trains scheduled; in fact, increasing the resources available, the number of trains can only remain constant or increase.

In order to understand which of the independent variables play a major role in the variation of the KPIs, numerous analyses with the help of the Design Expert software have been carried out. As regards the import cycle, Figure 3 shows that, for one of the three terminals, the number of rail tracks in the external rail park and the number of connecting rail tracks represent the most significant variables affecting the containers dwell time, with a contribution of 65,06% and 11,41% respectively, so representing the infrastructural bottlenecks of the considered terminal. The interaction "number of trailers-number of external tracks" has a low impact, affecting the cycle time for the import of 6,37%. The results are similar if we consider the cycle time.



Term	Stdized Effects	Sum of Squares	% Contribution
Intercept			
A- Number of RMG	-0.13	0.068	2.91
B- Number of trailers	-0.045	8.100E-003	0.35
C- Number of connecting tracks	-0.26	0.27	11.41
D- Number of external tracks	-0.61	1.51	65.06
AB	-0.042	7.225E-003	0.31
AC	-0.15	0.096	4.13
AD	-0.062	0.016	0.67
BC	-0.12	0.058	2.48
BD	-0.19	0.15	6.37
CD	-0.11	0.048	2.08
ABC	0.018	1.225E-003	0.053
ABD	-0.055	0.012	0.52
ACD	0.028	3.025E-003	0.13
BCD	-0.10	0.042	1.81
ABCD	-0.10	0.040	1.72
Lenth's ME	0.39		
Lenth's SME	0.79		

Figure 3. Dwell time trend by varying the number of connecting and external rail tracks (and maintaining constant the other variables)

However, it should be noted that the change in the import dwell and cycle times by varying the different independent variables is marginal; in fact, as shown in Figure 3 (graph below), the dwell time varies from about 10,1 to 10,9 days. This indicates that the increase in terminal infrastructural assets is not sufficient to significantly reduce the dwell and cycle times for the containers handled by rail transport: to do that is in fact necessary to act on the process times and, in particular, on the relative delays, for instance by parallelizing operations and/or by changing the organizational structure of the transport cycle (Table 2).

Table 2 – KPIs values by varying the decision variables in the case of no rail slots constraint

	# of SIMULATION	1	2	3	4	5	6	7	8	9	10
decision variables	# RMG	1	2	1	1	1	1	1	1	1	2
	# connecting rail tracks	1	1	2	1	1	1	1	1	1	2
	# binari terminal	5	5	5	8	5	5	5	5	5	8
	Maritime Agency time delay (day)	0,75	0,75	0,75	0,75	0	0,75	0,75	0,75	0	0
	Customs time delay (day)	0,75	0,75	0,75	0,75	0,75	0	0,75	0,75	0	0
	freight forwarder time delay (day)	1,5	1,5	1,5	1,5	1,5	1,5	0	1,5	0	0
	time for train composition (hr)	2	2	2	2	2	2	2	1	1	0
KPIs	dwell time IMP (day)	3,57	3,57	3,55	3,57	3,10	3,57	3,09	3,55	2,33	2,23
	cycle time IMP (day)	3,8	3,8	3,75	3,8	3,26	3,8	3,25	3,75	2,6	2,4
	trains left IMP	**2950**	**2950**	**2971**	**2950**	**3400**	**2950**	**3410**	**2970**	**4515**	**4720**
	dwell time EXP (day)	3,57	3,57	3,54	3,57	3,10	3,57	3,09	3,55	2,33	2,23
	cycel time EXP (day)	3,8	3,8	3,75	3,8	3,2	3,75	3,4	3,75	2,5	2,41
	trains arrived EXP	**2951**	**2953**	**2972**	**2951**	**3401**	**2951**	**3412**	**2971**	**4515**	**4720**

In fact, by analyzing the results in Table 2 (obtained as the average of a large number of simulation replications), the simulation scenario number 9, corresponding to a reduction of the main time delays, is the one that allows the highest increase in potential trains made in one year (4515).

Analogously for the import cycle, the same kind of assessments have been made for the export one and for all the other two container terminals, but in all the cases it emerges that the greatest benefits are achieved through a reduction in time delays. This indicates that the main bottlenecks of the systems are to be found in the organizational procedures.

3.2 "Technological" scenario

The technological scenario proposes to carry out a performance analysis of the terminals under consideration by assuming that the root of the internal rail parks is electrified, allowing to avoid the diesel shunting and, consequently, to speed up times.

Table 3 shows the results obtained, in terms of trains potentially carried out in the case of elimination of the shunting operation, dwell times and cycle times for import and export.

In all the three cases, the electrification leads to an increase in the potential rail capacity albeit in different entity for the three terminals (Table 3). However, it must be emphasized that the electrification imposes a stricter management and organization of the terminals: the external rail park, in fact, represents an important buffer also for managing emergency situations or criticalities.

3.3 "Dry port" scenario

The last scenario takes into consideration the presence of a dry port towards which it is possible to forward, in a quick way, containers from/to the port, so freeing the port areas. This is crucial especially for ports that are strongly embedded in the city fabric, so having strong limitation in their physical expansion, particularly necessary during freight volumes growth periods.

Of course a dry port can be considered as such if, in its areas, there can be realized all the activities normally carried out in a port: the scenario therefore considers the possibility of making the customs clearance in the dry port instead that in the port spaces.

Table 3 highlights the different results, in terms of maximum rail capacity and process times for the second and third scenarios. As it can be noticed, the "dry port" scenario has been evaluated only for the first terminal container, being the only one to be interfaced with such an inland terminal.

Table 3 – Maximum rail capacity and process times in the cases of electrification and utilization of a dry port

	Terminal 1	Terminal 2	Terminal 3
Annual maximum capacity (trains) WITH ELECTRIFICATION	**5723**	**5785**	**3320**
Dwell time IMP (day)	2,61	1,57	3,17
Cycle time IMP (day)	2,85	1,72	3,25
Dwell time EXP (day)	2,3	2,45	2,62
Cycle time EXP (day)	2,43	2,69	2,8
Annual maximum capacity (trains) WITH DRY PORT	**5483**	-	-
Dwell time IMP (day)	15,02	-	-
Cycle time IMP (day)	3,12	-	-
Dwell time EXP (day)	3,98	-	-
Cycle time EXP (day)	2,56	-	-

4 Conclusion

The present paper is devoted to the analysis of the port system, intended as a system of systems, considering the complexity of the subsystems composing it. In particular, the focus is put on the rail cycle, which has been assuming a stronger importance in the last decades thanks to its benefits in terms of environment pollution and congestion.

The rail cycle of three terminal containers have been analyzed by using the System Dynamics approach; moreover, three different scenarios have been implemented and tested with the goal of pointing out the bottlenecks of the systems and the potential capacity of the terminals in terms of trains to/from the port.

The obtained results highlight that, considering the strong impact of the diesel shunting times and costs over the whole operating rail port cycle, it is first necessary to provide for increasing the efficiency of railway shunting operations. From the infrastructural point of view, this can be made possible by rationalizing the tracks routes within the port, reducing crossings and extending the length of tracks so allowing longer trains.

From the organizational and processes standpoint, the attention must be focused on the lack of a proper synchronization and coordination among the various actors of the rail port chain, a weak process computerization and ICT integration. As a matter of fact, currently delay times, especially the ones related to the informative/document flow, play a determining role on the efficiency and effectiveness of the whole transportation cycle.

In general it emerges that an enhancement of the production factors (infrastructures and handling equipment) provides a limited benefit in terms of timing process unless accompanied by a restructuring of the entire system from the work organizational point of view. This implies, among other things, a detailed schedule of activities and workloads jointly made by the various actors involved (terminal operators, rail companies, shipping companies, etc.), so as to minimize the inefficiencies, the establishment of operational procedures aimed at streamlining the cycle, the identification – in advance and with certainty – of the land mode (road or rail) with which containers will be transported in the territory.

In summary, an integrated systems approach, both technological and organizational, is indispensable to enhance the efficiency of the rail port cycle, so increasing the flow of goods moved by rail.

Further research will be devoted to simulate the processes of the same terminal containers by using the Discrete Event Simulation methodology, so allowing to make a comparison among the results pointed out by the two different approaches and so making further evaluations.

References

[1] N. Karcanias and A. Hessami, "Complexity and the notion of System of Systems: Part (I): general systems and complexity" Proc. World Automation Congress, pp.1-7, 2010.

[2] N. Karcanias and A. Hessami, "Complexity and the notion of System of Systems: Part (II): defining the notion of System of Systems," Proc. World Automation Congress, pp.1-7, 2010.

[3] D. DeLaurentis and W. Crossley, "A Taxonomy-based perspective for System of Systems design methods", Proc. 2005 IEEE International Conference on Systems, Man and Cybernetics, pp. 86-91, 2005.

[4] A. Mostafavi, D. Abraham, D. DeLaurentis, J. Sinfield, "Exploring the dimensions of systems of innovation analysis: a System of Systems framework", *IEEE Systems Journal*, Vol.5, No.2, 2011.

[5] J. Simpson and C. Dagli, "System of Systems: power and paradox", Proc. IEEE International Conference on System of Systems Engineering, pp.1-5, 2008.

[6] M.Sha and X. Huang, "A system dynamics model for port operation based on time, quality and profit", Proc. International Conference on Logistics Systems and Intelligent Management, pp. 1669-1673, 2010.

[7] Briano E., Caballini C., M.Mosca, R.Revetria, "Modelling of Voltri Terminal Europe in Genoa Using System Dynamic Model Simulation", Proc. System Science and Simulation in Engineering, 2008.

[8] Munitic A., Bilic M., Dvornik J, "System Dynamics simulation model of the port-transhipment system manager expert-logical complex functions: sklok, sklow and sklos" Proc. International Conference on Simulation and Multimedia in Engineering Education/ Communication Networks and Distributed Systems Modeling and Simulation, pp.83-87, 2003.

Freeway networks as Systems of Systems: an event-triggered distributed control scheme

A. Ferrara, A. NaiOleari
Dipartimento di Ingegneria Industriale e dell'Informazione, University of Pavia, Italy
antonella.ferrara@unipv.it
alberto.naioleari01@universitadipavia.it

S. Sacone, S. Siri
Department of Communications, Computer and Systems Science, University of Genova, Italy
simona.sacone@unige.it
silvia.siri@unige.it

Abstract -*The objective of the present paper is the design of a control scheme for effectively managing congestion phenomena in freeways and interurban roadways. Such systems are typically made of several road stretches connected to compose a network and, of course, the dynamic behavior of traffic in each stretch influences the state of traffic in the overall network. In this sense a freeway network can be considered and analyzed as a "System of Systems". A distributed Model Predictive Control Scheme in which clusters of freeway cells are separately regulated is proposed in the paper. Moreover, the definition of an event-triggered scheme is included in the paper as well: in the proposed scheme the control action is not computed at each time instant as in the classical Model Predictive Control framework, but only when the system state fulfils specific conditions,.*

Keywords: Traffic control; distributed Model Predictive Control; event-triggered control.

1 Introduction

This paper deals with the design of a control scheme for managing congestion phenomena in freeways and interurban roadways. Such systems can be decomposed in several connected road stretches: the dynamic behavior in each stretch influences the state of traffic in the other stretches of the network. For this reason, a freeway network can be considered and analyzed as a “System of Systems” [1], [2].

Congestion phenomena in freeway networks cause not only the increase of travelling times, fuel consumption and hence pollution, but also the increase of car accident probability. Then, an effective management of these critical conditions improves the system level of service but also its security level and its environmental impact. One of the main causes of congestion is that the overall transportation demand is almost equal or even higher than the capacity of the available infrastructures. For this reason, the efficient utilization of the road system becomes more and more important and, to this end, the development of suitable modelling and control methods is crucial. In order to prevent and solve congestion phenomena in freeway networks, different traffic control measures have been proposed and implemented, such as ramp metering, variable speed limits, route guidance and vehicle-infrastructure integration systems [3].

The control scheme considered in this paper is based on the use of ramp metering as the adopted control action. Ramp metering, i.e. the regulation of traffic flows entering the freeway mainstream via the on-ramps, has been in use for some decades and it has been recognized as an effective way to reduce freeway congestion. The different ramp metering control schemes found in the literature can be distinguished between those relying on a second-order macroscopic traffic model, as for instance in [4], [5], and those based on the first-order cell transmission model, as [6].

In the present paper the CTM model will be used and reformulated as a mixed logical dynamical system, i.e. a system described by linear dynamic equations subject to linear inequalities involving both continuous and binary variables, according to the framework proposed in [9]. Based on this model, a Model Predictive Control(MPC) scheme is developed in which the considered finite-horizon optimal control problem has the objective of quadratically penalizing the deviation of the decision variables from a specific equilibrium point.

In [10], the authors have followed the same direction and some stability properties of the resulting control scheme have been discussed. Specifically, it has been proved that the control scheme yields the stability of the system when no disturbance is present and it guarantees the input-to-state stability of the system when it is affected by bounded disturbances. In the same paper, the MPC scheme has also been redesigned as an event-triggered control scheme in order to reduce the computational load of the proposed framework. Specifically, the event-triggered scheme is based on the fact that the control action is not computed at each time instant as in classical MPC schemes, but only when the system state fulfils a specific condition denoted as triggering condition. The concept of event-triggered strategies is quite new in the literature and is motivated by the difficulty, in some cases, to apply equidistant sampling for feedback control, due to limited communication resources or inadequate computation power. Instead, in event-driven control systems the sampling is event-triggered rather than time-triggered.

In this paper, the approach proposed in [10] is extended in two ways: first of all, clusters of freeway stretches, that is, sets of subsequent freeway cells, are separately modeled and controlled. Of course, in this case suitable communication mechanisms between cells are required in order to make them cooperate to achieve some desired performance of the overall freeway system. The second innovative aspect of the present paper with respect to [10] regards the definition of different triggering conditions to be used in the event-triggered MPC scheme. The effectiveness of the proposed approach will be shown in the paper with an extended simulation campaign based on real data.

The paper is organized as follows. Section 2 described the proposed event-triggered distributed control scheme. Section 3 reports some computational results, whereas some conclusive remarks are drawn in Section 4.

2 The proposed approach

This section is devoted to the proposed event-triggered distributed control scheme; firstly, the MPC framework is described, while the definition of the event-triggered approach and the distributed control architecture are presented later on.

2.1 The MPC scheme

The proposed approach adopts Model Predictive Control based on the first order cell transmission model (CTM) for the prediction. This model, introduced by Daganzo in the Nineties, is based on the subdivision of the freeway into cells with homogeneous traffic features and on the discretization of the time horizon(see [7] and [8] for a detailed model description). Let T denote the sample time and Δdenote the length of cells. Moreover, referring to each time step$k, k = 1, \dots, K$,let us define the following variables:

- $\rho_i(k)$ is the traffic density of cell i, in vehicles per space unit;
- $q_i^+(k)$ is the flow rate, in vehicles per time unit, entering cell i;
- $q_i^-(k)$ is the flow rate, in vehicles per time unit, exiting cell i;
- $r_i(k)$ is the flow, in vehicles per time unit, entering the freeway from the on-ramp between cell$i-1$ and cell i;
- $l_i(k)$is the queue length, in vehicles, in the on-ramp between cell $i-1$ and cell i;
- $d_i(k)$is the on-ramp demand, in vehicles per time unit, referred to the on-ramp between cell $i-1$ and cell i;
- $s_i(k)$ is the flow, in vehicles per time unit, of vehicles leaving the freeway from the off-ramp between cell $i-1$ and cell i.

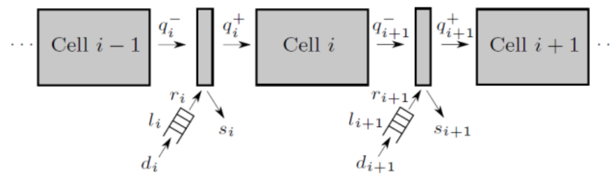


Figure 1. Sketch of the division of the freeway into cells.

The CTM model (given by dynamic discrete-time equations representing the time evolution of the state variables, i.e. the traffic density and the queue lengths) is nonlinear, due to the presence of some minimum functions in the equations. However, it can be seen as a mixed logical dynamical (MLD) system, i.e. described by linear dynamic equations and inequalities involving both continuous and binary variables, as proposed in [9]. In order to linearize the nonlinear equations, it is necessary to introduce some sets of auxiliary variables and some inequalities (refer to [10] for the mathematical details). Once the model has been linearized, the cell transmission model is defined in the form of a MLD system in which $\rho_i(k)$ and $l_i(k)$, $i = 1, \dots, N, k = 1, \dots, K$, are the state variables and$r_i(k)$, $i = 1, \dots, N, k = 0, \dots, K-1$, are the control variables.

As shown in [11], studying the behavior of the cell transmission model of a freeway, for each feasible demand and each corresponding on-ramp volume, there is a unique uncongested equilibrium in which free flow speed prevails in all sections. Let us denote with $\rho_i^e$the equilibrium density corresponding to the on-ramp volume valuer_i^e, for cell $i, i = 1, \dots, N$. The main difference between the model proposed here and the one reported in [11] is the presence of the queue lengths as state variables. Such state variables admit an equilibrium point in $l_i^e, i = 1, \dots, N$, which can be simply computed by considering the same feasible demands and on-ramp volumes$r_i^e, i = 1, \dots, N$. Analogous equilibrium values can be defined for the auxiliary variables.

On the basis of such considerations, a finite-horizon optimal control problem to be solved at the generic time step k, over a prediction horizon K_p,can be stated.The overall statement of the problem is not reported here for space limitation but the main aspects regarding the finite-horizon optimal control problem are briefly reported in the following. Given the initial conditions on the density and the queue, the flow entering the first cell, the flow exiting the last cell, the density measured before the first cell, the on-ramp demands and the flows exiting from off-ramps, this problem corresponds to finding the optimal control variables $r_i(h)$, $i = 1, \dots, N$, $h = k, \dots, k + K_p - 1$ in order to minimize a given cost function and to satisfy some constraints. Specifically, the objective function of this problem quadratically penalizes the deviations of the state, control and auxiliary variables from their corresponding equilibrium values, whereas the problem constraints include the systems dynamics (CTM model in the MLD form), as well as some lower and upper bounds on the state and control variables.

It is worth noting that this problem is a mixed integer quadratic programming problem that can be optimally solved by deriving, at each time step$k$, the optimal control sequence variables $r_i(h)$, $i = 1, \dots, N, h = k, \dots, k + K_p - 1$. In a classical MPC scheme, only the first element of the control sequence (that is, $r_i(k)$, $i = 1, \dots, N$) is applied to the system, and then the overall optimization procedure is repeated.

The solution to the finite-horizon optimal control problem requires the knowledge/estimation of a set of problem data relevant to the traffic flows entering the first cell and exiting the last cell, the on-ramp demands and the flows exiting through off-ramps. If such quantities are supposed to be known, it is possible to exploit the same reasoning lines already adopted in [9] and, then, to guarantee that the proposed control scheme makes the considered equilibrium state stable. Anyway, the assumption that the problem data are known or exactly estimated is not always realistic since quantities like the input flow of the first cell, the output flow of the last cell, the on-ramp demands and flows exiting from off-ramps are typically matter of thorough estimation procedures, but the resulting estimates are not always completely effective. Then, it is more proper to consider such quantities as bounded disturbances affecting the system behaviour. In this case, a further stability result referred to the input-to-state stability (ISS) property of the proposed MPC scheme can be formalized. Some more details on such stability properties can be found in [10].

2.2 The event-triggered control scheme

The MPC scheme theoretically represents a suitable way of deriving a ramp metering strategy for freeway stretches. The main drawback of the proposed approach regards the computational time required to solve the finite-horizon optimal control problem that, in the worst case, depends exponentially on the number of integer variables. Such a number depends on the number of cells considered and on the length of the control horizon K_p. To determine the control law in acceptable computation times it is necessary to find a compromise between having a good prediction (requiring a large control horizon) and considering a small number of integer variables.

A possible way to overcome this drawback stands in including the proposed control law into an event-triggered scheme in which the control law is not updated at each time step, but whenever the system state meets a predefined set of conditions, named *triggering conditions*. By calling *triggering time steps* the time intervals in which the triggering rule is met, the control law is defined in the following way. At $k = 0$, the finite-horizon optimal control problem is solved determining the control sequence and applying it in the subsequent time steps. At each time step $k, k = 1, \dots, K - 1$, the triggering conditions are verified; if they are not met, the already available control action is applied, otherwise time interval k becomes a triggering time step, the defined finite-horizon optimal control problem is solved and a new optimal control sequence is derived. The values of the control actions composing such a control sequence are applied to the system until the next triggering time step.

The triggering conditions can be of different types. In this work we consider, first of all, a triggering condition related to the change of operating condition in a given cell in a given time step, i.e.a regular cell becomes congested or viceversa. Moreover, other triggering conditions are related to the density error and the queue length error, i.e. to the absolute values of the differences between the real values of the traffic density and the on-ramp queue length and the corresponding predicted values found by solving the finite-horizon optimal control problem. Moreover, percentage deviations of the density (and the queue length) in a time step with respect to the previous time step, as well as the queue length in the on-ramps are considered to define appropriate triggering conditions. In Section 3 the triggering conditions adopted for the experimental tests will be described in detail.

2.3 The distributed control scheme

In order to face the computational complexity of the proposed control scheme, especially in case of large freeway networks, a distributed scheme is proposed. Basically, the idea is to divide the whole freeway section in some (cooperative) clusters of cells, to be controlled locally with the event-triggered control scheme described above. A simple sketch of this division is reported in Figure 2. The division of the freeway into clusters of cells makes the considered system associated with the class of “Systems of Systems”: in fact, each cluster is characterized by a specific dynamical model and by a given controller. Of course, the different clusters interact and possibly must be collaborative in pursuing the same objective, that is the minimization of congestion phenomena along the whole freeway.

The distributed control framework must be properly devised according to the considered system. First of all, an important aspect to be decided is the number and dimension of the clusters. A too high number of clusters

(clusters with few cells) has the computational advantage of "small" optimal control problems to be solved in each cluster but has the disadvantage of a too fragmented system. On the other hand, a small number of clusters can provide a large computational burden but may be more effective in the prediction of the dynamic behavior of the system.

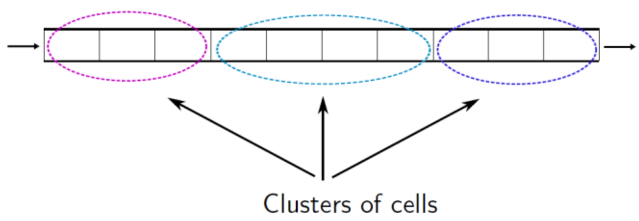


Figure 2. Sketch of the division of the freeway into clusters.

Another important aspect regards whether the clusters are a priori established or dynamically defined, depending on the traffic conditions in the freeway. The first option is of course easier to be adopted but could be not effective in capturing some dynamic traffic phenomena. The second option can be very complicated, because one possibility is to vary the clusters so that they contain at most a congestion front but this is not the only possible configuration. Therefore, the definition of clusters in a dynamic way requires an accurate study of the overall system.

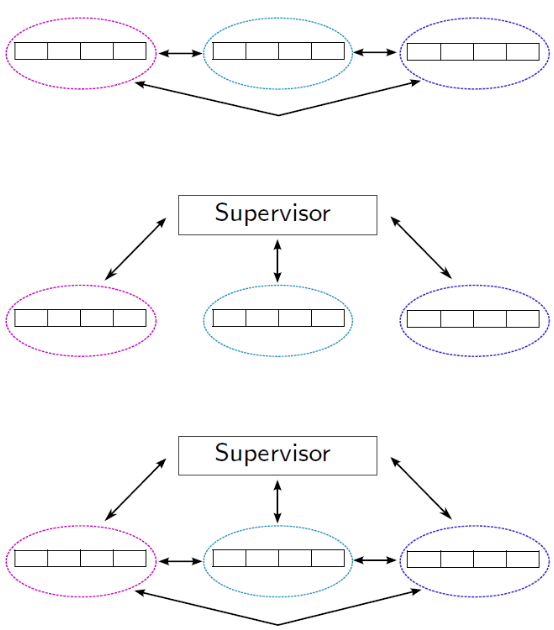


Figure 3. Sketch of different decentralized control schemes.

Finally, different decentralized schemes can be devised according to the way the clusters communicate. Three main schemes can be devised (for a sketch see Figure 3):

- all the clusters communicate with each other in order to find the "best" control action;
- there is a supervisor and the communication is possible only between a cluster and the supervisor; once received the information from the clusters, the supervisor chooses the "best" control action;
- there is a supervisor and all the communication channels are active, i.e. all the clusters communicate with each other and also with the supervisor; this is the most complicated, but probably the most effective, control scheme to be applied.

3 A real case study: some results

In order to demonstrate the effectiveness of the proposed control scheme, we defined a set of triggering conditions and we tested them on the model of a real freeway system. The model represents 5 km of a real highway, including one on-ramp and one off-ramp. In order to take into account the limited capacity of the system, we assume that the entering flow is controlled by a tollgate. The prediction model is a CTM in MLD form, while the simulation model is a complete nonlinear CTM.

The considered triggering conditions to be defined at a generic time step k are the following:

1. the density error in k (deviation of the actual values from the ones predicted by the MPC) is greater than a given threshold;
2. the queue length error in k(deviation of the actual values from the ones predicted by the MPC) is greater than a given threshold;
3. the difference between the present density value (in k) and the previous one (in $k-1$) is greater than a given threshold;
4. the difference between the present queue length value (in k) and the previous one (in $k-1$) is greater than a given threshold;
5. at least one cell, that was uncongested in the previous time instant (in $k-1$), becomes congested in k;
6. the queue length in k exceeds a maximum value.

It is worth noting that when the triggering condition is true, the finite-horizon optimization problem is solved again, which implies that the control vector is updated. This vector is used as a buffer when the triggering condition is false. When all the components of the buffer have been used, the control law is computed again, regardless of the triggering condition.

In order to evaluate the performances of the controlled system with the different triggering conditions, let us define an index to evaluate the computational load, i.e.,

$$Q(t_f) = \int_{t_0}^{t_f} N(\tau)\, d\tau \qquad (1)$$

where $N(\cdot)$is the length of the MPC optimization horizon, which, in our case, is a piece-wise constant function equal to 120 s when the control law is computed, and equal to 0 s otherwise, while $[t_0, t_f]$is the time interval in which the control is active. It is clear that this evaluation is based on the simplifying assumption that the computational burden of the MPC is strongly dependent on the number of timesthe optimization problem is solved and on the horizon length.

We also define an index of the overall performances of the system, as, i.e.,

$$\Gamma(t_f) = \int_{t_0}^{t_f} J(\tau)\, d\tau \qquad (1)$$

where $J(\cdot)$is the objective function of the MPC problem.

The initial set points for the system are the components of the uncongested equilibrium state, i.e. the equilibrium values for the control variables, state variables and auxiliary variables are calculated according to [11].

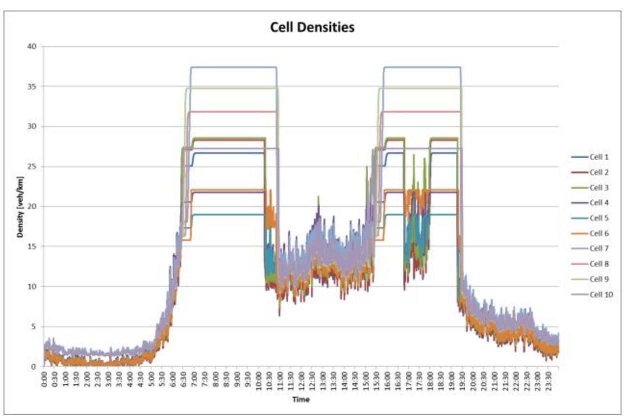


Figure 4. Cell densities versus time in the open loop case overthe 24 hours.

A preliminary open loop simulation, reported in Figure 4, shows that the most congested hours of the day are between 6:30 and 11:00, and between 15:30 and 19:30. We decided to evaluate the indicated triggering conditions between 15:30 to 16:30. In our tests we applied a single triggering condition at a time, to be able to better evaluate the effect of each condition on the overall performances. An example of the application of a triggering condition is shown in Figure 5 for the case of condition (4). The proposed control approach, in comparison with the open loop case, is able to remove the congested situation between 15:30 and 16:30 and to maintain a homogeneous density across the whole highway.Figure 6 shows the same pattern in the case of application of condition (4), only between 15:00 and 17:00.

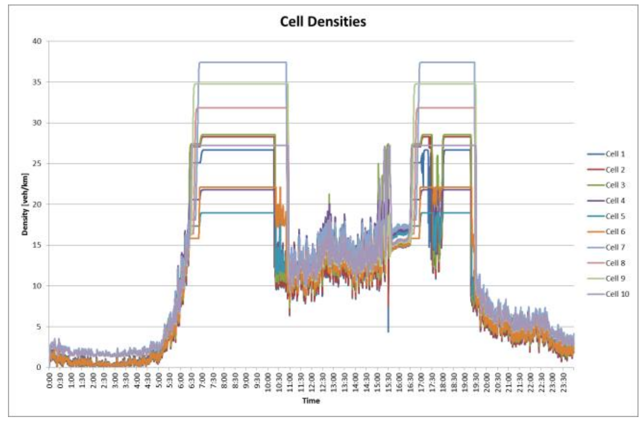


Figure 5. Cell densities versus time in the closed loop case, with condition (4) over the 24 hours.

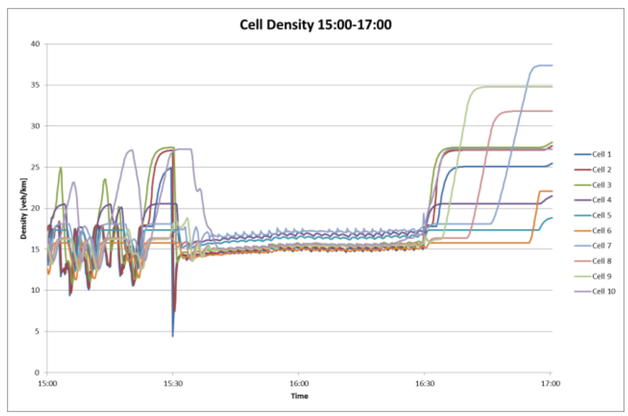


Figure 6. Cell densities versus time in the closed loop case, with condition (4) between 15:00 and 17:00.

By applying all the triggering conditions, it can be seen that the triggering condition (4) minimizes Γwhile conditions (1) and (2) minimize the computational load Q. Condition (5) is never satisfied during the considered time period, which means that the proposed control with condition (5) enables to keep the cells always in the uncongested state, by using the entire control buffer. Figure 7 provides a comparison of the computational cost generated by each triggering condition with the computational cost of the standard MPC control algorithm. We can see that in any case with the proposed approach the computational cost remains muchlower than the one obtained with the standard MPC algorithm, with a maximum reduction, over the interval $[t_0, t_f]$, of about 90%. In order to better capture the difference among the different triggering conditions, Figure 8 reports the same patterns of Figure 7, but without the case of standard MPC.

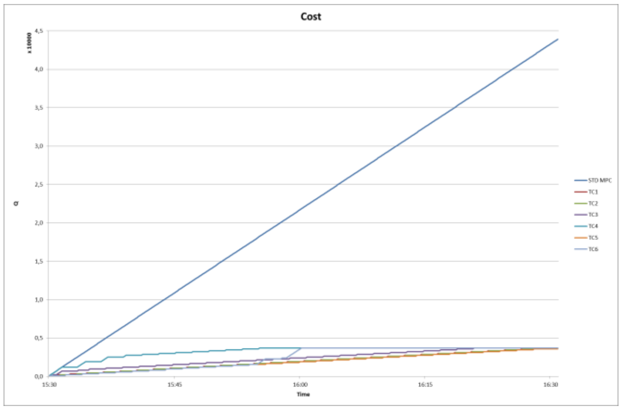


Figure 7. Computational cost with the different triggering conditions and the standard MPC.

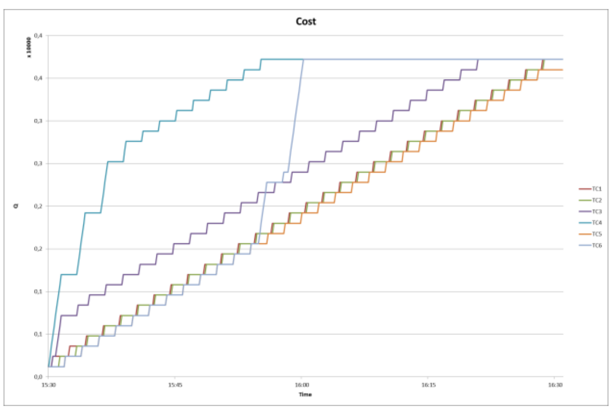


Figure 8. Computational cost with the different triggering conditions.

4 Conclusions

A freeway network is a very complex system, composed of different subsystems, that are interconnected freeway stretches. Moreover, the application of optimal control schemes to such systems can be very demanding, not only from the computational point view, but also as regards the measurements and communication aspects involved. In order to deal with such complicated "System of Systems", some innovative ideas have been proposed in the paper. First of all, an event-triggered concept has been introduced, with different triggering conditions whose effectiveness has been tested through experimental results. Secondly, a decentralized framework has been considered in which the overall freeway system is divided in different clusters of cells that are modeled and controlled separately.

References

[1] N. Karcanias and A.G. Hessami, "Complexity and the notion of System of Systems: part (I): general systems and complexity", Proc. World Automation Congress, 2010.

[2] N. Karcanias and A.G. Hessami, "Complexity and the notion of System of Systems: part (II): defining the notion of System of Systems", Proc. World Automation Congress, 2010.

[3] R.C. Carlson, I. Papamichail, M. Papageorgiou and A. Messmer, "Optimal mainstream traffic flow control of large-scale motorway networks", *Transportation Research Part C*, vol. 18, pp. 193-212, 2010.

[4] M. Papageorgiou, "Freeway Ramp Metering: An Overview", *IEEE Transactions on Intelligent Transportation Systems*, vol. 3, pp. 271-281, 2002.

[5] T. Bellemans, B. De Schutter and B. De Moor, "Model predictive control for ramp metering of motorway traffic: A case study", *Control Engineering Practice*, vol. 14, pp. 757-767, 2006.

[6] G.G. Gomes and R. Horowitz, "Optimal freeway ramp metering using the asymmetric cell transmission model", *Transportation Research Part C*, vol. 14, pp. 244–262, 2006.

[7] C. F. Daganzo, "The Cell Transmission Model: A Dynamic Representation of Highway Traffic Consistent with the Hydrodynamic Theory", *Transportation Research Part B*, vol. 28, pp. 269–287, 1994.

[8] C. F. Daganzo, "The Cell Transmission Model, Part II: Network Traffic", *Transportation Research Part B*, vol. 29, pp. 79-93, 1995.

[9] A. Bemporad and M. Morari, "Control of systems integrating logic, dynamics, and constraints", *Automatica*, vol. 35, pp. 407-427, 1999.

[10] A. Ferrara, A. Nai Oleari, S. Sacone, S. Siri, "An event-triggered Model Predictive Control scheme for freeway systems", submitted for presentation at the 51st IEEE Conference on Decision and Control, 2012.

[11] G. Gomes, R. Horowitz, A.A. Kurzhanskiy, P. Varaiya and J. Kwon, "Behavior of the cell transmission model and effectiveness of ramp metering", *Transportation Research Part C*, vol. 16, pp. 485-513, 2008.

A SelexSI Solution to Enable Distributed Decision Making in Vessels Traffic Management

A. Napolitano, D Di Crescenzo
SESM s.c.a.r.l.
Via Tiburtina, 1238 Roma,
Tel: +39 06 41505769,
e-mail: *anapolitano*@sesm.it

S. Gelli
Selex Sistemi Integrati S.p.a.
Via Tiburtina, 1238 Roma,
Tel: +39 06 41503808,
e-mail: *sgelli@selex-si.com*

Abstract – The surveillance and control of maritime traffic, along with the management of wide area, such as ports and costs, need of the cooperation and coordination of manifold actors. The capacity of efficiently putting in place proper activities allows improving the safety and security of maritime transport forecasting promptly dangerous situations, collisions and threats, such as crossing of the boats close to the coasts or bridges and carrying dangerous goods in protect areas; enhancing maritime traffic for increasing capacity of the harbours, the navigation efficiency, and routes utilization; and supporting the actors during the Search and Rescue phases. The proposed Vessel Traffic Management System attempts to reach the above aims enabling the cooperation decision making among all the actors involved in each activity. In particular, the solution lets on the one hand of sharing information among all actors, improving their situation awareness during all the operative phases, and on the other hand of making decision aids to the operators in both the planning of the mission and in tactical activities. The modularity and scalability gained from the solution architecture allow utilizing the VTMS solution in different and heterogeneous environments, such as national maritime territory, rivers, waterways, etc.

Keywords; monitoring and control, cooperative and distributed decision, vessel traffic management, system of systems.

1 Introduction

The *Vessel Traffic Service* (VTS) [1] [2] stands for the main system actually exploited for allowing surveillance and controlling of maritime traffics in open sea, near the coasts and in harbour area. In this wide environment, the key activities being required to VTS system are: i) improving the safety and security of maritime transport forecasting promptly dangerous situations, collisions and threats, such as crossing of the boats close to the coasts or bridges and carrying dangerous goods in protect areas; ii) enhancing maritime traffic for increasing capacity of the harbours, the navigation efficiency, and routes utilization; iii) monitoring of “Blue Border” for avoiding and detecting illegal activities as drug and weapons smuggling; iv) supporting the actors during the Search and Rescue (SAR) phases.

All these activities involve the cooperation and collaboration of manifold authorities and actors (e.g. Harbour Masters, Coast Guard Operators, shipping company, boat captain, towboat services, SAR units), which need to be suitably coordinated and managed within a wide geographical areas.

In fact, it has often been noted that actions undertaken by isolated units produce domino effect on the overall efficiency of the operations and maritime transport [3]-[5]. This is more evident during the *SAR* [6] and *Blue Border* [7] operations, in which timeliness, promptness, and cooperation in putting in action countermeasures are essential features for succeeding in the mitigation of the threats. The introduction of a distributed decision making approach could play a key role in coping with the aforementioned inefficiencies. The capability of sharing the information among the involved actors allows, indeed, to increase their awareness, and, in turn, taking decisions and doing actions in synergy way one another.

This has an impact on the quality and efficiency of the undertaken countermeasures in several aforementioned activities, which are:

- *Blue Border Control* for supporting the actors (e.g. military navy and local) in decision aid through threats automatic recognition and classification, scheduling of the countermeasures to put in action according with the kind of the threat, coordinating of the trajectories during the countermeasure actions, assessment of the actions, etc.;
- *Search A Rescue* for planning [8] [9] and coordinating the actions among the involved actors (e.g. SAR Coordinators, SAR Mission Coordinators, On Scene Coordinators, On Ship Reporting Systems), also according with their positions, and coordinating, further, the handover among different centre, i.e. between “*local”* (which detects the risk) and “*area*” centre (which orchestrates the countermeasures);
- *Environmental risks detection*: for informing the shipping of dangerous areas, monitoring and supporting in the planning of the new routes to avoid risks.

A novel solution, namely VTMS (Vessel Traffic Management System), for enabling the distributed decision making in vessel traffic management has hereinafter described. VTMS is, indeed, capable of: i) monitoring wide maritime zones through the integration and interoperability of several and heterogeneous sensors and systems; ii)

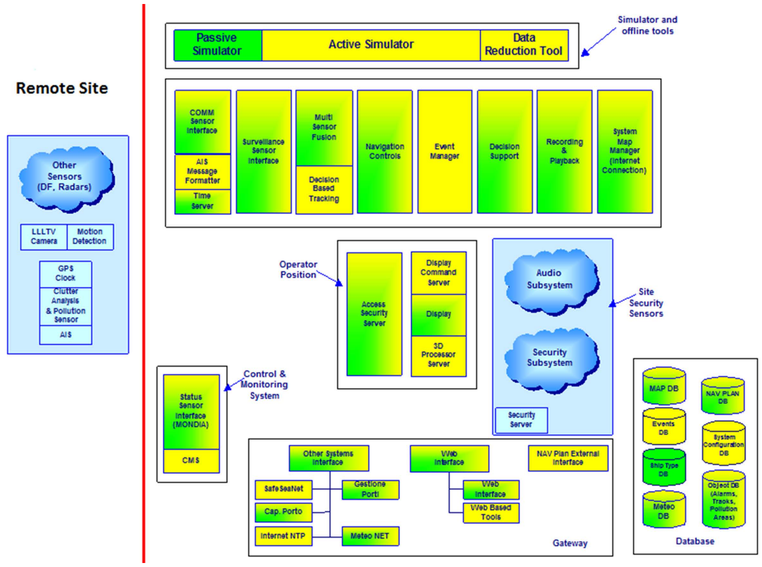


Figure 1 – VTMS Local Centre logical architecture

recognizing cooperative and non-cooperative vessels making use of a distributed multi sensor fusion technique; iii) analysing in proactive way the arising of critical situations and threats; iv) providing proper solutions for increasing the safety and efficiency of the maritime transport while granting maritime environmental protection; v) coordinating SAR operations allowing cooperation of the involved actors (i.e. SAR Coordinators, SAR Mission Coordinators, On Scene Coordinators, On Ship Reporting Systems).

The rest of the paper is organized as follows. Section 2 provides the description of the VTMS, as well as a comprehensive analysis of all the capabilities implemented by the solution. Afterward, Section 3 presents in details an experimental scenario, in which the VTMS has been utilized for showing its ability in providing decision aid to the operator in both planning and tactical phase, as well as execution.

2 The VTMS solution

A scalable and modular solution, namely VTMS, for allowing the surveillance and management of maritime transport, as well as the coordinating and orchestration of the involved actors for coping with threats and dangerous situations in the sea is hereinafter shown.

VTMS proposed solution relies on SOA architectures, which allows adapting itself to the extension of the area needs to be surveillance. Scalable and modularity properties are, indeed, assured form the VTMS solution in order to allow its operation in different contexts and environments. The basic element of the VTMS is the *Local Centre*, whose architecture is depicted in Figure 1. The main *Local Centre* functional blocks can be organized as follows:

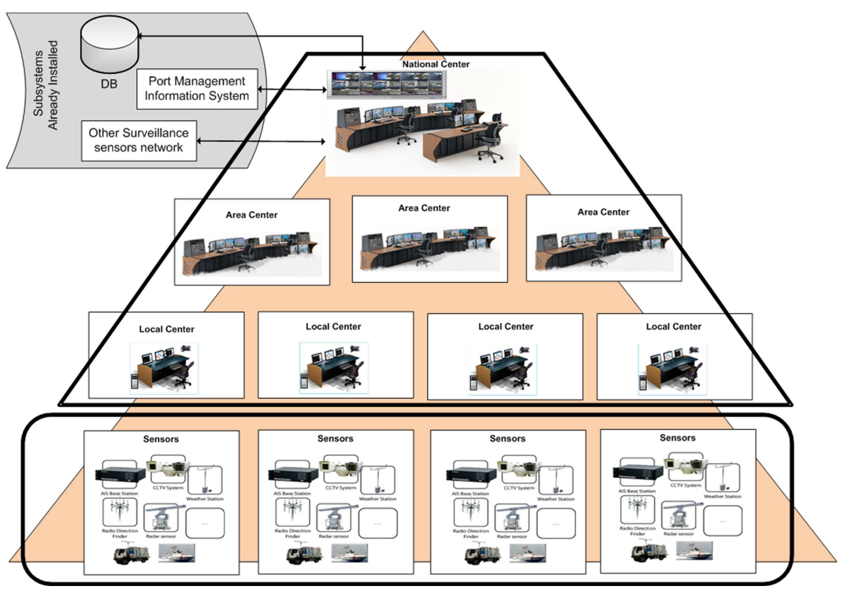


Figure 2 Typical hierarchical configuration of VTMS solution

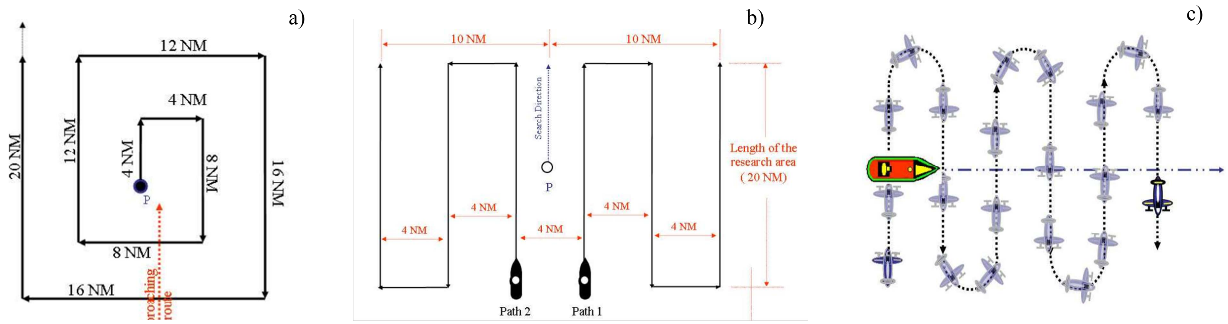


Figure 3 Search path typologies involving : a) single vessel unit, b) more than one vessel unit, and c) vessel and air units

Sensors: external sensors integrated in the system consisting of Radar, AIS, Weather, EOS (Electro-Optical Sub-System), and Direction Finder;

Processing Functions: including interfaces towards the sensors for command and control, communication management (i.e. AIS binary messages), and surveillance (Radar, AIS, DF etc) information. Processing functions also include Multi Sensor Data Fusion, Navigation Control for law enforcement purpose, and Decision Support block;

Auxiliary Functions: Recording and Playback of VTMS data, Map Distribution, and System Monitoring;

Presentation Functions: including user friendly graphical interface for the display of the integrated tactical picture both in 2D and 3D view and archived data;

Database: collecting all the information for statistical analysis;

Gateway: providing secure interface towards external systems or other available information sources on the internet.

Among all, Decision Support module is one of the core processing functions of the VTMS system, providing:

- Anti-pollution facilities to ease environmental protection;
- SAR Coordination functions according to national rules and operational procedures;
- Threat-Interception Function;
- Risk Factor calculation for each vessel according to a configurable Risk Model to allow prevention of dangerous situations.

Stemming from the *Local Center*, a hierarchy system can be built-on, which involves respectively the Area and National Centers, as shown in Figure 2.

In particular, a single Local Center is enough to manage and control small ports, whereas for wider area, such as high ports, many Local Center are needed, which are coordinated through the Area Center. Taking advantage form the modularity, the same strategy is utilized to monitor the naval traffic within a country national area. In this scenario, many Local and Area Centers will be placed on the national territory, and all the information can be managed through the National Center, which gathers all the information.

It’s worth noting that the VTMS solution allows managing and involving into its operations not only naval units but also air (i.e. helicopters and airplanes) and land units.

2.1 Anti-Pollution

Decision Support has the capability of automatically detect Oil Slicks on the sea surface making inference on data achieved by Radar and/or other surveillance systems (i.e. Anti-Pollution systems); manual creation of the polluted area is possible yet. Once the polluted area is detected, the Anti-Pollution functionality is able to continuously track the Oil Slick. Moreover, stemming from the stored data on sea and wind conditions, Slick Kinematics, and naval traffic in that area, it tries to identify the vessel responsible of the contamination. The detection of Oil Slick gives rise to mission planning for facing and mitigating, as best as possible, the emergency. In that phase, the module issues the list of the vessels suitable for coping with to the kind of environmental threat, sorting out by the minimum distance from the Oil Slick. For each of the vessel to be involved in the mission, the module calculates the minimum path to be followed for reaching the area, along with the estimated time of arrival, according with sea conditions.

Once the vessels are engaged for the mission, they routes to the Oil Slick are constantly monitored and the proper operations to be done are grant and coordinated by the system. The end of the mission is either manually or automatically established by the system when the last vessel involved reaches a parking area or a grounding point.

2.2 Search and Rescue (SAR)

Search and Rescue Missions are typically initiated after an SOS message is received and involve several operational procedures and actors that vary from time to time. Once the Emergency Point (i.e. EP) is communicated by the vessel(s) in distress, Decision Support Module continuously calculates the shortest paths to the EP for each involved vessel, and its possible drift according to the type of SAR activity to be performed (i.e. rescue of small boat, floating man, floating motor-scooter, etc.), local meteorological conditions and sea state. Stemming from this information, SAR Decision support module suggests to the operator the vessels, among all the available vessels, including allied means and in-transit, which in best way

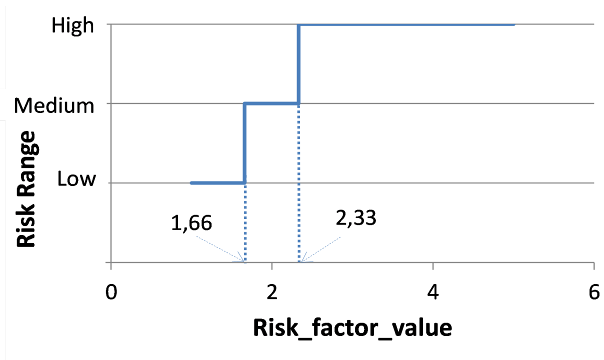


Figure 4 $Risk_{range}$ function

and minimum time can cope with the situation. Once this stage is accomplished, SAR Decision support module, before assigns the responsibility roles, such as On-Scene Coordinator and Local Coordinator, and successively orchestrates all the proper countermeasures to put in action. In particular, the module automatically assigns to each one of the selected vessels a sector of the Search and Rescue area and the searching path to be followed. Three different search paths are suggested according with the number and type of units entailed in the operations. In the case of single naval unit, due to its efficiency the *increased square search method* is put in place (Figure 3.a), in which the vessel moves around the estimated position and after it has cover a distance, it turns by 90 degrees so that it forms a square around the estimated position. Instead, if more than two vessel unit are available, the *increased square search method* is replaced by parallel course search method (Figure 3.b), in which the vessels cover parallel courses next to the EP. When the vessels reach the end of the research area, they come back covering other two parallel courses, in the opposite direction, with a bigger distance from EP. Finally, a coordinate ship-plane method is utilized (Figure 3.c) if the search activities entail also an aircraft. In this scenario, the plane crosses the ship's course, coming alternatively from its left and right side, in order to explore a wider range of sea.

On each search and rescue area, the system automatically active the following navigation controls:

- Speed limit;
- Collision avoidance;
- Transit in prohibit area;
- Anchorage in prohibit area.

2.3 Threat Interception Function

The Threat Interception Function allows to select a vessel as "hostile" (e.g. in anti-smuggling actions) and to orchestrate the hunting missions towards that "hostile" vessel. A vessel is classified as "hostile" if it either infringes the navigation control or grounds without any authorization. The target interception mission is typically based on the optimal route to meet the hostile vessel. To do

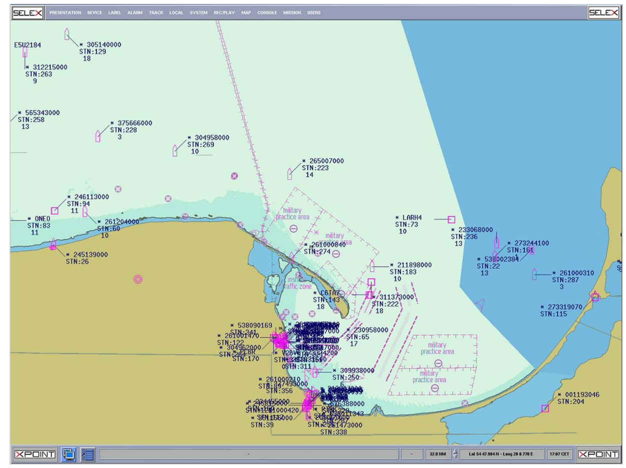


Figure 5 Maritime tactical picture

that, the Threat Interception module supports the operator performing in automatically manner the following operations:

- envisaged route of the target to be hunted will follow;
- list of the suitable units (according with sea and environmental conditions) for copying with the target, ranked by the lower time of engaged;
- generation of a "scenario", i.e. the set of interception trajectories, with the associated points and covering times;
- monitoring of the interception mission outcome. In particular, the target is defined as intercepted when the distance between the target and the interceptor is shorter than defined threshold (Mission Accomplished and Terminated), or the track data of the target to intercept is not received for a certain number of radar scansions (Mission not accomplished and terminated).

2.4 Risk Factor Calculation

Risk Factor Calculation allows estimating the likelihood that a vessel may encounter dangerous situations along its way in the area under VTMS monitoring. The function relies on a risk model identified through a weighted (i.e. c_i) average, expressed by (1), of the following inputs parameters:

- Ship type and age;
- Load type;
- Ship flag;
- Number of passenger;
- Wind intensity;
- Range of visibility;
- Historical data on the ship, such as accidents, and breakdowns within 6 months.

Each input is automatically detected by the VTMS system, and is normalized by means of a Risk Function ($Risk_n$)

$$Risk_{Factor\ value} = \frac{c_1 * Risk_1(Ship_{Type}) + c_2 * Risk_1(Ship_{Age}) + \cdots + c_n * Risk_n(Accidents)}{100} \quad where\ \sum_{i=1}^{n} c_i = 100 \qquad (1)$$

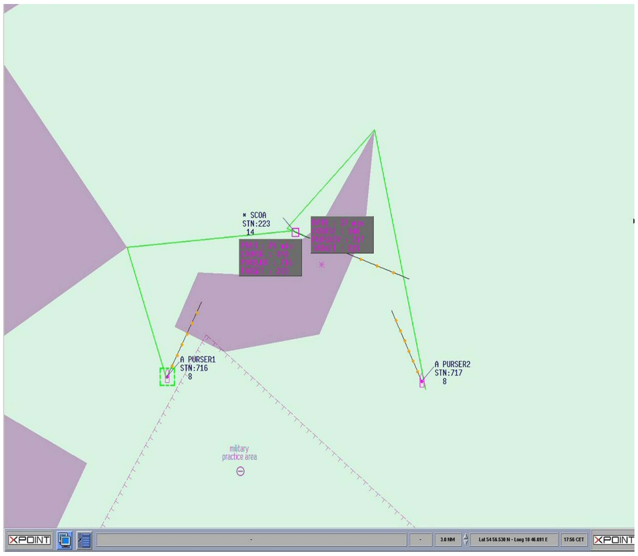


Figure 6 Threat interception function

according with the low ($T_{n,low}$) and high ($T_{n,high}$) thresholds. The $Risk_{Factor}$ value is automatically estimated anytime a new track is detected from the system or and/or identified by the operator and continuously updated to take proper countermeasures. The $Risk_{Factor}$ value is successively normalized by the stair heuristic function sketched in Figure 4, for ranking it (i.e. $Risk_{Range}$) into a three different classes; that is "High Risk", "Medium Risk", and "Low Risk".

3 Experimental scenarios

The VTMS solution comes with a user-friendly graphical interface to the operator for performing any allowed operations. In particular, the main window appearing to the operator is the log-in form, which allows profiling the operators within the system. As for as decision support module concerns, the main view to the operator, gained by the Operation console, is the Maritime tactical picture (MTP) (Figure 5). The MTP allows having a look on all the vessels are crossing the area under monitoring of a Local VTS Center, and for each vessel gathering more details such as kinematics Arrival port, identity details, and cargo details and so on. Operators acting directly on the MTP can activate any of the described operations/missions described in the previous section. As an instance, Figure 6 depicts how the system supports the operators in performing threat interception function. In particular, the systems has already chosen two units (i.e. purser1 and purser2) to be involved in the mission for intercepting the "hostile" target (i.e. SCOA). In the scenario the target is behind a forbidden area (violet block) and a wide military area can not be crossed. In this condition the system calculates (provided by continuos green line on the MTP) the shortest routes for the two pursuers. It's worth noting that the routes are calculated for assuring on the one hand, of course, the shortest path and on the other hand the higher probability of intercepting the target from the units, as little time as possible.

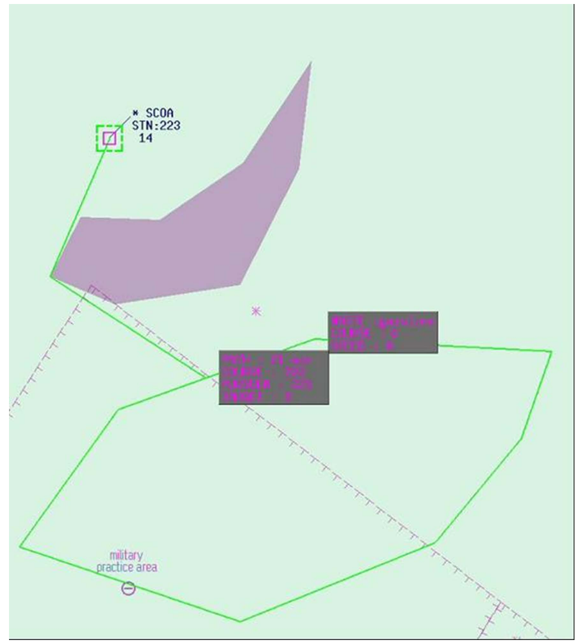


Figure 7 Anti-pollution function

Furthermore, the system estimates for each path also the interception time (i.e. PATH).

Decision support for anti-pollution operations is a further function in supporting the VTMS operators. Thanks to this functionality, the operator can see the pollution areas, which are automatically detected by either Radar Sensors or received by external systems; otherwise it can manually draw the area if no detection is performed. At the starting of anti-pollution mission, the system suggests a list of the most suitable allied means (referred as anti-pollution vessels) to the operator. The provided list is sorted out taking into account several parameters, such as distance from the polluted area, readiness state, equipment type in relationship with the pollutant, and time of arrival in the area. From the list, the operator can decide to select one or more vessels to enlist in the mission, and for each of them the shortest path is calculated and sketched on the tactical map by a continuous green line (Figure 7); the principle behind the shortest path calculation relies on avoiding obstacles and/or forbidden areas.

Finally, once the allied vessels reach the area and the operations are accomplishing, the systems allows periodically sending reports on anti-pollution mission situation in order to share tactical information about the mission. This information lets to increase the operator awareness on the mission state and better orchestrate the available resources.

4 Conclusion

The VTMS system, proposed in the paper, is a solution capable of supervising wide maritime area and coordinating actors in performing suitable countermeasures in different situations. The solution provides decision aid in several operating conditions supporting the operator in planning and executing actions during several kinds of missions, such as threats interception, Search and rescue, and anti-pollution. Moreover, its capability of monitoring wide maritime area and recognizing in transit vessels offers the possibility of calculating the map of risk associated with the monitored area.

The described solution is actually exploited in several countries for the vessel traffic management providing enough performance.

References

[1] Guidelines for Vessel Traffic Services. Resolution A.857(.20,1- IMO 1997.

[2] A. Border, "Vessel Traffic Systems, Year 1999," Proc. of Conf. OCEANS'90, pp. 464-469, 1990.

[3] V. Kharchenko, V. Vasylyev, "Decision making system for vessel traffic planning and radar control," Proc. of Conf. EURAD, pp. 61-64, 2004.

[4] V. Kharchenko, V. Vasylyev, "Application Of The Intellectual Decision Making System for Vessel Traffic Control," Proc. of Conf. Microwave, Radar, and Wireless Communication, pp. 639-642, 2002.

[5] P. Jaroslaw, K. Marek, B. Waldemar, W. Ryszard, "Integrated Vessel Traffic Control System," Proc. of Conf. International Radar Symposium, pp. 409-414, 2011.

[6] J. Peng; D. Yang, "Remote sensing monitoring system for maritime search and rescue," Proc. of Conf. Soft Computing and Pattern Recognition, pp. 101-106, 2011.

[7] Twinning Project PHARE, Project financed by European Union in 2006.

[8] J.T. Morris, V.I. Osychny, A.C. Turner, "Analytical techniques for the calculation of leeway as a basis for search and rescue planning," Proc. of Conf. OCEANS 2008, pp. 1-10, 2008.

[9] S. Sariel-Talay, E. Ergen, G. Avdan, C. Eroglu, "Sensor-based data storage for search and rescue," Proc. of Conf. Intelligent Sensors, Sensor Network and Information proceeding, pp. 7-12, 2008.

Operational and real-time team decision problems in the risk-averse transportation of dangerous goods by road

Claudio Roncoli, Roberto Sacile

Department of Communication, Computer and System Sciences
University of Genova
Genova, Italy
claudio.roncoli@unige.it

Michael G.H. Bell

Department of Civil and Environmental Engineering

Imperial College London
London, SW7 2BU, United Kingdom
m.g.h.bell@ic.ac.uk

Abstract – *Dangerous goods transportation (DGT) is a relevant safety and security issue that has been treated in many papers in the last decades. Considering the significant damages that a DGT accident can cause, a DGT routing strategy is to avoid the worst case accident scenario, following a risk-averse behaviour. Game-theory is very useful in this context. Taking into account a DGT flow on a road network and the technologies for vehicle/infrastructure communication which are nowadays available, the overall context can be modelled as a system of systems, where a team of road segments concurs to find an optimal strategy to minimise the maximum risk on the territory. Two problems related to DGT are presented and preliminary formulations, at operational and at real-time level, are described. A solving method based on linear matrix inequalities is shown and first results are shown.*

Keywords: *Dangerous good transport, uncertainty, game-theory, team decision problem, linear matrix inequalities.*

1 Introduction

Dangerous goods transportation (DGT) represents a critical issue in the present age. In fact, a DGT accident can produce damages such as fires, explosions, chemical leaks and environmental pollution, likely to affect a territory beyond the actual scene of the accident, even for a long time. For these reasons, several authors suggest the adoption of a risk-averse DGT strategy, aiming to minimise the maximum damages to population and environment [1, 2]. It has been demonstrated that a risk-averse dispatcher makes his choices considering a mix of routes instead of a single one [2], leading to hyperpaths, namely a set of paths any of which may be optimal. The use of hyperpaths is very useful while coping with uncertainty in transport networks, and from a mathematical viewpoint it could be described considering a game-theoretic framework [3]. In a DGT game-theoretic framework, the dispatcher could be represented by the team, trying to make his decisions minimising the possible damage, whereas the demon is the maximiser: this results as a zero-sum game.

In this respect, a class of problems are modelled as minmax problems, where a team of player must optimise the worst case scenario given a limited information of demon's decisions [4, 5]. From a mathematical viewpoint, this approach presents similarities with robust team decision theory. The aim of this paper is thus to present a mathematical formulation of the DGT game-theory problem, highlighting the common aspects with the minmax team decision theory, and obtaining a solving procedure based on Linear Matrix Inequalities (LMI). One formulation may be adopted at operational level, to decide the optimal percentage of flow to be sent on the different network links from one origin to one destination. Another formulation, which may be adopted in real-time, considers the possibility of monitoring and controlling the trip of a fleets of vehicles over a road network. This latter approach takes into account the distribution of DGT over a specified time windows, and the decision are thus taken dynamically, according to a Model Predictive Control (MPC). It is assumed that the routing instructions that are sent to vehicles cannot always be respected, and that other unforeseen DGT flows may be present generating a noise which may be hard to predict. The problem aims at minimising different objectives, among which the risk, taking into account the worst case. It is thus shown that this problem can be modelled as a dynamic team decision problem and, under some assumptions, solved in the same way of the static minmax problem.

2 Background

The following problem definitions are required as background to the models proposed in this paper.

Consider the following minmax team decision problem:

$$\inf_{u} \sup_{q\neq 0} \frac{J(u,q)}{\|q\|^2}$$

or equivalently

$$\inf_{u} \sup_{q\neq 0} J(u,q) - \gamma\|q\|^2 \quad (1)$$

subject to:

$$y_l = C_l q$$
$$u_l = \mathcal{F}(y_l)$$

where:

$$J(u,q) = \begin{pmatrix} q \\ u \end{pmatrix}^T \begin{pmatrix} G_{qq} & G_{qu} \\ G_{uq} & G_{uu} \end{pmatrix} \begin{pmatrix} q \\ u \end{pmatrix} \tag{2}$$

$$G_{uu} \succ 0$$

In [4], it is demonstrated that, if $J(u,q)$ is a quadratic cost function given by (2), the optimal decisions are linear (3), and they can be computed by solving the linear matrix inequality (4), where matrices C and K are defined as in (5).

$$u_l = \mathcal{F}(y_l) = K_l y_l \tag{3}$$

$$\min_{\gamma,K} \gamma$$

Subject to :

$$\begin{bmatrix} G_{qq} - \gamma I + G_{qu}KC + C^T K^T G_{uq} & C^T K^T \\ KC & -G_{uu}{}^{-1} \end{bmatrix} \preccurlyeq 0 \tag{4}$$

$$C = \begin{bmatrix} C_1 \\ \vdots \\ C_L \end{bmatrix} \qquad K = \begin{matrix} K_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & K_L \end{matrix} \tag{5}$$

3 Static minmax formulation in operational DGT management

The problem is formulated considering a logistic network $N(V,L)$ where the V vertices represent logistic junctions and the L links represent road segments. Its topology is represented by the incidence matrix E, where each element $e_{v,l}$ is defined according to:

$$e_{v,l} = \begin{cases} -1 & \text{if link } l \text{ leaves vertex } v \\ 1 & \text{if link } l \text{ enters into vertex } v \\ 0 & \text{otherwise} \end{cases}$$

Each link is characterised by r_l $[inh/km^2]$, that is the measurement unit for the exposure on link l, expressed by the population living in a given unitary area.

According to the concept that a risk averse dispatcher, assuming that there is no knowledge about the accident probability, prefers to choose a set of routes rather than a single one [2], the problem could be thus formulated considering the probability that a link is chosen for the shipment and the probability of an accident on that link, that are:

p_l probability of a DGT vehicle sent over link l
q_l probability of an accident on link l

In order to represent the origin and destination of a shipment, an additional variable is introduced, characterising the origin and destination vertices, that is:

$$f_v = \begin{cases} 1 & \text{value for the source vertex} \\ -1 & \text{value for the destination vertex} \\ 0 & \text{otherwise} \end{cases}$$

Considering decisions on a DGT flow, p_l can be taken into account as the percentage of DGT flow sent over link l, and q_l the strength of the accident on link l according to a given scenario. So, supposing that the area A $[km^2]$ involved in an accident follows a quadratic law, given by $A = \alpha(p_l q_l)^2$ - where α is a given constant supposed hereinafter as $\alpha = 1$, and $p_l q_l$ represents a measure of the linear extension of the accident - the expected loss can be modelled as:

$$R(p,q) = \sum_{l \in L} (p_l q_l)^2 r_l \tag{6}$$

Moreover, in order to respect the delivery plan, the following constraints must be respected:

$$\sum_{l \in L} e_{v,l} p_l = f_v \qquad \forall v \in V \tag{7}$$

This model could be considered in a distributed decision making context where the links are considered as independent cooperative subsystems, which can be modelled as a team of players aiming to optimise a worst case scenario against nature. Nature could be viewed as a demon, that tries to cause the maximum damage in the network. In this model, it is supposed that a subsystem has only the knowledge of the demon attack performed on itself. This entails that the decisions made by a subsystem l are related to the linear extension allowed to an accident evolving according to a given and known strength q_l:

$$u_l = \mathcal{F}(q_l) \tag{8}$$

Considering the assumption about the limited knowledge of a subsystem, the decisions can therefore be computed as :

$$u_l = p_l q_l \qquad \forall l \in L \tag{9}$$

Considering the problem briefly summarised in the background section, the problem could be simplified as:

$$\inf_u \sup_{q \neq 0} J(u,q) - \gamma \|q\|^2$$

subject to (7) and to (9),
where $J(u,q)$ is defined in (2) with:

$$G_{qq} = G_{qu} = G_{uq} = 0 \tag{10}$$

$$G_{uu} = \begin{pmatrix} r_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & r_L \end{pmatrix}$$

From this formulation, it arises that the computation of the probabilities p ($p = \mathrm{diag}(P_1, \dots, P_L)$) could be performed in a decentralised way, without the complete knowledge of q, by solving the following LMI problem.

$$\min_{\gamma,K} \gamma$$

subject to :

$$\begin{bmatrix} -\gamma I & P^T \\ P & -G_{uu}{}^{-1} \end{bmatrix} \preccurlyeq 0 \tag{12}$$

$$EPT + T^T K^T E^T - F \preccurlyeq 0$$

The last inequality is related to constraints (7), introducing matrices T and F, defined as:

$$\begin{aligned} T_{i,j} &= 1 && \forall i, j \in V \\ F_{i,j} &= -(f_i + f_j) && \forall i, j \in V \end{aligned} \tag{13}$$

4 Dynamic minmax real-time DGT control

In this section, a modelling proposal of minimax problems to be adopted in real-time risk averse DGT control is proposed. This second model considers a set of subsystems, representing specific territorial areas. The problem is formulated considering a logistic network $\mathrm{N(V, L)}$ where the V vertices here represent subsystems and the L links represents DGT flows from one subsystem to another. The dangerous good (DG) product within the subsystem at a specific time instant represents the state of the system. It is supposed that the DM may observe the state of the different subsystems, for instance through the knowledge of the GPS position of trucks, after a delay due to measurement issues. The DM take his decisions about the transfer of DG product – which is here supposed as continuum quantity – between two neighbouring subsystems, aiming to: respect the requested deliveries; respect the flow and speed constraints of a specific link; minimise the risk due to the DG presence in subsystems.

The control variables could be therefore defined as the recommended flow over a link to be respected in a time interval. This is sent as a signal from the DM to vehicles employed in the delivery. However, it is also supposed that not all trucks respect the DM requests; possible reasons could be that some signals are affected by communication problems, or the vehicle cannot respect DM directives due to technical problems. This could be modelled as a noise in the state of the subsystem. From a mathematical viewpoint, the system could be represented by following variables:

$x_v(k)$ the amount of DG in subsystem v at instant k
$u_l(k)$ flow assigned to l for time interval $(k, k+1]$
$\omega_v(k)$ noise in the state for subsystem v at instant k

The network is characterised by input values:

$\hat{x}_v^O$ product at the initial time instant in a vertex v
$\hat{x}_v^D$ quantity of product requested by a vertex v
$r_v(k)$ measurement unit for the risk at instant k
d_v extension of territorial subsystem v
$\hat{v}_l(k)$ planned speed related to link l

The dynamics of the whole system could be thus stated as:

$$\begin{aligned} x_v(k+1) &= x_v(k) + Eu_l(k) + \omega_v(k) \\ x_v(0) &= \hat{x}_v^O \end{aligned} \tag{14}$$

Supposing a delay in the measure of one time step, the observation of the system is described by:

$$y_v(k) = x_v(k-1) \tag{15}$$

The three objectives for the DM could be formulated, similarly to the approach utilised in [6], as:

- The difference between the DG product at the final instant and the planned one:

$$(x_v(K) - \hat{x}_v^D)^2$$

- The difference between the speed (obtained from the Speed-Flow-Density Relationship) and its planned value:

$$\left(u_l(k) d_v - x_v(k)\hat{v}_l(k)\right)^2$$

$$\forall l | E_{vl} > 0$$

- The risk at vertices, defined as:

$$(x_v(k))^2 \frac{(r_v(k))^2}{(d_v)^2}$$

The state variable $x_v(k)$ can be replaced by $\tilde{x}_v(k)$, according to:

$$\tilde{x}_v(k) = x_v(k) - \hat{x}_v(k) \tag{16}$$

Where:

$$\hat{x}_v(k) \begin{cases} = 0 & k < K \\ \hat{x}_v^D & k = K \end{cases} \tag{17}$$

Following these definitions, this problem could be stated as a dynamic minmax decision problem:

$$\inf_{u} \sup_{\omega} J(u,\omega)$$

subject to:

$$\begin{aligned} \tilde{x}(k+1) &= \tilde{x}(k) + Eu(k) + \omega(k) + \\ &+[\hat{x}(k) - \hat{x}(k+1)] \end{aligned} \tag{19}$$

$$\begin{aligned} \tilde{x}(k+1) &= \hat{x}^{O} \\ y_v &= C_v \tilde{x} \\ u_l &= \mathcal{F}(y_v) \end{aligned}$$

where:

$$\begin{aligned} J(u,\omega) &= x^T(K) Q_K x(K) + \\ &+ \sum_{k=1}^{K-1} \left(\begin{bmatrix} \tilde{x}_v(k) \\ u_l(k) \end{bmatrix}^T Q(k) \begin{bmatrix} \tilde{x}_v(k) \\ u_l(k) \end{bmatrix} - \vartheta \|\omega_v(k)\|^2 \right) = \\ &= \alpha \sum_{n \in N} \left(\tilde{x}_n(k)\right)^2 + \\ &+ \sum_{k=0}^{K-1} \sum_{n \in N} \left(\beta \left(u_l(k) d_n - \tilde{x}_n(k) \hat{v}_l(k) \right)^2 \right) + \\ &+ \sum_{k=0}^{K-1} \sum_{n \in N} \left(\frac{(r_n(k))^2}{(d_n)^2} + -\vartheta \|\omega_n(k)\|^2 \right) \end{aligned} \tag{20}$$

Introducing following variables:

$$\bar{x}_v = \begin{bmatrix} \omega_v(K-1) \\ \omega_v(K-2) \\ \vdots \\ \omega_v(0) \\ \tilde{x}_v(0) \\ 1 \end{bmatrix} \qquad u_l = \begin{bmatrix} u_l(K-1) \\ u_l(K-2) \\ \vdots \\ u_l(0) \end{bmatrix} \tag{18}$$

Then, according to [5], the problem (19) could be reformulated as a static team decision problem, allowing to solve it by using LMI (4).

5 Numerical example at operational level

In this section, a numerical example is shown, illustrating the behaviour of the first methodology at operational level presented in this paper. The network is made up of 20 vertices and 31 links; the network topology and the input values for r_l are shown in Figure 1.

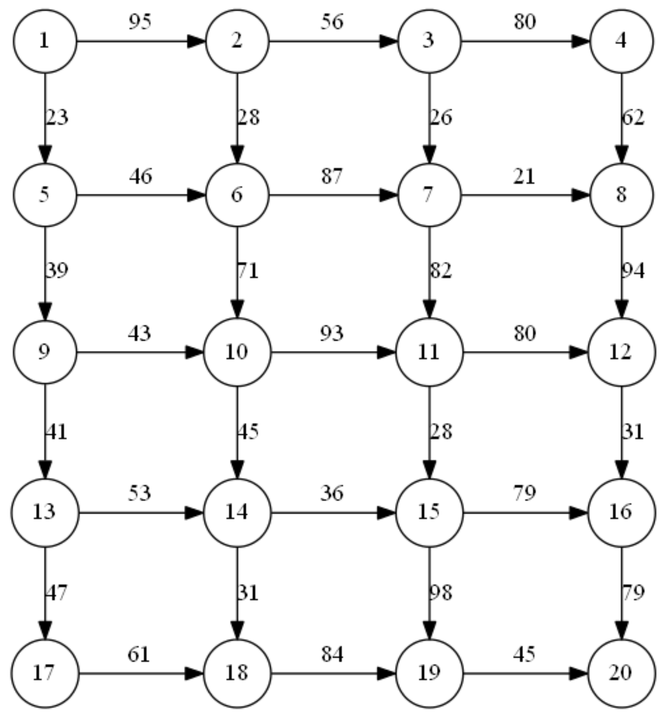


Figure 1. The demonstrative network. The number by the links are related to r_l values.

The model was implemented in MATLAB (MathWorks, Inc.); numerical results are presented in Figure 2, where links are characterised by the respective values of p_l. Evidently, the optimal solution involves the use of several routes from the origin to the destination.

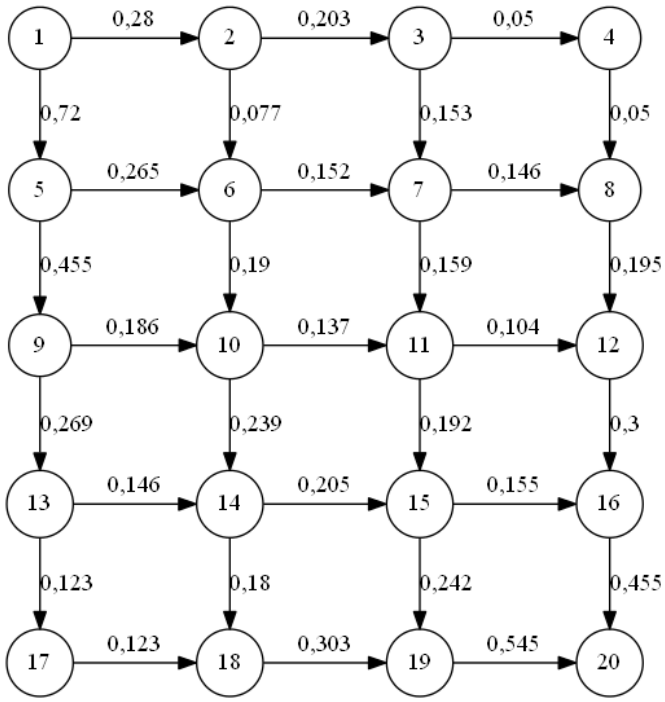


Figure 2. The numerical results. . The number by the links are related to p_l values.

6 Discussion and conclusions

In this paper two problems related to the DG distribution are presented, both formulated following a game-theoretic approach. The static minmax problem, despite its simple formulation and the possibility of solving it through other method (such as, under some assumptions, linear programming), is an interesting starting problem in this field. The dynamic minmax problem represents a more complex scenario when dealing with DG transport; however, the formulation as a team decision problem seems to present and interesting approach to deal with it, allowing an interesting solution approach.

References

[1] E. Erkut, A. Ingolfsson, “Catastrophe avoidance models for hazardous materials route planning”, *Transportation Science*, Vol 34, pp. 165-179, 2000.

[2] M.G.H. Bell, “Mixed Route Strategies for the Risk-Averse Shipment of Hazardous Materials”, *Networks and Spatial Economics*. Vol 6, pp. 253-265, 2006.

[3] J.D. Schmöcker, M.G.H Bell, F. Kurauchi, H. Shimamoto, “A Game Theoretic Approach to the Determination of Hyperpaths in Transportation Networks”, in W.H.K. Lam, S.C. Wong, H.K. Lo, *Transportation and Traffic Theory*, Springer Science+Business Media, pp. 2-18, 2009.

[4] A. Gattami, B. Bernhardsson, A. Rantzer, “Robust Team Decision Theory”, *IEEE Transactions on Automatic Control*, Vol 57, No. 3, pp. 794-798, 2012.

[5] A. Gattami, B. Bernhardsson, "Minimax Team Decision Problems”, *American Control Conference (ACC) 2007*, pp.766-771, New York, NY, 9-13 July 2007.

[6] C. Roncoli, C. Bersani, R. Sacile, “A decentralised control model for the risk-based management of dangerous good transport flows on road*” 6th Annual IEEE Systems Conference (SysCon)*, Vancouver, BC, Canada, 19 - 23 March 2012.

A Model-driven Approach for Configuring and Deploying Systems of Systems

Emanuela Barbi, Giovanni Cantone, Davide Falessi, Fabrizio Morciano, Marco Rizzuto, Vincenzo Sabbatino, Stefano Scarrone

SELEX Sistemi Integrati
(ebarbi, fmorciano, mrizzuto, vsabbatino)@selex-si.com

Simula Research Laboratory and University of Rome Tor Vergata
d.falessi@ieee.org

University of Rome Tor Vergata
cantone@uniroma2.it
stefano.scarrone@gmail.com

Abstract. *Configuration and deployment of systems for defense and air traffic control is often a complex task because a System of Systems (SoS) is always distributed on different geographic areas, composed by hundreds of components (e.g. applications, processes, services, hosts), running under multiple hardware constraints, on different resources, and subject to mission critical requirements. The configuration of such SoS or a part of it, involves the production of many configuration files describing the structure of the SoS in general, the configuration parameters of each component, and how each component has to interact with the others. Due to the considerable size and complexity of the configuration files (i.e. hundreds of lines of code), the adoption of a manual approach is clearly error prone. This work presents a model-driven approach for supporting the configuration of mission-critical SoS or part of it.*

Keywords: Model-driven engineering, automated software engineering, UML, system configuration and deployment.

1 INTRODUCTION

On the one side, the system of systems (SoS) concept, originally suggested as a method to describe the use of different systems interconnected to achieve a specific goal, has grown in its myriad of definitions and concepts [1-4]. Nevertheless, a common defining attribute of a SoS that critically differentiates it from a single monolithic system is interoperability, or lack thereof, among the constituent parts (i.e., systems). On the other side, model driven engineering has gained increasing interest in the industry as a way to express and analyze system requirements, specifications, architectures, and designs. Model driven engineering consists in developing models before the actual artifact products are designed and then transform the models to the actual artifacts in an automated way. Therefore, the application of model driven engineering is a perfect solution for defining and analyzing the communications among the systems constituting the SoS.

1.1 Context

Finmeccanica is a large Italian industrial group operating globally in the aerospace, defense, and security sectors, and is one of the world's leading groups in the fields of helicopters and defense electronics. It has revenues of 15 Billion Euros and invests 1.8 Billion Euros (12% of turnover) a year in R&D activities. SELEX Sistemi Integrati (also known as SELEX SI) is the Finmeccanica Company focusing on the design of SoS; it aims to be the European leader in the definition and integration of sensors and systems for defense, coastal/maritime surveillance, and air traffic management.

1.2 Study motivation

Configuration and deployment of systems for defense and air traffic control is often a complex task because a SoS is always distributed on different geographic areas, composed by hundreds of components (e.g. applications, processes, services, hosts), running under multiple HW constraints and on different resources, and subject to mission critical requirements. The configuration of such SoS or a part of it, involves the production of many configuration files describing the structure of the SoS in general, the configuration parameters of each component, and how each component has to interact with the others. Due to the considerable size and complexity of the configuration files (i.e., hundreds of lines of code), the adoption of a manual approach is clearly error prone.

1.3 Aim

This work presents a model-driven approach for supporting the configuration of mission-critical SoS. The proposed approach has been validated, in the context of SELEX Sistemi Integrati, via two benchmarks and a pilot study

2 APPROACH AND RELATED TOOL

2.1 Approach

In order to cope with the difficulties presented in Section 1.2, we analyzed the activity of configuration and deployment of SoS. In the context of SELEX Sistemi Integrati, most safety critical components of a large distributed system are developed on CARDAMOM and are intended to be part of a SoS [5]. CARDAMOM is a middleware platform enabling component based architecture for the implementation of safety and mission critical system, such as those for air traffic control and combat management systems. CARDAMOM is a framework to configure, deploy and execute near real-time and fault-tolerant distributed applications [5].

In this context, together with SELEX Sistemi Integrati professionals, some areas of improvement have been identified in the current approaches to configuration. The main points addressed are:

- In order to provide high configurability, the size of the XML configuration files is large. This makes the management of those files very hard.
- The lack of a general purposes supporting tool which would speed up the process and help integrators to create, adapt, and correct configuration files. This presumably could slow down the process and cause the release of both syntactic and semantic types of errors.

From these lacks we decided to develop an approach aimed to raise up the interaction language between practitioners and the configuration and deployment process. Eventually, we came to using a model-driven approach. Two graphical supported modeling languages were finally considered: the Unified Modeling Language, UML2.0 [6, 7], and the Graphical Modeling Framework, GMF [8]. We analyzed in deep these modeling languages by using lab case studies. Once this deep analysis was terminated, GMF was chosen as technology. In order provide a practical support to the reference company and to validate it, we developed a supporting tool and afterwards we ran a pilot study at SELEX Sistemi Integrati for a preliminary evaluation of the tool and its referred GMF model.

2.2 Metamodel

The starting point was to investigate and understand the structure, the components, and their relationships, of a safety critical CARDAMOM-based SoS. The aim was to define an abstract view with all the possible information used and needed in the configuration and deployment processes. The analysis of a safety critical system was enacted via an iterative process, as shown in Figure 1, and produced as a result a meta-model. This meta-model is a model representing all the possible correct configurations of a system. As a matter of fact, an instance of that meta-model is a model which represents a specific configuration.

The process showed in Figure 1 consists of three main phases:

- The analysis of the CARDAMOM documentation. This gave us an overview of the CARDAMOM environment.
- The analysis of a bunch of configuration files used in real projects of SELEX Sistemi Integrati. This allowed us to compare the theoretical aspects to the practical ones.
- Interviewing experts at SELEX Sistemi Integrati. This eliminated doubts and improved the understanding of CARDAMOM-based system.

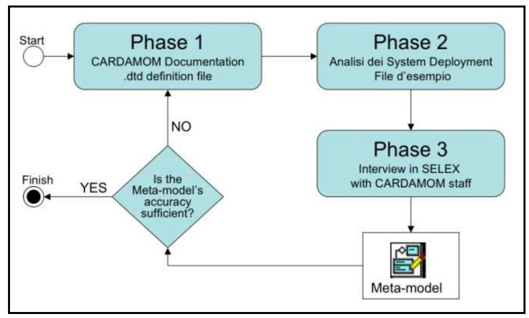


Figure 1: Meta-model definition process

As Figure 1 illustrates, we repeated the process many times, by incrementally redefining the meta-model, until it reached sufficient accuracy. Figure 2 reports a fragment of the meta-model. Considering the high level of abstraction of this analysis, we could not yet define the right technology to adopt and consequently the meta-model was implemented using an Ecore model. An Ecore model is a component of the Eclipse Modeling Framework that can be used to define meta-models [8].

We used the Object Constraint Language (OCL) to better define the meta-model and to complement and extend the power of the Ecore modeling language.

For the creation and management of a configuration, the tool can use a model-driven approach; in other words, we can integrate the meta-model inside the tool allowing the users to handle the configuration process through a model, without having to interact directly with the configuration files. This will reduce the proliferation of syntactic errors.

In addition, the adoption of a validation process can avoid the semantic errors. The graphical representation feature can be realized using diagrams extracted from the meta-model. Moreover, the adoption of a validation process can avoid the semantic errors. The graphical representation feature can be realized using diagrams extracted from the meta-model.

Our graphical solution is based on three main aspects, named Views [9, 10]:

- Structure View: it describes the structure of the components of a SoS with the relations and interactions between them.
- Deployment View: it describes the components distribution on the nodes (e.g. the A Process runs on the F Server).
- Activity View: it describes the boot and shutdown order of the system components.

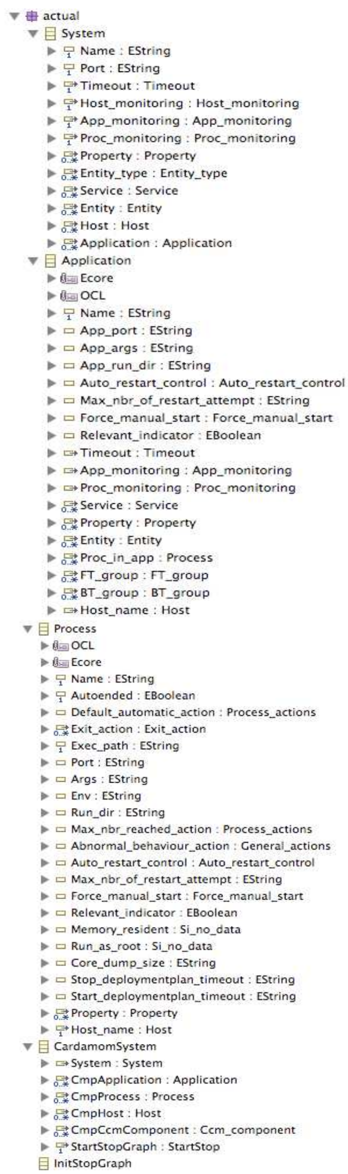


Figure 2: Abstract of the metamodel

The use of views aims to see the different aspects of the same concept (i.e. a process or a node) from different perspectives, thus allowing separation of concerns while assuring their consistency [11-14].

It was also clear from the beginning that the tool would provide an import/export mechanism. This to translate the model representing a system configuration to an XML file and, vice versa, to automatically generate the models from configuration files.

2.3 Tool architecture

Figure 3 describes the architecture of the tool ACTUAL (Automation of the Configuration and deploymenT of distribUted AppLications), showing the main components and their relations.

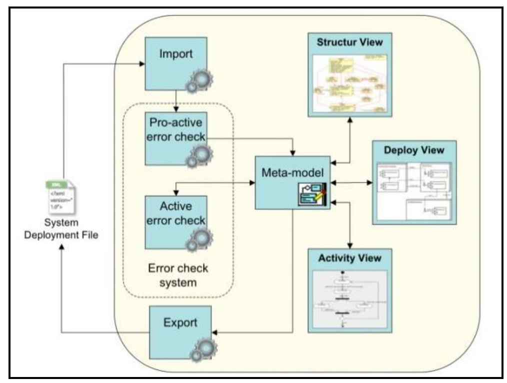


Figure 3: Architecture of ACTUAL

2.4 Use cases

The development of the tool ACTUAL is based on the Graphical Modeling Framework (GMF) [15]. As expected, this Eclipse framework actually allowed the automatic generation of both model and Views; additionally, thanks to the OCL constraints that we added to the meta-model, GMF facilitated the realization of the error control mechanism. Because an ACTUAL-model is managed by an eXtensible Stylesheet Language (XSL), the translation functionalities, from an ACTUAL-model to XML and vice versa, were implemented using the eXtensible Stylesheet Language Transformations, (XSLT). Figure 4, Figure 5, Figure 6, and Figure 7 show some usage examples of ACTUAL. Specifically, Figure 4 shows how ACTUAL allow the user to modify the property of a component of the system. In Figure 5, the user describes how an application is structured in processes and their relations. In Figure 6, the user allocates the entities (as defined in Figure 5) in specific physical nodes. Finally, in Figure 7 the user defines the start-up (green lines) and the shutdown (red lines) flows among the entities described in Figure 4.

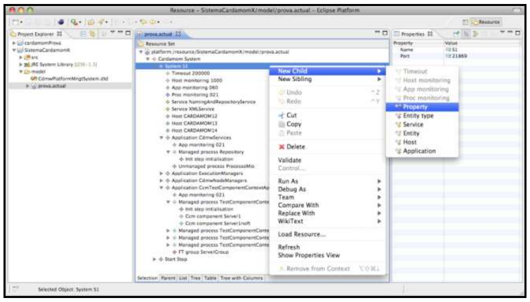


Figure 4: Specific usage of the model editor

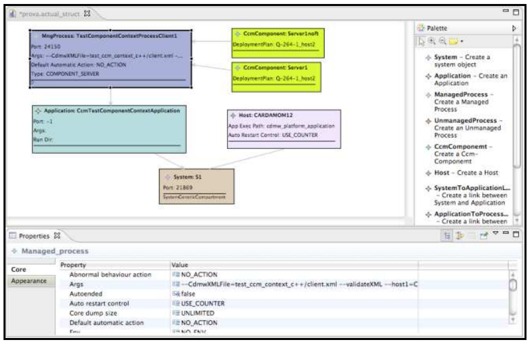


Figure 5: Specific usage of the Structure View editor

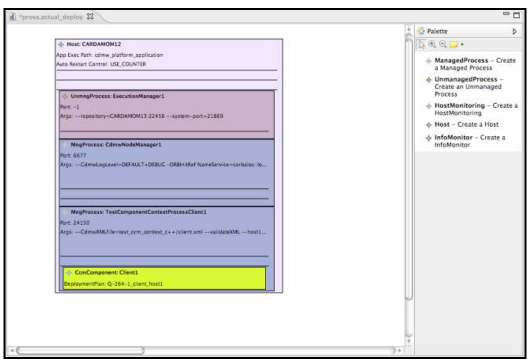


Figure 6: Specific usage of the Deployment View editor

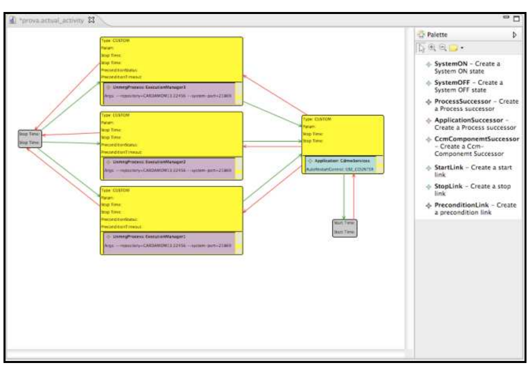


Figure 7: Specific usage of the Activity View editor

3 VALIDATION AND CONCLUSION

Our solution, comprising the adoption of GMF and the tool ACTUAL, has been validated in the context of SELEX Sistemi Integrati via two benchmarks and one pilot study related to the configuration and deployment of a system by adopting real industrial artifacts. The analysis of these preliminary results shows that the adoption of model-driven instruments seems really to help the organization in meeting its goals. Specifically, the tool resulted to detect all the injected faults in a configuration file and it resulted to provide an output without any defect.

The presence of a visual and an automated support to create the artifacts allows to hide the complexity of the system behind a simple and coherent user interface. Moreover, the same results seem to indicate that the adoption of model-driven tools saves effort and increases the productivity, if specific to the context and able both to support the design phase, and to cope with an high level of detail. The model driven approach, as supported by ACTUAL, related to the SELEX Sistemi Integrati problem, can be considered a useful, innovative and easy to learn support tool. Its adoption is expected to save effort and resources.

4 REFERENCES

[1] M. J. DiMario, "System of Systems Interoperability Types and Characteristics in Joint Command and Control," in *2006 IEEE/SMC International Conference on System of Systems Engineering*, Los Angeles, CA, USA, 2006.

[2] M. Jamshidi, "System-of-Systems Engineering - A Definition," presented at the IEEE International Conference on Systems, Man, and Cybernetics., Available at: http://ieeesmc2005.unm.edu/SoSE_Defn.htm, 2005.

[3] I. Jacobson, K. Palmkvist, and S. Dyrhage. (1995, May-June) Systems of Interconnected Systems. *ROAD*.

[4] D. Fisher, "An Emergent Perspective on Interoperation in Systems of Systems," *Software Engineering Institute Technical Report, Carnegie Mellon University,*, vol. CMU/SEI-2006-TR-003, 2006.

[5] THALES and SELEX SI, "http://cardamom.ow2.org/," 2005.

[6] J. Arlow and I. Neustadt, *UML 2 and the Unified Process: Practical Object-Oriented Analysis and Design*, 2005.

[7] Object Management Group (OMG), "Unified Modeling Language (OMG UML 2.4), Infrastructure," *http://www.omg.org/spec/UML/2.4/Infrastructure/Beta2/PDF*, 2011.

[8] R. C. Gronback, *Eclipse Modeling Project: A Domain-Specific Language (DSL) Toolkit*: Addison-Wesley Professional, 2009.

[9] I. P. ISO/IEC 42010, "Systems and Software Engineering — Architecture Description," work in progress, 2008.

[10] P. Kruchten, "The 4+1 View Model of Architecture," *IEEE Software,* vol. 12, pp. 45-50, November 1995.

[11] D. Falessi, G. Cantone, R. Kazman, and P. Kruchten, "Decision-making Techniques for Software Architecture Design: A Comparative Survey " *ACM Computing Surveys,* vol. 43, 2011

[12] D. Falessi, L. Briand, G. Cantone, R. Capilla, and P. Kruchten, "The Value of Design Rationale Information,"

ACM Transactions on Software Engineering and Methodology, vol. in press, 2012.

[13] S. Nejati, M. Sabetzadeh, D. Falessi, L. Briand, and T. Coq, "A SysML-based approach to traceability management and design slicing in support of safety certification: Framework, tool support, and case studies," *Information and Software Technology,* vol. 54, 2012.

[14] D. Falessi, M. Ali Babar, G. Cantone, and P. Kruchten, "Applying Empirical Software Engineering to Software Architecture: Challenges and Lessons Learned," *Empirical Software Engineering,* vol. 15, pp. 250-276, 2010.

[15] F. Seehusen and K. Stølen, "An evaluation of the graphical modeling framework (GMF) based on the development of the CORAS tool," in *4th international conference on Theory and practice of model transformations (ICMT'11)*, Heidelberg, 2011.

A System of Systems Approach to Near Miss Accidents in Dangerous Goods Road Transportation

Silvia De Nadai
Francesco Parodi
DIST - University of Genoa
Genova, ITALY
silvia.denadai@unige.it, francesco.parodi@unige.it

Domenico Pizzorni
ENI R&M
Genova, Italy
domenico.pizzorni@eni.com

Abstract - *The "near miss accidents" system arises from a wider based project by DELAB, a joint laboratory, between ENI, one of the most important Italian petrolchemical companies, and the University of Genoa. The objective is to identify the causes leading to truck accidents enhancing the knowledge base through the data collection of near miss accidents. A system of systems engineering approach was developed where the main system components relate to the driver, the truck and the external environment conditions. The methodology which was adopted to monitor the whole system and preliminary results are illustrated.*

Keywords: System of systems, Biometric parameters, Can Bus, near miss accident, behavioral parameters, biometric data.

1 Introduction

Car manufacturers have been interested in "driver monitoring" research for several years. The aim is to enhance the general knowledge of driver behaviour and secondly, to evaluate the functional state as it may influence drastically driving safety; such as distraction, fatigue, mental workload and attention, [1].

The transport of dangerous goods by road forms a substantial part of the Italian and the European economy. The research focuses both on trucks and drivers, with the objective of trying to understand which are the most important parameters that should be monitored during the hours of driving, and a possible interactions between them.

Dangerous goods are substances that, due to their physical or chemical properties, can be a serious hazard for the population and the environment. The risks related to the transport of dangerous goods by road have a complex scenario, due to their dependence on many variables, e.g., traffic density, weather conditions, road accidents, driver ability and level of fatigue.

The transport of dangerous goods comprises 6% of the total of goods transport on the Italian road network. ENI has a fleet of 1500 tankers that everyday travel to their service stations. Since 2002, Eni and the University of Genoa have been working on methodological and technological solutions to secure the transport of dangerous goods by road. Consequently, a web based software platform has been developed, including the tracking of tankers, staff training (with particular attention on drivers) and vehicle maintenance.

In this paper, the design of a new module of the platform devoted to the collection of near miss accidents (NMAs) according to system of systems (SoSE) approach is described. The aim is to find relationships in monitored data to understand when NMAs occurs and thus find preventive strategies.

The current research is developed in two stages: the first one is a simulation context, the second one is a real context. In both stages, the objective is find relationships in the data collected from the driver (e.g. biometric data), the truck (e.g. Can Bus data), and the surrounding environment (e.g. weather and traffic conditions). The research is briefly resumed in the following sections.

2 Material & Methods

2.1 SoSE architecture

In order to study driver's behavior, an in-vehicle system was developed which could record contextual driving information, drivers' actions and related physiological signals. The proposed SoSE model is composed of several distinct devices that are integral to the on board components:

- Wearable sensors system;
- Sensors and Can Bus
- Central on board unit
- Additional information from other sources such as road plans, traffic, weather, and cartography

Figure 1 shows the proposed SoSE architecture. At this initial stage, the research is focused on wearable sensors, in relation to the driver monitoring, and Can bus, as regards the vehicle monitoring. These two systems permit the collection of crucial parameters about the wellness and the behaviour of the driver . The main onboard unit is able to detect abnormal value trends in these parameters and send messages by GPRS to the central system, for database storage . Moreover, this unit can receive data from the base station for example to configure thresholds (customized for the driver) and to allow the making of some decisions independently.

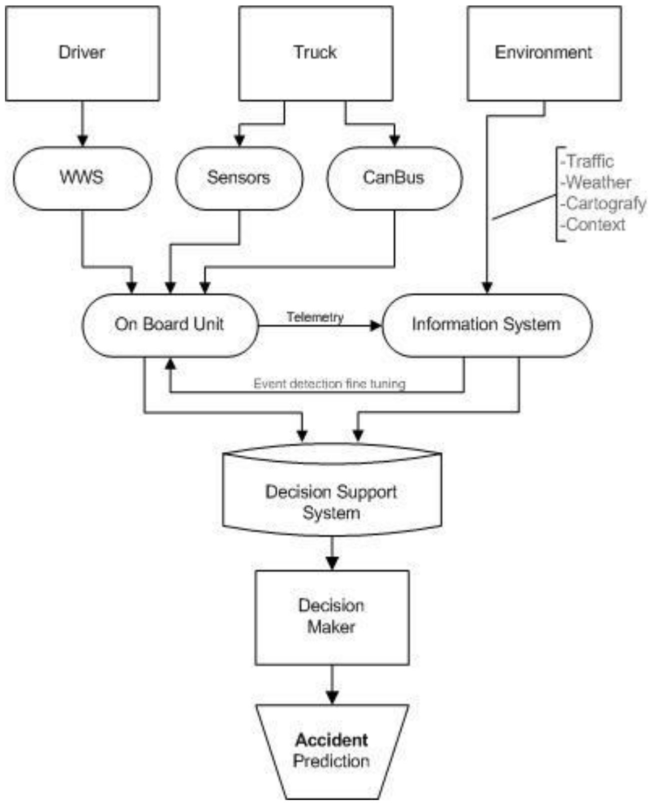


Figure 1: The SoSE architecture of the proposed Near Miss Accident Acquisition and Prediction System

2.2 Biometric data

The biometric data is acquired by means of a fully integrated wearable wellness system that consists of a particular type of t-shirt equipped with tree kind of sensors which are able to acquire a set of physiological parameters. These parameters are:

- Heart Rate (HR), that is the number of heartbeats per unit of time, typically expressed as beats per minute (bpm). The R wave to R wave interval (RR interval) is the inverse of the heart rate.
- Breathing Rate (BR), that is the number of respirations taken within a set amount of time, typically 60 seconds.
- Posture
- Movement

The HR and BR are monitored using sensors that are fully integrated into the fabric structure (see Figure 2), while Posture and Movement are controlled by a 3-D accelerometer is embedded in the electronic device called SEW3, by CSEM (see Figure 3).

Figure 2: Wearable Wellness System

Figure 1: SEW 3, by CSEM

The SEW3 integrates algorithms that pre-process raw data providing the following parameters:

- Electrocardiogram (ECG)

- Harte Rate Variability (HRV)
- Respiration Trend
- Breathing Rate
- Analysis of Movement Level

Data is stored internally, with a micro-SD, and/or transmitted via Blue-Tooth(R) directly to a PC. In the proposed architecture, the data is not transmitted to a PC, but to an on-board unit (OBU) equipped with a Blue-Tooth/Serial converter.

The currently monitored data is: the HR, the HRV and the BR. These samples are acquired at a specific frequency. The HR is sent to the OBU every five seconds, twelve samples per minute, which is expressed in beats per minute. HRV is sent every sixty seconds, one sample per minute. Finally, BR is sent every fifteen seconds, four samples per minute, which is expressed in respirations per minute.

An alarm message is given when a difference of 10 bpm, or more, between an HR sample and the previous one occurs.

2.3 Can Bus Data

The driver behavioral parameters are acquired with the use of the Can Bus interface. Can is a serial bus protocol which connects individual systems and sensors and allows automotive components to communicate on a single or dual-wire networked data bus up to 1Mbps.

In order to highlight a particular driving style and abnormal behaviour of the driver that can lead to a dangerous scenario, four parameters were identified:

- Abs / Ebs Status: these parameters express the state of activity, alert and intervention of the electronically controlled braking system.
- Pressure on Pedals: this information represents the use of brake and accelerator using the percentages of pressure on the two pedals.
- Wheel based speed: which is the speed (km/h) of the vehicle measured by the odometer.
- Speed of every wheel: using information provided by EBS the speed of each wheel is recorded with the precision of one km/h.

The OBU analyzes these four parameters sampled every 10 ms and detect events which must be sent in the form of telemetry. Every event has one or more threshold values, which were previously defined.

- precursor to rollover: this event has two possible cases: a warning threshold and an activation threshold;
- sudden breaking, that consists of the deceleration of at least 20 km/h per second of the wheel based speed;
- anomalous use of the steering wheel, detecting the sliding of the vehicle and measuring the difference of speed between left and right wheel
- anomalous use of pedals, that is an improper use of brake and accelerator pedals.

Specifically, four possible abnormal behaviours are detected: simultaneous pressure on brake and accelerator pedals, two or more repetitions of pressure and release of brake pedal in five seconds, two or more repetitions of pressure and release of accelerator pedal in five seconds, three or more alternate pressure and release of brake and accelerator in five seconds;

2.4 Experimental Context

The project included two different approaches. Firstly, a context based on simulation, and secondly the experimentation on real trucks.

In the first phase a driving simulator was set up, composed of physical devices representing a truck driving seat and a software simulator. This simulator allows the collection of biometrical information from the t-shirt, the can bus parameters and all the scenario’s details from the software.

This simulator is connected to a panel that replicates the devices usually placed on the ENI monitored trucks: such as sensors (temperature and pressure), and a GPS antenna, a GPRS module, a panic button, a touch screen and an onboard unit. Using this configuration the simulation station sends to the system all the information that can be received from a real truck; with the advantage of being able to create, with the simulation software, particular scenarios hard to replicate in the real world. An additional advantageous aspect is the possibility of placing the external events synchronized with the measured parameters.

The second phase consisted of collecting data directly from real trucks. In this phase two trucks were equipped with a blue-tooth antenna to allow the communication between the biometrical t-shirt and the onboard unit. In addition, a wired connection for reading data from Can Bus was also set up. Unlike the previous case, in which it was possible to correlate the instrumental data to the environment’s events, the experiments involved the manual collection of this information such as, street conditions, unexpected events, and weather .

2.5 Driving Sessions

In the case of both the simulation and of the use of real trucks, driving sessions were held reproducing the typical travel plan of a carrier of dangerous goods; that includes the departure from the base, the loading of the product and several stops to unload in the points of delivery. Therefore, in the simulation different road conditions, traffic and weather were simulated.

3 Conclusions

In conclusion, the last step is the analysis of the data; the biometric parameters (see Figure 4) and the behavioral data on a monitor which can be visually analyzed.

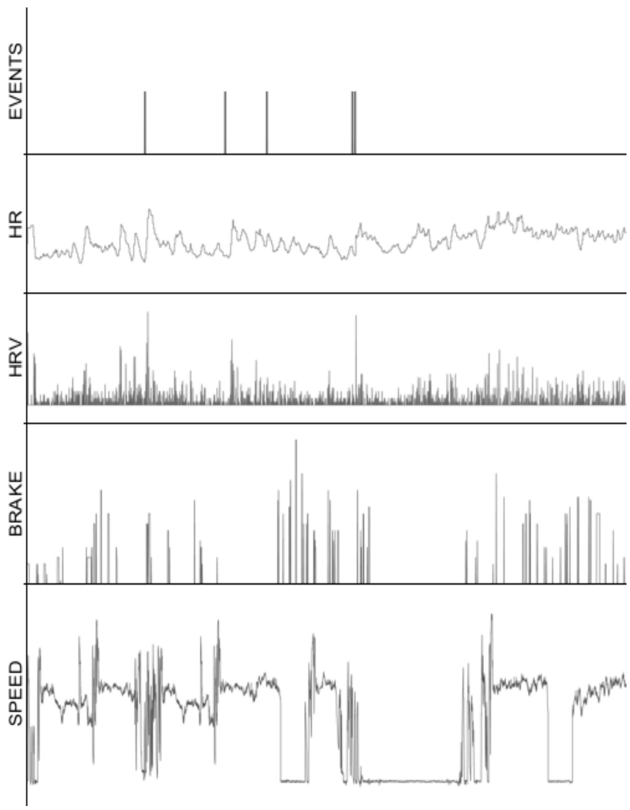


Figure 4: Parameters Recorded

The most important parameter seems to be the HRV. It is an established index of the sympathetic/parasympathetic balance and the intrinsic influence on heart rate in cardiovascular research and clinical routine [2] [3]. Changes occurring in selected frequency intervals of the HRV spectra are deemed expressive of specific functional changes [4]. Consequently, at the beginning the system must attune itself to the driver in order to recognize any anomalies. Therefore, a trial period is required for biometric parameters, as each person has their own physiological characteristics.

After the trial period some variations can be detected throughout the day; different plotting is to be expected before and after stops or loading and unloading points.

The same trend can be observed in the behavioral parameters; the data related to the pressure on the brake and accelerator pedals are of particular importance. During long driving sessions, the incidence of a gradual increase of anomalous uses of pedals was identified, often connected to an increasing in the reaction timing to external events. A further aspect, is the close relationship between the driver's fatigue and the incidence of sliding movement, of the recorded vehicle.

The last phase of the project will be the definition of a method for the prediction of near miss accidents. Accordingly, a dual strategy will be pursued; one approach is a strictly statistical study of values (inputs and outputs) of the system. This methodology is not based on the real physical meaning of the parameters so can highlight what information is really crucial in the identification of a near miss.

Alternatively, a future study will be realized focusing on the significance of the data, in which with the support of experts, aims to identify the combination of biometrical , behavioral and context patterns that can lead to a near miss accident.

References

[1] C. Ramon, A. Clarion, C. Gehin, C. Petit, C. Collet & A. Dittmar, "An integrated platform to assess driver's physiological an functional states" 30th Annual International IEEE EMBS Conference, Vancouver, British Columbia, Canada, August 20-24, 2008.

[2] Marianne B, Cess AS, Harm HVB, Peter CC, Volkert MC and Albert VGB, "Heart rate and heart rate variability as indexes of sympathovagal balance" Am J Physiol 1994; 266: 1565-71.

[3] Task force of the European society of cardiology and the North American society of pacing and electrophysiology, "Heart rate variability – standards of measurement, physiological interpretation, and clinical use", Circulation. 1996; 93(5):1043–65.

[4] Friedman BH. "An autonomic flexibility–neurovisceral integration model of anxiety and cardiac vagal tone", Biol Psychol 2007;74:185–99

Proc. of the 2012 7th International Conference on System of Systems Engineering, Genoa, Italy - 16-19 July 2012

Intelligent Transport Systems (ITS) applications on dangerous good transport on road in Italy

Benza[a] M., Bersani[a] C., D'Incà[a] M., Roncoli[a] C., Sacile[a] R., Trotta[a] A.

[a]DELAB –(DIST-ENI joint lab on safety and security in logistics and transport) Via Opera Pia 13 – 16145 Genova - Italy

Pizzorni[b] D., Briata[b] S., Ridolfi[b] R.

[b]ENI Eni Refining & Marketing, Logistica Secondaria, P.zza della Vittoria, Genova, Italy

Abstract - *In this paper the authors describe the architecture and the functionalities of the Transport Integrated Platform (TIP), an Intelligent Transport Systems (ITS) application on Dangerous Good Transport on road in Italy. The TIP represents a System of Systems (SoS) application because it aims to manage different kinds of sub systems (each module of the platform) and different kind of entities (electronical devices, DG vehicles, drivers). TIP manages the DG fleet planning routing and the real time vehicles' tours, the customer orders, and other support tools for drivers including training, resource management and advanced data mining. TIP application aims at promoting a safe management of DG freight transport providing timely information to truck drivers and fleet managers about weather, traffic, emergency or other critical situations on the road infrastructures.*

Keywords: Intelligent Transport Systems (ITS), Dangerous good transport (DGT).

1 Introduction

The share of Dangerous Goods (DG) in the total transport on road in 27-EU in 2010, for most countries, exceeded 4 %. All the major economies recorded figures in the 4 % to 6 % range. Some countries had a substantially greater proportion: Cyprus recorded almost 17 %, and Italia and Finland settled in the 6 % to 7 % range. The transport of DG in the EU-27 remained stable from 2009 to 2010 with more of 78.6 billion of tonnes-kilometres for the year 2010 [1].

To define the size of the DG transport by road, two different units of measure have to be introduced: [Mtkm/yr] which represents million of tonnes-kilometres for the year and [Mvehkm/yr] which represents million of vehicle kilometres for the year.

In Italy, more than 90% of its total DG transport on road is performed on its national territory (total transport of DG is 11.342 [Mtkm/yr] in 2010; National transport of DG is 10.504 [Mtkm/yr] in 2010; International transport of DG is 838 [Mtkm/yr] in 2010).

In Eu-27, the largest specific DG product group transported was flammable liquids, taking over a half of the total. Two other groups, gases (compressed, liquefied or dissolved under pressure) and corrosives, accounted for 13 % and 11 % respectively. This represents very little change compared with previous years when there was a very similar distribution between the product groups.

In Italy, the transport of flammable liquids by road represents the 70% of the national transport of DG in Mtkm/yr in 2010. Changing the unit of measure, the transport of flammable liquids by road in MVehkm/yr represents the 61% of the national transport of DG in 2010. This means that the distribution companies have invested in vehicle routing planning and optimization. Unfortunately, the routing optimization usually is based on the minimization of cost both covered kilometers and travel time [2]. Often the minimization of the risk related to DG transport has not taken into account by fleet managers.

On this framework, recently the European Commission emanates directives to impose the adoption of new Intelligent Transport Systems (ITS) in order to improve safety and security on road infrastructure. ITS, in fact, could significantly contribute to a cleaner, safer transport system improving transport safety and productivity, travel reliability, informed travel choices, environmental performance and network operation flexibility [3].

In this paper, the authors will introduce the legal framework for Dangerous Good Transport (DGT) by road and for ITS applications at European level. Then, the authors will explore all the functionalities of the Transport Integrated Platform (TIP): a ITS application developed in Italy exploring technological and methodological approaches applied to monitor and detect in real time DG vehicles.

2 Legal framework for dangerous goods transport

From a legislative viewpoint, the transport of dangerous or hazardous goods by road is subject to European legislation because of its specific risk factors. The general legislation is the European Agreement Concerning the International Carriage of Dangerous Goods by Road, known as ADR. The section "Security" in the ADR 2011 (chapter 1.10) underlines as security during loading, carriage, transshipment and unloading operations is imperative. Chapter 1.10.3.3 proposes the adoption of devices, equipment or arrangements to prevent the theft of the vehicle carrying high consequence dangerous goods (see Table 1.10.5) and its cargo. Those measures should be applied and taken to ensure that these are operational and effective at all times. This means that the DG carriers have to implement ITS applications to monitor in real time their vehicles in order to reduce risk for terrorism and accidents.

A new legal framework, Directive 2010/40/EU of the European Parliament, was adopted on 7 July 2010 to accelerate the deployment of these innovative transport technologies across Europe. This Directive is an important instrument to coordinate the ITS implementation in Europe. It aims to establish interoperable and seamless ITS services while leaving Member States the freedom to decide which systems to invest in.

Under this Directive the European Commission has to adopt within the next seven years specifications (i.e. functional, technical, organizational or services provisions) to address the compatibility, interoperability and continuity of ITS solutions across the EU. The first priorities will be traffic and travel information, the eCall emergency system and intelligent truck parking.

The EU Commission have recognized that innovation will have a major role to play in finding appropriate solutions for the Union and, in particular, Intelligent Transport Systems (ITS) represent advanced applications which aim to provide innovative services relating to traffic management and enable various users to be better informed and make safer, more coordinated and "smarter" use of transport networks.

This directive cited “ITS integrate telecommunications, electronics and information technologies with transport engineering in order to plan, design, operate, maintain and manage transport systems. The application of information and communication technologies to the road transport sector and its interfaces with other modes of transport will make a significant contribution to improving environmental performance, efficiency, including energy efficiency, safety and security of road transport, including the transport of dangerous goods, public security and passenger and freight mobility, whilst at the same time ensuring the functioning of the internal market as well as increased levels of competitiveness and employment”.

In detail, the European Commission started some initiatives to support ITS telematics application to freight transport and the use of EGNOS/Galileo, such as the Directive for the Implementation of the ITS Action Plan, the Freight Logistics Action Plan and related eFreight principles.

Since 2002, Eni (the major petrol Italian company) and DIST have been testing and developing innovative technological and methodological approaches to monitor DG vehicles in order to manage and control DG flow and prevent risk for people and environmental.

Recently, SCUTUM project, (SeCUring the EU GNSS adopTion in the dangeroUs Material transport), promoted by Italian Ministry of Infrastructure and Transports and Telespazio S.p.A (Italy and France), started on February 2010, has launched and pursued a concrete path for the use of EGNOS-based services for the dangerous goods transport market in Europe, implementing large scale trials to support standardization and harmonization at European level.

The EU-funded 'Scutum' project has delivered a crucial new document, laying out technical specifications to facilitate the development of products and applications based on EGNOS/EDAS.

The Scutum project is aimed at laying the groundwork for the operational adoption of EGNOS in the transport of dangerous goods.

The new CEN Workshop Agreement (CWA 16390), produced by Scutum, defines a standardised set of data output from mass market receivers, enabling application developers and service providers to easily build their own software solutions based on EGNOS and EDAS.

CWA 16390 specifies:

- The data (and relevant format) from GPS/EGNOS receivers that generate value-added services. An example of these services is the provision of EGNOS corrections via terrestrial networks and the calculation of position confidence level, i.e. maximum error;

- The type/format of the added value services produced by the software solutions.

- The technical specifications defined in CWA 16390 are architecture- and technology-independent and flexible.

Also in the framework of the Scutum project, Eni, a leading oil company, had the opportunity to test EGNOS versus GPS alone in the transport of dangerous materials, and to validate the relevant operational benefits in terms of higher safety and efficiency. Based on the results, Eni has decided to use EGNOS in real operations to track and trace its transport fleet throughout Europe

When the Scutum completed its work in December 2011, more than 300 Eni tankers transporting hydrocarbon and chemical products in Italy, France, Austria, Slovakia, Hungary, Romania, and the Czech Republic were being monitored using the EGNOS signal.

In 2012, Eni is planning to gradually extend the use of EGNOS to the transport of chemicals and aviation products, and to its operations in other European countries, including Germany and Switzerland.

In Italy, the National Ministry of Transports and Infrastructures is currently proposing to major companies involved in DG transportation to declare the whole daily trips planning and expected routes covered by their trucks. This might represent an interesting approach to re-allocate the trips during the day in base on the density of DG along roads [4].

2.1 ITS applications on dangerous good transport on road in Italy

Eni, R & M Secondary Logistics, from 2001, are working with DELAB to develop a ITS application to monitor and manage in real time its fleet which consists of more than 4600 vehicles. A part of the fleet (1600 vehicles) is dedicated to serve petrol product for automotive market to service stations (SS). Each days, each vehicle makes two or three tours and the mean distance travelled during each tour is 170 km. Currently, more than 560 vehicles are monitored in real time by the Transport Integrated Platform (TIP) developed by DELAB on the Italian territory.

The main objective of the TIP application is the adoption of technical and functional standards to manage different subsystems in order to provide environmental protection and safety for the health of the drivers and the people.

The ITS application consists of four sub systems:

- On board unit
- Transmission system
- Database
- GIS-based Applications

Those components are integrated in the TIP (Transport Integrated Platform) used by ENI to manage its fleets, to control transport services which are outsourced and to monitor in real time logistic processes and performances.

2.1.1 On board unit

On board, the system consists of a sub-systems located in different positions:

- on the trailer there are various type of device: the electronic counter, a collection of analog/digital sensors for the detection of information critical to monitor the single components of the vehicle and the satellite device called " concentrator" to which it is connected.

- on the tractor there is a distribution box installed, allowing the connection to the CAN (Controller Area Network) transmitting through the protocol FMS Standard, to the odometer, to the power supply, to an emergency switch in the cabin and in the distribution box in the trailer. In case that vehicles operates in explosive atmosphere the box must satisfy some specific requisites ruled by ATEX directive (94/9/CE).

Figure 1. Electronic counter on the trailer

The concentrator is equipped with a GPS antenna and a GPRS transmitter/receiver in order to detect in real time the position of the vehicle and all the parameters relative to the vehicle's operation and the state of the load. All the equipment and the wiring used must conform to the CEI law and be useable in areas with a possibility of explosion.

The concentrator performs activities related to data collection from all the on-board sources, data processing, addition of geo-referenced information (by a GPS receiver) and packing and sending messages on multiple queues.

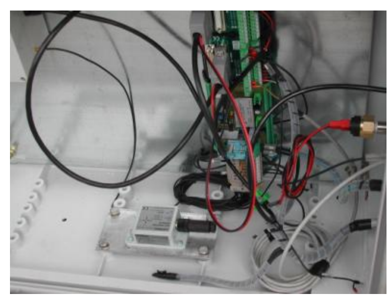


Figure 2. Accelerometer/inclinometer

Such a kind of information must be transmitted at predefined fixed time intervals (e.g. 5 minutes). The state of the truck is described by the following information: the position of the vehicle, speed and direction (by GPS and odometer), inclination with respect to the two axes, air suspension pressure, CANBUS information, amount, temperature and pressure of product, state of the load (data comes from the electronic counter).

In the transport of high risk DG, the ADR 2005 introduced the creation of safety plans, that had to include the detection of the dangerous goods and the evaluation of the critical operations and the risks to safety also derived from the rest breaks of the vehicles and/or pauses of the goods in the vehicles. With this aim, other information related to particular events detected with the device installed on board the vehicle is:

- Start / stop: raised when electronic meter has turned on/off
- Loading / unloading: these events contains relevant information about operation in depots and service stations
- Loading door opening / closing: raised at the start / end of loading and unloading operations
- Overturning Risk Alarm: raised contextually with threshold crossing for inclinometer sensor
- GPS signal loss: raised when the concentrator looses the GPS signal due to low coverage or HW anomalies
- Emergency buttons: raised when driver push the proper button in case of emergency

2.1.2 Transmission System

Messages should be sent by a GPRS module equipped with one or more SIM-cards. SIM redundancy copes with the frequent issue of the signal loss due to the GSM uncovering of some areas. Italian Department of Infrastructure and Transport is evaluating the realisation of a SIM card dedicated to DGT. It should work with every mobile operator and it should have transmission priority. Another experiment regards also the introduction of satellite transmission, that can increase the territory coverage. Different technologies can be used to perform transmission between concentrators and a remote server over the net: socket connection over UDP, Socket connection over TCP or a Web Services.

The Web Service is a software system that "exposes" a public interface (webAPI) to be called remotely to exchange data over the net. Messages are "encapsulated" inside a XML envelope (coded with SOAP standard) and transmitted using HTTP over TCP. The communication is bidirectional: the server can answer to the remote clients with a simple acknowledgment or more complex commands. Web Service grants a solid infrastructure and can work on interoperable platform but the high overhead increase the costs considerably.

One of the most relevant aspects in the client/server data transmission is the definition of a common data format for the messages packed and sent by the concentrators. Such kind of standardisation make the higher layers of the information system independent of different HW infrastructures sited on the trucks. A good data format must be scalable and consistent with older versions. It must code both the request and the response with an efficient diagnostics management. It must clearly define the rate of transmission distinguishing between standard-frequency and on-event messages. The first ones periodically give information about the truck state (telemetry data), the second ones notify a particular on-board event. The approach used in the DGT IS consists of a "fixed part" which does not depend on the class of data followed by a "variable part" whose length and format depend on the kind of data. The "fixed part" includes information about creation and transmission date, vehicle identifier, coordinates and kind of transmitted data.

2.1.3 Database

The approach to the database layer must include:

- Efficient data storage for real-time messages received from the remote vehicles
- Good management to front-end application requests (e.g. GIS –based application)
- Backup and recovery strategy

This database receives all the raw messages from the Web Service and stores data for the diagnostic activities. The high rate of transmission can lead to a danger of table locking. In order to avoid this issue it is suggested that messages are not immediately parsed: an external application could provide this operation at regular intervals (e.g. 1 minute), moving the unpacked data to the main DB server (M-DB). The M-DB server receives data from the T-DB via application server, provides data to the web applications, collects generic data (e.g. Trucks Registry, Users authentication) integrating them with the real time information and finally provides efficient mechanism to historicize daily data. The M-DB backups all data on a local storage device. The application server sends the backup files to the Backups-DB (B-DB). It can quickly substitute the M-DB in event of failure with minimum data loss and, periodically, the application server provides to delete oldest backup files (e.g. older than one month).

2.1.4 Geografical Information System (GIS)

A geographic information system (GIS) integrates hardware, software, and data for capturing, managing, analyzing, and displaying all forms of geographically referenced information. Considering DGT a GIS can assist to:

- Data presentation

- Risk analysis
- Decision support, i.e. in the definition of the vehicle routing optimization.

The proposed GIS-based application allows to use effective graphic interface, high scalability, a method to retrieve information from the interface and the ability to perform geographic calculation. More in details, available information includes maps, geocoding /reverse geocoding, routing, proximity researches.

The Web server operates in order to provide the information requested by the user and in particular it uses the procedures to receive the client GIS request, to call the Web map server to obtain the desired information and, if necessary, to process the received data and send the data back to the client.

2.2 Transport Integrated Platform (TIP)

The sections currently present in TIP are Control Room, Remote Monitoring, Trip Planning, Technical Management, Service Stations, Training, Accident Reporting, Document Management, Quality, Compliance / audit, Operational control, Performance Index.
Each section has specific features and is dedicated, but the data are shared in order to keep the control of each aspect of the transport services.

Control Room
The control room ensures a continuous, real-time monitoring of vehicles equipped with remote control, both during the day services both where the service is done at night. The control room is equipped with advanced tools for managing alarms and anomalies detected during transport.

Remote Monitoring
This module is equipped with real-time monitoring of transport through the representation of data in tabular form and on geo-referenced maps and interactive maps. This allows the users to constantly monitor the operations and to extract and export data about the position of the trucks and the quality and quantity of the products loaded and unloaded.

Trip Planning
The planning section is used to provide comparison in real time between the service planned and the service implemented. It's possible to compare the trip programmed with the data coming from the remote monitoring system installed on board.

Technical Management
All the technical controls of the trucks are included in this section and are constantly updated during each inspection. Are available, in addition to the controls, all the expiring dates (audits, insurance, fire extinguishers, certificates, etc...) In this section are also available the technical data of rail tank wagons.

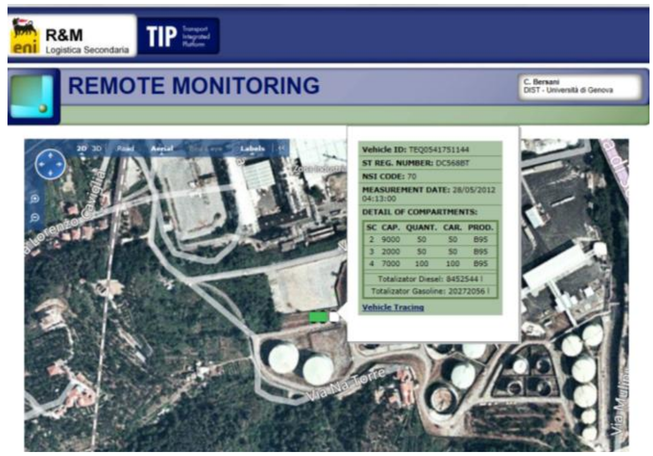


Figure 3. Remote Monitoring Module

Service Stations
This feature provides the data of the level of products that is in underground tanks at service stations. This makes possible to know the volume of product in the service sations in order to check the correctness of the unloading of the vtrucks and automatically generate orders for delivery.

Training
With the Training module is possible to create training sessions. After each course is concluded, the tests are automatically generated by the system and the results are stored in the data base. The training section is integrated with the service Safety Game, an innovative feature that permits to send to the users a daily question about the courses that he attended.

Accident Reporting
The section of accident reporting is designed to collect information on any anomalous event that has caused damage or injury. All event information is placed in the database and classified by type of service and product transported, in order to be analyzed, exported and used for statistical purposes.

Document Management
All documents relating laws, regulations, standards and business processes are classified and available in this module to ensure a continuous updating and better sharing of information with suppliers.

Quality
The management of information flows, procedures, definition of roles and responsibilities and verification of

operation of business processes are managed in a quality system. This module provides all the tools (non-compliance, improvement programs, complaints, corrective actions, etc...) to keep under control the Quality Management System in Secondary Logistics.

Compliance / audit

Continuous assessment of compliance with laws and procedures is the best way to detect abnormalities in business processes and ensure continuous improvement. This module is available to suppliers in order to make self-assessments and audits concerning: Environment, Fire Prevention, Safety (D. Lgs. 81/08), Relevant accidents (D. Lgs. 334/99), ADR, RID, IMDG, LPG Storage, Fuel Storage, Quality.

Operational control

The section of operational control is used in combination with technical checks to verify the proper implementation of the operation of the service, procedures and standards contained in the contracts of transport. The correct application of procedures for loading, transport and unloading of goods is the basis for good customer service and to prevent accidents and injuries.

Performance index

All modules above make it possible to produce "dynamic" performance indexes, which are generated in real time. In this way there is a continuous monitoring of performance and it's possible to make comparisons between different suppliers on the basis of objective parameters.

3 Conclusions and Future applications

The proposed Transport Integrated Platform (TIP) is based on dynamic, real time, as well as static data, and takes advantages combining the latest advanced telematics technologies, quantitative risk analysis and methodologies, optimization theories, and dynamic data collection systems, all incorporated into a integrated system.

Starting from TIP development, the main challenge in the next future on the dangerous good transport framework is to create different Regional Control Centres in order to monitor and manage a regional area by an integrated platform which gathers and processes in real time vehicle, traffic and environmental data and other important information provided and used by the different actors which are involved in the dangerous goods transport.

In the SECTRAM project (2010-2012), developed in the framework of the Interreg ALCOTRA programme, a first Transbordary Monitoring and Control Centre (TMCC) for the Alps-Mediterranean Euroregion has been realized. SECTRAM aimed at promoting a safe management of freight transport in the cross-border area between Italy and France and to supply objective information to competent bodies on the nature, number and routes of the vehicles.

Significant efforts should be dedicated by Public Institutions to involve other network of users (DG fleet manager, road infrastructure managers, local public authorities, etc..) to adhere to the TMCC. It is necessary to extend the monitored territory and to guarantee a real-time service in order to control traffic and provide data to truck drivers and road user. The sharing of real time information on weather, traffic, emergency represents the main element to reduce risk of accidents and minimize consequences for people and environmental.

Reference

[1] EUROSTAT, accessible to the website http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database. Last access 04/03/2012.

[2] Bersani C., Minciardi R., Tomasoni A.M., and Sacile R.. "Risk averse routing of hazardous materials with scheduled delays" in Security and Environmental Sustainability of Multimodal Transport, ED. M.G.H. Bell, S. H.Hosseinloo, U. Kanturska, NATO Science for Peace and Security Series C: Environmental Security, 2010, Part 1, 23-36, DOI: 10.1007/978-90-481-8563-4.

[3] Roncoli C., Bersani C., Sacile R., 'A decentralised control model for the risk-based management of dangerous good transport flows on road' 2012 6th Annual IEEE Systems Conference (SysCon), 19 Mar - 23 Mar 2012, Vancouver, BC, Canada.

[4] Roncoli R., Bersani C., Sacile R., Minciardi R. "Optimal traffic flow assignment for planned deliveries in dangerous good transportation". IEEE International Systems Conference 2011, 4-7 April 2011, in Montreal, Quebec, Canada.

Proc. of the 2012 7th International Conference on System of Systems Engineering, Genoa, Italy - 16-19 July 2012

Predictive Transportation Control for High Throughput and Short Turnaround Time

Yoshiyuki Tajima
Yokohama Research Laboratory,
Hitachi, Ltd.
yoshiyuki.tajima.hh@hitachi.com

Takashi Noguchi
Yokohama Research Laboratory,
Hitachi, Ltd.
takashi.noguchi.kh@hitachi.com

Takashi Fukumoto
Yokohama Research Laboratory,
Hitachi, Ltd.
takashi.fukumoto.kx@hitachi.com

Abstract - *In factory automation (FA) and laboratory automation (LA), an automation system alternates between transporting work-piece objects to an area that is appropriate for the purpose and then processing them. This kind of automation system needs high scalability. Therefore, the automation system is made up of various combinations of autonomous modules. That is, the automation system is a kind of system of systems (SoS). In the automation system, we often require high throughput (TP) and short turnaround time (TAT) for high priority objects. However, there is a tradeoff between these capabilities. In this research, we propose a queue-based model for defining congestion of the automation system. Then, we describe the difficulty in achieving high TP and short TAT simultaneously. Then, we propose a novel control method that is based on future congestion. A part of the effectiveness has been evaluated by simulations.*

Keywords: System of Systems, Complex Dynamical System, Transportation, Predictive Control, Turnaround Time.

1 Introduction

In factory automation (FA) and laboratory automation (LA), an automation system alternates between transporting work-piece objects to an area that is appropriate for the purpose and the processing them [1][2][3]. This kind of automation system needs the ability to grow depending on the scale of the task. To be scalable, the automation system should consist of autonomous modules. Therefore, the automation system is made up of various combinations of autonomous modules. That is, the automation system is a kind of system of systems (SoS). This belief is evident in this research [4]. In the automation system, we often require high throughput (TP) and short turnaround time (TAT) for a high priority object [5]. However, there is a tradeoff between these capacities [6]. For example, when TP is at a maximum value, TAT cannot be guaranteed. It is hard to control the relationship in a large and complex dynamical system that has some bifurcations or loop lines because the automation system has many delay elements. This means that the automation system cannot control the relation by only using information on the current state of the modules. This fact complicates the design of the automation system, making the decision-making of users difficult.

In this research, we show the relationship between TP and TAT in the automation system and propose a prediction-based transportation control method that allows a high TP and guarantees TAT. In this paper, first we describe the relationship between TP and TAT. Then, we propose a queue-based model and define congestion in the automation system. In the next section, we explain the results of a preliminary experiment. In this experiment, we describe efficiencies and challenges of simple control using information of congestion. Finally, we propose a new control method that is based on future congestion, and demonstrate that the method can achieve high TP and short TAT simultaneously.

2 Automation System

In this section, we describe a typical automation system, and explain the relationship between TP and TAT.

2.1 Typical System

In the automation system, a conveyor belt and an automated guided vehicle are used to transport objects. Especially, in the system using advanced automation, the conveyor belt is often used. Figure 1 shows a typical automation system using the conveyor belt.

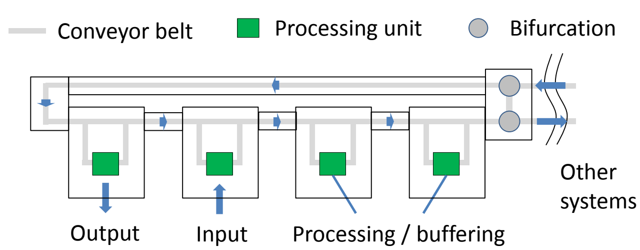


Figure 1. A Typical Automation System

In the automation system, each module should continue to process the object without a pause maintain high TP. However, each module often has a different processing time. When there are some time-consuming modules, the surrounding modules cannot continue to process the object. To deal with this, the module has some buffer areas for the objects as shown in Figure 2. The automation system can continue to work if it has adequate buffer areas.

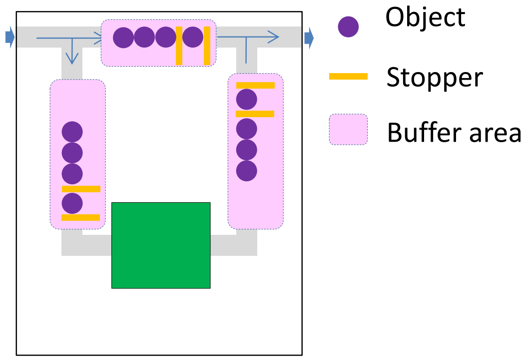


Figure 2. Buffer Area in a Module

2.2 Relationship between TP and TAT

When the system receives the objects in amounts greater than its capacity, the buffer areas become filled with the objects. This is good in terms of TP but not TAT. This is because the module always works, but a new object has to wait for the objects in the buffer areas to be processed. Figure 3 shows the relationship between TP and TAT. When input TP is sufficiently less than the processing capacity of the automation system, TAT does not change and is almost always very short. However, when input TP becomes close to the limit of the processing capacity, the relationship becomes a trade-off and TAT rapidly becomes longer. In a complex system, the limit is not clear. In addition, the system will be expanded and reduced with time. Therefore, high TP and short TAT are difficult to support simultaneously.

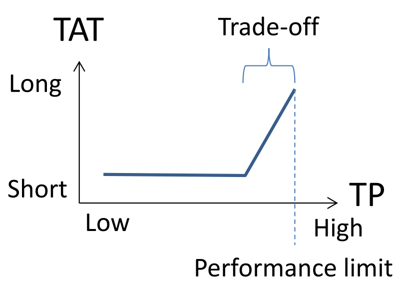


Figure 3. Relation between TP and TAT

3 Model for the Automation System

The method proposed in the later section is based on congestion of the automation system. In this section, we propose a queue-based model for the automation system and define its congestion.

3.1 Conventional Techniques

There are some techniques for modeling the automation system [7]. Queueing Theory [8] is often used to model the automation system [9][10][11][12]. The theory enables the system to be analyzed stochastically. When we can describe input arrival transport, and service time in a typical probabilistic model, we can use this theory easily. However, in the complicated system, they are difficult to describe with simple models. Unfortunately, it is hard to analyze the behavior of the automation system because we often cannot describe the model of a module's service time or an arrival distribution of objects with a typical stochastic model. Therefore, a queue based simulation is used. Some simulators can describe the behavior in programming languages.

3.2 Proposed Model

It is effective to control the number of objects in the automation system to curb an increase in TAT. In other words, stopping the spread of congestion is a good idea for achieving high TP and short TAT. We have to define congestion to control it. Therefore, we propose a queue-based model that can define congestion simply and naturally. Then, we focus on the state transition model of the object to model the automation system. Figure 4 shows the state transition model of the object. The object in the automation system has (a) a processing state and (b) a transporting state.

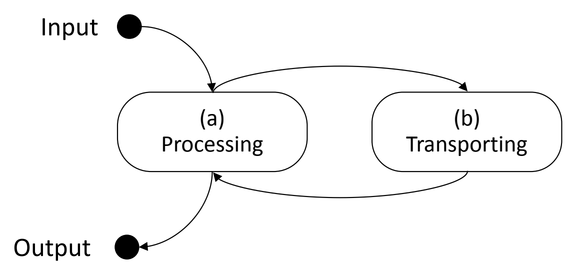


Figure 4. State Transition Model for an Object

The automation system often processes the objects in series. In this paper, on the premise that the module processes the object in first-in first-out (FIFO) manner, we model these states with two queues. The first queue is for the area that processes the object. The second queue is for the area that carries the object. We call the first queue "service area" and the second queue "transport area". We define a pair of these queues as a "logical module" and build the whole system using it. The simulator using this model can simulate the behavior of the automation system when we define the processing time of the service area and the transporting time of the transport area. For example, a module in Figure 2 is expressed like Figure 5 with three logical modules.

In this model, sojourn time in the service area corresponds to processing time of a stopper or a processing unit. In the same manner, sojourn time in a transport area corresponds to transport time between logical modules.

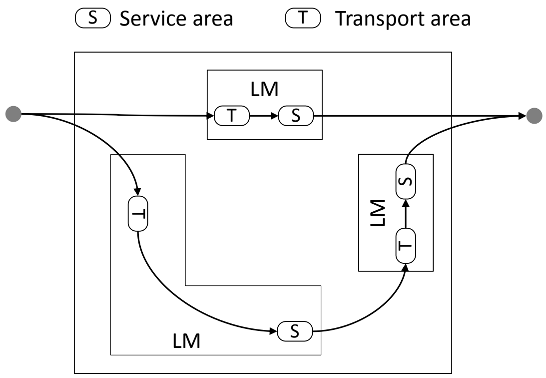


Figure 5. The Example Model of a Module

3.3 Definition of Congestion

In general, sojourn time in the transport area is shorter than sojourn time in the service area. The service areas should be filled by objects, and the transport areas should be kept a certain level of congestion. Therefore, we defined the congestion as the number of the objects in the transport areas. Expression (1) shows a definition of the congestion. Here, let I be defined as a set of logical modules for calculating the congestion, let $\mathrm{C_t}$ be defined as the number of the transport area in the i th logical module, and let $\mathrm{L_t}$ be defined as the number of the capacity of the transport area in the i th logical module.

$$Congestion = \frac{\sum_{i \in I} C_t(i)}{\sum_{i \in I} L_t(i)} \qquad (1)$$

4 Model Evaluation and Challenge Clarification

In this section, we evaluate our model first. Then, we explain efficiency and a challenge of current congestion-based input control.

4.1 Model Evaluation

We make an experiment with a simulator that is based on our model to confirm it. Figure 6 shows an example automation system for the experiment.

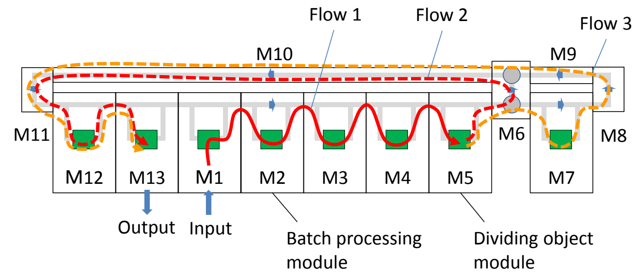


Figure 6. Example System

This system has several types of modules that include some bifurcations. Each module has enough buffer areas on the conveyor belts. M2 is a batch processing module. The cycle of the module is 5 minutes. M5 is a split processing module. All objects are put into M1 in this figure. An object goes along the line of Flow 1. When the object reaches M5, it splits into two objects. Then, one object goes along the line of Flow 2, and another object goes along the line of Flow 3. Finally, these objects are output at M13. TP of the modules are an average of 800 object/h except for M6. TP of M6 is an average of 1600 object/h because it becomes a confluence point of two flows. We add white noise to the processing time of each module.

In this experiment, we measured a time change in average TP and average TAT for the 30 minutes. Figure 7 shows the results with non-control. In this figure, the horizontal axis expresses time (min), and the vertical axis expresses TP (object/h) and TAT (sec). TP(M1) is input TP, and TP(M13) is output TP. TAT is the time of the object by way of flow1 and Flow 2. As a result, the simulator that is based on our model works well and recreates the situation of terrible congestion. In fact, during some period of later time, the input settled down to 400 object/h and the output was settled down to 800 object/h. TAT increased about 1200 sec. This depended on batch processing and a fluctuation of the processing time in which there is a wave in TAT.

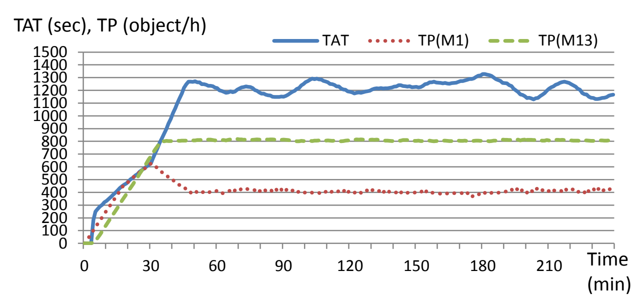


Figure 7. Non-Control

4.2 Efficiency and Challenge of Current Congestion-based Control

Input control is one solution to curb the objects in the automation systems. We did some experiments with current congestion-based input control to find out its impact. The input control is achieved by a method in which M1 takes in an object with the interval that is defined as an expression (2). In this expression, K_c is a constant term, K_p is a proportional gain, and Ref is a reference of the congestion that is defined by the expression (1). When the interval falls below the performance of M1, it is set at 4.5 because the maximum TP of M1 is 800 object/h.

$$Interval = K_c + K_p(Congestion - Ref) \qquad (2)$$

Figure 8 shows the first results with the input control. In this case, we set the following parameters: $K_c = 4.5$, $K_p = 50$, $Ref = 0.1$. The experiment shows that the automation system maintains high TP and curbs the increase in TAT. Figure 9 shows the second results with the input control. In this case, we set the following parameters: $K_c = 4.5$, $K_p = 50$, $Ref = 0.2$. The experimental results reveal that the automation system maintains high TP. However, TAT is a little longer than in the first results. Figure 10 shows the third results with the input control. In this case, we set the following parameters: $K_c = 4.5$, $K_p = 200$, $Ref = 0.2$. The experimental results reveal that the automation system maintains high TP. TAT is almost always short, but becomes oscillatory.

The automation system maintains high TP and curbs the increase in TAT when we set appropriate parameters for the controller. However, when the scale of the automation system becomes large, the state of congestion becomes oscillatory. Moreover, when we set parameters that are too small to control the number of objects, the automation system cannot prevent the congestion. On the other hand, when we set parameters that are too big to control the number of objects, some modules are idle. To summarize these points, the parameter turning was a big challenge in this control method.

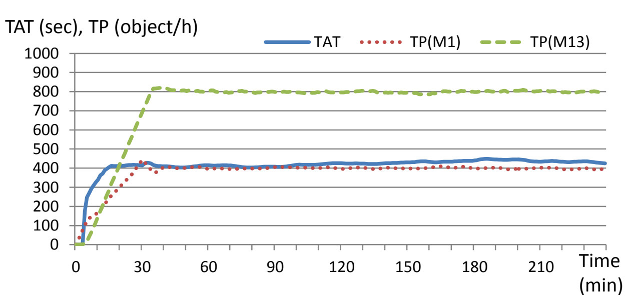


Figure 8. Proportional Control: Result 1
($K_c = 4.5$, $K_p = 50$, $Ref = 0.1$)

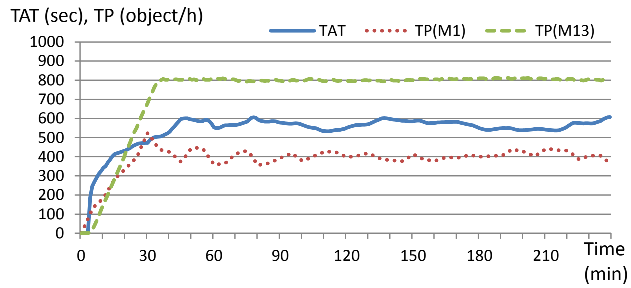


Figure 9. Proportional Control: Result 2
($K_c = 4.5$, $K_p = 50$, $Ref = 0.2$)

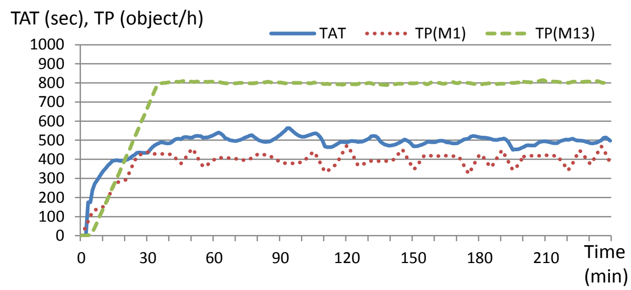


Figure 10. Proportional Contorl: Result 3
($K_c = 4.5$, $K_p = 200$, $Ref = 0.2$)

5 Proposed Predictive Transportation Control and its Evaluation

In this section, we propose a predictive control method based on future congestion and show the efficiency of the method.

5.1 Approach and Algorithm

A difficulty of the current congestion-based control is that a control order is reflected in TAT with a time lag, which is not negligible. The larger the automation system, the longer the time lag. Characteristics of the object also affect the time lag. The control using only the current congestion does not perform well. Smith's method [14] is known for the system that has a time lag. This method adds a compensator to PID controller to take account of time lag. However, it is difficult to apply the method to this problem because the method needs a transfer function of the system to design a controller. Another method is model-based predictive control [15]. This method can take account time lag by using the inner model of the system that has no need to be expressed as a transfer function.

In this research, we incorporate the model-based predictive control in the method previously described for the simultaneous pursuit of high TP and short TAT. The proposed method takes account of future congestion that is generated by the inner model. Figure 11 shows the algorithm of the proposed method. First, the automation system senses each object and sends this information to the simulator. Next, it makes some plans. Then, it calculates the effect of each plan by using future congestion that is predicted by the simulator. Specifically, the effect is defined as the difference between the future congestion and a reference. The reference is a steady-state situation that reaches the target for TP and TAT in a high-load operation. Finally, the automation system selects the plan with the smallest difference.

$u \in U$: Interval plan
U: Set of interval plan
τ: End time of prediction
$E(u)$: Evaluation of plan u
t, c, s_0, s : Temporal value
Δt: Step size of simulation

(1) $s_0 \leftarrow$ Current state (sensed by the system)
(2) For each $u \in U$:
(2.1) Initialize simulator
(2.2) $t \leftarrow 0, c \leftarrow 0, s \leftarrow s_0$
(2.3) Repeat until $t > \tau$:
(2.3.1) $s \leftarrow$ Next state of s (generated by the simulator)
(2.3.2) $c \leftarrow c + |\frac{\sum_{i \in I} C_t(i)}{\sum_{i \in I} L_t(i)} - Ref|$
(2.3.3) $t \leftarrow t + \Delta t$
(2.4) $E(u) \leftarrow c$
(3) Output $arg\ min_{u \in U} E(u)$

Figure 11 Future Congestion-based Predictive Control

5.2 Evaluation Experiment

First, we perform an experiment with same condition as the above experiments. In this experiment, let interval plan be defined as {4.5sec, 6.0sec, 7.5sec, 9.0sec, 10.5sec, 12.0sec, 15.0sec, 20.0sec}. End time of the prediction is 30 minutes later. Moreover, we set Ref in 0.2. Figure 12 shows the results of this experiment. The results reveal a good performance in terms of TP and TAT. Table 1 compares the results with those of other methods and parameters. TAT of the proposed method was shorter than proportional control in almost the same TP.

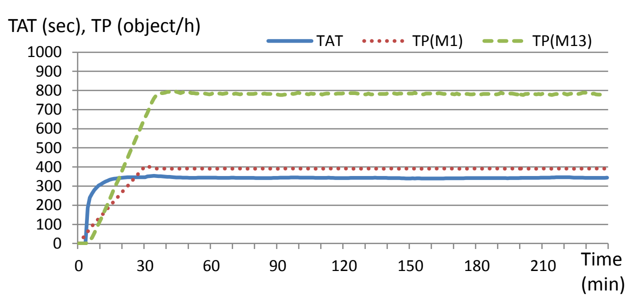


Figure 12 Proposed Control ($Ref = 0.2$)

Table 1 Average of TP & TAT in the Experiments

	Kp	Ref	TAT [sec]	TP(M1) [object/h]	TP(M13) [object/h]
Non - Control	–	–	1079	406	729
Proportional Control (Kc=4.5)	50	0.1	408	376	722
	50	0.2	532	383	724
	200	0.2	474	377	720
Predictive Contorol	–	0.2	334	363	705
	–	0.25	445	378	723
	–	0.3	481	381	724

Second, we perform an experiment with a processing change. In this experiment, M5 does not split the object in the first 2 hours. That is, terrible congestion does not occur. In the next 2 hours, M5 splits the object, meaning, terrible-congestion can occur. Figure 13 shows the results of non-control. Figure 14 shows the results of proportional control. Furthermore, Figure 15 shows the results of the proposed control. For the control, we choose each parameter in such a way that TP is almost 600 object/h. The simulator for the proposed control can predict change in M5. As a result, the proposed control can curb an increase in TAT.

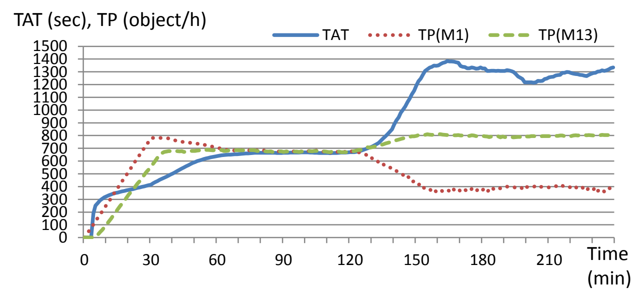


Figure 13. Non-Control in Variable Situation

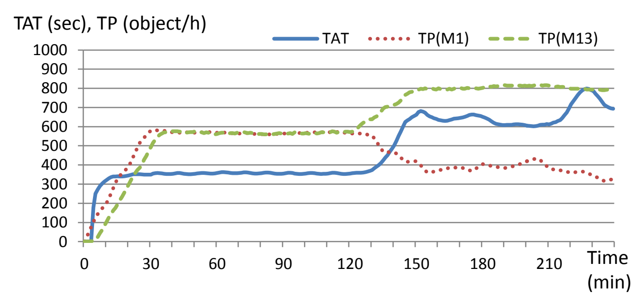


Figure 14. Propotional Control in Variable Situation

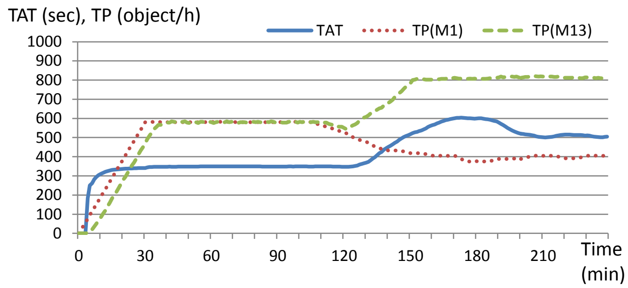


Figure 15. Proposed Control in Variable Situation

6 Conclusion

This research aimed to determine the balance between TP and TAT by transportation control. First, we described the relationship between TP and TAT. Then, we focused on the state transition of an object, and proposed a model for the automation systems. This model can define congestion naturally. Then, we explained the results of the preliminary experiment. In this experiment, we found efficiencies and challenges of simple control using information of congestion. Then, we foccused on the time lag in the automation system, and proposed a novel control method that is based on future congestion. The method can achives high TP and short TAT. It can also deal with change.

A future challenge is testing the effectiveness of the method in more detail. In addition, a learning function must be built to reduce the difference between the simulator and the actual system, otherwise the method will not work well.

References

[1] Richard S. Seaberg, Robert O. Stallone, and Bernard E. Statland, “The Role of Total Laboratory Automation in a Consolidated Laboratory Network”, Clinical Chemistry, vol. 46, no. 5, pp. 751-756, 2000.

[2] Agrawal, G.K. and Heragu, S.S.,“A survey of automated material handling systems in 300-mm SemiconductorFabs”, IEEE Transactions on Semiconductor Manufacturing, vol.19, no.1, February 2006.

[3] Peyman Najmabadi, Andrew A. Goldenberg, and Andrew Emili, “A Scalable Robotic-based Laboratory Automation System for Medium-sized Biotechnology Laboratories”, International Conference on Automation Science and Engineering, Edmonton, Canada, August 2005.

[4] Bo Zhou, Dvoryanchikova, A., Lobov, A., Minor, J., Martinez Lastra, J.L., “Application of the generic modelling method for system of systems to manufacturing domain”, IECON 2011 - 37th Annual Conference on IEEE Industrial Electronics Society, pp. 352 – 358, November 2011

[5] Hee-Jung Chung, Woochang Lee, Sail Chun, Hae-Il Park and Won-Ki Min, “Analysis of Turnaround Time by Subdividing Three Phases for Outpatient Chemistry Specimens”, Annals of Clinical & Laboratory Science, vol. 39, no. 2, 2009.

[6] Anne M. Spence and Douglas J. Welter, “Capacity planning of a photolithography work cell in a wafer manufacturing line”, IEEE International Conference on Robotics and Automation. Proceedings, vol.4, pp. 702 - 708, March 1987

[7] Mark F. Russo and Alan Sasso, “Modeling, analysis, simulation and control of laboratory automation systems using Petri nets: Part 1- Modeling”, JALA - Journal of the Association for Laboratory Automation, April 2005.

[8] Kendall,D.G., “Some Problems in the Theory of Queues”, J.Roy.Statist.Soc.,Ser.B, vol.13 no.2 pp.151-158, 1951

[9] T. L. Pap and L.Leisztner, “The use of queueing theory for planning automated analytical systems”, Journal of Automatic Chemistry of Clinical Laboratory Automation, Vol. 9, No. 2, pp. 87-91, April-June 1987.

[10] Hoshino S., Ota J., Shinozaki A. and Hashimoto H., “Optimal Design Methodology for an AGV Transportation System by Using Queueing Network Theory”, 7th International Symposium on Distributed Autonomous, pp.391–400 , June 2004

[11] Inoue T., Ishii Y., Takagi H., Muneta T. and Imaoka K. “Conveyer Belt Model to Analyze Cycle Time Conditions in a Semiconductor Manufacturing Line”, ISSM 2006 Conference Proceedings – 13th International Symposium on Semiconductor Manufacturing, September 2006.

[12] Kan Wu, L, Leon F. McGinnis and Bert Zwart, “Compatibility of Queueing Theory, Manufacturing Systems and SEMI Standartds”, Proceedings of the 3rd Annual IEEE Conference on Automation Science and Engineering Scottsdale, AZ, USA, September 2007.

[13] Mark F. Russo and Alan Sasso, “Modeling, analysis, simulation and control of laboratory automation systems using Petri nets: Part 1- Modeling”, JALA - Journal of the Association for Laboratory Automation, April 2005.

[14] Watanabe K. and Ito M., “A process-model control for linear systems with delay”, IEEE Transaction on Automatic Control, vol. ac-26, no.6, December 1981.

[15] Jan Maciejowski, “Predictive Control with Constraints”, Prentice Hall, June, 2000.

Modeling and Optimization of Aircraft Trajectories: a review

Maria Pia Fanti
Department of Electrical and Electronic Engineering
Polytechnic of Bari
Bari, Italy
fanti@deemail.poliba.it

Giovanni Pedroncelli, Gabriella Stecco, Walter Ukovich
Dept. of Industrial Engineering and Information Technology
University of Trieste
Trieste, Italy
giovanni.pedroncelli@phd.units.it
{gstecco,ukovich}@units.it

Abstract: *An important problem of the Air Traffic Management is the determination and optimization of safe aircraft trajectories in order to avoid hazardous weather regions and other aircrafts. This paper presents a review of the main approaches about the modeling and optimization of the aircraft trajectories pursuing two main objectives: i) resolving conflicts with other aircrafts; ii) avoiding hazardous weather.*

Keywords: Aircraft trajectories, Modeling, Optimization.

1 Introduction

As of today, airspace is more and more crowded because the air transportation flow is constantly increasing. The current Air Traffic Management (ATM) is unsuitable for the future expansion of the air traffic. Currently, aircraft routes are planned in advance, using fixed predefined waypoints: this approach simplifies the problem of conflict avoidance but it does not efficiently use the airspace. During the flight, the weather condition along the route is considered but the available forecast information cannot be used in an automated way to plan the aircraft trajectories.

This paper intends to present a review of the main contributions about the modeling and optimization of the aircraft trajectories with these two mentioned objectives: i) resolving conflicts with other aircrafts; ii) avoiding hazardous weather.

Considering the first problem, the main approaches can be found in the Free Flight research field. The Free Flight concept has been conceived as a new and better ATM system in which every pilot can select the optimal route for the aircraft in respect of safety constraints, avoiding conflicts with the other aircrafts. In this environment, automated mechanisms of trajectory generation for aircrafts are one of the key elements for the optimization of the use of the airspace and deal with the search of the optimal path from origin point to destination point which is conflict-free. Concerning the second problem, weather avoidance is one of the most important challenges of ATM because the weather-related delays increase as the volume of air traffic [1]. This is because an aircraft must remain separated from one another as well as from hazardous weather: a maneuver designed to avoid weather will affect neighbor aircrafts. The trajectory optimization in presence of hazardous weather phenomena deals with the search of the optimal route of deviation from weather encounters. In this paper, the main trajectory optimization models conceived for the weather avoidance issue are described along with the relative weather model.

2 Modeling and optimization for conflict detection and resolution

In the Free Flight environment, we consider each aircraft surrounded by two virtual cylindrical shapes, the protected zone and the alert zone. A conflict or loss of separation between two aircrafts occurs whenever the protected zones of the aircrafts overlap. The goal for the Conflict Detection and Resolution (CD&R) system is to predict that a conflict is going to occur in the future, communicate the detected conflict to a human operator and, in some cases, assist in the resolution of the conflict situation. In this framework, conflict detection can be thought as the process of deciding when action should be taken and conflict resolution involves determining how or what action should be performed. Interested reader can refer to [2] for an extensive review of CD&R methods, where the authors classified the contributions according to the different approaches used for the resolution of the conflicts, maneuvering options and management of multiple aircraft conflicts. Moreover, a specific review on decentralized methods is presented in [3]: the authors classify the approaches in optimization and stochastic methods and focus on conflict resolution in a decentralized environment.

From the model point of view, the Conflict Resolution modules are developed in different frameworks: some approaches consider the airspace as a multi-agent environment in which aircrafts are the agents occupying zones of the space. Some of these methods ([4], [5], [6]) use the principles of the game theory and usually assume the cooperation of the aircrafts which are considered as agents in negotiation. In these environments, the zones in which the airspace is divided are considered as resources contended by the agents. For example, [6] is a recent approach to the Conflict Avoidance problem, solved by means of the game theory. This method is agent-based, cooperative (negotiation-based) and decentralized: nearby-located aircrafts share their future intentions (as a limited part of their flight plans), providing a 3D description of their position in time; in this way conflict detection can be implemented. In the case a conflict is predicted, conflict resolution is necessary and the method presents two different algorithms: i) Iterative Peer-to-Peer Conflict Avoidance (IPPCA); ii) MultiParty Conflict Avoidance (MPCA). IPPCA works with multiple negotiation rounds until the conflict is solved while MPCA solves the conflict with a single algorithm run. Both approaches use a set of evasion maneuvers which are function of the sequence of waypoints which the aircraft intends to reach.

Other approaches ([7], [8], [9]) make use of the graph theory for the determination of the conflict resolution. In this case, the airspace is discretized as a grid with each node of the graph representing a zone of the space. These approaches usually do not assume cooperation between agents, that is, agents are greedy and no inter-agent communication takes place during the conflict resolution. For example, in [8] the airspace is presented as a discrete system partitioned in cells: the safety requirements are satisfied by guaranteeing that only one aircraft (an agent) can occupy a cell at any time. Therefore the airspace is modeled by an undirected graph where the vertexes represent the cells and the edges connect two adjacent cells. In this method, assuming that each agent will follow a determined path (a sequence of cells), a conflict is detected if at a certain time, two or more agents would occupy the same cell (disputed resource). The resolution of the conflict is performed by defining a set of rules of prioritization of the disputed resources.

Other CD&R methods are based on optimization of a cost function for the resolution of conflicts; this function is usually associated with a set of cost metrics. Each method defines an appropriate cost function which can minimize time of flight, fuel, projected separation, workload, etc. For example, in [10] two different problems are proposed: i) aircraft allowed to only change the velocity of flight (VC problem); ii) aircraft allowed to only change the direction of flight (HAC problem). For the mentioned problems the authors propose different cost functions.

Other methods for the Conflict Detection and Resolution problem in a Free Flight environment are based on a path planning approach: [11] presents a cooperative, decentralized path planning method which models as Mixed Integer Linear Programming (MILP) problems and makes use of the receding horizon strategy which will be further discussed in the following section. Regarding safety requirements, this method does not consider protected and alert zones around the aircrafts but defines the safe state: a vehicle is in a safe state, if from that state, there exists a known dynamically feasible trajectory to a state or sequence of states that is obstacle- and collision-free, and in which the vehicle can remain for an indefinite period of time. Considering a time horizon discretized in T steps, the cost function is the following:

$$\min_{x_k,u_k} J_T = \sum_{k=1}^{T} (q' \left| x_k - x_f \right|) - r(p_f - p_0)' v_k \quad (1)$$

where $\mathbf{p}_f = (x_f, y_f)$ is the destination position of the aircraft, $\mathbf{p}_0 = (x_0, y_0)$ is the initial one, $\mathbf{q}$ and r are weights, $\mathbf{x}_k$ and $\mathbf{x}_f$ are respectively the current (at time step k) and final state of the system, $\mathbf{v}_k = (\dot{x}_k, \dot{y}_k)$ is the current speed and $\mathbf{u}_k$ is the sequence of inputs $\mathrm{u}_{i,k+l}$ with $l = 1,\ldots,(T-1)$.

In the field of the trajectory modeling and optimization problem, we recall two additional contributions. In [12], the trajectory optimization problem is solved with the use of Dynamic Programming (DP), considering time constraints. For a single aircraft, the following are the equations of motion:

$$\begin{aligned} \dot{x} &= (V + V_W)\cos\gamma \\ \dot{h} &= V \sin\gamma \end{aligned} \quad (2)$$

where x is the range, h the altitude, V the airspeed, V_W the horizontal component of wind speed and γ the flight path angle, considering an aircraft with no vertical acceleration and a small γ. The objective function is the following:

$$J = \int_{t_i}^{t_f} (f + CI)dt = \int_{x_i}^{x_f} \frac{(f + CI)}{(V + V_W)\cos\gamma} dx \quad (3)$$

where $[(V + V_W)\cos\gamma]$ represents the ground speed of the aircraft, f is the fuel flow rate, and CI is the Cost Index, a fuel equivalent cost for the time cost component. The second equality in (3) is derived by the fact that time is a space-related variable (since time constraints are defined at specific points of space) while space is an independent variable. The problem is then solved using the DP approach after the discretization.

Finally, in the modeling of trajectories contexts, we recall paper [13] where the trajectories are considered in a 4D airspace and the cost function is then solved via pseudo-spectral methods.

3 Weather avoidance

Weather is one of the leading causes of delays for aircrafts and such a weather-related delays increase as the volume of air traffic increases. In addition to that, in a centralized air traffic management system, there will be an additional workload for the controller who will have to perform tactical decisions concerning the weather impact reduction. Weather can also cause damages to the electronic and navigational equipment of an aircraft leading to the pilot losing control and endangering passengers' lives; these problems are mainly caused by strong winds and low visibility. With the concept of Free Flight, the weather avoidance problem will become more complicated as pilots will have autonomous control of the aircraft. Efficient and robust weather avoidance algorithms will be needed to ensure safety and better fuel consumption economics resulting from optimal weather avoidance trajectories.

3.1 Weather models

The first aspect to be analyzed is relative to the modeling of weather events and the way of representing them in order to be avoided by the pilots. Some tools are developed in order to reduce the workload of the controller: there is the need of a fully automated system for decisional support with tools integrating flight information, trajectory modeling and weather forecasts to help the controller (or, in a Free Flight environment, the pilot) performing better decisions and to reduce workload. One of them is the Convective Weather Avoidance Model (CWAM) [14-16]. The CWAM generates estimates of pilot deviation probabilities in the case of weather encounter as a function of Vertical Integrated Liquid (VIL), echo top height and flight altitudes. More precisely, the VIL is the amount of liquid water (precipitation) that the radar detects in a vertical column of the atmosphere and is used to determine the severity of convection. High values are associated with strong convection that can be accompanied with heavy rain or hail. Echo top is the radar observed height of a convective system. These are both products of Corridor Integrated Weather System (CIWS) [17]: these products are deterministic weather information which CWAM translates into estimates of likelihood of pilot deviation. In the following versions of CWAM, new variables are introduced such as additional weather factors (vertical storm structure, vertical and horizontal storm growth, spatial variation in VIL and echo top fields and storm motion) [15] and additional information, including the location where the decision was made to deviate [16]. The products of CWAM are Weather Avoidance Fields (WAFs), 3D grids which represent, at each pixel, the estimates of pilot deviation probability in the encounter of weather. The CWAM outputs are used in some weather avoidance systems: for example, in [18], CWAM forecasts are represented as sets of polygons that define the contours of airspace containing convective weather. There are four subsets of polygons each having a different value of probability that the region of weather will be avoided (40%, 60%, 80%, 90%). These polygons, representing the estimates of probability of deviation, are integrated in the automated planning systems which routes flights. The systems check for weather encounters during the flight and re-plans the route until weather is encounters-free.

Another approach for the modeling of weather regions can be found in [19]: it makes use of the VIL product from the CIWS; the VIL measurements are quantized into 6 levels with level 3 and higher indicating a recommended no-fly zone. Weather maps are created starting from the region of interest, time, date and the VIL data (updated every 5 minutes) which are binary fly or no-fly zones. These regions are then split into separate storms and each storm is then enclosed in a minimum volume bounding ellipse which is an over-approximation of the weather. These elliptic approximations are not optimal but are computationally faster.

The method of trajectory optimization proposed in [20] provides a different kind of weather model which considers the stochastic nature of weather phenomena. This model takes into account weather forecasts with a time interval of 15 minutes and associates to each storm (or bad weather phenomenon) a Markov chain of 2 states where "0" corresponds to the state of having no storm in that region in a particular timeframe and "1" corresponds to the state of having storm in that region in a particular timeframe (for a single storm). If there are m storms, there will be a 2^m Markov chain. A transition matrix of dimension $2^m \times 2^m$ is associated to the Markov chain and it is used as an input of the optimization algorithm which will be discussed in the following paragraph.

3.2 Trajectory modeling and optimization in presence of hazardous weather

In [21], an optimal flight path search algorithm is provided, using the CWAM forecasts. Its purpose is the optimization of airspace throughput and efficiency in the presence of uncertainties that is weather encounters. This method makes use of a DP algorithm to find the shortest path from a starting point to an ending point. The airspace is represented by means of a graph: the nodes represent points of the airspace and the edges connecting the nodes represent the possible trajectories for an aircraft from one point to another. This method provides a provably global optimal solution when a feasible solution exists. The airspace is discretized in a Cartesian grid; therefore latitude and longitude of each point are transformed into Cartesian coordinates. The objective cost function in the analysis can be constructed by adding different components to represent obstacles or constraints. Each individual component can be added to or removed from the objective function depending on the application and available information. The position of the aircraft in the grid is denoted by $\mathbf{x}_{i,j}(t_{i,j})$ where index i denotes an arbitrary state and j denotes an arbitrary stage, i.e., a single iteration of the algorithm. Then, the global optimal solution can be obtained by minimizing the following objective function that is recursively calculated:

$$\mathbf{J}(\mathbf{x}_{i,j}(t_{i,j})) = \min_{i'}[\mathbf{D}_{i,j,t_{i,j}}^{i',j+1,t_{i',j+1}} + \mathbf{W}_{i,j,t_{i,j}}^{i',j+1,t_{i',j+1}} + \mathbf{C}_{i,j,t_{i,j}}^{i',j+1,t_{i',j+1}} + \mathbf{J}(\mathbf{x}_{i',j+1}(t_{i',j+1}))] \tag{4}$$

where $\mathbf{D}$ is the estimated fuel cost, $\mathbf{W}$ the cost of deviation due to convective weather, $\mathbf{C}$ is the cost associated with crossing a congested region, and i' denotes an arbitrary state at the next stage. Moreover, the formula for calculating the cost $\mathbf{W}$ is obtained by the predicted probability of deviation from the given flight path due to convective weather; this probability takes into account every probability of deviation because the pilot can encounter multiple hazardous weather zones. The probability of deviation is calculated for every path candidate and the one with the best probability of deviation is selected. One of the most favorable aspects of this approach is that additional cost components can be included in the equation (4), depending on the objective of the application or available information.

In [19] the trajectory of the aircraft in presence of hazardous weather is generated by means of the receding horizon framework. The trajectories are in two dimensions (on the horizontal plane) and altitude changes are not considered for the avoidance of weather. A reference trajectory (planned in advance) is considered and a cost function is determined to penalize the deviation from this planned trajectory. This cost is minimized in a non-cooperative manner (each aircraft tends to minimize its own cost rather than the sum of all aircraft costs). The equations of motion are the following:

$$\begin{aligned} \dot{x}_1(t) &= v\cos\psi(t) \\ \dot{x}_2(t) &= v\sin\psi(t) \end{aligned} \tag{5}$$

where x_1 and x_2 denote the position of the aircraft in the airspace, ψ denotes the orientation of the aircraft and v is the speed. Considering a set of N aircrafts, the cost function is defined as the L_2 norm of the tracking error of the reference path

$$J(u_1,...,u_N) = \sum_{i=1}^{N}\int_0^T \left\| x_i(t) - x_i^d(t) \right\|^2 dt \tag{6}$$

where $x_i(t)$ is the evolution of the state in the planning horizon T_f of the aircraft i, u_i denotes the input to the aircraft i over the time horizon of interest $[0,...,T_f]$ and $x_i^d(t)$ is the desired path (planned reference path). The optimization problem is then solved via the receding horizon technique: the planning horizon T_f is chosen as the discrete number of steps for which the trajectory planning is performed. After solving the problem in the defined horizon, the horizon is shifted by a certain number of time units (in this case that is 5 minutes, the time for which new weather forecasts are available). Every aircraft which reaches its destination is removed from the optimization in the subsequent iterations. The algorithm terminates when all the aircrafts reach their destinations.

The method proposed in [20] addresses the problem of a single aircraft facing a bad weather zone. The bad weather is modeled as a stationary Markov chain (as we recalled in Section 3.1) and a dynamic programming algorithm is provided for the minimization of expected delay when the aircraft may encounter bad weather. Firstly, a cost is assigned to every zone of the airspace, depending on the probability that this zone will be unusable in a certain time period, due to some sort of constraint (including weather). Then, the airspace is discretized in a rectangular grid and, at each stage, the solution of the problem is the definition of the point in the grid which the aircraft will occupy in the following 15 minutes. The expected cost function depends on the state of the system, the measure of probability of the random disturbance, the initial state and the cost associated with the current stage. The problem is finding a control law μ_k from the set of the admissible control laws $\mu = \{\mu_0, \mu_1, ..., \mu_{n-1}\}$ which minimizes the cost function.

In [22], an algorithm is presented for the path planning of multiple aircrafts with weather avoidance aircrafts: this method has the objective of solving both the weather avoidance and conflict avoidance problems, also considering the effects of strong winds. The aircraft motion is considered in a 2D airspace (horizontal plane only) and makes use of the concept of protected zones to avoid LOS (Loss of Separation) between aircrafts. The convective weather is represented through approximations of VIL values on the map. In this environment, for a single aircraft, the following are the motion equations:

$$\begin{aligned} \dot{x}(t) &= v\cos\theta(t) \\ \dot{y}(t) &= v\sin\theta(t) \\ \dot{\theta}(t) &= u(t) \end{aligned} \tag{7}$$

where u is the optimal input sequence for the aircraft, computed over a horizon T by minimizing a cost function obtained by solving Hamilton-Jacobi partial differential equation. Such a cost function contains the minimum travel point from a point in the airspace to the destination, also incorporating effects of weather avoidance and wind flow. The problem is then solved using Nonlinear Model Predictive Control, starting with the discretization of the continuous dynamics.
The cost function is the following:

$$\min_u J(u) = \min_u \sum_{i=1}^{N} V_i(x_{i,n}(u_i), y_{i,n}(u_i)) \tag{8}$$

where N is the total number of aircrafts, n is the number of steps, $x_{i,n}$ and $y_{i,n}$ denote the trajectory of aircraft i at time step n, that is the last time step (the destination point). In [23], another method for the aircraft trajectories optimization is presented: this method does not take into account weather encounters but aims at the minimization of emission: contrails form in the wake of the aircraft's trajectory because of the emission of water vapor and can have a severe impact on the climate. This algorithm calculates a wind-optimal trajectory for the aircraft while avoiding the region of airspace where it is more likely the formation of contrails. The motion of the aircraft is considered in the horizontal plane and it is described by the following equations:

$$\begin{aligned} \dot{x} &= V\cos\theta + u(x,y) \\ \dot{y} &= V\sin\theta + v(x,y) \end{aligned} \quad (9)$$

where x and y are the aircraft coordinates in the horizontal plane, V is the speed, θ is the heading angle and $u(x,y)$ and $v(x,y)$ are, respectively, the x and y components of the wind velocity. The cost function is the following:

$$J = \frac{1}{2}X^T(t_f)SX(t_f) + \int_{t_0}^{t_f}\left\{C_t + C_f f + C_r r(x,y)\right\}dt \quad (10)$$

where X is the state vector, S is the final state cost matrix, C_t is the delay cost, C_f is the fuel cost, f is the fuel flow rate, C_r is the cost of penalty areas and $r(x,y)$ is the penalty function.

4 Concluding remarks

We summarize the presented methods and approaches to model and optimize the flight paths in Table 1, according to their most important features.

In the first column there is the considered reference and the other columns show the following issues:

- *Dimensions* (Dim): the second column contains the variables used for the modeling of the airspace showing whether the state information used in the model involves the horizontal or vertical plane (2D analysis) or both. It is desirable to have a four-dimensional description of trajectories over time to have a complete observability of the airspace situation: that is the case of [13].
- *Discretization* (Discr.): the third column shows if the airspace is subject to a discretization for the resolution of the conflict or weather problem. If the airspace is discretized, then the optimization is performed by solving MILP [11] problems or applying DP [21] algorithms.
- *Single/Multiple Aircrafts* (Aircrafts): some methods are conceived for a single aircraft (Single) in the airspace: in such a case considering the possibility of conflict with other aircrafts is not considered. Other contributions consider the presence of multiple aircrafts (Multiple) and have to find solutions for the avoidance of conflicts. Such assumptions are pointed out by the forth column of Table 1.
- *Centralization* (Centr.): the last column of Table 1 shows if the weather or conflict problem is solved by a centralized approach (e.g. the airspace controller) or by distributing the task between the different aircrafts in the airspace. This feature is considered only for methods which optimize the trajectory in a multi-aircraft system. In [11], the resolution of the problem is distributed among the aircrafts in the airspace which act in a cooperative manner. This approach leads to the optimization of the available computation resources and is one of the central ideas of the Free Flight concept.

Table 1. Classification of the main trajectory models

Method	Dim.	Discr.	Aircrafts	Centr.
[11]	x,y	✓	Multiple	
[12]	x,y	✓	Single	
[13]	λ,φ,h		Single	
[21]	x,y	✓	Single	
[19]	x,y		Multiple	✓
[20]	x,y	✓	Single	
[22]	x,y	✓	Multiple	✓
[23]	x,y		Single	

Summing up in this paper we point out how the modeling approaches and the optimization techniques are applied in order to optimize the trajectories in air traffic management. In particular, the problem involves a number of potentially conflicting objectives such as minimizing deviations, weather avoidance, reducing emissions, minimizing distance traveled, and hard constraints like aircraft performance. Even if a large number of different solutions are proposed in the related literature, the problem is open and there is a deep gap between the technological strengths and the decision and optimization solutions. Future research should investigate about new modeling and optimization approaches in particular addressing, among the other tasks, sustainability objectives such as the reduction of emissions and pollution.

References

[1] W. Zhang, M. Kamgarpour, D. Sun and C.J. Tomlin, "A Hierarchical Flight Planning Framework for Air Traffic

Management," Proc. of the IEEE, Vol. 1, pp. 179-194, 2012.

[2] J. Kuchar and L. Yang, "A Review of Conflict Detection and Resolution Modeling Methods," IEEE Transactions on Intelligent Transportation Systems, Vol. 1, No. 4, pp. 179–189, Dec. 2000.

[3] D. Dimarogonasand and K. Kyriakopoulos, "Inventory of decentralized conflict detection and resolution systems in air traffic," National Technical University of Athens, Greece, Tech. Rep., Hybridge project, 2002.

[4] S. Wollkind, J. Valasek and T.R. Ioerger, "Automated Conflict Resolution for Air Traffic Management Using Cooperative Multiagent Negotiation," AIAA Guidance, Navigation, and Control Conference, 2004.

[5] J. Hill, R. Johnson, J. Archibald, R. Frost and W. Stirling, "A Cooperative Multi-Agent Approach to Free Flight," Int. Joint Conf. on Autonomous Agents and Multi Agent Systems, Netherlands, pp. 1083-1090, July 2005.

[6] D. Sislak, P. Volf and M. Pechoucek, "Agent-Based Cooperative Decentralized Airplane-Collision Avoidance," IEEE Transactions on Intelligent Transportation Systems, Vol. 2, No. 1, pp. 36-46, March 2011.

[7] S. Resmerita and M. Heymann, "Conflict Resolution in Multi-Agent Systems", IEEE Conference on Decision and Control, pp.2537-42, 2003.

[8] S. Resmerita, M. Heymann and G. Meyer, "A framework for conflict resolution in Air Traffic Management". IEEE Conference on Decision and Control, pp.2035-40, 2003.

[9] Y. Chiang, J. T. Klosowski, C. Lee, J.S.B. Mitchell, "Geometric Algorithms for Conflict Detection/ Resolution in Air Traffic Management" Proc. 36th IEEE Conf. Decision and Control (CDC '97), pp.1835 -1840, 1997

[10] L. Pallottino and E. Feron, "Conflict Resolution Problems for Air Traffic Management Systems Solved with Mixed Integer Programming," IEEE Trans. on Intelligent Transportation Systems, Vol.3 No.1, pp.3-11, 2002.

[11] T. Schouwenaars, J. How and E. Feron, "Decentralized Cooperative Trajectory Planning of Multiple Aircraft with Hard Safety Guarantees," AIAA Guidance, Navigation, and Control Conference and Exhibit, Providence, Aug. 2004.

[12] P. Hagelauer, F. Mora-Canino, "A soft dynamic programming approach for on-line aircraft 4D-trajectory optimization", European Journal of Operational Research, 107: pp. 87–95, 1998.

[13] Bousson, K., Machado, P. "4D Flight Trajectory Optimization Based on Pseudospectral Methods", The WASET Int. Conf. on Aeronautical and Astronautical Engineering, Amsterdam, The Netherlands, September 28-29, 2010.

[14] R. DeLaura and J. Evans, "An exploratory study of Modeling Enroute Pilot Convective Storm Flight Deviation Behavior," 12th Conference on Aviation, Range and Aerospace Meteorology, American Meteorological Society, Atlanta, GA, Jan. 2006.

[15] R. DeLaura, M. Robinson, M. Pawlak and J. Evans, "Modeling Convective Weather Avoidance in En Route Airspace," 13th Conf. on Aviation, Range and Aerospace Meteorology (ARAM), New Orleans, LA, Jan. 2008.

[16] R. DeLaura, B. Crowe, R. Ferris, J. Love and W. Chan, "Comparing convective Weather Avoidance Models and Aircraft-based Data," 89th Annual Meeting of the American Meteorological Society, Phoenix, AZ, Jan. 2009.

[17] D. Klingle-Wilson and J. Evans, "Description of the Corridor Integrated Weather System (CIWS) Weather Products," MIT Lincoln Laboratory Project Report ATC-317, Aug. 2005.

[18] R. Windhorst, M. Refai and S. Karahan, "Convective Weather Avoidance with Uncertain Weather Forecasts," 28th Digital Avionics Systems Conf., Oct. 2009.

[19] M. Kangapour, V. Dadok and C. Tomlin, "Trajectory Generation for Aircraft Subject to Dynamic Weather Uncertainty," 49th IEEE Conf. on Decision and Control, Atlanta, GA, USA, pp. 2063-2068, Dec. 2010.

[20] A. Nilim, L. El Ghaoui, M. Hansen and V. Duong, "Trajectory-based Air Traffic Management (TB-ATM) under Weather Uncertainty," ATM 2001, Santa Fe, New Mexico, 2001.

[21] H.K. Ng, S. Grabbe and A. Mukherjee, "Design and Evaluation of a Dynamic Programming Flight Routing Algorithm Using the Convective Weather Avoidance Model," AIAA Guidance, Navigation, and Control Conference, Chicago, Illinois, Aug. 2009.

[22] J. Pannequin, A.M.Bayen, I.M Mitchell, H. Chung and S. Sastry, "Multiple Aircraft Deconflicted Path Planning with Weather Avoidance Constraints," AIAA Conf. on Guidance, Control and Dynamics, South Carolina, 2007.

[23] B. Sridhar, H.K. Ng and N.Y. Chen, "Aircraft Trajectory Optimization and Contrails Avoidance in the Presence of Winds," ATIO Conference, Fort Worth, TX, Sept. 2010.

Causal Factors Behind the Failed FiReControl Project: a Large-Scale System-of-Systems

Cornelius Ncube,
Software Systems Research Centre,
School of Design, Engineering & Computing,
Bournemouth University,
Poole, Dorset, BH12 5BB, UK
cncube@bournemouth.ac.uk

Abstract—FiReControl was a large-scale system-of-systems that commenced in 2004 and was expected to be complete by October 2009. In 2007, the Department for Communities and Local Government (the Department) contracted a prime Supplier to design, develop and install the complex computer system underpinning the project. However, the project was subject to a number of problems and delays and costs escalated over its lifetime. In December 2010 the Department cancelled the project after concluding that it could not be delivered to an acceptable timeframe and quality. At the point cancellation, the Department estimated it had spent £245 million on the project and calculated that completion would take the total cost of the project to £860 million, more than five times the original estimate of £120 million.

Keywords*:* system of systems; large-scale projects; failure; emergent behaviour; complex systems

1 Introduction

FiReControl was part of the UK government's £1billion investment in the Fire and Resilience Programme to increase resilience, enhance capability and improve efficiency in England's Fire & Rescue Services (FRS) [9]. FiReControl was initiated in 2004 at an original estimated cost of £120m. FiReControl intended to abolish 46 local fire and rescue control rooms around the country and replace them with nine Regional Control Centres (RCC). However, it was eventually cancelled in Dec 2010 at a total estimated cost of over £860m after a series of problems and delays. The FiReControl project was singled out by Members of Parliament as the worst government IT failure in many years. The contract to implement the IT system linking the control centres was not even awarded until a full three years after the project started. The contract itself was poorly designed and awarded to a company without relevant experience. The IT system was never delivered and eight of the regional control centres remain vacant. It is estimated £342m more will be spent in rental costs on the project's purpose-built control rooms which the government is contractually locked into for the next 22-30 years.

In this paper, we report a post-mortem review of the causal factors that contributed to the failure of FiReControl. We conclude the paper by noting that the causal factors that contributed to failure are actually not unique to the FiReControl.

The rest of the paper is organised as follows. Section 2 introduces the FiReControl project. Section 3 describes our methodology. Section 4 describes our findings. Section 5 benchmarks FiReControl against other similar projects and Section 6 concludes the paper.

2 FiReControl: Background and History

This section gives a detailed background of FiReControl by summarising official documents from the National Audit Office (NAO) [4], the UK Parliament's House of Commons Public Select Committee [3] and Department of Local Government and Communities [9]. Although FiReControl was a system-of-systems (SoS), we do not intend to provide a definition, background and related work on SoS. Instead we refer to SoS as defined by Mair [1] and Jamshidi [2].

In 2000, the DLGC commissioned a review of efficiency options for England's Fire & Rescue Service (FRS) control rooms [9]. The review concluded by recommending the amalgamation of independent controls rooms to 21 sub-regional controls. At that time Fire and Rescue Service in England was operating with 46 individual control rooms in each Fire & Rescue Area (FRA).

In 2002, an independent review of the FRS highlighted modernisation of control room operations as a major area. In 2003 another independent review was conducted to determine if the original 2000 conclusions have changed as a result of new requirements for resilience and management of larger-scale incidents that had emerged since the 9/11 terrorist attacks in the USA. That review recommended amalgamation of 46 local control rooms to 9 Regional Control Centres (RCC). In early 2004, the FiReControl project was initiated to replace 46 local Fire and Rescue Service control rooms with a resilient network of nine purpose built Regional Control Centres.

In March 2007, a prime Supplier was contracted to design, develop and install the core IT systems. The Supplier was also contracted to procure and install IT hardware and software systems in nine Regional Control Centres, 46 Fire and Rescue Service headquarters, 1,400 fire stations and 3,400 fire engines and other equipment. As well as designing, developing and installing the core resilient IT system-of-systems, the Supplier was also required to supply operational support services, including fault repair, maintenance and data back-ups until 2015, with an option for a further three-year extension up to 2018.

However, the Supplier subcontracted the majority of the work to third parties and its main role was to integrate different packages from subcontractors to form the overall system-of-systems. The Mobilisation System was the key constituent system and required the integration and customisation of 50 pre-existing Commercial-Off-The-Shelf (COTS) software packages. The main role of the mobilisation system was to control and coordinate the chain of emergency activities: from taking and identifying an incoming emergency call, to dispatching the right resources to the incident site and keeping the crew updated with real-time information.

2.1 The Current System

Currently each FRS in England has its own local control room handling its emergency calls. The 46 control rooms are independent, each with differing levels of technology, different IT systems, business processes and different ways of responding to incidents. However, with FiReControl the nine fully networked RCCs will have access to the same modern technology and data, including local information. Calls will be automatically transferred to other RCCs if one becomes overloaded. Table 1 shows the key differences between the existing system and the future FiReControl system.

Table 1: FiReControl vs the Current System

Current set up	Under FiReControl
Forty-six stand-alone local control rooms that are not networked and use different technology.	Nine networked RCCs across England that have access to the same, compatible, modern technology and data sources.
Limited ability to back each other up in case of a large scale incident, high demand or failure.	The networked RCCs will be able to back each other up effectively and consistently.
Limited ability to mobilise appliances outside their county.	The RCCs will be able to mobilise resources locally, regionally and nationally.

2.2 Scope of FiReControl

The FiReControl project aimed to replace 46 local Fire and Rescue Service control rooms with a resilient network of nine purpose built Regional Control Centres (RCCs). The RCCs were to handle emergency 999 calls, mobilise resources and support the management of incidents. FiReControl involved the design and implementation of a complex system-of-systems that would provide:

- networked access to up-to-date information on the nearest and most appropriate fire appliances for any incident.
- a more resilient system that supports FRSs in responding to major emergencies.
- enhanced capability for dealing with high volumes of emergency calls.
- call handling, mobilisation and incident handling system to deploy the closest fire engine or other equipment to the scene of an incident.

2.3 Key Constituent Systems for FiReControl IT System-of-Systems

FiReControl was to provide a resilient network of nine regional control centres in England supporting the mobilisation of Fire and Rescue Service equipment and personnel to incidents. Its main IT system-of-systems has the following key constituent systems [9]**:**

- *Mobile Data Terminal (MDT) system* **-** using MDTs. data ranging from known risks and hazards in the buildings and locality, floor plans and access details, instructions on how to deal with different types of chemicals or other substances, as well as maps showing the quickest route to the incident would be provided to crews in the cab of a fire-fighting appliance.
- *Status Messaging system***-** allows fire-fighters and officers to send updates to the RCCs regarding their availability – for example changing their status from ‘mobile to incident’ to ‘arrived at incident’.
- *Automatic Vehicle Location System (AVLS)* **-** allows the exact location of all individual fire appliances to be identified, therefore enabling the nearest available appropriate appliance to be sent to an incident.
- *Satellite Navigation (Satnav) system* - provides fire-fighters with route planning information and directions to incidents.
- *The Enhanced Information Service for Emergency Calls (EISEC) and Automatic Location Service for Emergency Calls (ALSEC)* – these systems enabled control room operators to quickly confirm a caller’s location. EISEC displays the address where an emergency call was made using a fixed line, while ALSEC displays the location of a mobile caller on a map using GPS thus helping to identify hoax calls.
- *Geographic Information System (GIS)* – an electronic map that helps control room operator to visually determine the location of an incident if the caller cannot provide proper address details. It also enables crew members using MDTs to view information about the surrounding area, such as site-specific risks and hydrant data.
- *Full Premises Based Gazetteer* – an electronic database that uses the National Land Registry and Property Gazetteer (NLPG) information. The Gazetteer also contains the latest information on all

streets, motorways, towns, villages, hamlets and other features of interest to the FRSs.

2.4 Requirements Gathering Process

Figure 1 shows what the prime Supplier was contracted to deliver - mainly the Mobilisation System. The mobilisation system needed to meet the DLGC's demanding system performance requirements, including high system availability of 99.9999% and large data volume handling.

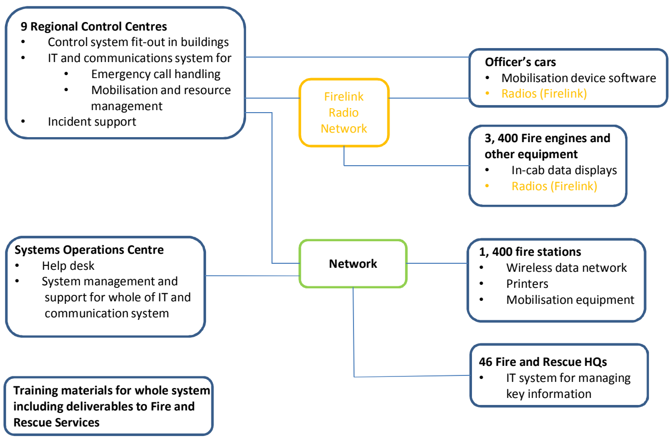


Figure 1. What EADS was contracted to deliver: Source: NAO.

To gather requirements, the Department held workshops with the fire and rescue services to identify and rationalise the business processes for FiReControl. Functional Requirements were then derived from the business processes. Non-Functional Requirements were then added to the functional requirements later. No system-level analysis was done but some form of already existing architecture was adopted. However, the Department did not actively involve the wider fire service in the derivation of the initial requirements neither did it involve key stakeholders and users. The assumption was that if the business processes were correct then the requirements would be correct too.

The FiReControl requirements document is more than 500 pages. Initially 2000 high-level requirements were identified which were later decomposed to 8000. A sample of some of the requirements extracted from the requirements document includes [10]:

- R1: The mobilisation and resource management system (MRMS) must be capable of managing and recording incidents, alerting and prompting for further information through initial call handling, mobilisation and incident management.
- R2: The MRMS must be capable of prioritising alerts by incident type, incident location and or any other combination as specified by FRSs. The MRMS prioritisation rules must be configurable.
- R3: In the event of an Integrated Communications and Control System (ICCS) failure at one or more RCC's, the ICCS architecture must allow ICCS equipment at the remaining Regional Control Centres (RCC's) to continue to operate.
- R4: ICCS terminal headsets must be compatible for use with fixed mobile radio terminals installed in RCC's by the Firelink Contractor. Connectors and / or adapters must be provided if required.
- R5: The list of recordings shall be listed by ICCS terminal ID, ICCS terminal user ID, date and time. Records shall be listed in descending date and time order, i.e. the most recent recording would be presented as the first recording in the list.
- R6: The continuous recording facility shall log faults for future retrieval in a chronological log.
- R7: The Management Information System (MIS) must hold up to a minimum of 6 years of data.
- R8: Where the MIS can only store a limited amount of data (minimum of 6 years), an archive facility must be provided.
- R9: If equipment supplied by FiReControl used to access the MIS is required to be connected to FRS systems, the Code of Connection must be adhered to.
- R10: The MIS must be able to report on equipment and Appliance repair times for access by FRSs and at the RCC.

3 Methodology Used to Gather Information

Table 2 shows the methods that were used to gather information about the FiReControl. It also shows the rationale for using each method.

Table 2: Methods and their rationale that we used to gather information about FiReControl

Method	Purpose
Interviews	To understand the Department's approach and rationale during the planning, design, delivery and cancellation of FiReControl. To discuss in more detail issues raised from the survey of Fire and Rescue Services.
Document review	To assess the impact of the cancellation of the project on the Fire and Rescue Service, and the Department's approach to project planning and management.
Benchmarking against other similar project	To compare the way in which FiReControl was procured, developed managed against other similar projects and draw parallels across government projects

We gathered information in several ways. *Interviews:* We conducted semi-structured interviews with the Department officials, National Audit Office senior officers, the Fire and Rescue Service representatives and RCC directors. *Benchmarking:* We compared the FiReControl project against other projects that failed or were cancelled and those that were successful; *Document review*: We examined the Department's procurement and planning documents, project evaluations and external reviews, operational and contractual information, performance monitoring information, and project closure documents, NAO reports, House of Commons Public Account Committee reports and requirements documents.

From the FiReControl perspective, some of the most relevant documents we reviewed were:

- FiReControl Project Concept of Operations, DCLG,
- NAO Memorandum for the House of Commons, Communities and Local Government Select Committee, February 2010
- Government IT Major Project Failures Report: Parliamentary Office of Science and Technology
- House of Commons Committee of Public Accounts, The failure of the FiReControl project, Fiftieth Report of Session 2010
- House of Commons Committee of Public Accounts, The National Programme for IT in the NHS
- The NHS's National Programme for Information Technology (NPfIT), A Dossier of Concerns by 23 prominent academics
- FiReControl Project Requirements Specification
- We interviewed key personnel at the National Audit Office who produced both the FiReControl and the NHS IT system audit reports for the House of Common Select Committees.
- We interviewed the FiReControl project Technical Director at the DLCG
- We interviewed Regional Control Centre managers
- We wrote our interview transcripts and sent them to NAO personnel and the FiReControl technical director for review and verifications.
- Due to legal and confidentiality issues we couldn't interview the key Supplier staff. Also due to the same reasons, we couldn't have access to the actual contract documents.
- Finally, we then benchmarked the FiReControl project firstly more general against other similar project to compare the way in which it was procured, developed and managed and to draw parallels across government projects. We also benchmarked FiReControl against the London 2012 Olympics IT system that is an equally large-scale complex system-of-systems that has been a success.

4 Key findings

Our analysis revealed several factors that contributed to the failure of FiReControl. We discuss some factor below:

a) Technical Problems in Developing FiReControl

In October 2007 the Supplier concluded that selected database and hardware products would not meet the high system performance and availability requirements as specified by the Department. By May 2009, the project had suffered further delays and the Supplier engaged another sub-contractor to develop a fall-back option for the mobilisation system. In December 2009, the Supplier terminated its contract with the first subcontractor and gave the contract to second subcontractor. However, the project was subject to a number of further delays and costs were escalating. In December 2010, the Department concluded that the project could not be delivered to an acceptable quality and timeframe and therefore cancelled it.

b) Ever Escalating Costs

At the point of cancellation, the Department estimated it had spent £245 million on the project and calculated that to complete the project would take the total cost of the project to £860 million, more than five times the original estimate of £120 million.

c) FiReControl Was Wrongly Concieved

Fundamentally, Fire Control was a change programme, NOT an IT project. It was a change programme and the mistake was to make it a project. It was essentially a system-of- systems that was wrongly conceived as a regular project.

d) Lack of user and stakeholder engagement

FiReControl was flawed from the outset because it did not have the support of the majority of those essential to its success – its users. Key stakeholders were not consulted about the requirements, which were derived from business processes by only one person. The project had a diverse range of stakeholder interests that should have been managed and balanced carefully.

e) High staff turnover

The project lacked consistent leadership and direction, and was characterised by a high staff turnover and over-reliance on poorly managed consultants. During the life of the project there were five different Senior Responsible Owners, four different Project Directors and five officers supervising the delivery of the technology. Only two senior managers worked on the project for its duration, one of whom, the project manager, was on contract from a consultancy.

f) Poor governance

Governance arrangements in the first five years of the project were complex and ineffective, which led to unclear lines of responsibility and slow decision-making. Additional layers of governance were created in response to emerging problems without clear lines of decision-making, accountability, responsibility, assurance, or internal challenge.

g) Weak requirements definition

There was no defined process for requirements elicitation and analysis. The Department set out approximately 2,000 requirements for the IT system in its contract with the main Supplier. These were further decomposed to 8,000 more detailed sub-requirements.

h) Underestimation of project costs

FiReControl was based on unrealistic estimates of project costs and expected local savings. These estimates did not include the costs of meeting local and regional implementation, or the costs of installing equipment, and overestimated the savings that could be achieved locally.

i) Underestimation of project complexity

The Department underestimated the complexity of designing a system to meet the needs of Fire and Rescue Services and overestimated the benefits of the project. The Department assumed that the development of the IT system would be straightforward, involving the integration of already customised components. However, in order to accommodate the wide variation in operational needs of the Fire and Rescue Services, key components required substantial modification.

j) Supplier Could Not Handle Project Complexity

The Supplier had no previous experience in developing similar systems and they could not handle the complexity of the system-of-systems.

k) Poor project management

A lack of project management in the early days of the project and poor financial management led to avoidable costs and delays in the procurement of the equipment, which presented significant risks to value for money.

l) IT Contract Badly Constructed

Poor contract design impeded the resolution of issues and the termination of the project at an earlier stage. A lack of interim milestones undermined the Department's ability to hold the Supplier to account for delivery. The payment schedule meant that the Supplier would be paid only once a key milestone for the building and testing of the system had been passed. The delays to delivery led to cash flow difficulties for the Supplier, which created further tensions in an already strained relationship.

m) No Syncronization Between the IT System Development and the Building of the RCCs

The Department identified two project streams – the Technical stream, i.e. the IT System and the Accommodation stream, i.e. RCCs. However, there was no synchronization between the two streams. Because of the Department's failure to manage the project as a whole, this resulted in the creation of empty regional control centres. The nine regional control centres were purpose-built to house the new computerised equipment and were designed specifically for that purpose. The Department's decision to prioritise the procurement of the centres over the IT system meant that the first RCCs were completed in June 2007, just three months after the IT contract had been awarded. All nine regional control centres were delivered before the cancellation of the project. By March 2011, the Department had incurred costs of £32 million in upkeep of the empty regional control centres.

5 Benchmarking FiReControl Against Other Similar Projects

In section 4 above we have identified many causal factors that contributed to the failure of the FiReControl project. The list of causal factors is by no means exhaustive. Furthermore, these causal factors are by no means unique to FiReControl. In this section, we benchmark FiReControl against similar projects in which the majority of them either failed or were cancelled. We conclude this section by benchmarking FiReControl against the London 2012 Olympics IT system that has been successful.

In the last 5 years, the UK government cancelled public IT system projects worth £27billion [5]. In February 2011 the UK's National Audit Office (NAO) published a report entitled 'ICT in Government – Landscape Review' [7]. The review report highlighted billions of pounds that would be wasted when it identified a significant number of government IT projects that would be cancelled, re-scoped or reviewed. The Landscape Review identified '*229 ICT projects under £50 million that were being cancelled with a further 193 up for review.* In addition, it found that *more than 80 IT projects over £50 million, budgeted to cost more than £28 billion overall, were being reviewed, with only a minority coming out unscathed. Just 26 projects, costing more than £4 billion, were being allowed to continue unchanged. Two IT projects, totalling nearly £2 billion, have been cancelled, while the remainder, more than 52 projects worth £22 billion, have been rescoped or were subject to ongoing reviews'*. Also these failures are not confined to the UK government departments alone. In November 2011, the Ombudsman in the Victorian Parliament in Australia examined 10 ICT-enabled projects that failed to meet expectations at an estimated cost of $2.74 billion - an additional cost of $1.44 billion over the original budget of $1.3 billion [6].

In March 2011, the UK's Institute for Government published a critically acclaimed report entitled 'System Error – fixing the flaws in government IT' [8]. In it, the authors observed that although there is a 'well-documented history of too many high-profile and costly government IT project failures' the underlying technology is rarely at fault. Rather the report states that '*policy complexity, late additions to already-long lists of requirements; inadequate change management processes; and a failure to bring users fully in to the picture'* all play their part as the main causal factors [8]. The report also identifies a number of symptoms that contribute to large-scale project failures, including FiReControl, such as:

- Chronic project delays; suppliers failing to deliver on their contractual commitments; not designing with the user in mind; the high cost of making even basic changes;
- Too many government IT failures occur on a massive scale and are only recognized as failures late into the process.
- Use of traditional linear development approaches such as the Waterfall and V-model which assume that the world works in a rational predictable way whereas in reality, priorities change rapidly and technological development is increasingly unpredictable and non-linear.
- Systems specifications are drawn up in advance, solutions are procured and delivery is managed against a pre-determined timetable. By the time projects are completed, if ever, what is delivered may be what users specified but is no longer what they need or want.
- Most government IT projects remain trapped in outdated models, which lock project requirements up-front and then proceed at a glacial pace. The result is repeated system-wide failures.
- The government spends over £1 billion on consultants and interim specialist staff per year and almost 60% of this on IT and project management. By outsourcing a large part of government IT, the public sector has also lost much of the knowledge and skills required for it to act as an intelligent customer.
- The single window for requirements leads business users to request any and all functionality that they think they might want at the outset. In a perfectly predictable world this approach would work well. However, in the real world, in which requirements, technologies and ministerial priorities are constantly evolving, this approach literally builds failure into the system.
- Most suppliers lack insider knowledge of government departments and this means that the requirements have had to be specified in a greater level of detail to try and prevent them being 'lost in translation'.
- Suppliers rarely have an incentive to question the validity of requirements. Additional complexity enables them to command bigger fees; the greater the specialisation of a system, the more likely suppliers' knowledge of the system will be called on to maintain or update the system.

Finally we benchmarked FiReControl against the London 2012 Olympics IT system that has been successfully implemented [11]. The aim was to identify key factors that contributed to the success of the London 2012 IT system. A key success factors was the experience of the Lead IT Integrator who have been the official Worldwide IT Partner of the International Olympic Committee (IOC) for nearly 20 years. This meant that the Supplier was not starting from scratch with the architecture and planning for London 2012 as this was their the sixth games it has been the official integration partner. There was already an existing architecture that works!

6 Conclusions and Future Work

The findings above show some key root causal factors that are common to almost all failed government IT projects: 'Stupendous incompetent' project management; Fundamentally flawed from conception; Lack of architecture design; Poor understanding of the project complexity; Lack of staff with system-of-systems engineering expertise; Lack of user and stakeholder engagement; Poor requirements gathering and analysis.

7 References

[1] Maier, M.W., "Architecting Principles for System of Systems," Systems Engineering, Vol. 1, No. 4, 1998, pp. 267- 154

[2] Jamshidi, M., "System-of-Systems Engineering - A Definition," IEEE SMC 2005, 10-12 Oct. 2005.

[3] The failure of the FiReControl project, House of Commons Committee of Public Accounts, Fiftieth Report of Session 2010–12, September 2011

[4] The failure of the FiReControl project, Report by the Comptroller and Auditor General, National Audit Offic (NAO),HC 1272, Session 2010-2012, July 2011

[5] Labour's computer blunders cost £26bn, http://www.independent.co.uk/news/uk/politics/labours-computer-blunders-cost-16326bn-1871967.html

[6] Own motion investigation into ICT-enabled projects, Victorian Ombudsman, Victorian Government, Australia, Session 2010 – 11, P.P. No. 68 November 2011, www.ombudsman.vic.gov.au

[7] ICT in government -Landscape Review, Report by the Comptroller and Auditor General, The National Audit Office, HC 757 Session 2010–2011, 17 February 2011 www.nao.org.uk

[8] Justine Stephen, James Page, Jerrett Myers, Adrian Brown, David Watson and Sir Ian Magee , System Error - Fixing the flaws in government IT, Institute for Government, UK, March 2011

[9] FiReControl Project Concept of Operations, Fire and Rescue Service (England) Concept of Operations for Regional Control Centres, Department of Local Government and Communities, Version 1.0, 27th June 2006

[10] FiReControl Project, Infrastructure Services Procurement, Exhibit C – Requirements, Department of Local Government and Communities, 29th March 2007

[11] Daniel Shane, 2011, Testing the game plan - Simulating the Olympic Games to stress-test London 2012's IT infrastructure, Information Age, 28 February 2011

An Enhanced Spatial Reasoning Ontology for Maritime Anomaly Detection

Arnaud Vandecasteele
MINES ParisTech
Centre for Crisis and Risk Research
Sophia Antipolis, France
arnaud.van_de_casteele@mines-paristech.fr

Aldo Napoli
MINES ParisTech
Centre for Crisis and Risk Research
Sophia Antipolis, France
aldo.napoli@mines-paristech.fr

Abstract: *Although originally conceived as a conceptual object for modelling knowledge, current ontologies do not make it possible to manipulate spatial knowledge. However, spatial knowledge is an essential component of any modelling specification. This problem provided the motivation for the creation of an expert system driven by an ontology. The system enables experts in the maritime domain to characterise abnormal ship behaviour based on formal semantic properties. Users are able to specify and execute spatial rules that are directly integrated into the ontology and a map interface linked to the ontology displays the results of the inferences obtained.*

Keywords: Spatial Ontology, Spatial Reasoning, Abnormal Maritime Behaviour

1 Introduction

Created in the early 1990s as a response to the various limitations identified in the development of expert systems, ontologies appeared as a new conceptual approach to knowledge modelling. This logical object made it possible to explicitly define not only concepts capable of describing the real world but also the rules governing the structure of these descriptions [1].

Nevertheless, the modelling and interpretation of spatial knowledge is still not sufficiently exploited in ontologies. Therefore, this paper proposes an extension to the initial capabilities of the Semantic Web Rule Language (SWRL) through the integration of spatial reasoning procedures. These procedures are incorporated into a Geographic Information System (GIS) driven by an ontology [2]. The approach is validated by the implementation of a platform capable of detecting abnormal ship behaviour.

We first describe the domain of ontologies and that of spatial ontologies (Section 2). Next, we discuss trajectory modelling and how this can be semantically enriched (Section 3). Finally, we present the prototype developed for the automatic detection of abnormal ship behaviour based on a spatial ontology (Section 4).

2 The contribution of ontologies to knowledge modelling

Although originally associated with the domain of philosophy, in this paper we examine the concept of ontology from the perspective of Artificial Intelligence (AI). Designed as a response to the problems posed by knowledge integration, ontologies appeared as a key paradigm in solving the problem of semantic heterogeneity and ensuring interoperability, as much between systems as between individuals [3].

This dual view of ontologies, which represent both a compendium of knowledge and an information object, is found in the definition given by Studer [4] for whom, “An ontology is a formal and explicit specification of a shared conceptualization.” *Formal* indicates that the conceptualisation and representation of the domain should be standardised and usable by an information system. *Explicit* specifies that the concepts used as constraints are defined declaratively. *Conceptualization* emphasises the fact that an ontology is only a partial abstraction of the real world. Finally, the notion of *sharing* implies that ontologies facilitate consensual knowledge.

The domain of ontologies therefore provided the necessary structures for knowledge modelling. Consequently, various ontology languages with different semantic capabilities were created.

2.1 Ontology languages

An ontology language makes it possible to express an interpretation of the world based on formal semantics and a precise syntactic structure. Since the 1990s there has been a proliferation of so-called traditional languages (Ontolingua, Cycl, Loom, etc.) coming from the field of AI [5], which it became necessary to standardise. Consequently, under the leadership of the OntoWeb group of the World Wide Web Consortium (W3C), standardised languages have been defined; examples include the Resource Description Framework (RDF) and the Web Ontology Language (OWL) [6]. OWL was implemented in the platform we have developed and we will briefly describe it here.

OWL has been the defined W3C standard for ontology creation since 2004. Based on the DAML+OIL language, OWL is founded on the basic primitives defined in RDF schemas. Nevertheless, far from being a simple extension of RDF, it provides all the semantics needed to describe knowledge such as mechanisms for the comparison of classes (concepts of equivalence, symmetry, etc.). Rather than defining a modelling language that was complex and difficult to use, the W3C decided to provide three increasingly expressive OWL sublanguages: OWL Lite, OWL DL and OWL Full. Each of these sublanguages is itself an extension of its predecessor [6]. In 2009, a new version of OWL (OWL 2) was proposed by the W3C, which aimed to be both an extension and revision of OWL. The motivation for the development of this new version came from the limited expressiveness of OWL, an overly complex syntax and the inability to annotate axioms [7].

Another limitation of OWL related to the absence of syntactic structures for rule creation. However, it is these structures that enable reasoning and the deduction of new facts from information contained in a knowledge base. Consequently, Horrocks [8] proposed the creation of the Semantic Web Rule Language (SWRL) that combines OWL DL and RuleML. Designed to support reasoning based on descriptive logic and Horn rules, the structure of a SWRL rule takes the form antecedent → consequent which is read as, “If the antecedent is true, then the consequent is also true”. Unlike OWL, SWRL only allows the addition of relationships and existing properties if they meet the rule. In addition to the OWL predicates, SWRL has supplementary ‘built-in’ functions. These functions extend the initial OWL capabilities; in particular they enable string comparisons and calculations. We have adopted this idea of built-in functions in order to integrate spatial capabilities into SWRL rules.

2.2 The spatial dimension of ontologies

The term ‘geographic ontology’ brings together two disciplines and worlds that employ different concepts; that of ontologies (which we have already discussed) and that of geography in the broad sense. However, as Agarwal [9] points out, the various existing studies that aim to bring together Geographic Information Systems and ontologies are essentially focussed on the strengths of each domain and do not form a true common discipline. For example (although the project is under consideration by the W3C) there is currently no standard for the representation of spatial data in ontologies or for spatial reasoning procedures in inference engines. As a result, various initiatives (e.g. SWEET[1], Ordnance Survey Ontologies[2], SWING[3], etc.) have attempted to define the necessary characteristics that must be implemented in spatial ontologies. However, these projects vary widely in terms of intended usage, the formalisation of representation and the rigor of the philosophical assumptions employed [10]. Consequently, the Geospatial Incubator Group (GeoXG), a W3C Working Group, is currently working on defining future directions for the integration of the spatial dimension into ontologies.

One of the first measures taken by the group was focused on the adoption of GeoRSS as a recommendation for the description of the geospatial properties of Web resources [11]. The result of the work of the GeoXG group, GeoRSS is inspired by GML while at the same time simplifying it in order to be as generic as possible. While the use of GeoRSS in the domain of ontologies offers the advantage of a simplified spatial representation, it nevertheless tends to suffer from an overly limited semantic. Moreover, beyond the formalisation of spatial entities, very little work has been carried out concerning the implementation of spatial reasoning procedures in ontologies [12].

3 Semantic modelling of trajectories

Although the study of moving objects is found in many domains, the principal contributions have come from the database community. These contributions particularly relate to the definition of new data types and operator-specific queries that can represent and query moving objects using a Database Management System (DBMS) [13]. The formalisation and conceptualisation of trajectories in an information system is a necessary but insufficient step in their understanding, use and analysis. Various approaches have been proposed in order to address these limitations including, notably, better integration of the semantics associated with trajectories [14]. In the context of this article, this is also the approach we have taken.

3.1 From trajectories to semantic trajectories

As defined by Spaccapietra [15], the concept of the semantic trajectory makes it possible to extract from raw data, trajectories whose components have been enriched by different types of information.

---

[1] Semantic Web for Earth and Environmental Terminology (http://sweet.jpl.nasa.gov/ontology/)

[2] http://www.ordnancesurvey.co.uk/oswebsite/ontology/

[3] Semantic Web Services Interoperability for Geospatial Decision Making (http://swing.brgm.fr/repository/ontologies)

A semantic trajectory is therefore defined as the movement of an entity in a geographic area, in a given time period. Therefore, it is marked by periods of movement and stops. The concept of the *stop* applied to a moving object means that trajectories can be defined, both temporally and spatially. The *move* is the time period between two stops during which an object is moving in space. It is thus defined by a sequence of temporally ordered positions. Each of these elements can be represented in ontological form [16] (Figure 1).

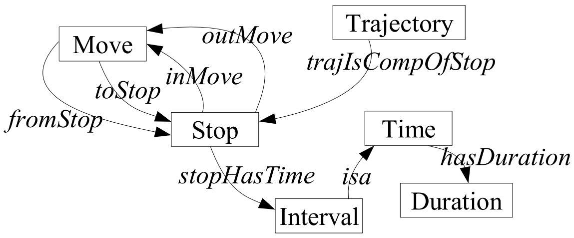


Figure 1. Ontological representation of the concept of semantic trajectories (based on [16])

As defined above, Figure 1 shows the two elements Stop and Move which delineate a trajectory both spatially and temporally. Therefore, in the approach taken by Baglioni et al. [16], the conceptualisation of a trajectory is represented by a stop sequence associated with a movement. This connection is made using one of four relationships: fromStop, toStop, inMove, outMove. In addition, each Stop takes place in a specific time dimension which is defined by the relation stopHasTime.

3.2 The semantic enrichment of trajectories

Using this ontological model, supplementary information can be added to the trajectory components. This semantic enrichment process generates, from the raw data, trajectories whose content has been enriched by the addition of related information – for example the geographical zone or domain of activity (Figure 2). Moreover, the user must be able to directly manipulate these semantic trajectories using an expressive language that can be understood by humans [16].

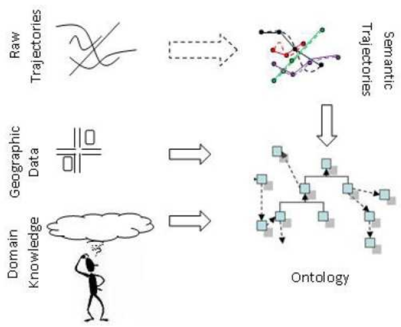


Figure 2. The process for the semantic enrichment of trajectories (based on [16])

Based on a similar logic of semantic enrichment, Yan [17] suggests a comprehensive architecture for the creation, management and analysis of trajectories. The architecture relies on a modular infrastructure which consists of three principal ontologies: a geometric trajectory ontology, a geographic ontology and a domain application ontology. These three ontologies are then combined to provide the semantic infrastructure necessary to describe the trajectory. We have adopted this approach for the creation of the ontology for the detection of abnormal ship behaviour.

4 The detection of abnormal ship behaviour using trajectory analysis

4.1 Prototype architecture

The principal idea of the prototype is to enable users to specify abnormal ship behaviour in relation to their trajectory. The prototype described in this paper is based on an architecture similar to that proposed by Yan [17]. The analysis of abnormal ship behaviour consists of four stages (Figure 3). First, databases are updated as new information arrives (step 1). The new information is added to the ontology (step 2). Then, the inference engine characterises the behaviour of each vessel taking into account the new information and the rules previously defined (step 3). Finally, a mapping module displays the result of the inference (step 4). Steps 2, 3 and 4 form the heart of our system and they are described in greater detail in the following paragraphs.

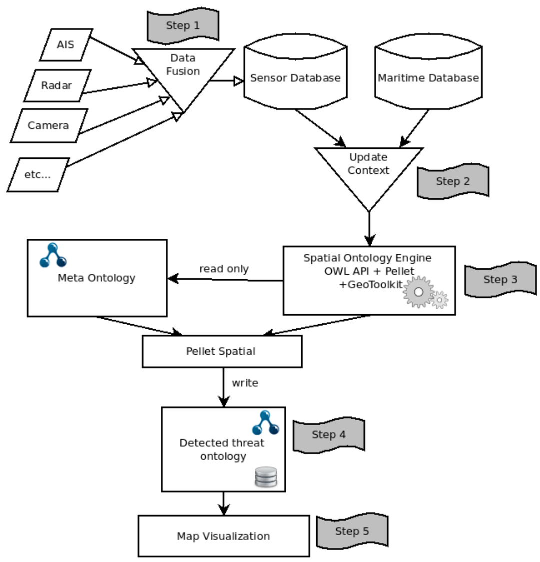


Figure 3. System architecture for the detection of anomalous ship behaviour

The first step is to update the ontology based on new information supplied by sensors. This step integrates dynamic and static data in the knowledge base. This is achieved through the OWL-API[4] Java library that provides a high-level programming interface for the creation and manipulation of ontologies.

Next, in order to be exploited, the ontology must be integrated into a reasoning engine. The purpose of this engine is to infer new facts from existing data. In this case, it will provide an evaluation of the situation taking into account the positions and characteristics of ships. From the various reasoners currently available, we decided to use Pellet[5]. This application has very good SWRL language support, it is compatible with the OWL 2 language and it integrates seamlessly with the JENA and OWL-API libraries. Above all, it is to our knowledge, the only application that enables the addition of new built-in functions in order to create custom SWRL rules [18]. Using class instances found in the knowledge base, the reasoner applies the various SWRL rules that have been defined. Ship behaviour can then be determined according to the outcome of the application of these rules.

Finally, once the ontology has been updated, the result of the inference can be displayed on a map interface. The map interface is directly linked to the ontology and provides a visualisation of the analyses using a customised graphical semiology. For example, the criticality level of a ship is indicated by graduated colours. Moreover, other information such as the vessel properties, its spatial characteristics and specified alerts can be displayed.

4.2 Experimentation and definition of detection rules

The experimental data consisted of records of ship positions for vessels equipped with an Automatic Identification System (AIS) in the Mediterranean Sea. It totalled more than 360,000 coordinates recorded during one week. Equipped with this dataset, the goal was then to detect abnormal ship behaviour. To achieve this, a set of rules was defined using both knowledge from domain experts and data contained in the scientific literature [19-21].

In the ontology, two types of rules can be distinguished: non-spatial and spatial. Non-spatial rules directly exploit SWRL reasoning capabilities. For example, it might be necessary to detect a vessel moving at an excessive speed for its type. The request, translated into SWRL, is shown below and reads as follows: "If the ship's speed (?vesselType) is greater than (greaterThan) the maximum speed for this type of vessel (?speedTypeVessel) then trigger an alert (Alert_Speed_HighSpeed)"

Vessel (?vessel),
Vessel_hasVesselType (?vessel, ?vesselType),
Vessel_hasSpeed (?vessel, ?speedVessel),
Vessel_hasSpeed (?vesselType, ?speedTypeVessel),
greaterThan (?speedTypeVessel, ?speedVessel) →
Vessel_hasAlert (?vessel, Alert_Speed_HighSpeed)

As we noted earlier, the analysis of abnormal ship behaviour must also take into account spatial dimensions. This element is one of the major contributions of our research. To achieve this, we extended the traditional functions of SWRL, creating custom built-in spatial functions (intersects, touches, etc.). The integration of these built-in functions was achieved using the expansion capabilities of the Pellet reasoning engine. The following is a concrete example:

Vessel(?vessel),
Vessel_hasPosition(?vessel, ?position),
Analysis_DataPath(Alert_Area_Restricted, ?geoData),
intersects(?geoData, ?position) →
Vessel_hasAlert(?vessel, Alert_Area_Restricted)

In this example, we perform an intersection (intersects) between the position of the ship (?position) and geographic data specified in the alert class (?geoData). Then, if the result is positive, we add an object property between the detected ship and the specified alert (Alert_Area_Restricted). Obviously, various spatial functions and other datasets are available. To model a different spatial function, it is simply a case of changing the keyword in the SWRL rule and for the data, specifying which data to use from the data property DP_Analysis_DataPath.

4.3 Examples of abnormal behaviour

To illustrate the operation of the ontology-driven GIS we chose the following two cases (Figure 4). First, we take the simple example of an intersection between the position of a ship and a restricted zone. The second case is more complex as it refers not only to the position, but also the trajectory of two ships.

In the first example, which relates to the restricted zone (Figure 4a) the inference engine has automatically detected an offence based on the ship's position and the geographic data specified in the ontology. The offending vessel is then reported as suspect. In addition, the alert criticality index automatically assigns a danger level to the ship.

In the second case (Figure 4b) there are two vessels on a parallel course. In the maritime domain, this could indicate a potential collision or a pair of fishing vessels. As before, the inference engine analyses the trajectories of the vessels and provides a geometric similarity index based on the mathematical formula of Haversine [22].

[4] http://owlapi.sourceforge.net/

[5] http://clarkparsia.com/pellet/

Depending on the outcome of this index, an alert is automatically added to the vessels involved. It should also be noted that (in addition to being on a parallel trajectory) the two ships are in a restricted zone. This particular gradation of dangerousness is highlighted by the orange colour that surrounds each of the two ships.

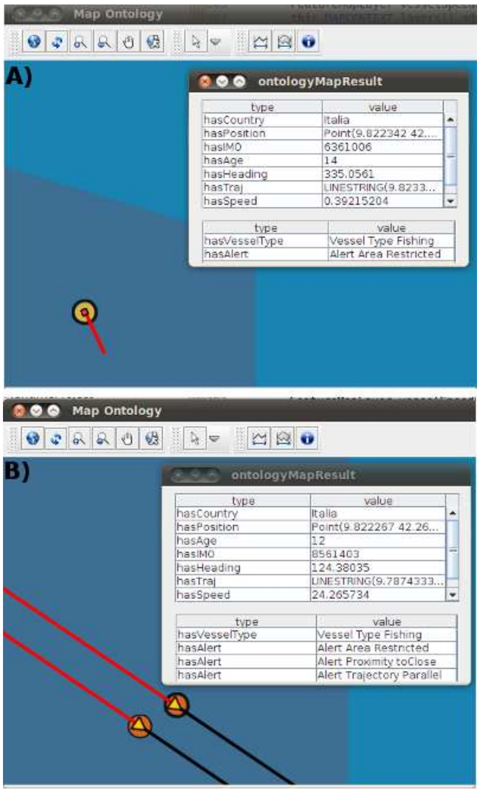


Figure 4. Examples of abnormal ship behaviour. A) Presence of a ship in a restricted area. B) Detection of parallel trajectories

5 Conclusion

This paper proposes the integration of the spatial dimension into an ontology in order to enable experts in the maritime domain to specify rules governing abnormal ship behaviour. To achieve this, we enriched the semantics of the SWRL language, which made it possible to define spatial functions. To validate our research hypothesis, the proposed solution was integrated into a GIS driven by an ontology. The prototype system aims to analyse ship positions and characterise their behaviour according to rules defined by experts.

The prototype described here is functional and can simultaneously analyse the behaviour of many thousands of ships. Nevertheless, the results of experiments carried out so far suggest that a useful improvement would be to facilitate the step of creating detection rules, which must currently be written using SWRL syntax. The creation of these rules can be a significant constraint for domain experts responsible for creating the model. Therefore, it may be useful to provide a graphical interface for rule creation along the lines of the Snoggle interface (Snoggle is a graphical, SWRL-based ontology mapper). The expert would then only have to draw abnormal behaviour, which would be translated into the SWRL language before being integrated into the ontology.

As we have demonstrated, the integration of the spatial dimension is an essential element in the structuring of knowledge. Nevertheless, it cannot yet be fully implemented in ontologies due to the lack of appropriate structures. Ultimately, it appears that it will be necessary to make use of spatial data types such as those currently found in geographic databases.

6 References

[1] F. Gandon, “Graphes RDF et leur Manipulation pour la Gestion de Connaissances (RDF Graphs and their Manipulation for Knowledge Management)”, Habilitation à Diriger des Recherches, INRIA, 2008.

[2] F. T. Fonseca and M. J. Egenhofer, “Ontology-driven geographic information systems”, in *Proceedings of the 7th ACM International Symposium on Advances in Geographic Information Systems*, New York, NY, USA, pp. 14-19, 1999.

[3] T. Mondary, “Construction d’ontologies à partir de textes. L’apport de l’analyse de concepts formels (The Construction of Ontologies from Texts. The Contribution of Formal Methods)”, Phd, 27 May 2011.

[4] R. Studer, V. R. Benjamins, and D. Fensel, “Knowledge engineering: Principles and methods”, *Data & Knowledge Engineering*, Vol. 25, No. 1-2, pp. 161-197, March 1998.

[5] V. Nguyen, “Ontologies and Information Systems: A Literature Survey”, Defence Science and Technology Organisation, Technical Note DSTO-TN-1002, June 2011.

[6] S. Bechhofer, F. van Harmelen, J. Hendler, I. Horrocks, D. L. McGuinness, P. F. Patel-Schneider, and L. A. Stein, “OWL Web Ontology Language Reference”. W3C, 10 February 2004.

[7] B. C. Grau, I. Horrocks, B. Motik, B. Parsia, P. Patel-Schneider, and U. Sattler, “OWL 2: The next step for OWL”, *Web Semantics: Science, Services and Agents on the World Wide Web*, Vol. 6, No. 4, pp. 309–322, November 2008.

[8] I. Horrocks, P. Patel-Schneider, H. Boley, S. Tabet, B. Grosof, and M. Dean, "SWRL: A Semantic Web Rule Language Combining OWL and RuleML". W3C Report, 2004.

[9] P. Agarwal, "Ontological considerations in GIScience", *International Journal of Geographical Information Science*, Vol. 19, No. 5, pp. 501-536, 2005.

[10] J. Ressler, M. Dean, and D. Kolas, "Geospatial Ontology Trade Study", *National Geospatial-Intelligence Agency*, Report No. HM1582-05-C-0014, 2007.

[11] J. Lieberman, R. Singh, and C. Goad, "W3C Geospatial Ontologies – W3C Incubator Group", W3C Incubator Group, Report, 2007.

[12] A. Karmacharya, C. Cruz, F. Boochs, and F. Marzani, "Use of geospatial analyses for semantic reasoning", in *Proceedings of the 14th International Conference on Knowledge-based and Intelligent Information and Engineering Systems: Part I*, Berlin, Heidelberg, pp. 576-586, 2010.

[13] V. Mallé-Noyon, "Modèle de vue relative et spatio-temporelle de trajectoires géographiques d'objets mobiles: application au contexte maritime (Relative and spatio-temporal model of the geographical trajectories of moving objects: application to the maritime context)", Ecole Nationale Supérieure d'Arts et Métiers, Phd, 2007.

[14] J. Malki, W. Mefteh, and A. Bouju, "Une approche ontologique pour la modélisation et le raisonnement sur les trajectoires. Prise en compte des règles métiers, spatiales et temporelles (An ontological approach for trajectory modelling and reasoning)", in *JFO 2009 3ème édition des journées Francofones sur les Ontologies*, France, pp. 157-168, 2009.

[15] S. Spaccapietra, C. Parent, M. L. Damiani, J. A. de Macedo, F. Porto, and C. Vangenot, "A conceptual view on trajectories", *Data & Knowledge Engineering*, Vol. 65, No. 1, pp. 126-146, April 2008.

[16] M. Baglioni, J. Macedo, C. Renso, and M. Wachowicz, "An Ontology-Based Approach for the Semantic Modelling and Reasoning on Trajectories", in *Advances in Conceptual Modeling – Challenges and Opportunities*, Vol. 5232, Springer Berlin/Heidelberg, pp. 344-353, 2008.

[17] Z. Yan, "Semantic Trajectories: Computing and Understanding Mobility Data", Phd, Ecole Polytechnique Fédérale de Lausanne, 2011.

[18] E. Sirin, B. Parsia, B. C. Grau, A. Kalyanpur, and Y. Katz, "Pellet: A practical OWL-DL reasoned", *Web Semantics*, Vol. 5, No. 2, pp. 51-53, June 2007.

[19] J. Van Laere and M. Nilsson, "Evaluation of a workshop to capture knowledge from subject matter experts in maritime surveillance", in *Proceedings of the 12th International Conference on Information Fusion, 2009. FUSION '09*, pp. 171-178, 2009.

[20] B. Idiri and A. Napoli, "Découverte de règles d'association pour l'aide à la prévision des accidents maritimes (The Discovery of Association Rules to Help in the Prediction of Maritime Accidents)", paper presented at the *Extraction et Gestion des Connaissances* (EGC), Bordeaux, 2012.

[21] X. Chaze, A. Bouejla, and F. Guarnieri, "The contribution of bayesian networks to manage risks of maritime piracy against oil offshore fields", paper presented at *the 17th International Conference on Database Systems for Advanced Applications (DASFAA),* South Korea, 15-19 April 2012.

[22] T. Devogele, "Système d'information géographique temporelle maritime; Des distances linéaires à l'analyse temps réel des trajectoires (Maritime Temporal Geographic Information Systems; Linear Real-time Analysis of Trajectories) ", Habilitation à Diriger des Recherches, Université de Bretagne Occidentale, 2009.

A New Embedded E-Nose System to Identify Smell of Smoke

Salahedin Sadeghifard
Instrument Maintenance Department
South Pars Gas Complex
Assaloye, Boushehr, Iran
s.sadeghifard@yahoo.com

Leili Esmaeilani
Instrument Maintenance Department
South Pars Gas Complex
Assaloye, Boushehr, Iran
l.esmaeilani@gmail.com

Abstract - *This work examines the important applications of modern electronic noses and focus on fire detection system due to advantages over classical method of detections. The three components of an electronic nose consist of sample handling; detection and data processing system are designed. These devices are typically array of sensors used to detect and distinguish odors precisely in complex samples and at low cost and capable of classifying smoke based on neural networks. The potential advantages of such an approach include, the ability to characterize complex mixtures without the need to identify and quantify individual components, Five commercial gas sensors (Figaro) with interesting cross sensitivity and low power consumption are used in sensor array; a micro-controller equipped with a compact flash memory assures data acquisition, analyzing procedures in real time. Signals from this sensor array have unique pattern and applied to the embedded system as inputs. The proposed method in this paper has 97.2% efficiency in smoke classification.*

Keywords: Neural network, Electronic nose, Fire detection, Gas sensor.

1 Introduction

One of the most important problems of human life is damages from fire, so improving the reliability of fire alarm systems is very important. House fire is a catastrophic phenomenon which kills thousands of people and injures more annually worldwide. In 2003, there were 388; 500 reported house fires in the United States, resulting in 3; 145 deaths, 13; 650 injuries and $5;9 billion in direct property damage [1]. Therefore, existence of a customized house fire alarm system with capability of classifying burning materials is very important for choosing a proper solution for suppression of fires. Existing fire alarm systems detect high temperature fire and its smoke and hydrocarbon gases. They monitor various locations to activate alarm signals as needed. Fire burns when temperature exceeds a threshold and lead to chemical activation. Results are temperature, flame, light, smoke, monoxide carbonic, and other components.

Depending on burning materials, different components are released in the space. Therefore, if a system senses the raised components and analysis them to find type of burning materials, it can use a proper solution for suppression of fires. Our work concerns the development of a portable system able to detect and identify fire in early stage by using artificial electronic nose system based on neural networks and it can active proper extinguishing system according to NFPA.[1] In this E-nose system, for measuring and classifying odor of fire, signals from sensors are collected and analyzed by ANN. The empirical results show high reliability and accuracy in early stage.

2 Artificial electronic nose system

The electronic nose is a relatively new analytical technology, and well-known as efficient analytic devices that are widely used for many applications including the food industry [2], perfumery, biotechnology, medicine [3], chemistry, and environmental sciences. This technology is based on the principle of mimicking the human odor recognition mechanism. Artificial electronic nose systems are used to detect and analyze the odor of materials. The main idea of these systems comes from human olfaction. E-nose has three main parts: sample handling; detection and data processing system.

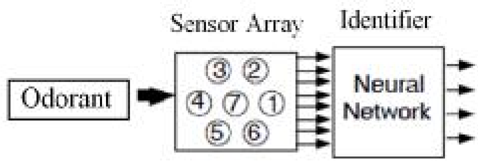


Figure 1. Basic structure of E-Nose

3 Metal oxide gas sensor array as odor sensor

Metal oxide gas sensors are widely used in various applications. The main part of the MOGS is the metal oxide element on the surface of the sensor. When this element is heated at a certain high temperature, the oxygen is absorbed on the crystal surface with the negative charge. The reaction between the negative charge of the metal oxide surface and deoxidizing gas makes the resistance of the sensor vary as the partial pressure of oxygen changes [4]. Based on this characteristic, we can measure the net voltage changes while the sensors absorb the tested odor. Every sensor has a unique response to different gases and the ability of identifying the odor will be improved if sensors are arranged together in an array. The sensing element of Figaro gas sensors is a tin dioxide ($SnO2$) semiconductor which has low conductivity in clean air.

In the presence of a detectable gas, the sensor's conductivity increases depending on the gas concentration in the air. A simple electrical circuit can convert the change in conductivity to an output signal which corresponds to the gas concentration. A major issue with gas sensors is their sensitivity to humidity. It is well-documented that water vapor affects measurements by electronic noses and manufacturers of these instruments have been forced to issue specific operating procedures but in this project it will be solved. Metal oxide gas sensors used in this work are listed in table 1.

Table 1.Figaro sensors used in this work

	Sensor Type	Detectable gases
1	TGS-2602	Air Contaminant
2	TGS-822	Organic Solvent vapor
3	TGS-825	Hydrogen Sulfide
4	TGS-813	Combustible gas
5	TGS-880	Cooking Vapor
6	SHT71	Humidity and

4 Configuration of fire alarm system

Schematic of intelligent fire alarm system, shown in Figure 3 consists of three parts:

- Sensing unit
- Signal processing and control unit
- Interface circuits and display

4.1 Sensing unit

Each odorant presented to the sensing system produces a characteristic pattern of the odorant. By presenting a mass of sundry odorants to this system a database of patterns is built up. It is used then to construct the odor recognition system. In this work sensing unit consists of 5 cheap and applicable MOG sensors, a temperature and humidity sensor. All the sensors are placed on an electrical board, providing also the electrical interface for the gas sensor temperature control and the sensor measurement conditioning. These Figaro type sensors require to be heated continuously at approximately 300°C in order to get the chemical operating point.

Electrical conduction of these sensors changes when they are exposed to gases and this conduction can be measured by circuit shown in Figure 2.R_L,V_C,V_H,V_{RL} are load resistance, biasing voltage of gas sensing element, biasing voltage of heating element and output voltage respectively.

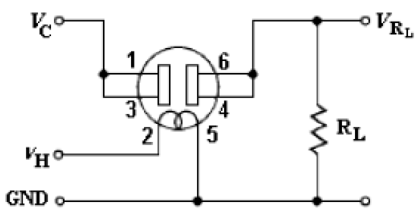


Figure 2.MOGS bias circuit

It is important to know that none of these sensors can measure a specific mixture of gases absolutely [2].In addition, each of them designed to measure a special gas but actually they have some responses to other gases [4].While some metal oxide gas sensors are put together in an array, the ability of detecting odor will be improved because odor is a combination of different gases. Each smell (especially gases made by fire) exposing to sensor array makes a unique pattern that is a combination of different sensor response to specific smell in gas combination.

4.2 Humidity Control

As discussed in section 3, most MOGS gas sensors are sensitive to humidity. Therefore, if two identical samples with a different humidity are measured, the results can be different. In our work, we solved the humidity problem is a software-based approach to achieve maximum accuracy. A mathematical model describing the resistance of each gas sensor at different humidity level can be calibrated to subtract the humidity signal from the total signal. Samples other than the smoke of burning material can also be used with this algorithm. Consideration relative humidity [%RH] was varied from 30% to 80%. Mathematical models for the sensors' response to humidity can be fitted via the following formulations [8], equations of TGS813, TGS822, TGS880, TGS2602, and TGS825 are:

$$R_{S_{813}} = 86682\exp\left(\frac{-[\%RH]}{29.05}\right) + 55063.48 \quad (1)$$

$$R_{S_{822}} = 24931.58\exp\left(\frac{-[\%RH]}{37.48}\right) + 9054.41 \quad (2)$$

$$R_{S_{880}} = 90496.88\exp\left(\frac{-[\%RH]}{36.33}\right) + 55135.22 \quad (3)$$

$$R_{S_{2602}} = 6958.22 + 129.172[\%RH] - 0.9788[\%RH]^2 \quad (4)$$

$$R_{S_{825}} = 5646.63 + 103.26[\%RH] - 1.34[\%RH]^2 \quad (5)$$

4.3 Signal processing and control unit

Most of the electronic noses developed are implemented in a Personal Computer (PC) based platform which, due to cost, size and power requirements, limits their use in day to day life. This study discusses the development of an AVR family microcontroller based embedded system and implementation of ANN in the embedded system for fire classification. The autonomous control part, corresponding to the brain of the electronic nose figure 3, 4 uses an 8 bits AVR microcontroller (ATMEGA32). The microcontroller assures measurement control, data acquisition, analyzing treatments and data transfer or storage. In this unit, signals coming from sensor array processed and classified according to type of burning material and a particular output signal set on to actuate extinguishing system, matched with NFPA will be produced.

For accurate operation of neural network, it is necessary to calculate weights and biases and use them in the form of a program in the microcontroller. For this purpose, signals from sensors are simultaneously sent to PC via microcontroller. These signals are sample in each second and these data make a matrix with dimension 5 (number of sensors) × number of sampling times, then mean value of data in duration 20 to 40 second chose as proper data, the criteria for choosing this duration is the time where sensors are stabled and saturated. The materials that selected and burnt for smoke sampling in this work are listed in table 2.

Figure 3:Embedded E-Nose system

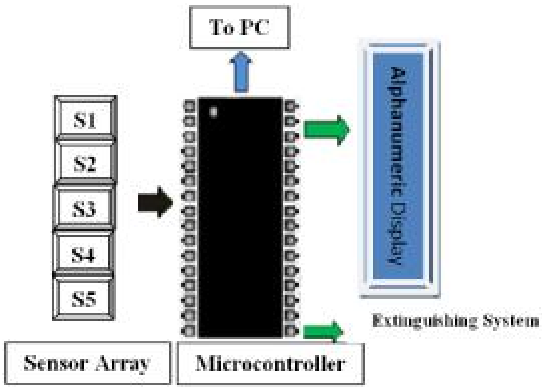


Figure 4.Hardware configuration

Table 2. The codes assigned to the materials.

Material	Assigned Code
Fresh Air(Normal)	[000000001]
paper	[000000010]
Wood	[000000100]
Cotton	[000001000]
Carpet	[000010000]
Plastic	[000100000]
Oil	[001000000]
Methane	[010000000]
Incense	[100000000]

We should take into consideration that in lack of smoke, air is the dominant gas, so we should add the signal of air to our data, as the base line. According to sensors structure, an internal heater should warm up the sensor and it takes time to make sensor stable, this time is about 50 to 60 seconds according to figure 5.

After sensor stabilization, it is time for sampling. Each source of smoke was sampled and tested 20 times to achieve high repeatability of the system.

To form the response pattern, the mean value of sensors output is calculated and considered as the effective value of the sensor during 20 to 40 seconds. figure 6 is an example pattern formed by sensor signals from burning of paper.

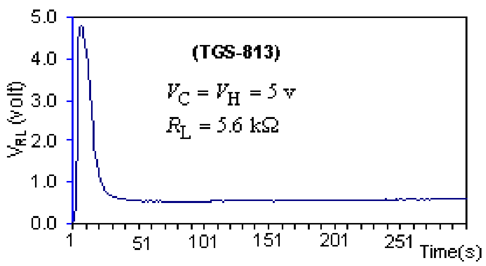


Figure 5.Response of sensor “Figaro-TGS813” in free air

For recognition the pattern and make relation with the smoke of specific material, we need a pattern recognition system. According to good background of neural network (MLP) with back propagation algorithm to specify the complicated relation between the combination of different gases made from burning different material [5] [6] [7] and the ability of classifying pattern an MLP network is used. A neural network which is designed to detect and classify fire is able to adjust weights and biases to reach the best response according to experiences achieved. This system has high potential to classify the material burning in the first stage.

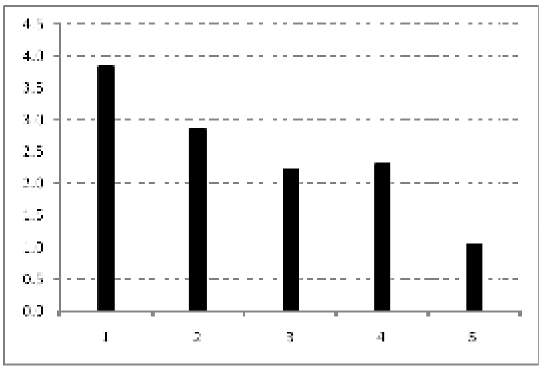


Figure 6.Sensors response pattern for paper (y axis is voltage)

MLP used in this ANN consists of a 5 input layer (proportional to sensors number), a hidden layer of 10 neurons and a 9 output layer (proportional to burning tested material). For training this network, efficient values of sensors in patterns formed from free air and 8 smoke samples (result of burning different material) are choose as input vector and detected materials are chose as desired outputs. Different transfer functions were tested in network and “Logsig” function had the best response with the least error while it is used in hidden and output layers of network. According to behavior of “Logsig” function in network data should be normalized between 0 and 1 .It is done by equation (6).

$$X_{nor}=\frac{X_i-X_{min}}{X_{max}-X_{min}} \qquad (6)$$

After training network with samples, the weight and bias of ANN was transfer to Microprocessor and accurate operation of network was tested in Real time operation of embedded device with smoke from burning material and results are shown in table 3.

Table .3 Practical results

Material	Test	Correct	Incorrect	Validity
Plastic	9	9	0	100%
Wood	9	7	2	77.7%
Paper	9	9	0	100%
Incense	9	9	0	100%
Methane	9	9	0	100%
Oil	9	9	0	100%
Carpet	9	9	0	100%
Cotton	9	9	0	100%

We should take into consideration that it is very important how sample of smoke, effect on different sensors, depends on many parameters. After detecting the kind of burning material, an output signal will actuate fire extinguishing system according to NFPA.

In NFPA, paper, wood and cotton are classified in extinguishing group “A” and methane is in group “B” and materials such as incense just detected and no output is activated.

5 Experiment

In this section, we report performance of the system after final assembly. To do the experiments, we turn on the system and combust the materials in the environment and record the detections of the system. We test implementations of the system using ANN Table 3 shows confusion matrix of the experiments with ANN pattern recognition algorithm. It can be seen that, except a wood instance that is classified as paper, all other instances are correctly classified. So the accuracy of the system in these experiments is more than 97 percent. For sending information to PC and calculating proper biases and weights, an interface circuit is used between microcontroller and PC meanwhile a single display is used to show data.

6 Conclusion

In this paper, we have presented the reliability of a new EN system designed from various kinds of MOGS as intelligent fire alarm and extinguishing system. Electronic noses have been proposed as fantastic instruments which could solve almost any problem concerned with odor of fire, The EN has the ability to identify various sources of burning material in the early stage with more than 97.2% of accuracy in the BP case. The operation of artificial electronic nose was shown just by 5 sensors. It can be concluded that the EN is suitable for detecting the early stage of fire and we can improve accuracy, reliability and ability of system by using more sensors. High reliability of this sensors combination can make our fire alarm system safe to false alarms.

References

[1] US Home Structure Fires, 2010 Annual Report, Fire Analysis and Research Division, NFPA

[2] .Vestergaard, J.S.; Martens, M.; Turkki, P. Application of an Electronic Nose System for Prediction of Sensory Quality Changes of a Meat Product (Pizza Topping) During Storage. LWT 2007, 40, 1095–1101.

[3] Chan, H.P.; Lewis, C.; Thomas, P.S. Exhaled Breath Analysis: Novel Approach for Early Detection of Lung Cancer. Lung Cancer 2009, 63, 164–168.

[4] "Technical Information on Usage of TGS Sensors for Toxic and Explosive Gas Leak Detectors", http//:figarosensor.com.

[5] Toru Fujinaka, Michifumi Yoshioka, Sigeru Omatu, Toshihisa Kosaka, "Intelligent Electronic Nose Systems for Fire Detection Systems on Neural Networks," advcomp,pp.73-76, The Second International Conference on Advanced Engineering Computing.

[6] B. Charumporn, M. Yoshioka, T. Fujinaka, and S. Omatu, “An electronic Nose System Using Back Propagation Neural Networks with a Centroid Training Data Set”, Proc. Eighth International Symposium on Artificial Life and Robotics, Japan, pp. 605–608, 2003.

[7] Roland Linder , Siegfried J. Pppl, "Food Quality Assurance Applying a Sophisticated Neural Network to Olfactory Signals”, 6th Sensometrics, 2002.

[8] Chatchawal Wongchoosuk , Mario Lutz and Teerakiat Kerdcharoen, "Detection and Classification of Human Body Odor Using an Electronic Nose”, Sensors 2009, 7234-7249; September 2009.

A System of Systems for air quality decision making

C. Carnevale, G. Finzi, E. Pisoni, M. Volta
Universitá degli Studi di Brescia, Dipartimento di Ingegneria dell'Informazione
Viale Branze 38, 25123, Brescia
Email: enrico.pisoni@ing.unibs.it

Abstract – *The paper presents a system of systems approach to tackle the air quality planning problem through a multi-objective optimization methodology. In this context the different axes of the air quality planning issue (economic, environmental, social) are described and modeled in a unique framework and in an integrated/systemic way allowing for the selection of policies that optimally consider the three aforementioned perspectives. In the formalization of the proposed system approach two objectives are optimized: the Air Quality Index (related to health effects due to high concentrations of secondary pollution, environmental and social impacts of the problem) and the Cost Of Policy Index (related to the cost of emission reduction implementation, and so modeling the economic dimension). A case study for a Northern Italy domain is shown and discussed, stressing how a multi-Air Quality Index framework has to be developed to manage conflicting objectives.*

Keywords: Environmental systems, Air quality control, Artificial Neural Networks, optimization.

1 Introduction

In literature different approaches have been formalized and implemented, with the aim of improving air quality evaluating alternative emission reduction policies. These approaches are mainly based on scenario analysis [10], cost-benefit analysis [17], cost-effectiveness analysis [2], multi-criteria and multi-objective mathematical programming [9]. This paper focuses on the most recent approaches, as cost-effectiveness and multi-objective optimization ones, also defined in literature as Integrated Assessment Models (IAMs).

IAMs are based on the concept of ”system of systems”, as they pool models and capabilities of different systems together to create a more complex system, that can provide improved capabilities/performances in comparison to the simple sum of the constituent systems. Using Integrated Assessment Models, Decision Makers are supported in the selection of the technologies to be applied to reduce emissions (precursors of air pollutants) so that human and ecosystem exposure to pollutants are improved in a cost-effective way.

One of the most outstanding examples of these kind of tools is the RAINS/GAINS [19] model, applied since a number of years at International/European level to determine cost-efficient policies to reduce emissions and achieve target for given air quality indicators (e.g. acidification, eutrophication, tropospheric ozone, primary and secondary particulate).

Apart from European Scale, some national IAMs exist in Italy [6], Netherlands [1], Finland [16], UK [13], Belgium [7]). Based on similar approaches to GAINS, these models can then be used to optimize emission reductions within a given country at the regional level. At the local/urban scale few IA models have been developed and applied [12]; these have generally been used for non-reactive species, and so these kind of approaches have limitations if used for secondary (reactive) pollutants, as ozone and PM.

The aim of this paper is to present and discuss the application of an IAM developed at regional scale.The main goal of this tool is to identify the most efficient mix of local policies required to reduce tropospheric ozone and particulate matter, in compliance with National and International air quality regulations (e.g. EU directives), accounting for local peculiarities in terms of emissions, meteorology and technological, financial and social constraints. A case study on Northern Italy is proposed, focusing on the three interlinked dimensions (economic, environmental and social) that are modeled and integrated. Also, a focus on conflicting objectives for the proposed case study application is shown and commented.

2 Formalizing the approach

The proposed methodology implements and solves a multi-objective problem, for selecting effective policies to control population exposure to primary and secondary pollutants. To do so, the methodology requires a) current and prospective emission reductions technologies and related costs (derived by GAINS [19]); b) regional activities and emission data (from the regional emission inventory); c) source-receptor (S/R) models, developed for the specific regional environment (that is to say, the surrogate of a larger and more detailed chemical transport model, as described in [4]). In the following Sections the procedure is described in detail.

2.1 The decision problem

The decision problem can be formalized as follows:

$$min\,[P(E(\theta))\;\;C(E(\theta))] \tag{1}$$

where E represents the precursor emissions; $P(E(\theta))$ is the air pollution index considered; $C(E(\theta))$ represents the implementation costs of pollution reduction measures, and both objectives depend on precursor emissions through a set of decision variables θ. The solutions of the multi-objective problem are the efficient emission control policies, in terms of air quality and emission reduction costs.

2.2 The decision variables

The total emission reduction for a pollutant p, due to the application of a set of technologies, can be calculated as the sum of the emission reductions over all the macrosector-sector-activity triples (these triples, defined as in the GAINS model nomenclature [19]):

$$E_p = \sum_{ijk} E_{ijkp} \tag{2}$$

The decision variables (defined more in general as θ in the previous equation) are the application rates of the emission reduction technologies. The reduced emissions are computed as follows:

$$E_{ijkp} = \sum_{t \in T_{ijk}} A_{ijk} \cdot ef_{ijkp} \cdot eff_{iktp} \cdot X_{ijkt} \tag{3}$$

where:

- E_{ijkp} are the emissions [kton] of the pollutant p, in the macrosector, sector, activity, ijk triple, remaining after the application of a set of technologies.
- T_{ijk} are the technologies that can be applied in the macrosector, sector, activity ijk triple.
- A_{ijk} are the activity levels of a macrosector, sector, activity ijk triple.
- ef_{ijkp} represents the unabated emission factor [kton/Act.Unit] for a macrosector, sector, activity ijk triple, for a particular pollutant p.
- eff_{ijktp} is a measure of the efficiency of technology t. More precisely, it is the fraction (between 0 and 1) of pollutant p remaining after the application of a particular technology t, to the activity ijk.
- X_{ijkt} represents the application rate (between $\underline{X}_{ijkt}$ and $\overline{X}_{ijkt}$, respectively minimum and maximum value) of a macrosector, sector, activity, technology $ijkt$ quadruple.

It is important to note how the spatial dimension is considered in the emission reductions computation. Even if the control variables X_{ijkt} are not space dependent (i.e. given an optimal control variable value, this is then applied constantly all over the domain), the resulting E_{ijkp} is space dependent, being that the activity levels depend on the considered cells.

2.3 The Air Quality objectives

The relationship between emission precursors and ozone or PM concentrations can be given by the simulation of deterministic 3D modeling systems. Such models require so high computational time that are not of practical use in an optimization problem. For this reason simplified source-receptor models have been identified processing simulations performed by a deterministic modeling system.

2.3.1 Deterministic approach

PM concentrations are typically simulated by three-dimensional deterministic modeling systems. In this work the Gas Aerosol Modeling Evaluation System (GAMES) [18] has been used. It consists of different modules, as (a) the multi-phase Eulerian 3D model TCAM; (b) a meteorological pre-processor; (c) the emission processor POEM-PM [5].

The core of the modeling system is the multi-phase photochemical model TCAM [3]. It is an nonlinear distributed parameters three-dimensional model with hundreds of state and input variables, that solves, time by time, a PDE system (considering equations for horizontal transport, vertical processes, chemical reactions), for each cell and pollutant species. TCAM implements the COCOH chemical mechanism, including 95 species involved in 187 chemical reactions. The gas phase chemistry is solved by means of the Hybrid Implicit-Explicit (IEH) solver [14]. The model considers the main aerosol processes (condensation/evaporation, nucleation). The particles are described by means of a *fixed-moving* approach and they are chemically characterized in 21 species and split in 10 dimensional classes.

2.3.2 Source-receptor approach

Simplified models based on neural network approach are identified through the processing of input and output of several runs of the GAMES deterministic modeling system.

Neural networks are composed by simple connected elements (neurons) operating in serial/parallel. Each element is characterized through a function (usually non linear) relating inputs and outputs (activation function). During the identification phase the weights of the connections between the different neurons are adjusted in order to define a particular function between the network input and output. In particular, the feed-forward neural network has been used in this study. This network computes a vector function $f_{NN} : \mathbb{R}^Q \to \mathbb{R}^L$ where Q and L are the dimensions of the input and output vectors of the net respectively; the l-th element of the vector function f_{NN} is defined as (M is the number of the neurons in the hidden layer):

$$f_{NN}(v) = af_2(\sum_{m=1}^{M}(OW_{l,m} \cdot a_m) + g_l) \tag{4}$$

where:

$$a_m = af_1(\sum_{q=1}^{Q}(IW_{m,q} \cdot v_q) + b_m) \tag{5}$$

where af_1 and af_2 are real continuous functions, called activation function of the hidden layer (af_1) and of the output layer (af_2). The matrices IW ($M \times Q$) and OW ($L \times M$) are the input and output matrix respectively, and b ($M \times 1$) and g ($L \times 1$) vectors are the bias terms. Neural networks learn on a training data set, tuning the parameters IW, OW, b and g by means of a back-propagation algorithm.

The Air Quality objectives considered in this work are a) the annual mean PM25 concentration, b) the Years Of Lost Life (YOLL) as a proxy of the PM25-related health effects and c) the SOMO35 (sum of the differences between maximum daily 8-hour running mean concentrations greater than 35 ppb, and 35). For PM25 and SOMO35, the relationship between the decision variables and the indexes is directly modelled by Artificial Neural Networks (see also [4] for more details on this step). For YOLL computation the PM25 ANNs is used and then, through the approach presented in [19] (that consider among others life expectancy, PM25 relative risk and population data), the YOLL are derived.

2.4 The Cost Objective

The Cost Objective is calculated as follows. For each activity i in sector j, macrosector k, the cost of applying all technologies is computed as:

$$C_{ijk} = \sum_{t \in T_{ijk}} C_{ijkt} \cdot A_{ijk} \cdot X_{ijkt} \tag{6}$$

where:

- C_{ijk} are the abatement costs [Meuro] for macrosector, sector, activity, ijk triple.
- C_{ijkt} are the unit costs [Meuro] of application of technology t.
- A_{ijk} and T_{ijk} and X_{ijkt} are the Activity Level (A) and the set of technologies (T) that can be applied for a certain sector-activity.

So the total costs [Meuro] are:

$$C = \sum_{ijk} C_{ijk} \tag{7}$$

2.5 The constraints

The problem constraints are the following:

- technology feasibility (control variables are constrained to remain between minimum and a maximum value):

$$X_{ijkt}^{CLE} \leq X_{ijkt} \leq \overline{X}_{ijkt}, \; \forall ijkt \tag{8}$$

- emission conservation (for each sector-activity, and for each precursor, you can apply emission reductions to a maximum of 100 % of available emissions):

$$\sum_{t \in T_{ijk}} X_{ijkt} \leq 1 \tag{9}$$

2.6 Solution of the problem

To compute the solution of the problem, the two-objective optimization problem is tackled following the ϵ-Constraint Method [8]: the Air Quality objective is minimized using the Sequential Quadratic Programming approach [11], while the Cost objective is included in the set of constraints with a parametric threshold, i.e.:

$$\min_{X_{ijkt}} P(X_{ijkt}) \tag{10}$$

$$C(X_{ijkt}) \leq L, \; 0 \leq L \leq \overline{L} \tag{11}$$

where $\overline{L}$ is the cost of a full application of all the available technologies. This is the same form of a standard cost-effectiveness analysis: a problem that the Decision Maker may be interested to solve, when the budget L is known.

3 Case study

3.1 The Air Quality Index training and validation

The methodology has been applied on a Northern Italy domain [15], characterized by high level of pollution concentrations both in winter (pm) and summer (ozone). The region is a basin, with low wind circulation that brings to accumulation and reaction of secondary pollution.

Artificial Neural Networks (ANNs) have been identified to link control variables (through emissions) to concentrations on the study domain. Starting from more than 20 simulations performed with a Chemical Transport Model (CTM), ANNs have been identified and validated, so that the surrogate of the full CTM can then be used in the optimization methodology. The whole procedure has been already presented in detail in [4]. After having tested different ANNs configurations, the best ones (in terms of mean squared error) has been selected, for each of the Air Quality Index considered in this work. In particular the identified artificial neural networks apply a ”logsig” transfer function and 20 neurons in the hidden layer, and a ”purelin” transfer function in the output layer. A validation has been performed on the 20% of domain cells (not used in identification), comparing the CTM modeled values (x-axis) and the ANNs ones (y-axis). Both Figure 1 (related to mean PM25) and Figure 2 (related to mean SOMO35) show good agreement between the two models, ensuring that ANNs can be used as a surrogate of deterministic CTM. This allows for fast computations of different emission scenarios in the optimization procedure (not feasible directly using the CTM).

3.2 The Decision Making problem

In this Section the results of the optimization procedure are analyzed, stressing how the three interlinked problem dimensions (economic, environmental and social dimensions) are taken into account in the proposed approach. More in detail, the three dimensions are described as:

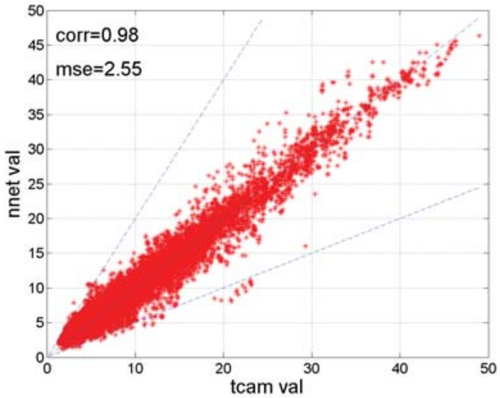


Figure 1: PM25 validation scatter plot, comparing CTM modelled values (x-axis) and the ANNs ones (y-axis).

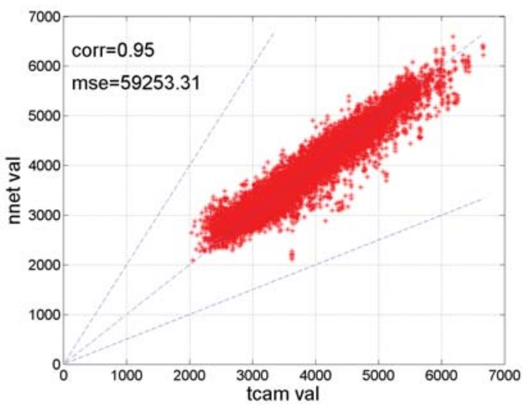


Figure 2: SOMO35 validation scatter plot, comparing CTM modelled values (x-axis) and the ANNs ones (y-axis).

- economic dimension, through the cost of implementation of a policy;
- environmental dimensions, through the evaluation of air pollution concentrations of mean PM25 and mean SOMO35;
- social dimensions, through the assessment of PM25-related Years Of Lost Life.

In Figure 3 an optimization is performed, to compute the Pareto curve considering two conflicting objectives: Cost Of Policy Implementation and Years Of Lost Life. The resulting Pareto curve is depicted in red (with triangles) showing that, starting from a base level of roughly 20 months of lost life (due to the elevated level of pollution in the case without further reductions) it is possible to reduce of more than 4 months this mortality value (to 16 months) spending 500 Millions of Euro per year.

As a side-effect of this optimization, two further AQIs are computed: the mean PM25 and the mean SOMO35. Mean PM25 (in red, squares) ranges between 24 and 19, and shows the same trend of YOLL, due to the fact that the two AQIs are roughly related to the same target (that is to say, reducing PM exposure). SOMO35 (in black) is instead an ozone AQI, and is in competition with the PM25 ones. In fact, as shown in the left part of the Pareto curve, an improvement of YOLL (or PM25) values is associated with a worsening of SOMO35 values (even if this trend is slightly mitigated for expensive policies, depicted to the right of the Pareto curve). This behaviour is due to the fact that emission precursor reductions required to improve PM25 and ozone AQIs are different (this aspect is further analysed in Figure 5 and Figure 6), and to the high nonlinearities involving ozone and PM formation and accumulation.

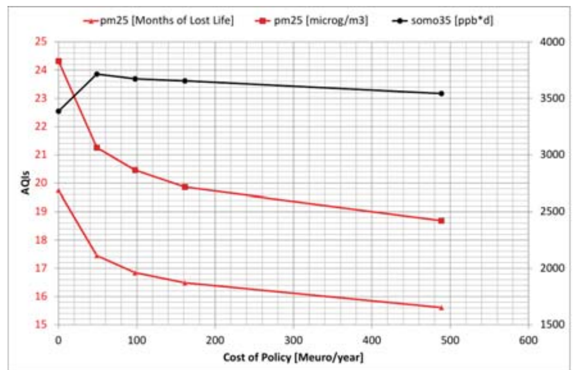


Figure 3: Pareto curve computed considering YOLL as optimization AQI. The x-axis is related to the cost of implementation of the policy, while y-axis is related to the AQIs values (considering PM25 and YOLL in the left axis, and SOMO35 in the right axis).

To emphasize the trade-off between AQIs, a second problem has been solved (Figure 4), optimizing the two conflicting objectives Cost Of Policy Implementation and SOMO35. In this case the results show that the more one spends in terms of policies, the more SOMO35 is improved (black curve); but no improvement is achieved for PM25 and YOLL indexes. In this case it is possible to notice that the SOMO35 policies are not worsening (as in the previous case) the other AQIs. This is mainly due to the fact that ozone emission reduction policies act on a subset of precursors required to reduce PM25 related AQIs.

Figure 5 focuses on emissions to be removed (per macrosector-pollutant) to obtain the second point of the Pareto curve, in the YOLL optimization case. This Figure suggests that optimal emission reductions have to be mainly in traffic macrosector (7 and 8) for NOx and VOC, and in agriculture macrosector (10), for NH3. Figure 6 shows similar results, but for the case of SOMO35 optimization. Here the removed emissions priorities are quite different and

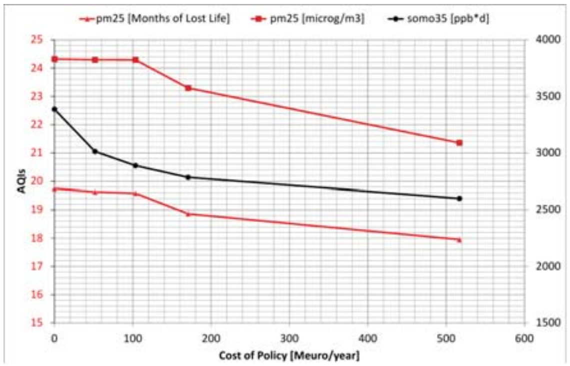


Figure 4: Pareto curve computed considering SOMO35 as optimization AQI. The x-axis is related to the cost of implementation of the policy, while y-axis is related to the AQIs values (considering PM25 and YOLL in the left axis, and SOMO35 in the right axis).

are mainly concentrated on solvent use macrosector (6), for VOC. The comparison of Figure 5 and Figure 6 explains why PM25 and SOMO35 optimal policies are in conflict (different priorities of precursor emission reductions, to face the different AQIs) and so that the only way to tackle the problem in a more comprehensive way is to switch to a multi-AQI framework.

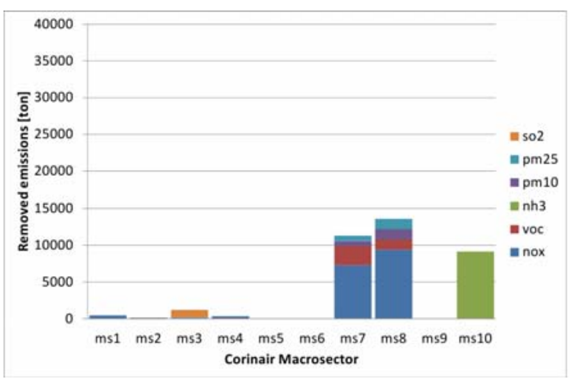


Figure 5: Removed emissions (beyond CLE) to obtain the second point of the pareto curve considering YOLL as optimization AQI.

As an example of the spatial dimensions of the decision problem, in Figure 7 and Figure 8 the PM25 maps over the optimization domain are shown, for the less expensive (Figure 7) and most expensive (Figure 8) points of the Pareto curve, showing the spatial minimum and maximum values (range of variability) of the obtainable optimal PM25 concentrations.

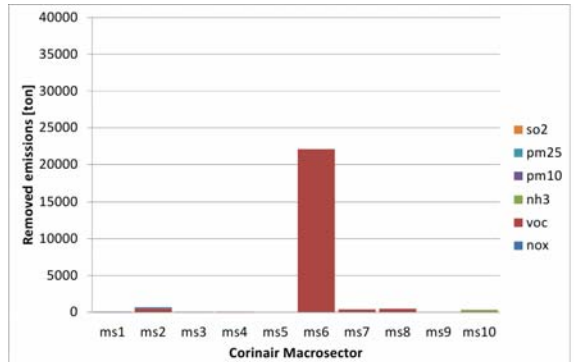


Figure 6: Removed emissions (beyond CLE) to obtain the second point of the pareto curve considering SOMO35 as optimization AQI.

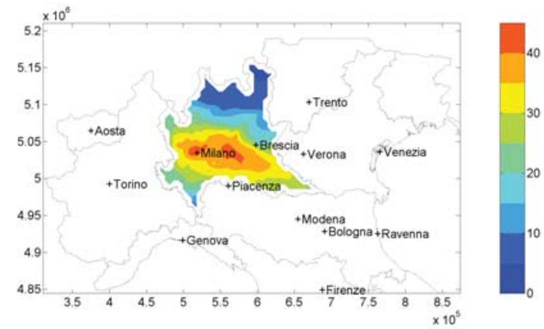


Figure 7: Spatial map of PM25 concentrations, for the less expensive point of the Pareto curve (extreme left of the Pareto curve, in Figure 3).

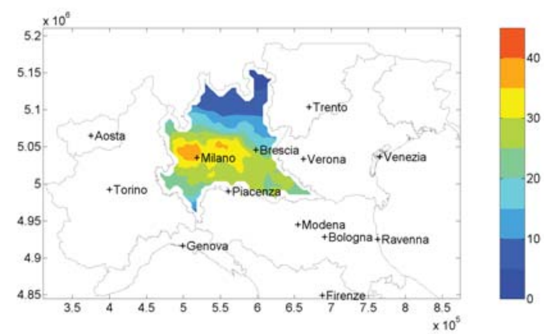


Figure 8: Spatial map of PM25 concentrations, for the most expensive point of the Pareto curve (extreme right of the Pareto curve, in Figure 3).

4 Conclusions

In the frame of a multi-objective optimization approach, this paper presents the comparison of two possible configurations of the problem. The correspondent ANNs have been identified and validated, and two optimizations have been performed. The results show how the percentage possible improvement on ozone Air Quality Index are higher than the one that can be obtained on PM Air Qualilty Index. Also the paper shows how the percentage emission reduction priorities differ between the ozone and PM optimizations. This last aspect stresses the conflict between ozone and PM emission reduction measures, and underlines how it is important to extend the proposed optimization approach to deal with multi-AQIs objectives, to properly look for a trade-off among different air quality measures.

Acknowledgments

We acknowledge Philippe Thunis (JRC-IES), Giorgio Guariso (Politecnico di Milano), Giuseppe Maffeis (TerrAria) for their valuable contribution. We acknowledge MAG-IIASA (Mitigation of Air Pollution & Greenhouse Gases, International Institute for Applied Systems Analysis) for data sharing. This work has been also supported by Regione Lombardia and CILEA Consortium through the LISA Initiative (Laboratory for Interdisciplinary Advanced Simulation) 2010 grant [link:http://lisa.cilea.it].

References

[1] J. Aben, J.-P. Hettelingh, and W. Schopp, "Rains-nl: An integrated assessment model to support dutch air quality policy making," in *Proc. 19th Intern. Conf. Informatics for Environmental Protection*, 2003, pp. 513–517.

[2] A. Altman, M. Amann, G. Klaassen, A. Ruszczyski, and W. Schpp, "Cost-effective sulphur emission reduction under uncertainty," *European Journal of Operational Research*, vol. 90, no. 3, pp. 395 – 412, 1996.

[3] C. Carnevale, E. Decanini, and M. Volta, "Design and validation of a multiphase 3d model to simulate tropospheric pollution," *Science of The Total Environment*, vol. 390, pp. 166–176, 2008.

[4] C. Carnevale, G. Finzi, G. Guariso, E. Pisoni, and M. Volta, "Surrogate models to compute optimal air quality planning policies at a regional scale," *Environmental Modelling and Software*, vol. 34, no. doi:10.1016/j.envsoft.2011.04.007, pp. 44–50, 2012.

[5] C. Carnevale, V. Gabusi, and M. Volta, "POEM-PM: an emission model for secondary pollution control scenarios," *Environmental Modelling and Software*, vol. 21, pp. 320–329, 2006.

[6] I. D'Elia, M. Bencardino, L. Ciancarella, M. Contaldi, and G. Vialetto, "Technical and non-technical measures for air pollution emission reduction: The integrated assessment of the regional air quality management plans through the italian national model," *Atmospheric Environment*, vol. 43, pp. 6182–6189, 2009.

[7] F. Deutsch, C. Mensink, J. Vankerkom, and L. Janssen, "Application and validation of a comprehensive model for pm10 and pm2.5 concentrations in belgium and europe," *Applied Mathematical Modeling*, vol. 32, pp. 1501–1510, 2008.

[8] M. Ehrgott, *Multicriteria Optimization*. Springer, 2000.

[9] J.Hugh and Ellis, "Multiobjective mathematical programming models for acid rain control," *European Journal of Operational Research*, vol. 35, no. 3, pp. 365 – 377, 1988.

[10] J. Lumbreras, R. Borge, J. M. de Andrs, and E. Rodrguez, "A model to calculate consistent atmospheric emission projections and its application to spain," *Atmospheric Environment*, vol. 42, no. 21, pp. 5251 – 5266, 2008.

[11] MathWorks, "Optimization Toolbox for Use with Matlab," The MathWorks, Tech. Rep., 2010.

[12] A. Mediavilla-Sahagun and H. ApSimon, "Urban scale integrated assessment of options to reduce pm10 in london towards attainment of air quality objectives," *Atmospheric Environment*, vol. 37, pp. 4651–4665, 2003.

[13] T. Oxley and H. ApSimon, "Space, time and nesting integrated assessment models," *Environmental Modelling & Software*, vol. 22, pp. 1732–1749, 2007.

[14] S. P., C. D.P., and W. S.L., "An Implicit-Explicit Hybrid Solver for a System of Stiff Kinetic Equations," in *Proc of Air & Waste Management Association* 87^{th} *Annual Meeting*, Cincinnati, US, 1994.

[15] E. Pisoni, C. Carnevale, and M. Volta, "Multi-criteria analysis for PM10 planning," *Atmospheric Environment*, vol. 43, pp. 4833–4842, 2009.

[16] S. Syri, N. Karvosenoja, A. Lehtil, T. Laurila, V. Lindfors, and J.-P. Tuovinen, "Modeling the impacts of the finnish climate strategy on air pollution," *Atmospheric Environment*, vol. 36, pp. 3059–3069, 2002.

[17] C. Vlachokostas, C. Achillas, . Moussiopoulos, E. Hourdakis, G. Tsilingiridis, L. Ntziachristos, G. Banias, N. Stavrakakis, and C. Sidiropoulos, "Decision support system for the evaluation of urban air pollution control options: Application for particulate pollution in thessaloniki, greece," *Science of The Total Environment*, vol. 407, no. 23, pp. 5937 – 5948, 2009.

[18] M. Volta and G. Finzi, "GAMES, a comprehensive Gas Aerosol Modelling Evaluation System," *Environmental Modelling and Software*, vol. 21, pp. 587–594, 2006.

[19] F. Wagner, M. Amann, and W. Schoepp, "The GAINS optimization module as of 1 February 2007," International Institute for Applied Systems Analysis, Laxenburg, (AU), Tech. Rep., 2007.

Technology Contribution Rate: Concepts, Framework and Case Study

Hanlin You
Department of Management
National University of Defense Technology
Changsha City, Hunan Province, China
hlyou1990@hotmail.com

Qingsong Zhao
Department of Management
National University of Defense Technology
Changsha City, Hunan Province, China
zqszqr@163.com

Yanjing Lu
Department of Management
National University of Defense Technology
Changsha City, Hunan Province, China
yanjinglu_nudt@yahoo.com

Jiang Jiang
Department of Management
National University of Defense Technology
Changsha City, Hunan Province, China
jiangjiangnudt@hotmail.com

Abstract - *Technology Contribution Rate (TCR) evaluation of Weapon System-of-Systems (WSoS) is an essential question of WSoS technology evaluation , a key basis of WSoS architecture optimization and WSoS technology development planning. TCR studies the contribution of a technology made for the WSoS to accomplish its mission. TCR evaluation includes two phases: the contribution of a technology made to a system and the contribution of a technology made to the WSoS. Relative concepts and framework to evaluate the TCR are introduced. The Analytical Network Process and the Grey Target theory are applied to build the framework of solution. A field defense WSoS is studied to validate the efficiency of the theoretical framework.*

Keywords: *WSoS*, *TCR*, *ANP*, Grey Target theory

1 Introduction

With the increasing of the complexity of Weapon System of Systems (WSoS), more and more advanced technologies need to be developed for application. Considering the limitation of resource, the effective WSoS architecture optimization and WSoS technology development planning is needed to maximize the overall efficiency of WSoS with acceptable cost. As a result, the research of evaluating the technologies of WSoS has become more and more important.

WSoS requirements related research has drawn a lot of attentions [1]. The development of technologies is the essential forces in driving major innovations and revolutions in military field. There is a very pressing demand for the contribution of a technology made to a WSoS to fulfill its mission. Therefore Technology Contribution Rate (TCR) evaluation is of great significance in WSoS requirements study [2].

In this study, a method of evaluating the influence of technologies to WSoS by calculating TCR is proposed to obtain the priorities and compare the TCR of different technologies to analyze the influence factors. The relative concepts and framework of calculating TCR are proposed in Section 2. A case of a field defense WSoS is studied to validate the efficiency of the solution framework in Section 3. In Section 4, relative conclusions are drawn.

2 Relative concepts and framework

This study is based on the WSoS hierarchy derived from the WSoS requirements related research, which dictates that there is a “mission-tasks-capability-system-technology” requirements chain [3]. The mission of the WSoS is decomposed to be certain tasks. Tasks require capabilities. Capabilities are carried out by systems and finally systems are supported by technologies. As a result, a complete hierarchy of WSoS can be drawn by analyzing the technology supporting the functions of weapons, as shown in Figure.1

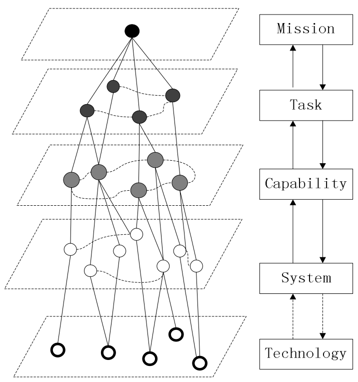


Figure.1 hierarchical relationships of WSoS

Relative concepts applied in this study include: **System Military Value (SMV)** is defined as the contribution of each weapon system made to complete the missions of WSoS [4]. **Technology Support Rate (TSR)** is defined as the contribution of a technology made to support the functions of the system. **TCR** is defined as the contribution of a technology made to complete the mission of WSoS. The analysis of TCR is based on TSR and SMV.

The **Technology Influence Matrix (TIM)** [5-7] is introduced to define the relationship between technologies and tactical & technical indices, which is applied in TSR evaluation. The **Grey target theory** is introduced as a theory of grey correlation analysis to handle mode sequences [8-14], which is applied in TSR evaluation. The steps to calculate TSR are as following, as shown in Figure.2:

Step 1: Obtain the system list from WSoS and select the system sys_j to be analyzed.

Step 2: Obtain tactical & technical indices list from relative technology reports based on selected system sys_j and determine the tactical & technical indices $index_k$ $k \in (1,\dots,K)$ related to the system sys_j.

Step 3: Build the TIM and determine the technology $tech_i$ $i \in (1,\dots,n)$ related to the tactical & technical index $index_k$.

Step 4: Obtain the support rate of each technology to each tactical & technical index and calculate the influence rate of each technology to each tactical & technical index $value(tech_i, index_k)$.

Step 5: Calculate the influence rate $value(index_k, sys_j)$ of the tactical & technical index $index_k$ to the system sys_j using Grey target theory.

Step 6: Calculate the contribution rate $TCR(i,j)$ of the technology $tech_i$ to the system sys_j by combining the results of *Step4* and *Step5*:

$$TCR(i,j) = \sum_{i=1}^{n} value(tech_i, index_k) \bullet value(index_k, sys_j)$$

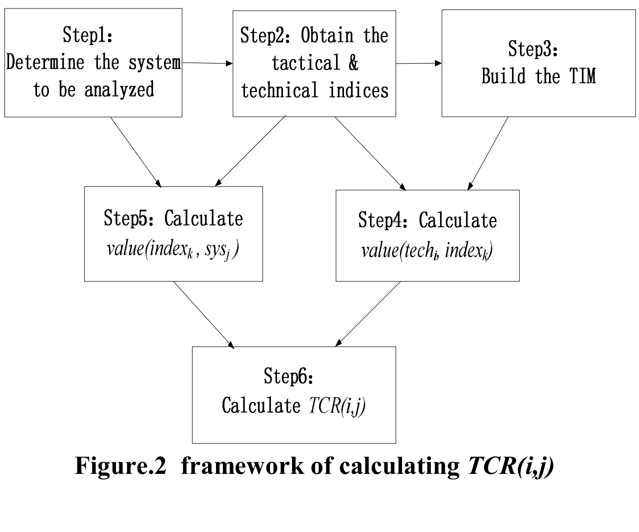


Figure.2 framework of calculating *TCR(i,j)*

The Analytical Network Process (ANP) is applied in **SMV** evaluation [4]. The "mission-tasks-capability-system" hierarchy architecture model can be built by analyzing the mapping relationships. The contribution of each system to complete the mission of WSoS can be calculated using the decision software SD (super decision) based ANP theory developed by Rozann W. Saaty and William Adam [15].

Suppose there are N systems and M technologies in a WSoS, the steps to calculate TCR are as following, as shown in Figure.3:

Step 1: Build the WSoS hierarchical;

Step 2: Map the relationship of index-system from the index list of each system;

Step 3: Build TIM to obtain the influence rate of each technology to each index;

Step 4: Calculate contribution of each index to the systems using the mapping relationship obtained in *step2* using Grey Target theory. This step will be repeat N times.

Step 5: Build the mapping matrix of systems and technologies, and obtain the support relation between system n_i $(0<n_i<N)$ and technology $tech_i$ $(i \in (1,2,\dots,M))$.

Step 6: Calculate the contribution of each system n_i, in which there is mapping relation based on the result of *Step4* and *Step5*. Repeat this calculation step n_i times to obtain $TCR(i,j)$ of $tech_i$ $(i \in (1,2,\dots,M))$ to $sys_j (j \in (1,\dots,n_i))$.

Step 7: Calculate the $SMV(j)$, $(j \in (1,\dots N))$ of each weapon systems sys_j using ANP[4];

Step 8: Calculate the $TCR(i)$ of technology i

$$TCR(i) = \sum_{j=1}^{n} (TCR(i,j) \bullet SMV(j)) .$$

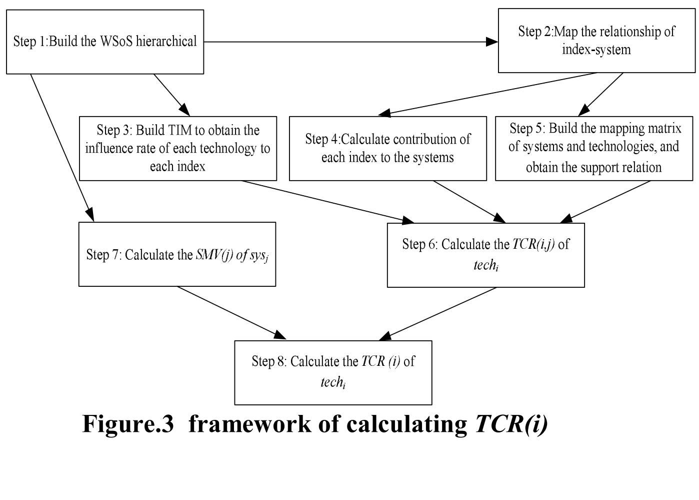


Figure.3 framework of calculating *TCR(i)*

3 Case study

A field defense WSoS is proposed as a case study. TCRs of the technologies of the WSoS are calculated to verify the feasibility and validity of the proposed framework and method. The results are analyzed by comparing the TCR of different technologies.

3.1 Backgrounds

The background of the case study is demonstrated as following: because of certain historical reasons, country *N* and country *S* are in confrontation. In a period of time in the future, country *S* may get support from a third country and invade. In order to defend the territory's integrity, country *N* plans to build up a ground defense WSoS, which aims at aggressively defense of the territory of country *S*. As the basis of the building planning of the ground defense WSoS, the scenario of future combat is drawn as Figure.4

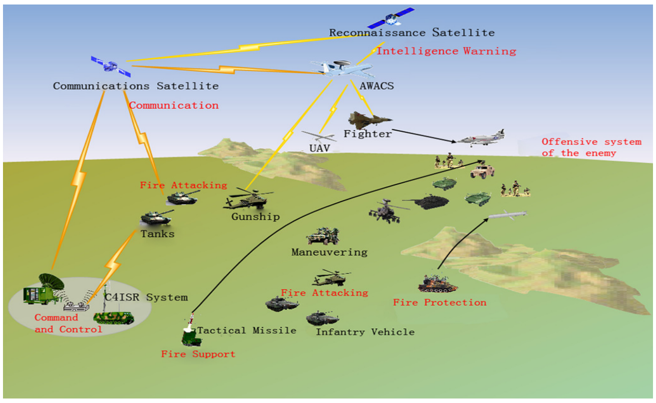


Figure.4 scenario of future combat

As shown in Figure.3, the ground defense WSoS includes a "mission-tasks-capability-system-technology" equipment system, which is demonstrated in Table.1.

Table.1 Mission, task, capability and weapon systems of WSoS of ground defense

Project	Details
Mission	Attacking the land-based invasion of country S effectively and attaining the goal of positive defense.
Task	Finding enemy targets, information defense, fire attacking, moving counterattack, fire support and surveillance
Capability	Moving capability, defense capability(air-defense and information defense), fire attacking capability, surveillance capability, command and control capability and support capability
Weapon system	Tactical missiles(artillery), tanks, infantry combat vehicles,C4ISR systems, wired communication, wireless communication, UAV, reconnaissance satellites, support vehicles, emergency rescue vehicles

In the section, tanks and infantry combat vehicles are selected from weapon systems as an example to demonstrate the solution framework of TCRs.

3.2 TCR result

The overall performance of tanks should be considered to analyze the TSR. With the limitation of accessibility, objectivity, quantifiable and comparability, the tactical & technical indices such as fire performance indices, motor performance indices, protection performance indices and overall control indices are selected to analyze the support of each technology to tanks. The support rate of each technology to each tactical & technical index is shown by TIM in Table.2.

Table.2 TIM of tanks

Technology / Tactical & technical indices	body design	ammunition	material	engine	moving-control	fire-control	defense technology
weapon weight	10%		-20%				
caliber	20%						
ammunition rounds	10%						
ammunition muzzle velocity	10%			10%			
effective range	10%	10%				5%	
rate of fire							
thickness of punctured armor in 2km		20%					10%
unit power				50%	25%		
maximum power				20%	10%		
cross-country average speed				50%	10%		
highway maximum speed				50%	10%		
maximum stroke					10%		
maximum grade ability	10%				5%		
height of crossing	30%				5%		
width of crossing	20%				5%		
depth of diving	30%				5%		
armor types			50%				
equivalent thickness of best armor			20%				

The influence rate of each technology to each tactical & technical index is calculated by the TIM and the influence rate of the tactical & technical index to tanks is obtained using the Grey target theory as steps shown in Section 2. Then TSR of each technology to tanks is resolved, shown as Table.3.

Table.3 TSR of each technology to tanks

Technology	Index	Value ($index_j$, sys_k)	TSR
body design	weapon weight	0.655	0.659
	caliber	0.755	
	ammunition rounds	0.621	
	ammunition muzzle velocity	0.755	
	rate of fire	0.579	
	maximum grade ability	0.67	
	height of crossing	0.676	
	width of crossing	0.645	
	depth of diving	0.596	
ammunition	effective range	0.558	0.591
	thickness of punctured armor in 2km	0.607	
material	weapon weight	0.655	0.735
	armor types	0.706	
	equivalent thickness of best armor	0.727	
engine	ammunition muzzle velocity	0.755	0.697
	unit power	0.723	
	maximum power	0.722	
	cross-country average speed	0.704	
	highway maximum speed	0.643	
fire-control	effective range	0.558	0.558
defense technology	thickness of punctured armor in 2km	0.607	0.667
	equivalent thickness of best armor	0.727	
moving-control	unit power	0.723	0.661
	maximum power	0.722	
	cross-country average speed	0.704	
	highway maximum speed	0.643	
	maximum stroke	0.635	
	maximum grade ability	0.67	
	height of crossing	0.676	
	width of crossing	0.645	
	depth of diving	0.596	

SMV of tanks is resolved by the software SD based on ANP theory and TCR of each technology related to tanks is obtained by the steps shown in Section 2. Meanwhile, TCR of each technologies related to infantry combat vehicles is obtained using the same method.

The target tank in this case is the ZTZ99, whose referenced systems include MIA2sep, Merkava Ⅳ, Leclerc, T90 and Leopard 2A6. The target infantry combat vehicles are the BMII-3, whose referenced systems include BMII-2, M2, and AMX-10. According to the TSR and SMV of tanks and infantry combat vehicles, conclusions about TCR are calculated, shown as Table.4. In the "technology" column, the word "technology" is omitted and in the system column, "vehicle" stands for the "infantry vehicle".

Table.4 TCR of the ground defense WSoS

Technology	System	SMV	TSR	TCR
body design	tank	0.214	0.659	0.176
	vehicle	0.046	0.756	
ammunition	tank	0.214	0.591	0.146
	vehicle	0.046	0.426	
material	tank	0.214	0.7348	0.183
	vehicle	0.046	0.553	
engine	tank	0.214	0.6971	0.183
	vehicle	0.046	0.723	
moving-control	tank	0.214	0.661	0.173
	vehicle	0.046	0.681	
Fire-control	tank	0.214	0.558	0.147
	vehicle	0.046	0.589	
defense technology	tank	0.214	0.133	0.067
	vehicle	0.046	0.832	
water-surface moving	vehicle	0.046	0.622	0.029

3.3 Discussions

SMV has large influence on TCR. For instance, water-surface moving technology only support the water-moving function of infantry combat vehicle, and its TSR is 0.622, it ranked 4th in all 9 support technologies. However, if the SMV (0.026) is taken into account, its TCR is only 0.029, which ranked at the bottom of the list.

The technologies which support more weapon systems have bigger TCR. Comparing water-surface moving technology and fire-control technology, the TCRs are 0.029 and 0.027 if we just take the support to functions of infantry combat vehicle into account (the two TSRs are 0.622 and 0.589). However, if the support to functions of tanks of fire-control technology is also taken into account (its TSR is 0.558), its total TCR reaches 0.146.

4 Conclusions

The evaluation of TCR is a part of assessment of WSoS, which is useful for stakeholders to obtain a better knowledge of technology. The technologies with bigger TCR should be developed with higher priority and substitutes are demanded to cover the functions of the technologies lack of TCR. Analysis of TCR is applied for structure-optimization of WSoS and development plan of technologies. The method to build WSoS with lower cost, shorter duration and higher overall efficiency is still needed to research in future work.

References

[1] Z. R. Ni, Y. P. Wang. “Architecture and technical framework of weapon system of systems laboratory,” Military Operational Research and System Engineering, Vol 3, pp. 28-31. 2011.

[2] Leilei Chang. “Method Research on Technology Contribution Rate Evaluation of Weapon System-of-Systems,” Changsha: National University of Defense technology, 2010.

[3] Qingsong Zhao. “Research on Methods of Weapon and Equipment SoS Requirement Description and Analysis Based on Semantic Model,” Changsha: National University of Defense technology, 2008.

[4] L. Liu. T. Jing, X. Y. Wu. “Research on Evaluation of Armament System of Systems Evolution,” Journal of System Simulation, Vol 18, pp. 621-624, Aug. 2006.

[5] M. R. Kirby. “A Methodology for Technology Identification Evaluation and Selection for Complex System,” PowerPoint Presentation, 2003.

[6] Kirby M R . A Methodology for Technology Identification，Evaluation，and Selection in Conceptual and Preliminary Aircraft Design. Georgia Institute of Technology，2001.3.

[7] Kirby M R.，Rupp J.，Mavris D N. An Approach for the Strategic Planning of Future Technology Portfolios. World Aviation Congress，Reno，NV，2004.11.

[8] J. L. Deng. “The Control Problems of Grey Systems,” Systems＆Control Letters, Vol.1, No.3, pp. 288-294, 1982.

[9] Liu Si-feng. “Emergence and Development of Grey System Theory and Its Forward Trends,” Journal of Nanjing University of Aeronautics & Astronautics, Vol.36, No.2, Apr.2004.

[10] Deng Julong . The Control Problems of Grey Systems[J] . Systems & Control Letters ， 1982 ， 1(5)： 288-294.

[11] Chandan Chakraborty，Debjiani Charkraborty. A fuzzy clustering methodology for linguistic opinions in group decision making. Applied Soft Computing，2007，7(3)：858-869.

[12] Wu Desheng. Supplier selection in a fuzzy group setting ： A Method using grey related analysis and Dempster-Shafer Theory . Expert Systems with Application，2009，36(5)：8892-8897.

[13] Wang Yujie. Combining grey relation analysis with FMCGDM to evaluate financial performance of Taiwan container lines. Expert Systems with Application，2009，36(5)：2424-2432.

[14] Xu Zhongxiang etal . The sliding grey relational method for delineating reqions containing oil & gas. The Journal of Grey System，1996(3)：275-282.

[15] Rozann W Saaty. Decision making in complex environment: The Analytic Hierarchy Process (AHP) for Decision Making and The Analytic Network Process (ANP) for Decision Making with Dependence and Feedback. 2003.2.

Technology System of Systems: concepts and hierarchical structure

Leilei Chang
Department of Management
National University of Defense Technology
Changsha City, Hunan Province, China
leileichang@hotmail.com

Yanjing Lu
Department of Management
National University of Defense Technology
Changsha City, Hunan Province, China
yanjinglu_nudt@yahoo.com

Qingsong Zhao
Department of Management
National University of Defense Technology
Changsha City, Hunan Province, China
zqszqr@163.com

Jiang Jiang
Department of Management
National University of Defense Technology
Changsha City, Hunan Province, China
jiangjiangnudt@hotmail.com

Abstract - *This study aims at planning for the development and standardization of military related technologies. From the perspective of military requirements, the Weapon System of Systems (WSoS) related research is referenced. The definition and characteristics of Technology System of Systems (TSoS) are discussed. TSoS related research is reviewed. Types of capability requirements, as well as technologies classification, are studied and compared. A TSoS construction method is proposed using the tactical and technical indices as bridges between capability/system requirements and technologies. A showcase to construct a TSoS is studied to validate the efficiency of the proposed method using sixty technologies derived from the US weapon acquisition for the 2012 fiscal year.*

Keywords: technology system of systems; weapon system of systems; hierarchical structure

1 Introduction

The development of technologies is the essential forces in driving major innovations and revolutions in military. However, the connections between technologies with capabilities and system requirements have been overlooked.

To define the Technology System of Systems (TSoS), the concept of Weapon System of Systems (WSoS) [1] needs to be introduced first, especially the WSoS requirements related research. WSoS is the collection of systems those are connected, interoperated and coordinated to accomplish certain combatant tasks and fulfill certain mission.

WSoS requirements related research demonstrates that there is a “requirements chain” in the hierarchy of WSoS [2]-[5], as shown in Fig 1.

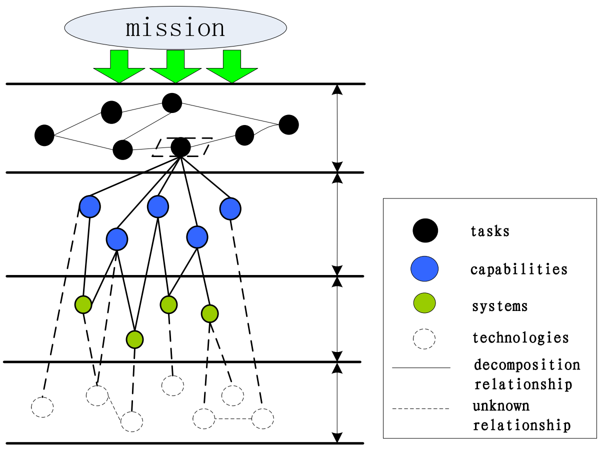


Figure 1. The hierarchy of WSoS requirements

On top of the hierarchy of WSoS is the mission, which depicts the goal of the WSoS. The mission of the WSoS is mostly recapitulative and serves as the guideline for the WSoS. Driven by the mission are the tasks which are mostly time-sequenced. In the third level, there come the capabilities. Each capability is demanded by one or more tasks and each capability is fulfilled by one or more weapons, equipment or platforms, which leads to the system (weapon/ equipment/ platform) level. However it is novel that there is a missing part in the hierarchy of WSoS---the technologies.

TSoS studies the relationship between technologies, Results of TSoS related research can be applied but not limited to: (1) extend the WSoS requirements analysis to a fundamental level; (2) provide a scientific description on the WSoS related technology framework; (3) provide principles for WSoS planning and decision making; (4) serve as a key influential factor to risk analysis of WSoS construction.

2 concept, construction of TSoS

2.1 Concept and purpose of TSoS

The TSoS is the family of technologies those are required by all systems and capabilities of the WSoS. TSoS studies the connections between technologies which support WSoS to fulfill certain function and accomplish its mission. The description, modeling and evaluation of the TSoS structure are included in the TSoS related research as well.

Maier summarized that there were five characteristics for a System of Systems (SoS) [6]. Following the five characteristics and the definition of TSoS, the characteristics of a TSoS are pointed out: (1) technology attributes are similar because they share the same attributes such as level, type, TRL, risk and et al. This characteristic separate TSoS from WSoS essentially. (2) The structure of a TSoS is a hierarchy since capability requirements and system requirements, in the upper level than technologies, are hierarchy. The clearer the requirements are, the more layers in the TSoS hierarchy and vice versa; (3) Characteristics of technologies emerge to be the characteristics of the TSoS, which can only be recognized and studied in a SoS level; (4) TSoS is evolving in both the technology level and the SoS level.

Since technologies are in the fundamental level of the WSoS hierarchy, TSoS related research is of great significance in military and especially in the WSoS requirements study.

2.2 TSoS related research

Currently there is no an independent subject on TSoS. Related research is resided in military or technical subjects. Few literatures on TSoS are quite superficial and hardly touch the core of the TSoS.

In military subjects, DoDAF [7] and MoDAF [8] are the official documents for the defense architecture framework issued by the US and UK. From the multi-view perspective, DoDAF and MoDAF provide guidance on principles and methodologies for WSoS construction. However there are only two Technology Views, or the Standard Views (StdVs), StdV-1 and StdV-2, which is not even close to solving the complicated challenges of TSoS. Other military related studies [3] [9] [10] provide research results from WSoS requirements, military requirements, experiments & evaluation and war deduction. These studies did not come to the technical aspects.

In technical aspects, the most widely-accepted method is the Technology Readiness Level (TRL) [11] which classifies the stages of a technology’s life time into 9 scales and each scale is linked to one or two phases of the corresponding life stage of a product or weapon acquisition (depending on what subjects it is applied to). Another popular method is “technology prediction” [12]. Japan has been the most successful in implementing technology prediction. Technology map is carried out by China [13] in 2007 for the first time. However, these three methods are subjective and hardly produce any quantitative results.

Liu [10] discussed the evaluation for the capabilities of WSoS when there is technology infusion. Ren [4] extends the TRL assessment to both the system and the SoS level. Chang [5] discussed the contribution made by a technology for the system/platform to equip its capability and the WSoS to fulfill its mission.

It is noted that there is still no a systematic subjects on TSoS. To step out the first step, we need to determine how to find all the technologies those belong to a TSoS.

2.3 How to get technologies of a TSoS

The basis of TSoS related research is to obtain the technologies those belong to a TSoS. Technologies in a TSoS originate not only from system requirements but also capability requirements. Technologies are abstract and essentially obscure. Trying to tie a technology to a system or a capability is difficult. Thus a framework is proposed to mapping a TSoS.

Any technology required by a WSoS must be utilized in a system or a capability, which means that technologies are driven by system or capability requirements. This leads us to the embodiments of technologies in the systems and capabilities: indices, to be more specific, the tactical indices and technical indices.

Tactical indices are required by capabilities. For example, the mobility of “200 kilometers within two hours” requires certain kind of engine technology. Technical indices are required by systems. For example, the speed of a tank should be over 100km/h, which requires a particular kind of tank engine technology.

However, under most circumstances, the capability requirements need to be transferred into system requirements first. This is because capabilities are carried out by systems. Still following the previous example, if a tank is used, then this would be a very high demand on a tank engine technology; if a military truck is used, this would be a very low demand. Fig. 2(a) shows this common means to obtain most technologies.

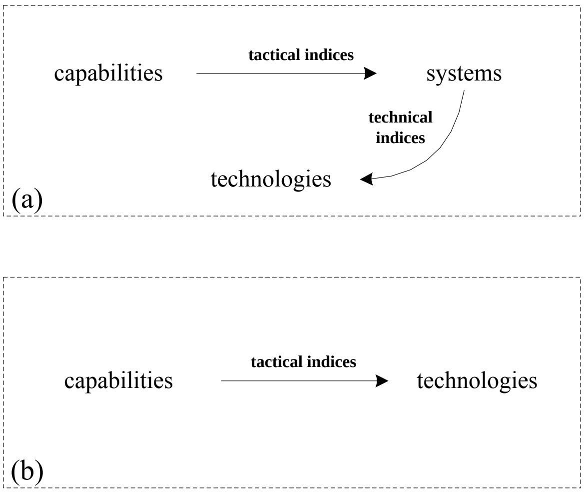


Figure 2. The derivation of technologies for a TSoS

There are also some exceptional conditions. Certain capabilities are very abstract and conceptual, e.g. the one-hour “prompt global strike” of US air force [14]. Mostly these are strategic capability requirements, which make it even harder to determine what the corresponding system requirements are. Therefore, the previous “capability-system-technology” means is no longer applicable. Under this condition, capability requirements are directly tied to technologies, as shown in Fig. 2(b).

3 TSoS hierarchical structure

Technologies derived from different means vary as well. The “capability-system-technology” derived technologies are mostly clear and realistic, existing, developing or within a developing scheme. These technologies are either to be improved or in a scheme to replace a retiring technology.

The technologies directly derived from capabilities, on the other hand, are mostly obscure, theoretic, nonexistent and unpredictable. These technologies are mostly still in its laboratory stage with a very low TRL [11].

There are three kinds of technologies, named as the “3S” as shown in Table 1. Characteristics and classification of the technologies are shown as in Table 1. The characteristics of the TSoS constructing means are included in Table 1 as well.

Table 1. Construction means and classification of TSoS

TSoS construction means			Technology	
illustration	graphics	capability requirements characteristics	characteristics	types
capability-system-technology	Fig. 2(a)	clear, explicit, can be tied to a specific system	required only by one system	system technology
			shared by more than one system	shared technology
capability-technology	Fig. 2(b)	clear, explicit, cannot be tied to a specific system	platform independent	
		obscure, implicit, cannot be tied to a specific system	exploring, innovative, dynamic	strategic technology

The system technologies that are required by a specific system, the shared technologies that are shared by more than one system and the strategic technologies that are required directly by the capability.

We would also like summarize and propose a four-step framework to construct a TSoS:

Step 1: group all technologies that belong to a certain system requirement. Since the capability and system requirements are hierarchical, the technologies derived in this step are in hierarchical structure. Additional, the very first technology tree is derived;

Step 2: Delete the technologies those belong to more than one technology groups. Add the technologies those do not belong to a certain system but are derived from capability requirements. Note that although the two kinds of technologies are both shared technology, they are derived in different ways.

Step 3: Add the technologies that meet the future challenges, known as the strategic technologies. The strategic technologies are abstract and notional;

Step 4: construct the TSoS by summarizing the 3S technologies;

The four-step framework, as shown in Fig. 3, is designed for two purposes: (1) to make sure that the constructed TSoS meet the current demand from WSoS capability and system requirements; (2) to make sure that constructed TSoS does not fall behind the strategic requirements of national defense in a certain period.

It should be noted that the system technologies would be dominant in numbers for the TSoS since they are the main resource of the TSoS while the shared technologies and the strategic technologies are the key supplements. The hierarchy structure of the technologies depends on the clearness of the capability requirements. The clearer the capability requirements are, the clearer the system requirements are and the more we understand about the hierarchy of the technologies. System technologies are never just one level (usually three) and so are the system-shared technologies since they are derived from system technologies; cross-platform technologies are mostly with fewer layers because they are needed by more than one system and therefore already subdivided; strategic technologies are so abstract to be divided.

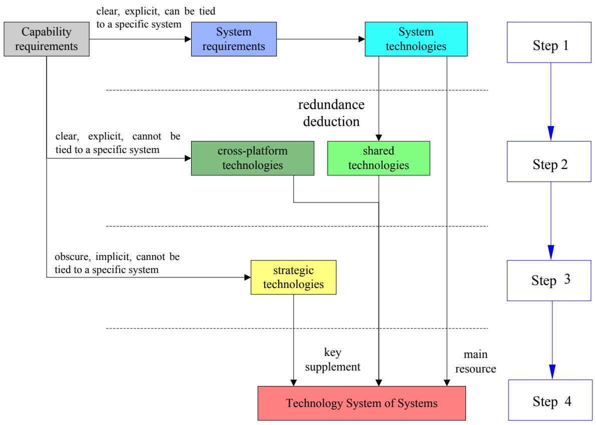


Figure 3. Steps to get a TSoS

4 Case study

The data of this case study is derived from the US weapon acquisition for the 2012 fiscal year [15], of which 60 key technologies are selected. With the 60 technologies, the TSoS is constructed to verify the efficiency of the proposed approach. Fig. 4 shows this process. Names and the classification of 60 technologies are in the Appendix.

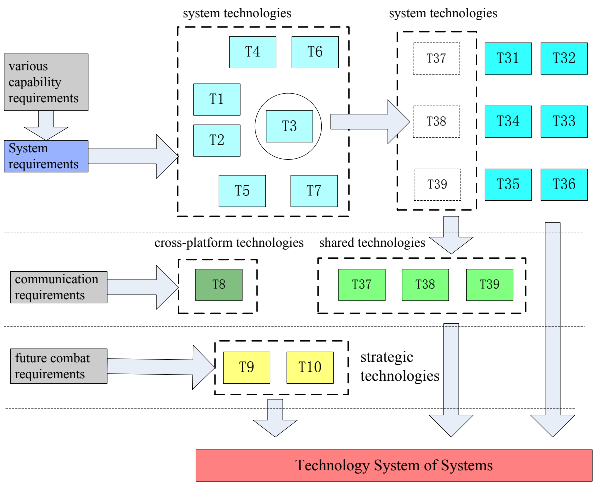


Figure 4. The derivation of a TSoS

Follow the four-step framework in Section 3 and construct the TSoS:

Step 1: Driven by capability requirements in various combatant conditions, seven main technologies are derived, including the shipbuilding and maritime technology (T1), space based and related technology (T2), aircraft technology (T3), C4I technology (T4), ground program technology (T5), missile defense technology (T6) and missile technology (T7).

Step 2: integrate those technologies that required by many systems/platforms, “material technology (T37)”, “(T38)” and “(T39)”; “communication technology （T8）” derived from communication capability requirements.

Step 3: strategic technologies, including “nanotechnology (T9)” and “infrasound technology (T10)”, are derived from future combat requirements.

Step 4: construct a TSoS by summarizing the 3S technologies;

5 Conclusions

The concepts, characteristics, structure and derivation of TSoS are discussed. Driven by different capability requirements, a TSoS construction method is proposed using the tactical and technical indices as bridges.

Compared with other SoS related research, e.g., health, transportation and et al., TSoS related research is still in its very preliminary stage. The concept of TSoS is not clear yet the demand is quite urgent, which make TSoS related research even more exciting and promising since it means that there are more fields to explore. We would like to point out the future work that needs to be done: the practical study. Without practical background, TSoS related study would stay superficial and cannot be validated.

References

[1] Z. R. Ni, Y. P. Wang. “Architecture and technical framework of weapon system of systems laboratory”, *Military Operational Research and System Engineering*. Vol 3, pp. 28-31, 2001.

[2] U.S, Department of Defense. “*DoD Dictionary*”, http://www.dtic.mil/ doctrine/jel/doddict, Dec. 2011.

[3] Q. S. Zhao. “*Research on methods of weapon and equipment SoS requirement description based on semantic model and its application*”, Changsha: National University of Defense Technology, 2008.

[4] C. S. Ren. “*An evaluating approach for technology maturity of weapons system of systems*”, Changsha: National University of Defense Technology, 2010.

[5] L. L. Chang. “*Method research on technology contribution rate evaluation of weapon system of systems*”, Changsha: National University of Defense Technology, 2010.

[6] M. W. Maier. “*Architecting Principles for system of systems*”, Proceedings of the 6th Annual Symposium of INCOSE, 1996.

[7] DoDAF Working Group. “*DoDAF (version2.0)*,” Washington: Department of Defense, May. 2008.

[8] The MODAF Development Team. “*MODAF handbook version1.2*”, London, Ministry of Defense, 2008.

[9] D. A. Brown, “*Quantitative Technology Assessment in AFRL*”, Washington: the Department of Defense M&S Conference，2006.

[10] X. L. Liu. “*A methodology for impact evaluation of technology infusion on the capability of weapon system of systems*”, Changsha: National University of Defense Technology, 2009.

[11] J. C. Mankins, “*Technology readiness levels, A White Paper*”, NASA, Washington, DC, 1995.

[12] T. Kuwahara. “*Technology forecasting activities in Japan*”, Technology forecasting and social change, Vol 60, pp. 5-14, 1999.

[13] National technology prediction group, “*Suggestions on mapping the national technology*”, Technology Forum in China, Vol 5, pp. 3-6, 2008.

[14] E. Grossman. “*Air force proposes new strike missile*”, military.com, Feb. 2012.

[15] Office of the under secretary of defense (Comptroller)/ CFO. “*Program acquisition cost by weapon system*”, Washington: Department of Defense, Feb. 2011.

Appendix

TSoS comprised of 60 technologies derived from 2012 US weapon acquisition

types of technologies	Technologies (the word "technology" is omitted)	
system technology	maritime (T1)	surface combatant (T11), submarine combatant (T12), support ships (T13), high speed voyage (T14), complex environment adaption (T15)
	space based and related (T2)	launch (T21), satellite (T22), Support (T23), narrow band Ultra High Frequency (UHF) (T24), Ultra High Frequency (UHF) (T25), GPS related technology (T26) space-based infrared (T27)
	aircraft (T3)	UAV (T31), day-night Full Motion Video (FMV) (T311), signal intelligence (SIGINT) (T312), Synthetic Aperture Radar (SAR) (T313), data links (T314), ISR (T315), combat aircraft (T32), Short Take-off and Vertical Landing (STOVL) (T321), supersonic and long distance (T322), cargo aircraft (T33), oversized payloads lifting (T331), short distance light payloads (T332), training aircraft (T34), helicopter (T35), Pilot Night Vision (PNV) (T351), vibration reduction (T352), Target Acquisition Designation Sight (TADS) (T353), heavy payload (T354)
	C4I (T4)	theater combat C3 (T31), automation (T32), base communications (T33), information security & assurance (T34), support (T35)
	ground program (T5)	medium tactical vehicles (T41), light tactical vehicles (T42), heavy tactical vehicles (T43), weapons (T44), support (T45)
	missile defense (T6)	ballistic missile defense (T51), ballistic missile defense ground support (T52), tactical missile defense (T53)
	missile (T7)	strategic missiles (T71), tactical missiles (T72)
shared technology	cross-platform technology	communication technology (T8)
	shared technologies by	warning aircraft (T36), material (T37), stealthy technology (T38)
strategic technology	nanotechnology(T9), infrasound(T10)	

Research on Evolving Capability Requirements oriented Weapon System of Systems Portfolio Planning

Zhou Yu
College of Information System and Management，
National University of Defense Technology，Changsha
Hunan 410073, P. R. China
zhouyu19830627@gmail.com

Ke-wei Yang
College of Information System and Management，
National University of Defense Technology，Changsha
Hunan 410073, P. R. China
kayyang27@hotmail.com

Yue-jin Tan
College of Information System and Management，
National University of Defense Technology，Changsha
Hunan 410073, P. R. China
yjtan@nudt.edu.cn

Zhen-yu Yang
College of Information System and Management，
National University of Defense Technology，Changsha
Hunan 410073, P. R. China
zyyang@nudt.edu.cn

Abstract – *The capability based planning has been gradually accepted by top planners in military, but it lacks quantitative and operational methods to implement the development and planning of weapon system of systems. The evolving capability requirements oriented portfolio planning problem of weapon system of systems is introduced from the perspective of operational research. A quantitative description and its evolving types of capability requirements are proposed. A portfolio planning model of weapon system of systems is constructed aiming at the description; an intelligent optimization algorithm is designed to solve the portfolio planning model based on differential evolution. A portfolio planning example of intelligence, surveillance, reconnaissance assets is analyzed to illustrate the application of the model and algorithm. The research results can support the decision making on the planning and development of weapon system of systems.*

Keywords: Weapon system of systems, portfolio planning, capability requirements, differential evolution.

1 Introduction

The planning and programming of weapon systems in top layer involves more different type weapon systems which should be given overall considerations. On the other hand, the weapon systems alternatives from diverse force pursue overlapping or similar functions and that is becoming more serious. These two respects make it harder for top planners to select and plan various weapon systems as a system of systems, namely Weapon System of Systems (WSoS) [1]. Therefore, it is urgent to propose scientific approach which can be applied in decision making on the planning of WSoS under the restriction of budget, risk, time and so on. Furthermore, the WSoS planning must be constantly updated and optimized to satisfy various and uncertain missions in future with minimal costs.

Aiming at the aforementioned requirements, The U.S. Department of Defense proposes the capabilities based planning (CBP) to instruct the planning and programming of weapon systems. They think that the military threats are uncertain and unimaginative in future cause of various missions, but the evolvement of capability requirements (CR) is relatively stable and easily descriptive, and they expect that the evolving capability requirements can reflect the variation of missions and guide the planning of WSoS. But current published literatures on capabilities based planning almost are qualitative and procedural research results, and lacks quantitative approach based on operational research. For some examples, Yi provides some clarity for capability, networked enabled capability and so on in the through life capability management [2]. Krieg introduces the concept of applying CBP to three levels of decisions, especially in top-level decisions to help manage risk and allow senior officials to balance investments across capability areas [3]. Todor analyzes security, scenario and mission based on national interests, proposed two stage processes for developing capability requirements [4]. Elena et al explore the implications of a capability-based conceptual approach on the development of the systems engineering (SE) disciplines [5]. Ben defines the essential elements for CBP, and the generic processes required to linking these elements [6].

In this paper, we transform aforementioned realistic requirements to a problem of weapon system of systems portfolio planning (WSoSPP) from the angle of operational research, in order to implement CBP for supporting the decision making on planning the type, quantity and development time of weapon systems. The rest of the paper is organized as follows: section 2 proposes two kinds of key problems on WSoSPP through the integration of qualitative and quantitative methods. Section 3 describes the capability requirements quantitatively and analyses its evolving types. Section 4 presents WSoSPP model aiming at first kind of key problems. Section 5 designs solving algorithm combined with Differential Evolution (DE). Section 6 validates the model and algorithm with a case study. Finally, conclusions are drawn in Section 7.

2 Problem Description

From the perspective of operational research, the problem of WSoSPP can be descripted as the brief form: given the constraints of budget and risk, at the time t=0, K kinds of weapon systems is selected from **M** kinds of weapon systems in order to maximize satisfying **N** kinds of CR; at the time t>0($t \in \mathbf{N}^+$), the decision variables of K kinds of weapon systems are adjusted with minimal cost in order to adapt the evolvement of the CR. Considering the reality, two kinds of key problems are hard to be solved in the evolution of WSoS, CR and missions, as shown in Figure1.

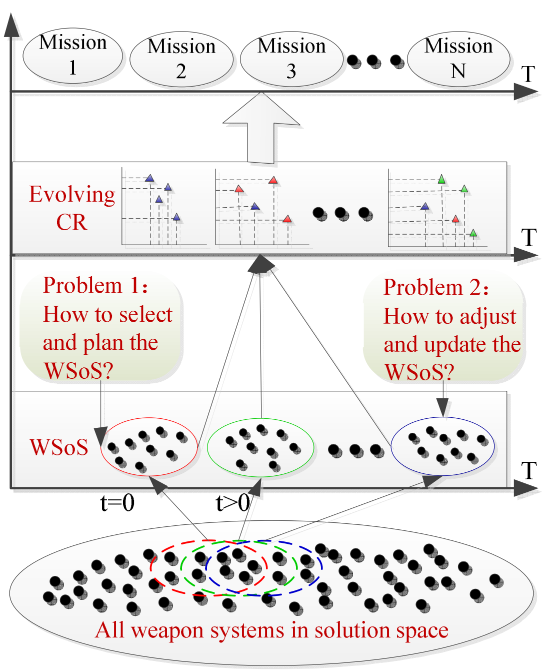


Figure 1. The evolution of WSoS-CR-Mission

Problem 1: How to select and plan the WSoS? It is complex to construct and solve the model of WSoSPP. Each kind of CR usually needs more than one supporting weapon systems, each kind of weapon systems almost supports more than one kind of CR. The type, quantity and time of development on each kind of weapon systems directly affect the satisfactory degrees to CR, meanwhile, the approach which can realize comparing all portfolio planning alternatives must be proposed. Furthermore, considering the scale of solution space and all constraints, the modeling and solution of WSoSPP themselves are complex.

Problem 2: How to adjust and update the WSoS? It is complex resulted by the evolvement of CR. In the development of WSoS, the kind, numerical value and required time of CR evolve with the variation of missions. To satisfying the evolving CR, it is necessary to consider the robustness at initial stage and adaptability in process of the development of WSoS. Besides, the ways and trends of evolvement of CR can be forecasted and described with a certain probability, but cannot be completely specified. Therefore, the abovementioned factors to be considered bring more difficulties on modeling and solving the problem of WSoSPP.

3 Analysis of Evolving CR

To solve the problem of WSoSPP, the quantitative descriptions of CR and its evolving types must be analyzed to support the modeling of WSoSPP. Therefore, the following hypotheses are put forward: at $t=t_0$, there are **N** kinds of CR, each kind has two requirements, namely numerical value a_i and required time t_k. The quantitative description of CR is shown in Figure 2. For the sake of analysis, every kind of CR was converted to the uniform form as follows: let each kind of CR be C_i, which should be reached the numerical value a_i at the time t_k, $0< a_i<1$, $t_k >t_0$.

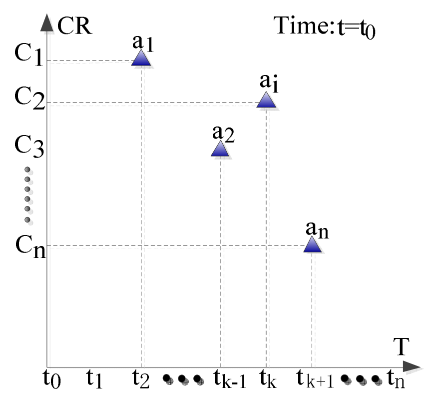


Figure 2. The quantitative description of CR

In essence, the evolvement of CR means that CR dynamically adapts itself at one or more discrete decision point in future following the variation of missions. It can be concluded as two points: firstly, the variant of CR is discrete at a finite number of time points, but not continuous; secondly, at time $t=t_0$, the evolving types of CR in future could be deduced and forecasted, but not confirmed completely cause of the uncertainty of missions. At time $t>t_0$, the kind, numerical value and required time of CR are of independent and combinatorial evolvement. The detailed types of evolvement are shown in Figure 3. There are four independent types of evolvement: ①Increase of the kind of CR, ②Decrease of the kind of CR,③Variation of the numerical value in any kind of CR, ④Variation of the required time in any kind of CR. The remaining types of transformation are any combination of aforementioned four independent ones.

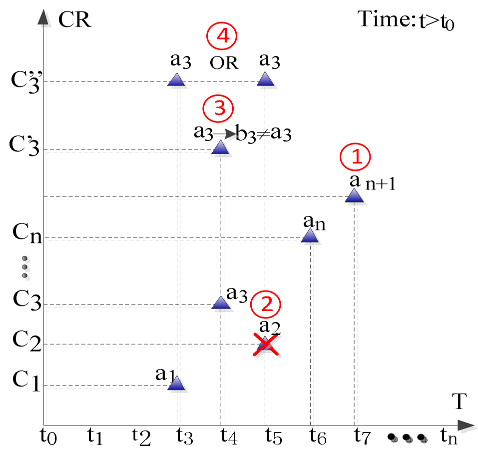


Figure 3. The evolving types of CR

4 Formulation of WSoSPP

The CR is constructed as an N dimensional ideal point based on Technical for Order Preference by Similarity to Ideal Solution (TOPSIS) [7], the weighted Euclidean distance between the degrees of one WSoSPP alternative satisfying CR and the ideal point dividing the cost the alternative needed is taken as objective function value. Therefore, the formulation of WSoSPP can be expressed as follows at the upper right in this page.

Decision variables are v_j, x_j, y_j, v_j is a binary variable, if weapon system j is selected, v_j is taken value 1, 0 otherwise; x_j and y_j are the quantity and time of development of each kind of weapon systems in sequence respecctievely. S is the Synthetic Satisfactory Degree of Capability Requirements (SSDoCR) by one alternative of WSoSPP, C is the cost for implementing this alternative; R_i is one Single Kind of Capability Requirements (SKCR) ,namely CR i; I_{ij} is the degree that weapon system j satisfies CR i; Cap is the constraint of SSDoCR, F is the constraint of cost; if weapon systems j is selected to supports CR i, the binary variable u_{ij} is taken value 1, 0 otherwise; K_i is the risk that one WSoSPP alternative satisfies CR i, it is relative with the development time of each type of weapon systems; H_i is the constraint of risk of satisfying CR i; P_i is the constraint of Satisfactory Degree of Single Kind of Capability Requirements (SDSKCR), for avoiding generating redundant or poor abilities.

Maximaize: $Z = S(u_{ij}, x_j, y_j) \big/ C(v_j, x_j, y_j)$

subuject to:

$$S(u_{ij}, x_j, y_j) = \left(\sum_{i=1}^{n} \left(\sum_{j=1}^{m} u_{ij} I_{ij} - R_i \right)^2 \right)^{1/2},$$

$$I_{ij} = e^{w_1 c_{1ij} (a_{1ij} x_j^{k_{1ij}} + b_{1ij} x_j^{k_{2ij}}) + w_2 c_{2ij} (a_{2ij} y_j^{k_{3ij}} + b_{2ij} y_j^{k_{4ij}}) - 1},$$

$$C(v_j, x_j, y_j) = \sum_{t=1}^{m} v_j (h_{1j} x_j + h_{1j} t_j e^{-h_{2j}(x_j - 1)} + h_{3j} h_{1j} y_j),$$

$$S(u_{ij}, x_j, y_j) \geq Cap, \quad C(v_j, x_j, y_j) \leq F,$$

$$w_1 + w_2 = 1, 0 < c_{1ij}, c_{2ij} < 1,$$

$$k_{1ij}, k_{2ij} > 1, k_{3ij}, k_{4ij} < 1, 0 < R_i < 1,$$

$$0 < h_{2j}, h_{3j} < 1, x_j \leq t_j, y_j \leq s_j, x_j, y_j, t_j, s_j \in N^+,$$

$$K_i = \frac{\sum_{j=1}^{m} u_{ij} (d_{1ij} e^{-d_{2ij}(x_j - 1)} + d_{1ij}(1 - d_{3ij} x_j))}{\sum_{j=1}^{m} u_{ij}}, K_i \leq H_i,$$

$$\left| \sum_{j=1}^{m} u_{ij} I_{ij} - R_i \right| \leq P_i, P_i < 1, 0 < d_{1ij}, d_{2ij}, d_{3ij} < 1$$

5 DE based Solving Algorithms

The problem of solving the formulation of WSoSPP belongs to nonlinear integer programming; its simplest form corresponds to the classical NP-Hard problem-knapsack problem, which can be solved effectively by intelligent optimization algorithms [8]. Differential evolution algorithm is a kind of new evolutionary computation technology [9], its simple mutation operation and one-on-one competition as a survival strategy, reduces the complexity of genetic operations with strong global convergence and robustness. Therefore, the solving algorithm was designed based on DE. Specifically as follows:

Step 1-Producing initial population: Multiple-dimension real-numbered vector is adopted as the coding method to denote the development time and quantity of weapon systems, the development time bit taken value 0 represents that this weapon system is not selected, as shown in follows:

$$\overbrace{x_1, x_2, \ldots x_i, \ldots, x_m}^{\text{the bits of development time}} \qquad \overbrace{y_1, y_2, \ldots y_i, \ldots, y_m}^{\text{the bits of developement quantity}}$$

Step 2-Caculating the value of objective function: First of all, the value of function C and S are programmed and calculated, and then the values of objective function of initial population are calculated, the value of global optimization is recorded last.

Step 3-DE based mutation operations: There are three DE operators to be adopted in solving process as follows [9, 10, 11]:

$$r' = r_3 + rand \times (r_2 - r_1) \qquad (1)$$

$$r' = \lambda r_3 + (1-\lambda) r_{gbest} + rand \times (r_2 - r_1), \quad \lambda = (T-t)/T \qquad (2)$$

$$r' = r_t + (r_{gbest} - r_t) + rand \times (r_2 - r_1) \qquad (3)$$

r_1, r_2 and r_3 are the randomly chosen individuals; r' is the temporary individual from mutation; r_{gbest} is the individual of global optimization, T is the overall iteration times, and t is the present iteration times; r_t is the present population.

Step 4-DE based crossover operation: Each bit of previous generation is selected with certain probability to replace the corresponded bit of present generation.

Step 5-Operating infringing constraints: The bits of development time and quantity are randomly generated to replace the ones which infringe the constraints in each individual for satisfying the constraints.

Step 6- Surviving strategy based on Competition: If the previous individual and present individual both satisfy the constraints, the individual with maximum value of objective function is selected as new individual; if one of previous and present individual infringes the constraints, the individual with not infringing the constraints is selected; if the previous individual and present individual both infringe the constraints, the individual with minimum infringe the constraints is selected.

Step 7-Repetitions iteration: The value of global optimization is updated. Steps 3 to 6 are repeated until the maximum number of iterations or the desired solution is obtained.

6 Case Study

The scenario is that the intelligence, surveillance, reconnaissance weapon system of systems (ISRWSoS) must be planned and developed for satisfying the capability requirements and missions in future. Each kind of CR and the kind, type, development time and quantity of weapon systems which can be selected are shown in Figure 4.

The arrow denotes the supporting relationship between the kind of weapon systems and the kind of CR. Ty, Ti and Q denote in sequence the type, time, and quantity which can be selected to develop ISRWSoS respectively, the 'year' is taken as an unit for development time. RT and RNV denote in sequence the time and numerical value of CR respectively. Airborne Warning And Control System, Unmanned Aerial Systems, Phased Array Warning System, Command Control, Information Transition and Distribution in turn are abbreviated as AWACS, UAS, PAWS, C2, ITD. The scale of solution space is more than 5^{27} without considering constraints due to the development time and quantity of all weapon systems selected.

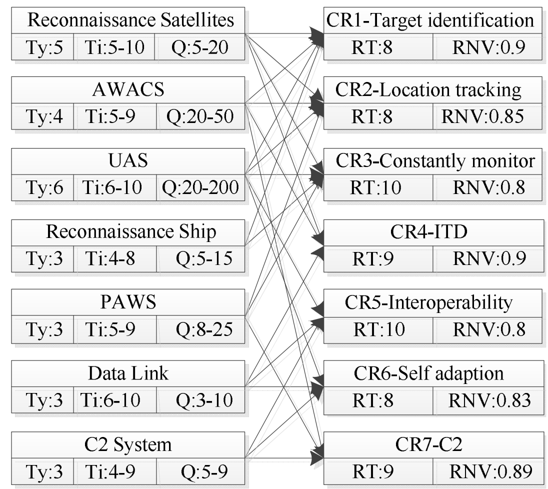


Figure 4. The description of ISR planning problem

Assuming that all kinds of parameters are known, the WSoSPP model of ISR is constructed based on the formulation in section 4.The three mutation operators are adopted in turn to simulate and solve the model. The diverse optimized curves based on the three mutation operator are shown in Figure 5 with parameter set at Cap=0.95, F=20 (unit: billion dollars), H_i=0.3, and the limiting value of SSDoCR is 1.

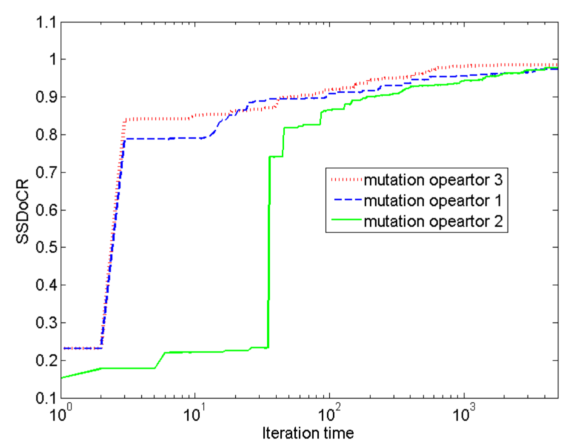


Figure 5. Three mutation operators based optimized curves

As shown in figure 5, operator 1 has quick constringency speed which is equal to operator 3 in early iteration, but its capability of local research is inferior to operator 3 in later iteration; operator 2 has the advantages of local research which is equal to operator 3 in later iteration, but obviously in early iteration, its global constringency speed is slower than operator 3. Therefore, the mutation operator 3 based solving algorithms is selected to executed 30 times repeatedly, as a result, the final planning solution of ISRWSoS is shown in table 1.

Table 1 The final solution of ISR WSoSPP

Kind	Ty	Ti	Q
Reconnaissance satellite	Type—1	8	9
	Type—2	8	11
AWACS	Type—3	9	21
UAS	Type—2	10	22
	Type—4	9	177
	Type—5	10	138
PAWS	Type—2	7	20
Data Link	Type—3	10	10
C2 System	Type—1	10	5

The planning solution in table 1 not only acquires the maximum synthetic degree of 0.9848 on satisfying the CR, but also never generates redundant and poor abilities, because of using TOPSIS in the model of WSoSPP, as shown in Figure 6.

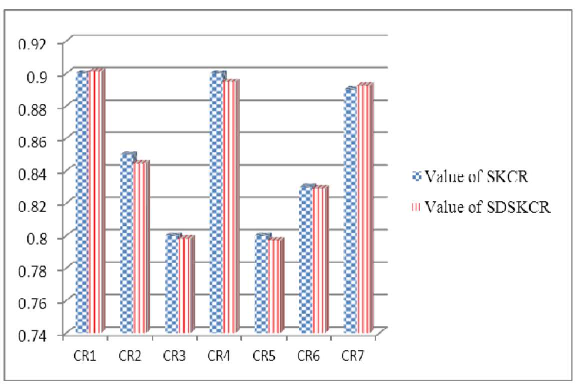


Figure 6. Contrast between SKCR and SDSKCR

Each kind of CR has distinct value, for considering each kind matching with each other. All ISR weapon systems be selected in solution can only establish the operational capabilities of system of systems, if each kind of CR is configured at specific value.

7 Conclusions

One important aspect of WSoS research is supporting top planners to plan and program the kind, type, time and quantity of development in the constraints of defense budget, risk, and so on. Meanwhile, Capabilities based planning is an indispensable part of WSoS research [12], and the evolvement of capability requirements and system of systems exist objectively. Therefore, one of the research thoughts is planning WSoS based on combinatorial optimization and portfolio planning to satisfy CR and complete various missions. The other one of research thoughts is planning WSoS based on uncertain programming and multi-stage planning to strengthen the robustness and adaptability of WSoS.

In this paper, the preliminary research of evolving capability requirements oriented WSoSPP is carried out based on aforementioned thoughts. The capability requirements and its evolving patterns are proposed, the model and algorithm of WSoSPP are constructed and designed, which can support decision making on planning and programming weapon systems in long term. The future research is how to apply the methods of uncertain programming and multi-stage planning to WSoSPP for strengthening the robustness at initial stage and adaptability in process of the development of WSoS.

References

[1] Ben Cheng, Xu Tan, Yue-jin Tan, “A Novel Bi-level Programming Model for Capabilities-based Weapon System of Systems Planning”, Computational Intelligence and Design (ISCID), 2011 Fourth International Symposium, Hang zhou, China, pp. 266-269, October 2011.

[2] Yi Yue, “An Holistic View of UK Military Capability Development”, [EB/OL], [2011-11-13], https://dspace.lboro.ac.uk/dspace-jspui/handle/2134/4392

[3] James Bexfield, Lisa Disbrow, Corrina Ross-Witkowski, et al. “Capabilities Based Planning: The Road Ahead” [R] Arlington, Virginia, USA: MORS Workshop， 2005.

[4] Todor TAGAREV, “Capabilities-based planning for security sector transformation,” *Information & Security*, Vol. 24, pp. 27-35. 2009.

[5] Elena Irina Neaga, “The Influence of the Concept of Capability-based Management on the Development of the Systems Engineering Discipline,” 7th Annual Conference on Systems Engineering Research, Loughborough, Britain, April, 2009.

[6] Ben Taylor: “Guide to Capability-Based Planning” [EB/OL].[2009-1-8] http://www.rto.nato.int/abstracts.asp

[7] Hwang C L, Yoon K S. *Multiple Attribute Decision Making*, Springer Verlag, Berlin: 1981.

[8] Wang Ding-wei, Wang Jun-wei, Wang Hong-feng, et al, *Intelligent Optimization Methods*, Senior Education Press, Beijing, 2007.

[9] Rainer Storn, Kenneth Price, "Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces," *Journal of Global Optimization*, Vol. 11, pp. 341-359, 1997.

[10] K. Price, R. Storn, and J. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization*, Springer Verlag, Berlin, 2005.

[11] Wu Liang-hong, Wang Yao-nan, Zhou Shao-wu, et al, "Differential Evolution for Nonlinear Constrained Optimization Using Non-stationary Multi-stage Assignment Penalty Function," *Systems Engineering-Theory＆Practice*, Vol.30, NO.1, pp.167-172, 2010

[12] Defense Acquisition University: Defense Acquisition Guidebook [EB/OL].[2011-9-22]. https://dag.dau.mil

Using Indicators for System Complex Safety

Tullio J. Tanzi
CNRS - LTCI UMR 5141
Centre de Recherche sur les Risques et les Crises,
Mines-ParisTech
Sophia Antipolis - France
tullio.tanzi@mines-paristech.fr

Raoul Textoris
L'Oréal Clichy
Centre de Recherche sur les Risques et les Crises,
Mines-ParisTech
Sophia Antipolis - France
raoul.textoris@loreal.com

Abstract - *In our modern societies, technological systems are taking on a large part in numerous domains such as automatic control, calculation, communication, information technologies, etc. They are put in place in more and more fields e.g. production, defense, national security, space, etc. These very important developments are offering new possibilities such as distributed cooperative and concurrent decision making based on complex dynamic systems or on advanced simulation capacities. To facilitate decision making in various fields such as transport, energy or even risk management, it is necessary to define indicators generated by such systems in order to deliver engineers or managers an image of the considered object and it's evolution. This image must be coherent, reliable and sustainable in order to participate at the decision in a complex sociotechnical environment.*

The aim of the article is to present our approach to define this category of new indicators.

Keywords: Safety assessment, indicators, system complex, risk management.

1 Indicators

Modern systems are usually and inherently very complex because of the simultaneous integration of different technics and technologies. Using such systems for safety offers new perspectives but at the same time clearly raise the question of the relevance of the information that feeds the decision making process.

With these critical conditions, information that is erroneous (totally or partly), obsolete, inadequate, etc. is leading quickly to wrong decision. The cause of the deviation can be due to the way the system is working itself (design level) but also due to the fact that it does not report relevant information in all cases.

This can be due to errors of functioning, but also due to the right way of working itself (from system perspective) but unexpected (from the user standpoint). To avoid this kind of deviation you can define indicators that clearly give the right picture of the way the system is working, but also give a relevant picture of the target of the system and its evolution.

An indicator is a measure that enables to assess the efficiency of the considered object in order to predict the information that is susceptible to have an impact on its performance goal. An indicator can be an individual measure or a set of measures and its associated analysis that can predict performance before the goal is fully achieved. The performance of the considered object can be an indicator to measure its performance in its environment (system, process,…).

Indicators allow a good control based on a good understanding on performance itself and its evolution. Predictability of the future is not always taken into account by measurement process. Without the right indicators, it is difficult to assess probability to conduct up to the end a complex activity meeting constraints such as frame, calendar, quality and budget, etc.

A classical measure (a conventional measure) gives information on historical and real data. An indicator must rely on trend on conventional measures or demonstrated correlations that can give a predictable analysis. An indicator could rely on the evolution of a list of constraints to predict the future behavior of a process.

Although we use same data, a fundamental difference is that indicators (compared to conventional measure) have an objective to meet information needs that can be either predictive or prospective. Even if indicators seem to be similar to existing measures and use same basic data, the difference lies on the way this information is collected, assessed, interpreted and used to give information and knowledge on the future.

Indicators are supposed to be used to enlarge the set of all the existing measures that are already in place into the organizations. To optimize efficiency, indicators must be put in place via a measurement structure of the organization (generally based on

CMMI[1] principles), that enables to automatize the way to gather, analyze and interpret data.

We also have to note that indicators often mean we have to use empirical data to define scheduled objectives and the thresholds that are used to analyze and interpret. When these data are not available, we often use an expert judgment in order to define initial goals and thresholds until a good historical can be gathered.

A qualitative or quantitative indicator belongs to one of the 3 types below:

- Lagging indicator
- Leading indicator
- Coïncident indicator

The concept of the 3 types of indicators (lagging, leading, coïncident) is a long story. It started in 1938 with a book written by Mitchell, Wesley C., and Arthur F. Burns. Statistical Indicators of Cyclical [1]. Since that date, numerous articles demonstrated the interest of such indicators. The evolution of the last years make it easier to use them considering alert needs in order to go from a reactive control mode (reaction to a lagging indicator) to a proactive control mode (action on leading indicator).

Some of these indicators give information on past performance. They are called indicators of result (lagging indicator). As an example, for a ship, the logbook registers the distances. Another kind of indicator gives information on the ongoing performance that can have an effect on future performance. They are called indicators of action or advanced indicators or piloting indicators or alert indicator (leading indicator). This is the case for the anemometer or the radar that allow the pilot to be alerted of a potential hazard, to anticipate any phenomenon with the right picture. Coïncident indicator put in evidence events almost happening at the same moment.

2 Limits of the conventional approach

We are going to illustrate the limits of the conventional approach by taking an example using indicators for car traffic control.

To control the traffic on a one way composed by three high speed lanes, we put in place a device that can count each vehicle and for each gives its speed. The speed is regulated. Maximum authorized speed is 110 km per hour and minimum authorized speed is 70 kilometers per hour.

With this information, we put in place indicators with the objective to facilitate decision making for personnel in charge of car traffic control on both fluidity aspect and safety concern.

The number of vehicles driving on one lane gives its yield of use, and so it's possible saturation. The speed gives a picture of safety, for example, by comparing the individual speed of each vehicle with the capacity of the infrastructure. Combined with flow[2] it gives more precision on the yield of use of the lane.

Very quickly we can assess the limits of the use of individual data. First step we have to put in place aggregation function then merging function of individual data in order to build synthesis.

In term of flow, a first aggregation of individual data over a time frame of one or six minutes, for example, makes sense. The calculation remains simple: We sum individual measurement on a period of six minutes.

The yield of use of the road (composed by three lanes) will be calculated by using the sum of the flows of each lane compared to the sum of their respective global capacity. If we want to have an idea of an average flow over a period of six minutes, a simple arithmetic average can be used.

The aggregation of individual speed, without introducing any bias, raises a difficult problem. Let us consider individual data.

1. We have a table of values T(v)[90, 50, 160] that represent measured speeds. The arithmetic average corresponds to a speed of 100 km/h for a maximum authorized speed limit of 110, however 60 % of vehicles are outside the specifications and can be considered as dangerous[3]. The possible aggregations on a given period of time will introduce big bias.

By using a more complex aggregation function, it is possible to deliver a result on the safety level of the road depending on speed measures.

Let us define a weighting table Ct(v,i) that will be initialized taking into account individual speeds such as :

$$\textit{if } T(v,i) \in [70,110]\, \textit{then } Ct(i) = 1, \text{ifnot } Ct(i) = 2$$

It gives for our example weighting values [1,2,2]. The speed coefficient is calculated, for example, with the following algorithm:

$$Coeff = \frac{Echantillon}{\sum_{i=1}^{echantillon} \left(\frac{T(v,i).Ct(i)}{100} \right)} \qquad (1)$$

In our example, we calculates *Coeff = 3/5.1 = 0.58.* This value can be round up to 0.6. This clearly demonstrates that 60 % of speeds are out of specification².

[1] CMMI : Capability Maturity Model Integration

[2] Number of vehicles on prefixed period of time.

[3] Mobile zigzag for one and high speed for the other one.

3 Toward new indicators

New indicators must include a management dimension, which corresponds to classical indicators, but equally a piloting dimension that is still to be defined. Difference lies on the fact that information to pilot is directly linked to how to drive action, while management information is dedicated to information structure of the company. In order to facilitate how to use them, so how to interpret them, they can be organized in synthetic scorecard. Research from Kaplan and Norton on the notion of "balanced scorecard" [2] [3] is a key contribution to our field of study.

More precisely : "a performance indicator that can help a manger, at an individual or more often at a team level, to pilot the action up to the objective or that can allow to assess the result …" [4].So it is not an "absolute" measure, a characteristics of the measured phenomenon independently from the observer. It is built by the actor. [5].

As a consequence it is a sophisticated management tool with some specific features. For example:

- The strategic objective to which it is linked, its targets with timeframe and measurable features , the relevant references,
- The clear identification of who is in charge to deliver them, and the one in charge of its performance,
- Frequency and follow up. Son mode de suivi : budgété, réel, historique, …
- Technical definition : formula and calculation convention, sources of information, …
- Segmentation modes to decompose aggregated form : geographical data, type of product, center …
- Presentation (ex : numerical data, tables, graphics, …) and communication list.

Such an indicator is composed by two different functions depending on how it is located compared to the action (se figure 1).

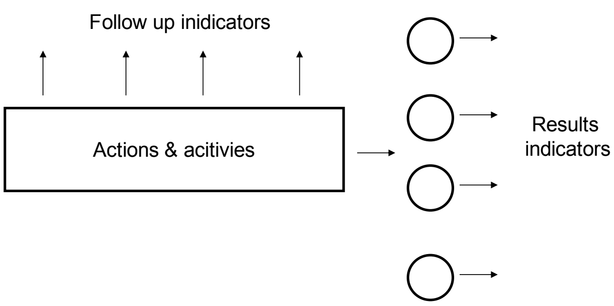


Figure 1: indicator of result or follow up [4]

It can be an indicator of result. In that case, it gives an assessment of the final result when the action is completed.

But it can also be a follow up indicator. It allows to anticipate or to react on time. By definition, the result indicator comes too late to shift the action.

The way it is located compared to the structure of power and responsibility gives it also a final duality (fig 2). The corresponding reporting gives an indication of the percentage realization of the objectives, which can be considered as a control a posteriori, and the piloting whose objective is to adapt actions in progress.

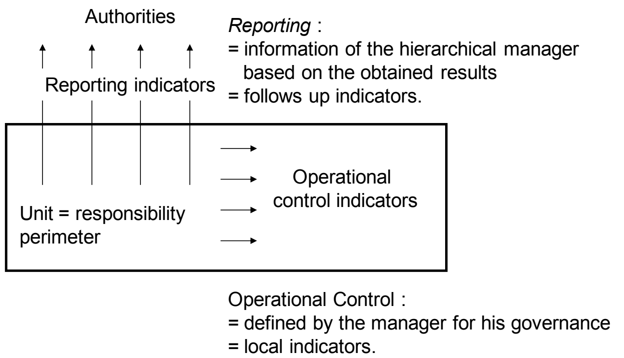


Figure 2: Leading indicator or reporting [4]

The composition of such an indicator must take into account aspects linked to operational relevance such as, for example, combining indicator / action, the question of "controllability", and the impact of levers on actions.

It is also necessary to take into account some aspects linked to strategic relevance such as, for example, the association of indicator / objective [6], [7], the measure of the completion of the results (Indicator of result), and data on how actions are implemented (Leading Indicator).

This reflexion must be completed with another dimension concerning the cognitive efficiency. Indeed, these indicators are used by the actors in a given context. They influence the action and the way it is understood. It is so necessary to define how to read them, to understand and to interpret as soon as the indicators are designed. It is the only condition to set a frame to take into account the context of the actor, and that is easy to use.

Some questions are rising when we want to define indicators. Do we want to use financial indicators, or non financial, or use a mix of both? If we define non financial indicators, is it better to valorize the stakes? What is the right number of indicators to get a clear and coherent picture?

It appears necessary to dissociate management indicators and piloting indicators [8].

The way indicators are organized within a scorecard, for example as for balanced scorecard, makes it possible to have both types of indicators financial and non financial. Indicators are organized in four parts: learning, process, customers and financial aspect. Inside the scorecard, indicators are linked with a causal model.

4 Proposal to define a methodology to build indicators

To build new indicators, on the scheme above, it is necessary to put in place lagging indicators, of a classical type, and leading indicators, that are still to be defined. An important problem is to take into account, at initial design level, cultural differences that can exist through the same organization (different jobs and state of the art, level of education, social origins,…) or in the various locations in different countries of an international company.

A specific focus must be put on the models which, although they are at the same time part of the indicator, are necessary to determine it (calculation process) and how to interpret it. The mathematical approach and the following modeling phase have a direct impact on the relevance of the indicator (pertinence, reliability, « easy to use”,…). Handrails such as, for example, the definition of functioning segments, restrictions, …, allows to put in place real time control process. Steps of aggregation and consolidation of data allow ensuring that the dynamics of information within the organization (geographical or organizational grouped together, reporting to the upper level, …) do not introduce bias, so do not destroy its coherence.

While defining indicators and the way they are organized, it is necessary to define a test and validation protocol that will be based on a set of selected data. The protocol must represent the way the organization is working, and the set of data has to work properly for the main expectable cases of use.

The approach exposed in our article is only a first step that needs to be developed. Application domains are numerous especially for safety people management and associated indicators such as frequency rate and severity rate [9]. The objective is to determine a methodology to define and put in place indicators able to control on one hand the way the system is working and on the other hand to give a coherent picture of the observed phenomenon and its evolution.

References

[1] Wesley C. Mitchell, Arthur F. Burns. Statistical Indicators of Cyclical. Geoffrey H. Moore, ed. 1938

[2] Kaplan Robert, Norton David. The Balanced Scorecard – Measures that Drive Performances, in Harvard Business Review. Jan-Feb 1992.

[3] Kaplan Robert, Norton David. The Balanced Scorecard. Harvard Business School Press, Boston, USA. 1996.

[4] Lorino Philippe. Méthode et pratiques de la performance. Le pilotage par les processus et les compétences. Editions d’Organisation. Paris, 2003.

[5] Lorino Philippe. Comptes et récits de la performance. Editions d’Organisation. Paris, 1995.

[6] Kerr Steven. On the folly of rewarding A, While hoping for B. Academy of Management Journal. 1975, 769-783.

[7] [Epstein 1998] Epstein Marc, Manzoni Jean-François. Implementing Corporate Strategy / From Tableaux de Bord to Balanced Scorecards. European Management Journal, vol. 16, avril 1998, 190-203.

[8] [Hopkins 2007] Andrew Hopkins .Thinking about Process Safety Indicators. 2007.

[9] F. Juglaret, JM. Rallo, R Textoris, F. Guarnieri, E. Garbolino. The Contribution of Balanced Scorecards to the Management of Occupational Health and Safety. European Safety and Reliability Conference: Advances in Safety, Reliability and Risk Management, ESREL 2011, Troyes: France.

[10] Abideen Tetlay. Through Life Capability Management Perspective for Framework Development for Assessing and Measuring System Maturity, System Readiness and Capability Readiness using Architecture Frameworks. 5th IEEE International Conference on System of Systems Engineering (SoSE 2010). June 22nd – 24th 2010, Loughborough University (UK)

[11] Harding, A., Mollett, J. and Touchin, M. (2009), “A structured approach to planning and managing systems engineering capability evolution in a complex through-life business space”, in: 7th Annual Conference on Systems Engineering Research 2009 (CSER 2009), 20th – 23rd April 2009, Loughborough University (UK)

Proc. of the 2012 7th International Conference on System of Systems Engineering, Genoa, Italy - 16-19 July 2012

Advancing the Defense in Depth Model

Stephen Groat
IT Security Office and Lab
Bradley Department of Electrical and Computer Engineering
Virginia Tech
Blacksburg, Virginia 24061
Email: sgroat@vt.edu

Joseph Tront
Bradley Department of Electrical and Computer Engineering
Virginia Tech
Blacksburg, Virginia 24061
Email: jgtront@vt.edu

Randy Marchany
IT Security Office and Lab
Virginia Tech
Blacksburg, Virginia 24061
Email: marchany@vt.edu

***Abstract*—Systems and network defenses currently implementing a Defense in Depth (DiD) strategy frequently slow attackers' progress but do not act as a secure barrier. These systems of network defense methods are primarily comprised of static defenses focused on preventing attacks from entering a network by enabling the features of blocking access, requiring authentication, or analyzing traffic. To adapt to the ever-changing threat profile of network attacks, the DiD model must be adapted to be symmetric and focus on new vectors for defense instead of authenticating, blocking, or analyzing all traffic. Instead of a focusing on feature-centric network defense requirements, the DiD model should be redesigned to be a functional or capability focused model. Symmetry in the DiD model allows for the network defense system to recognize the insider threat, preventing data exfiltration and allowing attacks to be stopped at the originating network instead of being defended by the attacked network. Dynamic defenses must also be enabled, which change attack surfaces to proactively defend a network. New vectors, such as dynamic network addressing, enterprise computing resources, and network architectures, must be used by the DiD model to prevent attacks from reaching network, consuming attackers often limited resources, and securing networks in their design and architecture.**

***Index Terms*—Information security, Defense in Depth, Dynamic defense, Symmetric defense**

I. Introduction

Computers and networks are constantly under threat from attackers attempting to steal data or disrupt access. To combat the cyber security threat, different system and network defenses have been created. Firewalls are a common network and host protection system used to block network access for certain programs. Intrusion Detection Systems (IDSs) and Intrusion Prevention Systems (IPSs) are used to detect and stop advanced threats, including application layer exploits. Other network defense systems have tailored applications, such as web application firewalls that analyze traffic for specific web application exploits or virtual private network (VPN) technologies that offer secure access and authentication to restricted networks. Since most networks are composed of many different types of systems, security systems must be layered to offer holistic network protection as a system of systems.

The DiD model uses layers of different network protection devices to create a secure network [1]. By having different pieces of the security system address disparate network security vulnerabilities and distribute the analysis load between multiple systems, a network can be secured with minimal latency and maximum availability. Yet, the network security tools deployed to protect a network often are plagued by the same problems as the hosts being protected. Exploits due to programming errors are not as common in security tools as in common applications, but still occur. Security systems can also be misconfigured, unknowingly allowing an attacker access to sensitive systems or information. Also, threats coming from within a trusted system or network are often difficult or impossible to block, leaving security professionals vulnerable to the insider threat. By layering the security systems, the DiD model looks to address the potential vulnerabilities by offering multiple levels of protection.

Though the DiD model is still a important architecture for network defense, new network and system defense concepts must be implemented to secure systems instead of only offering minimal protection by slowing attackers down. By refocusing the DiD model on capabilities important for network defense instead of features, network defense can be advanced and improved. Implementing the concept of symmetry into the DiD model allows for each network to provide inbound and outbound security, preventing unknowingly compromised systems from being used as attack relays. Also, symmetry in the defense model allows for nefarious activity inside a network to be detected, securing a network from the insider threat. Dynamic defenses must also be incorporated into the DiD model to improve security, reliability, and confidentiality. By constantly changing the attack surface through network dynamic defenses, attackers are forced to expend the majority of their time in the planning stage of their attack and are never able to execute. New dynamic resource tools, such as cloud computing, can also be used to absorb attacks, preventing standard Denial-of-Service (DoS) attacks from being effective.

To secure networks, the DiD model must be viewed as a system of systems and updated with current network defense strategies. First, some background about the DiD model and how it impacts information security is presented in Section II. Section III addresses some previous work relating to advancing information security architecture. Some examples of implementation of the new design are discussed in Section IV.

Section V analyzes the impact of symmetry and dynamics on the state of network defense followed by some future directions for this work in Section VI. Concluding remarks are provided in Section VII.

II. Background

DiD is an Information Assurance (IA) strategy developed by the National Security Agency (NSA) that involves multiple layers of defenses for networked electronic and systems security. Historically, a military defender would build a series of defensive positions and fall back as the attacker advanced, eventually defeating the attacker. This same defensive strategy was also applied to computer systems. Different types of network security tools, including firewalls and IPS/IDS systems, are layered to integrate different network security strategies on a single network. To assure system and network security, all the different variables that could have an effect on system security, including physical security, policy and procedure, and electronic security, are addressed in the model.

In the electronic security model, the security of networked systems has evolved to a well-accepted model for Internet Protocol version 4 (IPv4). To help to secure systems in an enterprise, systems which must be globally accessible put in a demilitarized zone (DMZ). Workstations that do not require global accessibility are often placed in a private network. The lack of addresses in the IPv4 has led to the use of Network Address Translation (NAT), allowing multiple systems to use a single address. Yet, NAT creates significant network and configuration issues with emerging technologies. Technologies such as Voice over IP (VoIP) require public addresses to work seamlessly and are broken by NAT. The large number of address available in the Internet Protocol version 6 (IPv6) fixes the issues of limited address in IPv4. In IPv6, all systems, including workstations, are expected to have public accessibility. To adapt to this dramatic change in network architecture, the DiD model must be updated to address changes in security.

Networks of all types are constantly under attack, including attackers ranging from individuals to organized crime and nation states. In 2012, Trustwave approximated a 90% year over year growth in attacks on commercial targets [2]. Even highly restricted government networks, such as the United States Air Force (USAF) Predator and Reaper unnamed aerial vehicle (UAV) or drone console networks, have been infected by malware looking to record keystrokes [3]. Those these attackers were ineffective due to their inability to exfiltrate any of the collected data, the infection shows the potential to attack any network, including protected, restricted networks.

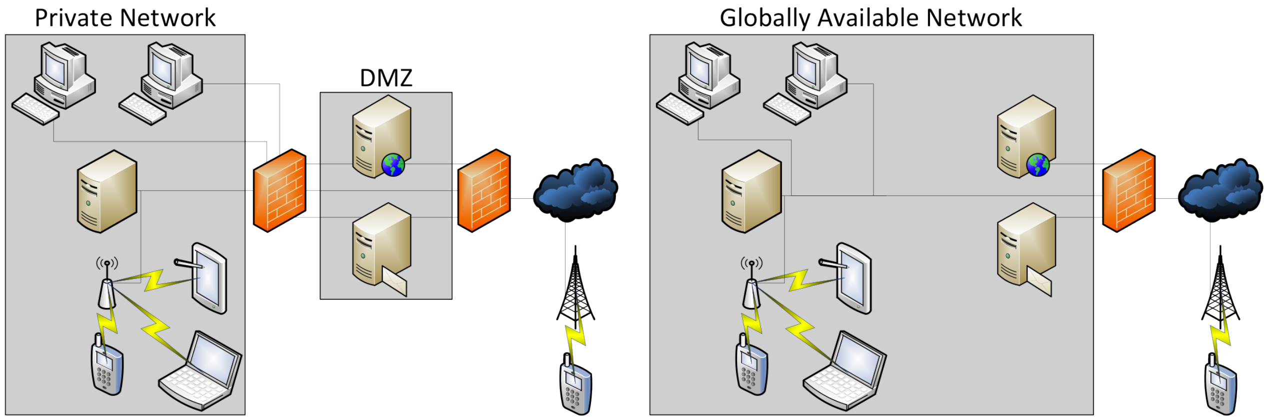
While computer and network security models have not changed in many years, the feature-focused DiD model has become ineffective against the current threat profile that exists for systems, both networked for global available and in restricted networks. For openly networked systems, the known features, protection, and weaknesses of network security products allow for attackers to test defensive systems and find entry points before attempting attack on a network, often making the first attempt successful and producing a minimal amount of anomalous traffic. The static addresses used in these globally available networks networks to allow for connectivity, but also offer attackers a vector to easily find and exploit targets. Though restricted, non-networked systems are more secure due to their limited availability, security threats to these systems still exist in the form of unknown, zero-day security holes that can be exploited and are often harder to detect due to the assumed security of the systems and the new, unknown attack vectors. For restricted networks, the insider threat is still a significant security risk that can only be addressed by analyzing data within the restricted network. A new DiD model must incorporate defenses for both open and closed networks, increasing security for both types of systems.

In addition to using the DiD model to stop attackers from entering a network, a DiD model should include capabilities to prevent attackers from targeting networks. In the current DiD model, this is primarily achieved by addressing the operational security (OPSEC) of the organization. Currently, approximately 95% of attackers' overall time is spent in the planning phase [4]. Much of this time is often spent in research about the target and social engineering additional information about the target. By adding security through a dynamic defense, the DiD model can render research less valuable and force attackers to continue to reacquire targets, holding them in the planning phase and preventing them from executing their attack.

III. Related Work

Other work has been done to validate the DiD model. Bass and Robichaux [5] verified the DiD model against large and complex network operations. By analyzing how the DiD model scales to large networks, the work verifies that feature-based DiD concepts can scale to large network operations. This work differs by analyzing how DiD features can be adapted to DiD functions, while assuring the model continues to scale. Applying the DiD model to new systems, such as supervisory control and data acquisition (SCADA), Lippmann et al. [6] used the model to help identify attacks and restore functionality to targeted control systems. SCADA is an important, self-contained network to analyze how DiD principles applyed in a restricted network environment. Stylz [7] attempted to apply the network-centric DiD model to software engineering and applications. Since the mistakes made in software engineering are often the threats that network security systems must defend, attempts to apply the DiD model earlier in the system development process has the potential to avoid dangerous security issues. This work focuses on looking on deploying a DiD model that focuses on new, functional capabilities to networks where threats already exist.

Specifically, some work has been done analyzing the possible functional advancements of symmetry and dynamic defenses. Keweley and Bouchard [8] examined the effectiveness of dynamic defenses in live network simulations. By showing the security offered by dynamic defenses in live network exercises, the study shows how essential kinetic defenses are successful network security. Yet, this type of network security



(a) Current Network Architecture in IPv4

(b) Future Network Architecture in IPv6

is not enumerated by the DiD model. Dunlop et al. examined the feasibility of using a network and transport layer moving target defense, focusing their efforts in IPv6 [9]. These types of dynamic and moving target defenses are essential to enabling a functional DiD model. Touch et al. [10] looked to apply a dynamic defense to prevent DoS and distributed denial of service (DDoS) attacks for private networks. In the current network environment focusing on global connectivity using the large address space of IPv6, this system is outdated, but shows the potential for dynamic defenses. Other significant work has been done to create systems to block DoS and DDoS attacks. Keromytis has proposed multiple systems [11]–[13] to alleviate the threat of DoS attacks, but the systems added too much complexity to the overall network defense architecture to be practical. Symmetric defenses have been used in radio frequency (RF) and wireless applications to ensure signal security. Mingyan et al. [14] used symmetric defenses to identify and mitigate jamming attack in sensor networks. This work looks to apply the same principles of symmetric defenses used in the RF space to system and network defenses.

IV. Design of a Functional DiD Model

In designing a new, functional DiD model, many of the current feature-based defenses are still used and deployed, but new capabilities are added to increase the security of the system. Since the DiD concept is focused on the layering of defenses to achieve security, the current security systems and features still must be deployed as seen in Figure 1(a). Systems such as firewalls, IDSs and IPSs are still used, but layered with new devices that provide different new capabilities to the network defense system. New capabilities must also be deployed to prevent these static defenses from being continuously probed and to detect attack internally without creating excessive complexity within the system. Also, the new paradigm of global availability in networks offered by IPv6, as shown in Figure 1(b), must also be account for. Though layering more feature-based network security systems on each other may lead to slight increases in the security of the network, the additional complexity leads to a higher probability of service disruption and misconfiguration. Instead, adding new capabilities allows for networks to use the feature-based defenses already deployed more effectively. Since the majority of network cannot take offensive measures against attackers to stop penetration attempts, the additional capabilities deployed in the functional DiD model must be defensive as well. An example of a network deploying a security system designed from a functional DiD model is shown in Figure 1.

For globally available networks, a dynamic defense facing the outward edge, as shown on the right in Figure 1, is the most essential piece to effective functional DiD protected systems. The dynamic defense enables the current feature-based defenses, such as firewalls, IDSs and IPS, and malware analysis systems, to be harder to detect, test, and penetrate. Currently, these feature-based security tools are easy to penetrate due to the fact that they are static and easy to target. Once an attacker finds a network defense protecting the targeted network or system, the attacker has an unlimited amount of time and attempts to test the security of the system, eventually finding a hole. By placing a dynamic defense in front of these static, feature-based defenses, the existing defenses become more difficult to find and test and, therefore, provide more protection. As the attack surface continuously changes, an attacker is forced to use large amounts of time and resources to reacquire the target. Even if an attacker can find the target, they have a limited amount of time to test the feature-based defenses behind the dynamic defense before they are forced to reacquire the target.

Symmetric defenses, shown on the right in Figure 1, are also essential to increase the complexity of the defenses as they appear to an attacker while maintaining the DiD layering principle. Though firewalls, IDSs, and IPSs are ineffective network security systems when deployed by themselves, layering them provides additional protection. Also, if these systems are allow to analyze and act on internal traffic as well as external traffic, new insight into the network can be gained to increase security. Data exfiltration, the act of moving data out of a contained area, such as a corporate network, can be detected as an anomalous large outbound transfer of data. Employing symmetry is network defenses also helps to detect the insider threat, offering protection for restricted networks. Blocking attacks before the leave the external edge of the network helps to protect other networks as well, increasing the overall security of hosts using the Internet.

Inside the network, dynamic, scalable computing resources are also an important piece of a functional dynamic network defense (Figure 1). These resources may be physically located within the network and dynamically repurposed to perform day-to-day tasks or consume network attacks. Enterprise computing resource can also be outsource to large, cloud computing providers to provide on-demand, salable computing resources without the overhead of physical system hosting. The new ability for networks of any size to take advantage of large computing resource platform has enabled a new form of dynamic defense.

V. Analysis of a Functional DiD Model

To advance the current DiD model, a functional DiD model should be focused on the capabilities of symmetry and dynamic defenses. Current network defenses are designed around the features of specific network defense tools, such as identifying malware, blocking packets, or analyzing network events. These defenses are effective against specific attacks, but cannot holistically defend networks. Implementing symmetric defenses allows for protection against insider threats, both through data exfiltration and network attacks originating within the host network. By deploying proactive, dynamic defenses, functional network defenses are further enabled by limiting the scope of possible attackers the the attackers' amount of time to penetrate before reacquiring the target.

A. Symmetric Defenses

Instead of attempting to prevent inbound attacks and blocking specific forms of outbound traffic, a functional DiD model should look to deploy defenses that are symmetric. By deploying the same defenses on the internal network as on the external edge, the network can secure itself and other networks. By analyzing internal traffic through the use of symmetric defenses, a functional DiD model can detect and stop data exfiltration. Also, by detecting and stopping attacks at the originating network, symmetric defenses contribute to the global network security.

Symmetry allows for defenses to stop data exfiltration as well as network infiltration, securing the network against the insider threat. Many networks currently analyze traffic, both inbound and outbound, to detect anomalous activity and stop any detected threats. Yet, usually the threats detects are either inbound network attacks or outbound callbacks to bot controllers. One of the most significant threats to networks is not inbound attack, but the transfer of sensitive data out of the network. Symmetric defenses would detect the anomalous outbound transfer of a large amount of data as unauthorized data exfiltration and allow for that data transfer to be stopped, protecting the network and the sensitive data.

By stopping outbound attacks at the originating network instead of the destination or target network, symmetric defenses also help to protect other networks and reduce the overall network attack traffic. Current network defenses are focused on stopping inbound attacks while very few networks look to stop outbound attacks. For example, the majority of network firewalls look to stop the common Transmission Control Protocol (TCP) [15] SYN flood attack. Yet, these same network firewalls currently do not have the capability to detect and block this attack outbound from their network. Though stopping outbound attacks must be done carefully to avoid impacting legitimate traffic and could been seen as encroaching on users' privacy, the potential to protect networks outweighs the risks. Many of the current network defense systems currently have the computing power and resources to block these attacks, but are not implemented properly to block these attacks. Symmetry in defenses can help protect other networks by stopping outbound attacks before they leave their host network.

In many cases, already implemented network defenses blocking inbound attacks could easily be made bidirectional to block outbound attacks as well. In implementing bidirectional defenses on existing functional network defenses, normal, non-offending network traffic must not be impacted. To ensure comparability, the feature-based network defenses must have the capability, both in design and processing power, to handle the additional traffic and analysis.

B. Dynamic Defenses

Dynamic defenses can be enabled both through dynamic computing platforms and dynamic network addressing. New advances in cloud computing allow for users to rapidly scale provisioned computing resources to consume DoS and DDoS attacks. Many of these platforms also allow users to pay for the resources on demand, decreasing continuing costs. By changing network addresses, dynamic network defenses secure network by forcing attackers to continually reacquire targets. Dynamic defenses scale and adapt to threats, offering the ability to consume or avoid potential attacks.

The decrease in the cost of computing power allows for many DoS and DDoS style attacks to be successfully consumed and absorbed at a lower cost than blocking without any impact on service. As cloud computing becomes widely deployed, users looking to ensure connectivity and service are deploying resources to the cloud to scale with demand. Some networks have locally hosted virtualized resources that can be quickly reprovisioned from consuming to attacks to providing necessary businesses services. Since the resources are not necessary locally hosted, networks can also offload their cloud computing to large, public clouds to consume computing resources on-demand without having to pay overhead costs associated with the equipment. As attackers launch DoS and DDoS against users' resources, the public cloud resources can scale to meet the demand of an attack on an ad-hoc basis, only charging the network for the resources consumed. By enabling computing resource to consume non-invasive

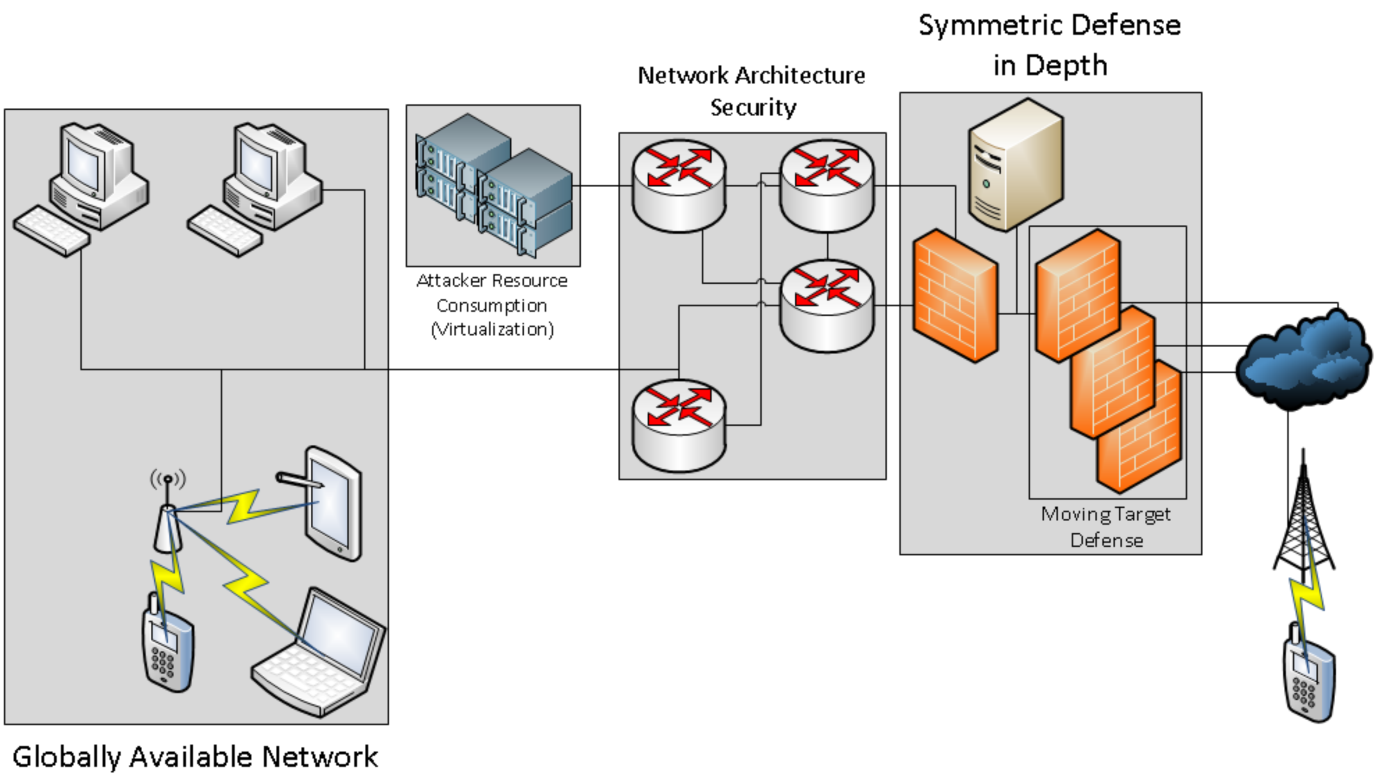


Fig. 1. Example of Functional DiD model implementation

attacks, defensive resources and systems can be focused on protecting valuable assets, such as information in databases, instead of stopping computing resources from being consumed. Currently, the vast majority of attacks at around 90%, are focused on obtaining and exfiltrating sensitive data instead of DoS. Therefore, focusing valuable network defense systems on sensitive data is more prudent.

Dynamic network addressing is another form of dynamic network defenses that proactively prevents attackers from exploiting network systems. Current network addressing systems use static addresses that allow users to easily identify systems. These static addresses, though, allow for attackers to easily target and exploit systems. By using dynamic addressing and frequently changing network addresses, networks can avoid attacks by constantly changing the attack surface. This defense is also effective since it creates a dynamic element at a very low layer of the Open Systems Interconnection (OSI) model. By securing the network and transport layer, dynamic addressing provides inherent security for the higher layers, including the application layer. Since the attacker must continually reacquire the target as the dynamic defense continues to change, the attackers must use more resources and time to find the system. By increasing the cost of network attacks, network administrators and network security professionals can avoid the need to analyze an excessive amount of basic, weak attacks launched by individuals and focuses defenses and efforts of dangerous threats perpetuated by well resourced attackers, possibly nation states. Conversely, static network systems are essentially free for attackers to find. Also, by forcing attackers to remain in the planning and target acquisition phase of an attack, targets can avoid exploitation.

While advancing the DiD model through dynamic defenses will increase security and confusion for attackers, the auditability and usability of the network must be maintained. Dynamic defenses will provide additional security, but logs and records connecting users to network identities must be maintained to provide historical insight for audits. Without this insight, the insider threat will become more effective due to a network administrator's inability to log and track user actions. Also, in creating dynamic and symmetric defenses for networks, usability must be maintained. If too many security systems are put in place that jeopardize the business processes necessary for the organization to succeed, the enterprise will fail. Maintaining the ability to effectively audit and use a network is essential to adopting a new, more complex DiD model.

VI. Future Work

The functional additions of symmetric and dynamic network defenses add additional overhead to existing security systems. To understand the complete impact of a functional DiD model, the overhead of symmetric and dynamic defenses must be calculated. For example, deploying symmetric defenses must cover double the bandwidth of traffic than normal, increasing the load on the system versus systems that only are required to analyze inbound traffic. Dynamic defenses, which are proactive in their network protection, are particularly resource intensive. By analyzing the impact of the functional DiD model, the impact of the new capabilities can be evaluated to ensure the system are cost effective.

To determine if the additional overhead of a functional DiD model is beneficial versus a feature-centric DiD model, the protection offered by the functional DiD model must also be measured. Creating a framework or metric to evaluate capabilities within the DiD model would allow for system administrations to perform a cost-benefit analysis before a system is implemented. Also, capabilities in the new model could be analyzed for their specific return on investment, identifying the most important and cost-effective capabilities for network security. By creating a framework for evaluating DiD models, network security capabilities could begin to be evaluated.

As dynamic defenses add stealth to the currently available services of detection, in the forms of antivirus and IDSs and IPSs, new systems and capabilities must be created to add the ability to mitigate, survive, and recover from attacks. Dynamic defenses add stealth capabilities to the current DiD model. Yet, new inventions are needed to expand the DiD model to provide defensive capabilities similar to those available in physical attacks. Jammers and chaff are used to mitigate attacks in air attacks; honeypots could be deployed on networks to provide similar services.

VII. Conclusion

Current feature-based DiD models are ineffective in blocking network attacks. When attackers can target an entity's network defenses and have an unlimited amount of time to penetrate with little investment of resources, they maintain an unfair advantage. To increase network security and the cost for attackers, a functional DiD model that implements symmetry and dynamic network elements increases cost and complexity for attackers. Threats have to penetrate not only the target's security systems, but also any transit participating network. The dynamic network elements require attackers to reacquire targets, increasing the cost of an attack and decreasing the amount of time an attacker has to exploit threats.

Acknowledgment

The authors would like to thank Science Applications International Corporation (SAIC) for supporting this research.

References

[1] N. S. A. (NSA), "Defense in depth: A practical strategy for achieving information assurance in todays highly networked environments," Available at http://www.nsa.gov/ia/_files/support/defenseindepth.pdf.

[2] T. . S. Labs, "2012 global security report," https://www.trustwave.com/downloads/Trustwave_WP_Global_Security_Report_2012.pdf, 2012.

[3] N. Shactman, "Exclusive: Computer virus hits u.s. drone fleet," http://www.wired.com/dangerroom/2011/10/virus-hits-drone-fleet/, 2011.

[4] D. Levin, "Lessons learned in using live red teams in IA experiments," *DARPA Information Survivability Conference and Exposition*, vol. 1, p. 110, 2003.

[5] T. Bass and R. Robichaux, "Defense-in-depth revisited: qualitative risk analysis methodology for complex network-centric operations," in *Military Communications Conference, 2001. MILCOM 2001. Communications for Network-Centric Operations: Creating the Information Force. IEEE*, vol. 1, 2001, pp. 64 – 70 vol.1.

[6] R. Lippmann, K. Ingols, C. Scott, K. Piwowarski, K. Kratkiewicz, M. Artz, and R. Cunningham, "Validating and restoring defense in depth using attack graphs," in *Military Communications Conference, 2006. MILCOM 2006. IEEE*, oct. 2006, pp. 1 –10.

[7] M. Stytz, "Considering defense in depth for software applications," *Security Privacy, IEEE*, vol. 2, no. 1, pp. 72 – 75, jan.-feb. 2004.

[8] D. Kewley and J. Bouchard, "Darpa information assurance program dynamic defense experiment summary," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 31, no. 4, pp. 331 –336, jul 2001.

[9] M. Dunlop, S. Groat, W. Urbanski, R. Marchany, and J. Tront, "Mt6d: A moving target ipv6 defense," in *MILITARY COMMUNICATIONS CONFERENCE, 2011 - MILCOM 2011*, nov. 2011, pp. 1321 –1326.

[10] J. Touch, G. Finn, Y.-S. Wang, and L. Eggert, "Dynabone: dynamic defense using multi-layer internet overlays," in *DARPA Information Survivability Conference and Exposition, 2003. Proceedings*, vol. 2, april 2003, pp. 271 – 276 vol.2.

[11] A. Keromytis, V. Misra, and D. Rubenstein, "Sos: an architecture for mitigating ddos attacks," *Selected Areas in Communications, IEEE Journal on*, vol. 22, no. 1, pp. 176 – 188, jan. 2004.

[12] D. Cook, W. Morein, A. Keromytis, V. Misra, and D. Rubenstein, "Websos: protecting web servers from ddos attacks," in *Networks, 2003. ICON2003. The 11th IEEE International Conference on*, sept.-1 oct. 2003, pp. 461 – 466.

[13] N. Nagashima and A. Keromytis, "A repeater encryption unit for 1pv4 and 1pv6," in *Networks, 2005. Jointly held with the 2005 IEEE 7th Malaysia International Conference on Communication., 2005 13th IEEE International Conference on*, vol. 1, nov. 2005, p. 6 pp.

[14] M. Li, I. Koutsopoulos, and R. Poovendran, "Optimal jamming attacks and network defense policies in wireless sensor networks," in *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*, may 2007, pp. 1307 –1315.

[15] M. Duke, R. Braden, W. Eddy, and E. Blanton, "A Roadmap for Transmission Control Protocol (TCP) Specification Documents," RFC 4614 (Informational), Internet Engineering Task Force, Sep. 2006. [Online]. Available: http://www.ietf.org/rfc/rfc4614.txt

Threshold Design for Low Cost Wave Sensors through Statistical Analysis of Data

Maricris C. Marimon
Graduate School of Information Science
Nara Institute of Science and Technology
Ikoma City, Nara, Japan
maricris-m@is.naist.jp

Kenji Sugimoto
Graduate School of Information Science
Nara Institute of Science and Technology
Ikoma City, Nara, Japan
kenji@is.naist.jp

Abstract - *Wave monitoring is an essential activity in building offshore infrastructures and in producing weather forecasts. In this study, a technique for wave sensors used in wave monitoring is introduced. This technique utilizes the statistical parameters – kurtosis and skewness as threshold value generators. The values generated by this technique serve as baseline for normal wave conditions in the area of interest. Sensor deployments on five different areas are done to test this technique. This technique aims to boost the sensor's potential in determining the severity of the ocean waves. Also, this can be used in wave alarm systems that inform coastal communities of impending disasters.*

Keywords: Wave monitoring, sensor, remote sensing, accelerometers, threshold.

1 Introduction

Wave monitoring is an important activity done by researchers and organizations who are interested in the studying the dynamics of the ocean waves [1]. The data collected from this activity is significant in generating prediction models for ocean waves. Through these models, designs for marine structures and alarm systems can be fitted according to the model parameters. Also, these will increase the robustness of these structures and the accuracy of the alarm system.

There are several existing wave monitoring systems that utilize different methods of measuring and assessing ocean wave conditions ranging from point buoys to wide range satellites. Using these sophisticated measuring tools such as satellites and radars provides a wider coverage however it also comes with a high maintenance cost. For applications that just monitor a smaller area, using high cost equipment seems impractical, thus deploying point buoys are more manageable especially for independent researchers. For this study, point buoys with sensors are deployed to gather data.

Wave monitoring processes are complex since dealing with an unpredictable phenomenon is tricky and difficult [2]. It has been known that ocean waves are unpredictable due to its behavior [3]. In order to properly investigate the random data, statistical analysis is applied. Most monitoring systems utilize statistics in determining the dynamics of the parameters that they are monitoring. Wave monitoring systems calculate significant wave height to determine the present wave conditions [3]. For this study, statistical analysis is utilized however calculation of significant wave height is not performed. Instead, a proposed utilization of statistical parameters – kurtosis and skewness as indicators for ocean wave condition severity is presented.

These statistical parameters will compose the threshold design for effective sensors of a wave alarm systems. Further explanation about these two statistical parameters is presented in 2.

To test this design, the sensors are deployed in five (5) locations with different coastal characteristics. Through the series of experiments, the threshold points will be determined and these points will be implemented to discriminate wave conditions. With this approach, the sensors are intended to be more sensitive to abnormal deviations of wave values.

It is important to create a good threshold design for the sensors because this will determine the accuracy of the system. Also, this enables the deployed sensor to be autonomous because it can utilize this design to preprocess the data before sending the information to the system.

2 Analysis of Ocean Wave Behavior

There are two ways of understanding waves. First, ocean waves are assumed as regular waves that follow linear wave motion. However, this assumption is insufficient. Regular waves cannot be attained in actual seas due to existence of multiple waves. Ocean waves are composites of multiple waves, which have different amplitudes and wavelengths (Figure 1) given by the equation:

$$\mathrm{A(t)} = A_0 + \sum_{i-1}^{n} A_n \cos(\omega t + \varphi) \qquad (1)$$

where: A_n is the amplitude of the waves
A_0 is the mean water depth
ω is the angular frequency
φ is the phase angle

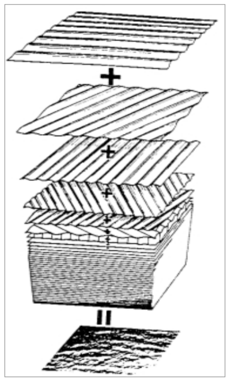


Figure 1. Superposition of multiple waves [2]

These waves are generated independently and randomly. There are several factors that influence the generation of these multiple sinusoidal waves. Wind mainly generates the waves. Within the wave generation area, waves have bigger amplitudes. As they propagate out from the generation area, the waves experience interferences from other waves, which are either constructive or destructive. These interferences cause the waves to be more nonlinear and chaotic.

Due to this nonlinearity, it is practical to introduce another method, the statistical method, to describe wave characteristics. Through the statistical parameters – kurtosis and skewness, the wave condition can be identified as Gaussian or non-Gaussian. Gaussian or Normal Distribution is a continuous probability distribution that has a bell-shaped probability density function. Normal wave conditions follow a particular statistical pattern however abnormal wave conditions deviate from this pattern. This can be observed in the statistical values.

Kurtosis and skewness are good indicators of non-Gaussianity of the distribution of wave heights. Note that assuming that a certain wave condition has a Gaussian distribution can be inaccurate since the probability of outliers is under-predicted [4].

Skewness is the measure of vertical symmetry of the wave field. Positive values indicate that the wave field is skewed above the average wave height.

$$\text{Skewness} = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^3}{(N-3)s^3} \qquad (2)$$

where $\bar{Y}$ is the mean, s is the standard deviation and N is the number of data points.

Kurtosis is a measure of whether the distribution for the wave field is peaked or flat and defines the contribution of large waves to the wave field. High kurtosis indicates that data heavily deviate from the average and are outlier-prone.

$$\text{Kurtosis} = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^4}{(N-1)s^4} \qquad (3)$$

where $\bar{Y}$ is the mean, s is the standard deviation and N is the number of data points.

The occurrence of the outliers signifies the severity of the wave conditions. There is also a possibility that within these outliers, freak waves may occur. These high amplitude waves are dangerous since they can have destructive impact on structures and marine vessels. The Gaussian distribution generates skewness and kurtosis values of 0 and 3, respectively.

The behavior of waves is dependent on the water depth. As water depth decreases, the nonlinear wave interactions become more prevalent. The circular path followed by water particles is deformed near coasts (Figure 2). This is observed as sharpening of wave crests and flattening of wave troughs which influence non-Gaussianity. For waves in shallow water [4], it is expected that skewness > 0 and kurtosis > 3.

To recall, the ratio between the water depth (h) and the wavelength (λ) in different depths is expressed in the following:

- Deep water: $h > \lambda/4$
- Transitional depth: $\lambda/25 < h < \lambda/4$
- Shallow water: $h < \lambda/25$

The most noticeable indication of these circular path deformations is the occurrence of white caps on shores. The waves near the coasts usually appear taller than their counterparts in deep waters. As the wave approaches the shore, its height will also increase. This is due to the conservation of energy. Since the group velocity decreases and the wavelength decreases, the energy in wavelength must increase in order to conserve the energy until the effects of friction becomes significant.

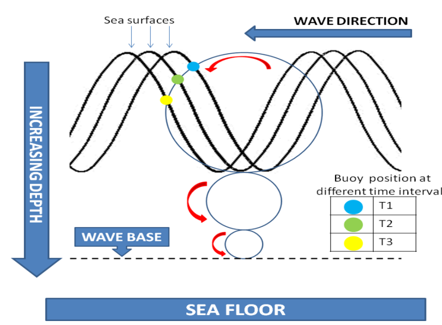
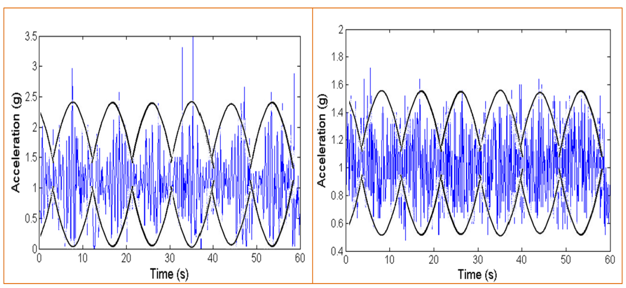


Figure 2. Vertical path or orbital movement of water particles at different time intervals

3 Evaluation of Threshold Design

The effectiveness of the sensor is dependent on its data processing ability and deployment location. It is then important to test the threshold design technique for the sensor and deploy the sensor in different locations to test the technique.

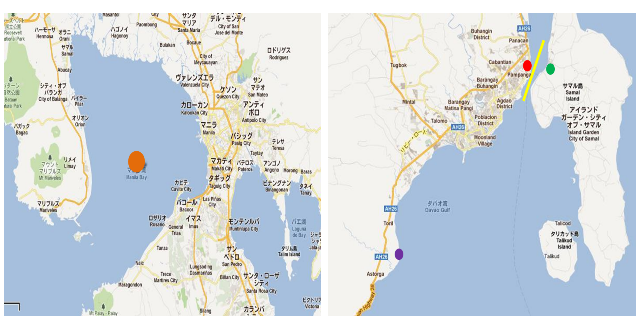
The sensor utilized for the experiments is a MEMS-based tri-axial accelerometer. This type of sensor is inexpensive and reliable which make it suitable for deployments. The accelerometer connects via bluetooth to its main board, which houses the data logger. It gathers data for 60 seconds at an interval of 0.1s. The calibration process is similar in previous works [5]. The sensor is calibrated in two ways. In the first set-up, the sensor is mounted on a double pendulum and in the second, the sensor is mounted on a buoy that is floating on seas. The results are shown in Figure 3. According to [5], the motion generated by the double pendulum is similar to the motion made by the buoy while it is floating on seas. It can be observed in the figure that the generated waves in both setups are similar to a modulated wave with distinctive envelope like appearance. From this, it can be said that the sensor has the potential in detecting ocean waves.



(a) (b)

Figure 3. Comparison of double pendulum generated waves (a) and ocean waves (b)

Determining the deployment location is essential in evaluating the effectiveness of the threshold design hence the chosen locations have different characteristics. The research locale of this research is the Philippines.



(a) (b)

Figure 4. Maps of the five (5) locations: [a] Manila Bay (orange); [b] Coast of Davao near IGACOS (red), Strait between Davao and IGACOS (yellow), Coast of IGACOS near Davao (green), and Coast of Davao far from IGACOS (violet)

Five different locations in the Philippines are considered namely, considered for the testing are Manila Bay Coast, Strait located between Davao and IGACOS, Davao Coast near IGACOS, IGACOS Coast near Davao and Davao Coast far from IGACOS.

Since this study aims to make a threshold design, the wave conditions must be relatively fair. Since wave conditions are dependent on weather conditions, the weather during the deployments must be relatively fair.

For each deployment location, only one accelerometer is utilized to keep the complexity low thus, keeping the power consumption of the device to a minimum [5].

Since the accelerometer utilized is tri-axial, it can measure the acceleration in the three axes. However, the orientation is an issue since the deployed sensors are mounted on a buoy and the buoy freely moves according to the ocean waves. To make it practical, it is best to take out the orientation. Figures 5 to 9 shows the graphs of acceleration data from the five different locations.

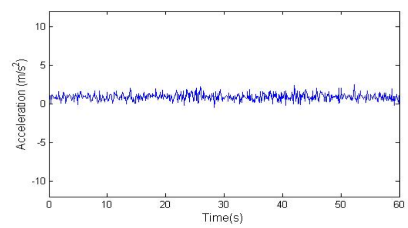


Figure 5. Acceleration data from the sensor deployed in Manila Bay

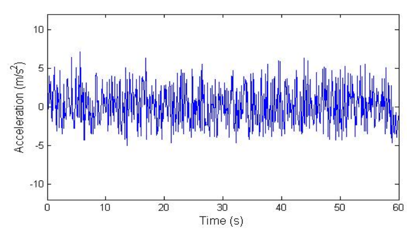


Figure 6. Acceleration data from the sensor deployed in the Coast of Davao near IGACOS

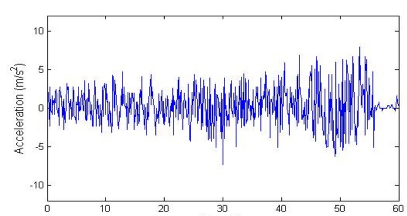


Figure 7. Acceleration data from the sensor deployed in the Strait between Davao and IGACOS

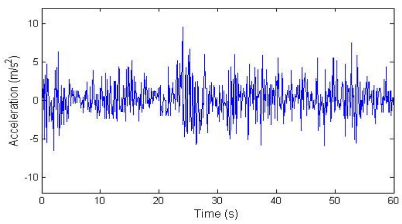


Figure 8. Acceleration Data from the Sensor Deployed in the Coast of IGACOS near Davao

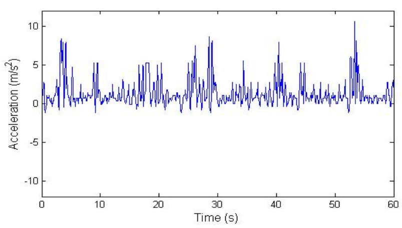


Figure 9. Acceleration data from the sensor deployed in the Coast of Davao far from IGACOS

The acceleration data is then converted to wave height similar to the process done in [5]. The calculated wave height (H) is then added with the Mean Water Level (MWL) value of each particular location to get posMWL (Refer to Table 1).

$$\text{posMWL} = \text{H} + \text{MWL} \tag{4}$$

Table 1. Name of the five locations and its corresponding Mean Water Level (MWL)

	LOCATION	Mean Water Level
A	Manila Bay	17m
B	Coast of Davao near IGACOS	10m
C	Strait between Davao and IGACOS	29m
D	Coast of IGACOS near Davao	10m
E	Coast of Davao far from IGACOS	10m

Histograms of wave height values from each location are shown in figures 10 to 14.

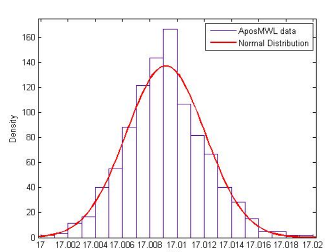


Figure 10. Histogram of wave height data from sensor deployed in Manila Bay (AposMWL)

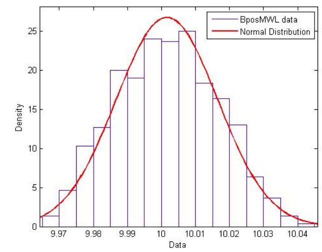


Figure 11. Histogram of wave height data from sensor deployed in the coast of Davao near IGACOS (BposMWL)

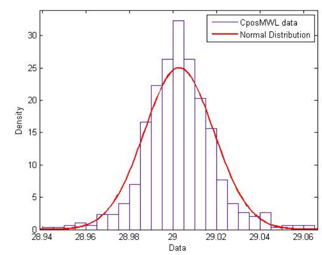


Figure 12. Histogram of wave height data from sensor deployed in the strait between Davao and IGACOS (CposMWL)

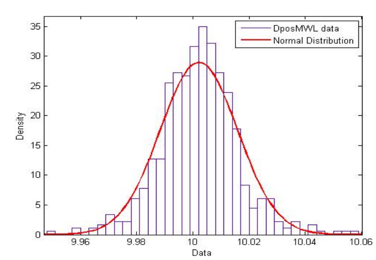


Figure 13. Histogram of wave height data from sensor deployed in the coast of IGACOS near Davao (DposMWL)

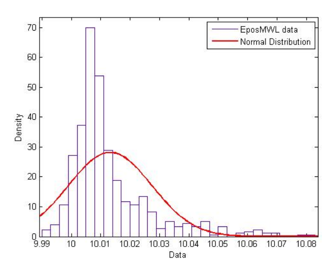


Figure 14. Histogram of wave height data from sensor deployed in the coast of Davao far from IGACOS (EposMWL)

Through these graphs, the fluctuations around the MWL can be observed. The normal distribution fit line is plotted over the histogram. The wider the fit curve, the higher will be the occurrences of outliers. Flatter peaks signify that the values fall near the mean water level. The statistics of the wave data gathered are generated and shown in Table 2.

Table 2. Statistics of the wave height data from five locations

DATA	KURTOSIS	SKEWNESS
1. AposMWL	3.2106	0.1889
2. BposMWL	2.0859	0.0357
3. CposMWL	3.1035	0.2073
4. DposMWL	3.7276	0.1704
5. EposMWL	8.5712	1.9068

Most of the locations show kurtosis > 3. As mentioned before, shallow waters generate kurtosis > 3 and skewness > 0. This is verified in the results of the experiments showing that locations 1, 3, 4 and 5 have kurtosis above 3. Note that these locations are within coastal waters. It is interesting to observe that Location 2 has a kurtosis lower than 3. This location happens to be in deeper waters. Wave heights in deeper waters appear to have a flatter distribution.

Location 5 has interesting characteristics since it has high kurtosis value and positively skewed. Note that this location is on a coast of an open sea. Open seas are prone to more random values since there are bigger chances for wind -generated waves from nearby storms to enter the area. This is in contrast to a sheltered bay (Location 1), which has coastal formations that limit the entry of these waves. Locations 3 and 4 are on opposite coasts of a strait. From the results, they generated similar values. This shows that waves have a tendency to reflect back and forth to these opposite coasts.

Results from these experiments present that using the statistical parameters – kurtosis and skewness, help determine the wave condition in the particular area. Using these values as baseline for the wave conditions in the area can be an efficient way in discriminating severe wave conditions.

To further test the capability of this threshold design, the generated statistical parameters are compared to three sets of data. The first set of data is simulated based from the concept that ocean waves are composites of multiple sinusoids (equation 1). The second set is simulated based from the concept that ocean waves are composites of nonlinear waves. These waves are usually modeled using Stokes' waves given by the equation:

$$A(t) = a\cos(kx - \omega t) + \frac{1}{2}k\,a^2\cos 2(kx - \omega t) + \frac{3}{8}\,k^2a^3\cos(kx - \omega t) \qquad (5)$$

where: a is the amplitude
k is the wavenumber
ω is the angular frequency
x is the displacement

The third set is an actual data, which is taken during a severe wave condition. The first two sets are simulated through Matlab with parameters set to simulate bad wave conditions. These conditions usually have frequencies within the range of 0.01 to 0.1Hz and amplitudes from 7m and above according to the Beaufort scale. Table 3 shows the statistical parameters generated from these sets and the their differences from the thresholds generated from the five locations.

Table 3. Differences of the statistical parameters generated from the new three sets of data and from the data of the five different locations

		KURTOSIS (K)	SKEWNESS (S)
STRONG GALE (LINEAR) K = 2.7708 S = 0.0183	A	-0.4398	-0.1706
	B	0.6849	-0.0174
	C	-0.2427	-0.189
	D	-0.9568	-0.1521
	E	-5.8004	-1.8885
STRONG GALE (NONLINEAR) K = 3.0638 S = 0.9146	A	-0.1468	0.7257
	B	0.9779	0.8789
	C	-0.0397	0.7073
	D	-0.6638	0.7442
	E	-0.5074	-0.9922
ACTUAL WAVE DATA (SEVERE) K = 4.7376 S = -0.86597	A	1.527	-1.05487
	B	2.6517	-0.90167
	C	1.6341	-1.07327
	D	1.01	-1.03637
	E	-3.8336	-2.77277

Based from table 3, it can be observed there are big differences in the kurtosis values on the third set of data. This signifies that if this severe wave condition occurs in the five locations, the sensors from the five locations will be able to identify that this is an abnormal wave condition. Note that the skewness value of the third set of data is negative and indicative that it is negatively skewed hence the differences in the skewness values are all negative. The first two data sets generated within the normal wave conditions hence the differences are not so large.

Through this test of the threshold technique, it can be said that this threshold technique is an improvement from previous works [6], which only utilizes counting mechanism to evaluate the severity of wave condition. This provides a more in depth approach since it refers to the baseline before it decides on the severity of the wave condition. The difference from the baseline measures how severe this particular wave condition is.

4 Application of threshold design in the system

After evaluating the mechanism of the threshold design, its application and integration in the system is important. Figure 16 and 17 describe the flow chart of the system where this threshold design is integrated within the system's sensor nodes.

Figure 16 shows the initial process wherein the sensor nodes gather data for the threshold values. Figure 17 shows the normal process to be done by the sensor nodes when they are in operation.

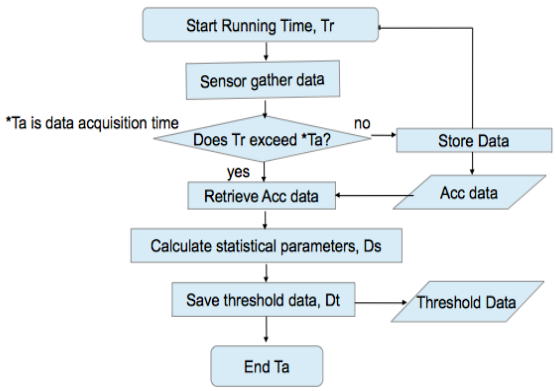


Figure 16. Initial process of sensor nodes to gather threshold data

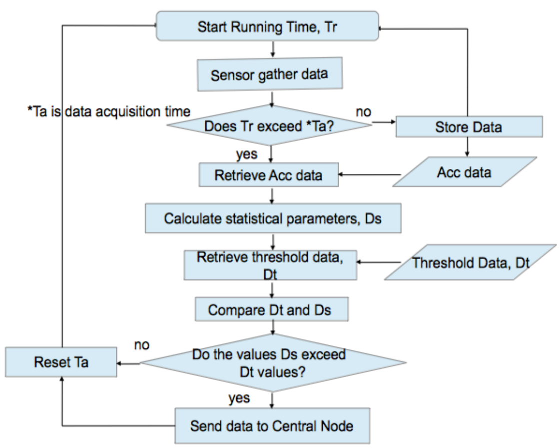


Figure 17. Process done by sensor nodes in normal system operation.

This threshold design offers autonomy of the sensor nodes from the system. It will process first the data within its node before sending data to the central node. This lessens power consumption in the transmission of data. This decreases the need to constantly connect to the central node.

Also, this threshold design can be applied to a bigger system with multiple subsystems where each subsystem is tasked to report if it has reached a specific level of threshold.

This design can be utilized to any applications that monitor phenomenon where specific conditions are significant.

5 Conclusions

In order to have a good wave sensor, it should have a good threshold design that makes it sensitive to abnormal deviations of wave conditions. This study introduced a threshold design that utilizes statistical parameters – kurtosis and skewness, to determine the severity of waves. To test this design's potential, the sensors are deployed on different locations and the data collected are given the statistical treatments. From the experiments, the sensors are able to distinguish the difference of these locations. Through the data gathered from the experiments, a set of threshold values are generated for each of the location. This is important because it makes the sensors more sensitive to a particular location. Note that a threshold for one location may not be similar to others. If a wave alarm system is to be constructed within the area, the wave sensor must consider these threshold values. If the sensor detects wave condition that has exceeding values, it can publish a warning that this current wave condition is abnormal. This threshold design makes the sensor smarter in analyzing wave conditions.

References

[1] M. Marin-Perianu et al., "Wave Monitoring with Wireless Sensor Networks," in *4th International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, Sydney, Australia, 2008.

[2] *Guide to Wave Analysis and Forecasting - Second Edition,* World Meteorological Organization, WMO-No.702. Geneva, Switzerland, 1998.

[3] R. Stewart, *Introduction to Physical Oceanography*, Department of Oceanography, Texas A & M University, Sep 2008.

[4] E Pelinovskiĭ, *Extreme Ocean waves,* Springer, London 2008.

[5] J. Kuperus, "Wave Monitoring using Wireless Sensor Nodes," M. Eng. Thesis, University of Melbourne, Melbourne, Australia, 2009.

[6] M. Marimon et al., "Low Cost Sensor System for Wave Monitoring", in *7th International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, Adelaide, Australia, Dec 2011.

Shutdown Reduction Methods for Compressors in Condensate Stabilization Units

Elaheh Esfandiari Jahromi
Instrument Engineering Department
South Pars Gas Complex
Assaluyeh, Boushehr, IRAN
elaheh.esfandiari@yahoo.com

Leili Esmaeilani
Instrumentation Department
South Pars Gas Complex
Assaluyeh, Boushehr, IRAN
l.esmaeilani@gmail.com

Salahedin Sadeghifard
Instrumentation Department
South Pars Gas Complex
Assaluyeh, Boushehr, IRAN
s.sadeghifard@yahoo.com

Alireza niknam
Instrumentation Department
South Pars Gas Complex
Assaluyeh, Boushehr, IRAN
niknam_ar@gmail.com

Abstract - *In this paper, the procedure of determination, analysis and rectification of repeatable trips causes of off-gas compressors located in condensate stabilization unit of South Pars Gas Complex 2nd refinery, that have instrumentation and control system sources, are presented. Besides decreasing of production rate and increasing acid gas flaring, each trip results serious and inevitable damages on environment, human health, equipments and their accessories in consequent. Significant reduction in number of trips has been achieved by finding proper solutions for all mentioned problems and applying needed modifications that could be listed as changes in temperature transmitter settings, changing the types of them, upgrading Human Machine Interface, addition of new signals to PLC software to be shown in monitoring system and DCS for enhanced trip analysis and also correction of some control system program parts such as determination of First out alarm. In addition the above cases; security enhancement in the monitoring system is much considerable.*

Keywords: Shutdown, DCS, PLC, monitoring System, temperature transmitter.

1 Introduction

According to the shutdown operation reports, increasing the number of trips on two similar off gas compressors in stabilized condensate unit was investigated in the second refinery of south pars gas complex. Sometimes the number of these trips was exceeded over 27 per year as long as it could damage the device seriously. It can be mentioned that these compressors are reciprocating type which are produced in Italy by Nuovo Pignone Company. The PLC controller is MODICON of Schneider family, quantum series and programming software is concept and also the monitoring system software for this unit is complicity. To make a solution for this problem, all the alarm lists in each shutdown were checked in both DCS system and HMI (compressor monitoring system), but in most cases, no special reason indicating the main trip signal was found. Therefore as a first step, the analyzes and possibility of alarm observation were checked in monitoring system logger and realized that the main problem is because of too much faults in PLC software programming which was delivered by vendor and compressor manufacturer in commissioning time. Then, wrong parts of logics were modified .More details are described in the second chapter of this article.

An important note in controlling independent packages is sending an alarm list to DCS. This list must contain all important data such as trip signals and pre trip alarms. In the third chapter of this article all modifications on these signals and new alarm list to DCS will be described. Receiving first out alarm in DCS is also considerable to be a trip factor which was not sent correctly to DCS because of some mistakes in PLC logic. Rest of this chapter is dedicated to these modifications.

Also according to the trends that were logged in HMI, it was found that one of the main reasons derivates from a failure in temperature transmitters which are installed on the compressor. To rectify this problem, all transmitter settings were checked and modified that are completely explained in chapter 4.In the rest of this chapter, it will be discussed about the wrong wiring from transmitters to analogue input modules of Modicon PLC. This wrong wiring caused a non recognition fault at the time of connection failure to PLC module. The result on HMI gets a random value.

At last in the fifth chapter, optimization and upgrading of monitoring system and related software will be explained.

1.1 The possibility to observe and analyze all alarms in monitoring system

An Investigating the alarm lists during all compressor trips, it was cleared that "MCC FAULT" alarm has been the first signal trip as a common factor. This alarm is a digital input signal to PLC from electrical equipment which is a feedback signal of the main motor stop. This signal would be appeared on alarm list because of different reasons that can be ; loop problem from motor to PLC or motor stop itself. In these two cases "MCC FAULT" acts a first trip signal but in most cases this alarm is as a trip consequent and different shutdown reasons cause main motor stop, then this alarm would be logged as a secondary factor. [1,2]. This loop and main motor signals were exactly checked by the assistant of electrical personnel; but no fault was found. It can be realized that "MCC FAULT" can't be a first trip signal while it is a trip consequence signal. This result shows that main trip signal does not log in alarm list in most shutdowns, then all PLC logics regarding to alarm generating and sending blocks must be checked and analyzed.

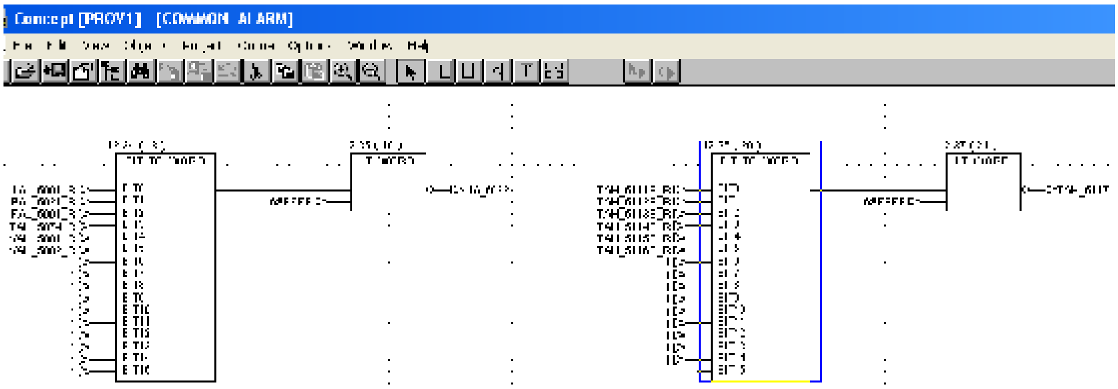


Figure 1:First Logic

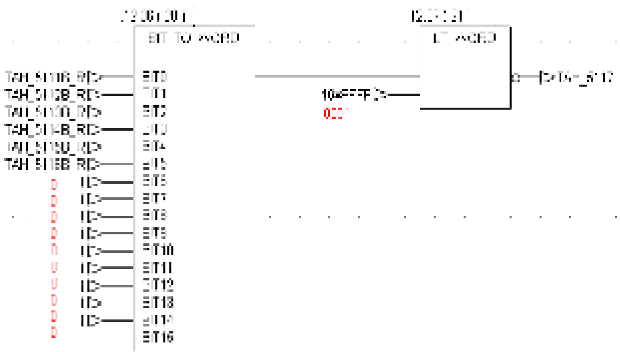


Figure 2. Modified Logic

During above checking, many errors in PLC programming were detected and necessary modification was applied to PLC by control system personnel. In the logic; some alarms will be sent together to HMI as a common alarm, which represent group of signal as "one". According to Figure.1, "bit to word" and "LT_WORD" blocks are used. If each "TAH-5111B" or "TAH-5116B" gets faulty, then "TAH-5117" must be generates. But unfortunately, wrong programming prevents to generate any fault and send "TAH-5117" to HMI unless necessary changes apply to PLC according to Figure.2.

This type of modification was applied for several alarm groups such as "UA-5022" which is indicated in Figure.2. In this case if one of the input blocks actives, then the output would gets "1" state; while in the primary condition, the outputs could never be activated in different situations. Moreover the above case, during PLC logic checking, another problem in software programming was detected. As software programming philosophy was such a way that all trip signals goes to two redundant Modicon Modules in order to prevent any hardware failure to avoid trip the compressor unless this signal be in alarm condition. For all mentioned signals a common block is that is shown in Figure.3 is existed. [2,3]

But unfortunately in such a way of programming; if one of the inputs disconnects or have module problems, the other input also will not send data either. In Figure.3; "err-1" is channel 1 failure and "err-2" is channel 2 failure. The existence logic behaves correctly, if both channels are healthy otherwise the "enable" port of "select" block and "out" get inactive. The correction was applied to the logic as shown in Figure.4 and downloaded to PLC.

1.1 Alarms Modification that were sent to DCS

One of the main problems of this compressor is non sending pre trip alarms to DCS.As the compressor operator is not present at the station permanently; considering the need of controlling this critical compressor from central control room, a complete alarm list was prepared by the help of operation team and applied to the system, all signals were replaced with twelve spare locations which had been left in primary style without any change in network traffic. Then illustrated modification according to Figure.6 was applied to logic.

After every shutdown, it is necessary for operation team to realize first trip signal to rectify consequence trip problems soon.

In logic; an analogue signal has been sent to DCS to show the first trip factor; if this value is not zero, DCS logic will demonstrate this received value with a "tag name" in DCS alarm list. Again because of software programming mistakes, this analogue signal was sent wrong data to DCS like is shown in figure 4. These logic errors were corrected and downloaded to PLC as Figure.5..As in figure.5 is shown, enable port of comparator blocks in first out page would be activate one on ,one off frequently, in non alarm situation, then some of these blocks be active and do comparator jobs. Therefore in some condition, first out alarm does not be generated and secondary alarm would be sent by mistake.

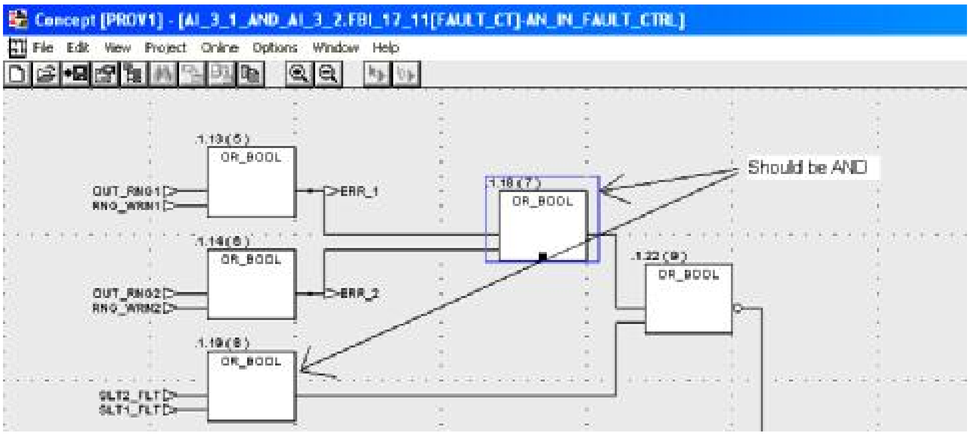


Figure 3. Modified logic to distinguish healthy channel

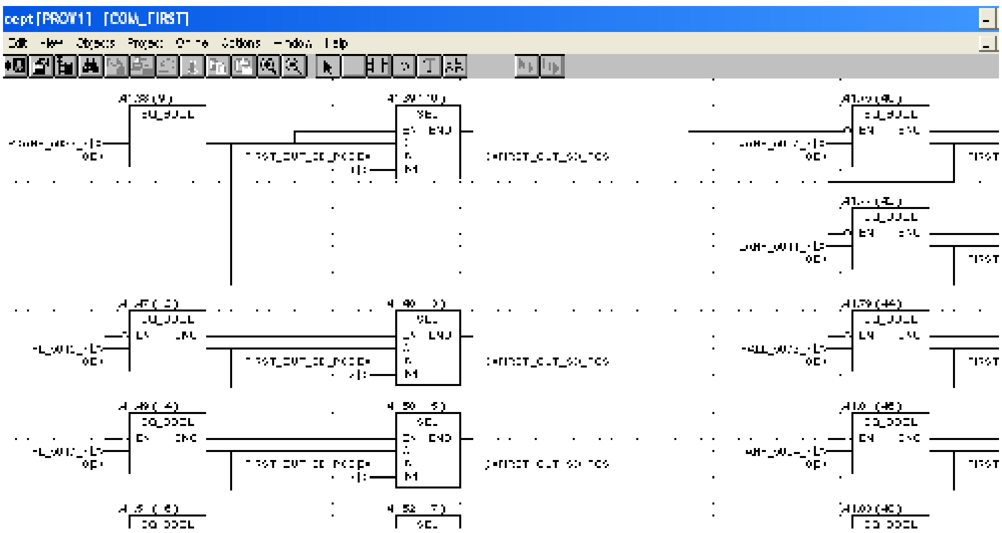


Figure 4. First logic of first out alarm

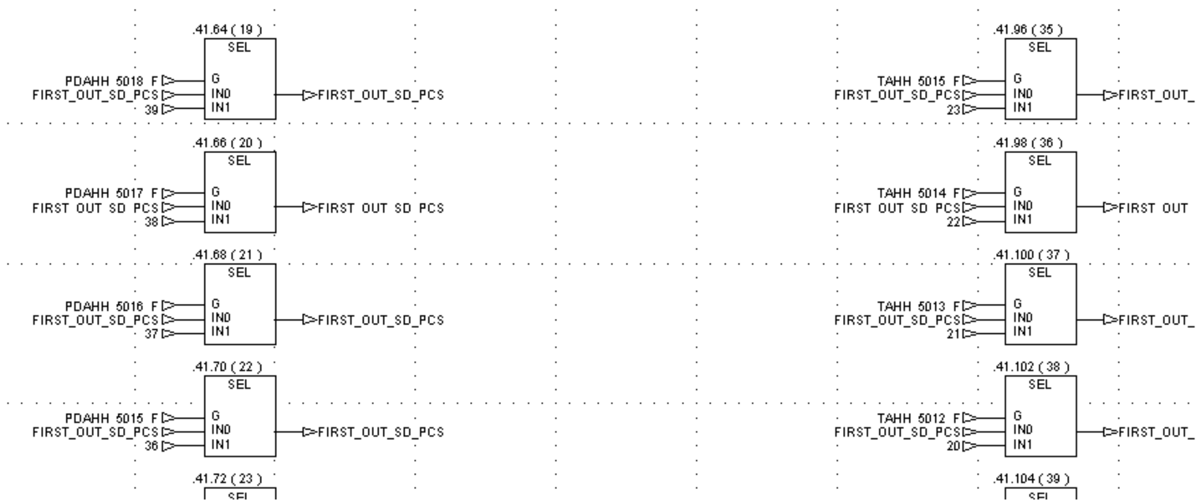


Figure 5. Corrected logic to distinguish first out alarm

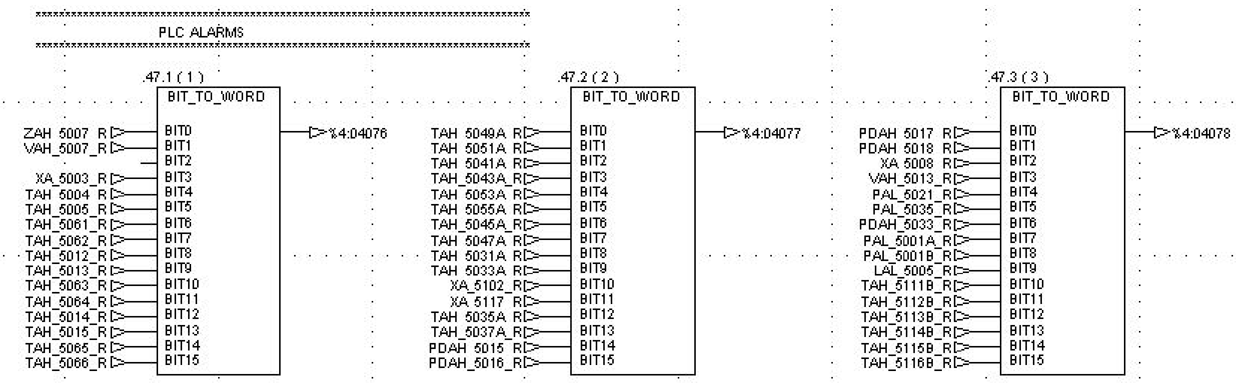


Figure 6. Applied modification to send more signals to DCS instead of spares

1.2 Temperature transmitter adjustment and hardware problem rectification

By statistical studies on the number of unwanted shutdowns, it was cleared that most of these trip reasons are because of temperature transmitters fail which the source is "RTD" sensors tip Brocken. These transmitters as is shown in Fig-9 are usually installed at bearings or body of the compressor with a high range of vibration, finally will trip the compressor in high rate of temperature when reach to alarm value rate.RTD disconnection will cause false increasing in temperature value indicated in HMI; while transmitter value will pick suddenly to the maximum value and finally trips the compressor. Although RTD broken in temperature transmitter led to reach maximum value in trend menu, so it recognized as false alarm.

Smart transmitters which are installed outside the control room were modified by changing the output settings via a hart communicator. In this condition, as soon as sensor fails, the value sets downscale instead of upscale. Also for the other type of smart ones located in control room cabinets, by installing P+F software on PC via an interface cable; the same settings applied to them. These settings can prevent many trips. In order to demonstrate transmitter failure and operator awareness. Background color of the mentioned transmitters would be changed into purple color in HMI; as soon as the value of each changes to zero reaching to a minimum value.

In addition to above case; analogue input values to controller often gets "FALSE" values in the monitoring system. Any disconnection in each analogue input loop cannot be recognized and would get random value in HMI. During different tests on this loop; it was cleared that any disconnection in analogue input loop won't get zero value because of internal charge in the loop, but remain as a random value around normal. So after discharging FALSE voltage via ground; this value changed into zero, and input channel fault was detected by highlighting "F" LED on the module. By investigating on the test results and search in technical documents, it was realized that; by putting a pull down resistance with the value 2mega ohm between negative and positive input channel, the possibility of fault detection for disconnection input signals to analogue modules are provided. This resistance value is less adequate of the channel internal resistance, so it won't make considerable error for the value. As well this resistance in parallel with the channel internal resistance has a result more than the necessity value for two barrier types. These two barrier types are KFD2-UT-EX1-1 and KFD2-STV1-EX1; 4-20mA to 1-5v converter manufacture by PEPPERL+FUCHS.

1.3 Changing the types of temperature transmitters

Existing type for temperature transmitters in this unit do not have the capability to sense sudden disconnection & connection of RTD tips. While happening this problem, no alarm reports to HMI (monitoring system) and the responsible operator can not aware of the loop problem, The continues disconnection and connection of the loop led into completely separating the RTD tips of this loop ,then at last trip the compressor.

After many investigation on temperature transmitter specification of various manufactures, replacement the existing type of the transmitter with a new model of ENDRESS HAUSER products; was the best solution.

The previous type was FISHER ROUSMOUNT; while was checked with all other manufactures such as old model of ENDRESS HAUER and ABB. As soon as changing these types, first for two high risks transmitters, no trip on compressor have been reported(around 6mounths)

1.4 Monitoring system upgrading

Following problems and using old version of windows NT encouraged us to replace the hardware and upgrade the HMI.

- Unavailability urgent components in market during hardware problems and un compatibility new hardware with old operating systems.
- Impossibility of any modification in monitoring system because of using runtime revision of Cimplicity software without development license.
- Non industrial computers which were delivered from manufacture company and non applying components with long time operation.
- Not being user friendly of old operating systems and time consuming maintenance and troubleshooting.
- Impossibility installation new antivirus on monitoring system
- Not automatic start up the computer after a sudden power off/on and restarting, then necessity instrument personnel presence at the time of restarting to enter password to run Cimplicity services.
- Continues locking the system because of software problems and security weaknesses.

To rectify above problems; following steps was executed:

- Suitable hardware and new industrial computer selection with long work components.
- Win2000 installation as well as proper antivirus software.
- SQL software installation, also SQL Enterprise in order to troubleshooting possibility and data registration in data base in different formats for more investigation.
- Installation a complete Cimplicity version as well as development license.
- Investigation to realize all executive and "dll" files regarding to communication network between monitoring computer and the controller, also necessary changes in windows registry to identify to Cimplicity
- Data base description and adjustment and specification in order to suitable communication between Cimplicity software and register and reload data from SQL.
- Transferring the previous project to this new upgraded system and necessary modification such as temperature transmitters color changing into purple and creating new diagrams according to operation request.
- Creating printer queue bottom in monitoring system such a way the operator be able to delete the print.
- Security enhancement:
- To limit the user's accessibility via some modification in windows registry adjustment and writing proper script. Nowadays, security is highlighted in industrial network .many points such as entering various, intentional or

unintentional using of network facilities by irresponsible people, has prepared a safe environment for the monitoring system applicant.

- Some of these securities was applied to the monitoring system by using *.scr instruction which is executed in KIX32 software.

- Auto restart and auto log on capability until run all services automatically of monitoring system, in case of any sudden power off or power on and without instrument personnel presence and password.

1.5 Software upgrading

To achieve the best state of HMI performance, software upgrading is inevitable, so below steps are necessary for upgrading previous software environment with windows XP service pack2.

- Install Win2000
- Install W2KSP4
- Install SQL7.0
- Install Cimplicity V4.01
- Register Cimplicity
- Install Cimplicity SP10
- Copy Unit 103 Project folder to: (094_99)
- C:\CIMPLICITY\HMI\projects
- Copy tony folder files to:
- C:\CIMPLICITY\HMI\exe

2 Conclusion

In this article, the procedure of determination, analysis and rectification of repeatable trips causes of off-gas compressors were presented. Achievement of these modifications was resulted in easiness troubleshooting, quick maintenance, permanent monitoring possibility from far from the compressor station without maintenance personnel and security enhancement of monitoring system.

Considerable amount of these unwanted trips which were because of sensor faults or loop problems were .reduced Also production team can make proper precautions before and during any trips according to new created alarm list and first out alarm which are sent to DCS. It worth to mention that analyzing shutdown causes are being progressed considering the possibility to monitor the system faults from control room in addition to upgrading HMI and creating new signals in logic by a great deal of modification in logic to find out trip reason. Finally optimization the monitoring system by hardware and software upgrading as well as applying security settings, installing updated antivirus and creating printer queue bottom enhanced the security system level without the necessity of maintenance personnel presence.

References

[1] "SCENETRON SYSTEM – UCP Wiring Diagram", SOP 6050482/3, GE-Nuovo Pignone Maintenance manual

[2] "Gas Reciprocating Compressor – Functional Logic Diagram", SOP 6050479/3, GE-Nuovo Pignone Maintenance manual

[3] "Local Panel Wiring Diagram",", SOP 6050592/3, GE-Nuovo Pignone Maintenance manual

[4] "Manual KFD2-UT-EX1 Convertor", Pepperl+Fuchs,

[5] "PACTware Process Automation Configuration Tool user manual", PACTware Consortium e.V., Karlsruhe, Deutschland, 2010

[6] "Vidio Graphic Pages", SOP 6050489/4, GE-Nuovo Pignone Maintenance manual

[7] "Modicon Quantum 140AVI03000 Analog Input Module Data Bulletin", 8000HO0460, North Andover, MA USA, 2004

[8] "Manual KFD2-STV1-EX1 Barrier", Pepperl+Futchs, 2002

[9] "GEMIS Tools for Cimplicity – Configuration Tools for Using Cimplicity HMI in GEMIS Equipment", GEH-6372D, General Electric USA, 1998

[10] "ODBC--Open Database Connectivity Overview", Microsoft Support, http://support.microsoft.com/kb/110093

[11] "Cimplicity HMI for Windows Based System User Manual", GFK-1180F, GE Fanuc Automation, 1997

[12] "Cimplicity HMI Trend and XY Chart Operation Manual", GFK-1260F, GE Fanuc Automation, 1999

[13] "Command Line Printer Control in Windows 2000 / XP", http://www.robvanderwoude.com/2kprintcontrol.php

[14] Jialong He, "Kixtart Quick Refrence", http://www.kixtart.org

[15] "Registery training book";Hossein Sedighi

A Framework for Enabling an Integrated and Proactive Decision Making In Airport Systems

A. Napolitano, M. Loffreda
SESM s.c.a.r.l.
Via Tiburtina
1238 Rome, Italy
{anapolitano; mloffreda}@sesm.it

D. Di Crescenzo
SESM s.c.a.r.l.
Via Circumvallazione Esterna, loc. Pontericcio
80014, Giugliano in Campania, Italy
ddicrescenzo@sesm.it

Abstract - Airport systems stands for one of most important and essential asset within the transport system of a country. The air transport liberalization has promoted the air traffic growth with a consequential cost reduction to the passengers and air carriers. In turn, this growth is moving the main airports towards the saturation, affecting the performance and quality of the services provided to the passengers. Some analyses, indeed, foresee that the capacity of the airports, in the management of the aircrafts and passengers, will likely be the bottleneck of the air transport growth. Hence, airport capacity needs to be improved. The common and agreed view for succeeding to this goal moves to the optimization and orchestration of the available resources within the airport. To do that, the interoperability and cooperation among stakeholders, operating in the airport, often with different purposes, has to be rethought, along with the development of novel technological systems enabling and supporting this process. To this aim, the paper proposes a new framework that enabling the information interoperability on the top of classic data interoperability among heterogeneous stakeholders (often legacy systems) operating both in airside and landside area and provides a net-centric orchestration service capable of forecasting and promptly optimizing the choices of the stakeholders with the aim addressed to the airport performance as a whole.

Keywords: complex dynamic systems; cooperative and distributed decision making; service oriented approach, airports.

1 Introduction

European air transport liberalization has been a success in terms of traffic growth and costs reduction [1]; airports are now considered as central and essential assets for the functioning of the international transport system. In this context, the envisaged bottleneck will likely be the limited capacity of the airports in the management of the flights and passengers.

Hence, airport capacity needs to be improved operating through two complementary ways; that is either building new airport infrastructures, such as runways, taxiways, gates, security checks, etc., or optimizing and better orchestrating the available resources [2]-[5].

Of course, the former is more expensive than the latter and needs of onerous investments from the airport manager, which tries to avoid this if not strictly necessary. Consequently, the latter is catching more interest because it requires less economic investments, which are mainly focused on the acquisition and use of novel technological subsystems, modifying as little as possible the airport infrastructure.

In turn, this approach impacts on the airport subsystems architecture already in use, as well as their internal behaviors/operations, which need to be rethought and adapted towards a cooperative context involving all airport stakeholders.

It's worth noting that this is not a simple task to be accomplished because the airports are extremely complex systems, being composed by several subsystems, processes and operations, where a multifaceted and fragmented network of stakeholders operates with different and sometimes conflicting objectives, strategies and procedures. A typical instance comes form the air carriers and airport manager. The former, indeed, would operate as fast as possible in the airport because their billing depending on the waiting time for airport services (i.e. refueling, cleaning, etc.) on the apron area, whereas the latter would grow the revenue delaying as much as possible the passengers in terminal area and air carriers on the apron area.

Several activities have been undertaken at European level for improving the stakeholders\subsystems collaboration with respect to a common goal, i.e. the airport performance as whole. A-CDM [6] (Airport Collaborative Decision Making) project, or other research activities financed in FP6 program, SWIM-SUIT [7], and TAM [8] (Total Airport Management) projects, or as SESAR-JU [9] (Single European Sky ATM Research Joint Undertaking), stand surely for the main activities that have been nowadays undertaken.

Some of them (i.e. A-CDM, SESAR, and SWIM-SUIT) try to solve the problem providing some innovative solutions for enabling an interoperability (i.e. data interoperability) process among the stakeholders, allowing a collaborative decision making. Nevertheless, they involve in the decision processes only the stakeholders operating in airside, completely neglecting those operating in the landside of the airport, which can affect most of the choices undertaken in *airside or apron area*; e.g. delay in check-in processes, because of passengers caught in a queue, may lag the boarding process modifying the allocation of the stands and turnaround scheduling.

Moreover, the only capability of exchanging information among the stakeholders, if on the one hand allows increasing the single awareness of each one of them, improving its internal decision process, on the other hand it does not reflect often on the optimization and improvement of the airport performance. An orchestration service capable of driving the stakeholder is hence needed, which takes advantage form the actual data, coming from the stakeholders systems, for forecasting the airport performance and identifying the presence of drawbacks.

Thanks to that, the arising of envisaged drawbacks that affect the airport performance, such as longer turnaround processes, delay in the stand allocation, longer waiting time for taxi allocation, could be promptly managed advising the involved stakeholder's, which could put in action suitable countermeasures. This approach needs to add on top of data interoperability the stronger concept of *information interoperability*, which allows not only receiving information but also properly interpreting it.

Stemming from this analysis and actual lacks, the paper presents a new framework capable of addressing the aforementioned challenges, that is:

- enabling the *information interoperability* [10] on top of classic data interoperability among heterogeneous stakeholders (often legacy systems) operating both in airside and landside area;
- designing a net-centric *orchestration service* capable of forecasting and promptly optimizing the choices of the stakeholders with the aim addressed to the airport performance as a whole.

The rest of the paper is organized as follows. Section 2 introduces the actual state of the art on the interoperability and integrated decision making issues, providing a description of the actual techniques and solutions available. Section 3 presents in details the proposed frameworks, along with a description of the architecture, whereas Section 4 describes an experimental scenario in which part of the proposed framework has been applied.

2 The Airport challenges

2.1 Interoperability

The interoperability of large systems and/or of System of Systems is something not new in many domains. Air Traffic Management (ATM) with SESAR is just one of the latest examples, but the same issue is faced in other international initiatives not related to the ATM like NCOIC (Network Centric Operations Industry Consortium) and NATO ATCCIS (Army Tactical Command and Control Information System) just to mention a couple of examples. All of those examples are dealing with the definition of network centric and information centric infrastructures that should enable a large number of heterogeneous systems to interact each other with complex (and, in general, not defined a-priori) patterns.

SESAR is, among those mentioned, the programme where airport systems are already considered. It is defining technical solution(s) and standards for enabling interoperability among ATM systems. In particular, SESAR considers SWIM [6] (System Wide Information Management) as the key enabler for the effective and seamless interoperability of the future ATM systems.

SWIM definition encompasses both the technical infrastructure (i.e. the software infrastructure, the technologies and the associated technical standards that will have to be put in place in order to enable such interoperability) than the "exact" definition of the data and services that have to be exchanged among ATM systems. The latter implies the definition of how data have to be represented, how services are defined, which is the associated behaviour and the SLA (Service Level Agreement) among a service provider and a service consumer. In short, SESAR is therefore trying to define both the semantic of the information to be shared and the interactions among systems via "services", than the technical infrastructure (i.e. the technical solutions) that should be put in place to enable such interaction and information sharing.

As far as airport "domain" concerns, SESAR SWIM is currently focused mainly on the "airside" rather than the "landside" since, so far, the interest is on defining how heterogeneous systems (ATC centres, AOCs, Airports, CFMU/NOP [11], etc.) can interact and share information rather than detailing how each system could be supported by SWIM in order to allow the interaction among its sub-systems. Moreover, since each airport can be seen in the little (compared to the whole ATM system) as a system of systems, on which different stakeholders operate through different systems and according to different business models, it is quite simple to envisage that the same SWIM principles, approaches and ideally solutions could however be applied within an airport itself, both landside and airside, for making easier the system interoperability and integration. This approach will find, of course, the interest of the stakeholders, which have already accepted the concept provided in SESAR through SWIM.

2.2 Integrated collaborative decision making

Following the idea that collaboration is the only mean to improve airport efficiency and performance, EUROCONTOL actually recommends to all airport the use of Airport-Collaborative Decision Making (A-CDM) concept, to gain information and situation awareness among airport stakeholders. The main envisaged objective of A-CDM is to increase airport performance enabling efficiency and punctuality (Figure 1) of the airport, with respect to the merely interest of the single airport stakeholders.

A-CDM brings substantial benefits to all stakeholders by improving the quality of information on which decisions should be made. On top of A-CDM implementation, and as part of this, further improvements are achieved by

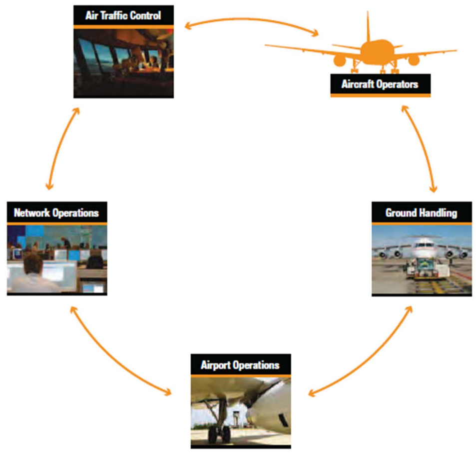


Figure 1. Information flow in Airport Collaborative Decision Making concept

implementing the Milestones Approach. Milestones are critical events in the progress of a flight, allowing a more accurate prediction of subsequent events. The main objective of this functionality is to further improve the common situational awareness of all stakeholders by defining Milestones against which flight progress is monitored.

The application of A-CDM implies that, due to different and sometimes conflicting interests, airport stakeholders have to be orchestrated in order to establish a collaborative planning. Hence, it should be accompanied with rules for negotiation or an entity capable of orchestrating them. The current implementation of A-CDM lacks of such an orchestration approach. A-CDM recommendations define only general rules for sharing the data and the kind of data to be exchanged, without providing any coordination among the stakeholders involved in such complex process.

A further limitation of the A-CDM approach concerns its scope. In fact, it just foresees the involvement of airside operations and some other systems supporting operations surrounding the aircraft (i.e. turnaround operation), neglecting those operating in landside which can take advantage from information sharing for prompt resources optimization and planning.

There is therefore the need to re-design the communication flow, introducing new modalities and rules among the stakeholders, as well as a orchestrator module capable of coordinating the resource allocation decisions in order to understand if choices made by one subsystem could have deteriorating effects on othersand on the overall airport performance.

A first attempt of introducing an orchestrating approach related to the integrated total airport system comes out from TAM project. In fact, the innovative principle introduced by TAM stands for creating an environment enabling airport stakeholders to maintain a joint plan, thus making decisions towards dynamically agreed goals mainly related to the overall airport performance. The scope of TAM is the airport as a whole,

Figure 2. High level architecture of proposed framework

monitoring and guiding airside and landside operations while taking into account additional information available through SWIM – (e.g. from departure airports). Expected benefits of TAM are an optimum use of available resources, increased passenger satisfaction due to situational awareness of predicted events, better understanding and more transparency of co-operative negotiation and decision making for all airport stakeholders together with conflicts resolution with better acceptance by the different stakeholders.

3 Proposed Framework

The proposed framework relies on the concept of *integrated total airport system*, in which each stakeholder/service/system is interconnected with the others for taking decisions in a cooperative way thus having no more several and fragmented stakeholders/services/systems independent from one another.

The proposed framework provides suitable solutions for making possible the realization of an *integrated total airport system*, enabling information interoperability among the stakeholders/service/subsystems (both *airside and landside*) and providing a novel decision orchestration service able to promptly forecast the inefficiencies in the airport and driving the stakeholders in cooperative decision making for their mitigation. This will allow preventing unexpected reduction of airport performance, which are mainly due to countermeasures undertaken in isolation by the different stakeholders.

A High level architecture of the framework, named SES-BOX, which meets such aspects, is depicted in Figure 2.

SES-BOX architecture relies on a Service Oriented Approach in which "Interoperability service" has in charge the task of integrating and enabling interoperability among subsystems, "Decisions Orchestration service" provides the rules and strategies for making decisions collaboratively and aiming to increase the total airport performance, whereas the API (Application program interface) stands for the standard interface for exploiting the internal services by mean of the external systems, such as Control Working Position (CWP), Flight Data Processing systems (FDPS), Ground Management System (GMS), Departure and Arrival Management Systems (DMAN and AMAN), Conflict Detection, etc.

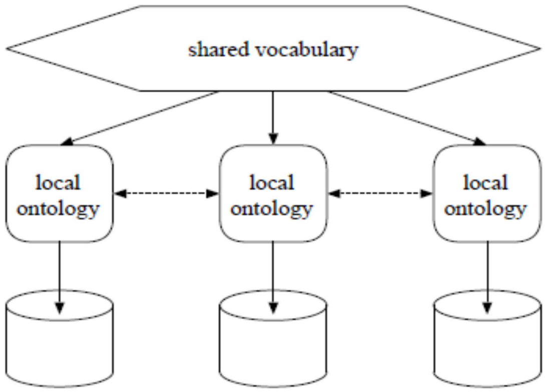


Figure 3. Information interoperability

3.1 Interoperability service

Interoperability service (IS) mission is twofold. On one hand it has to ensure the *integration* of the airport systems and on the other hand it has to enable the *interoperability* among the stakeholder's systems. As far as interoperability concerns, both data/object and information interoperability is reached through the IS service. The former allows only the physical exchange of the data or structure of data, whereas the latter introduces the capability of properly interpreting the received data. Thanks to the information interoperability, systems can understand in unambiguously the data thus enabling machine computable logic and inference on the data.

The approach to be exploited in IS service for succeeding in this purpose is in the definition of a vocabulary shared among the systems and the definition of several ontologies, one for each of class of systems (referred hereafter as *Data Domain*) (Figure 3). It's worth noting that a class of systems group together systems operating in the same domain of interest; e.g. all systems operating in Air Traffic Control.

The shared vocabulary allows expressing and describing the information in a common way, whereas the ontologies define the concepts/meanings glued to the

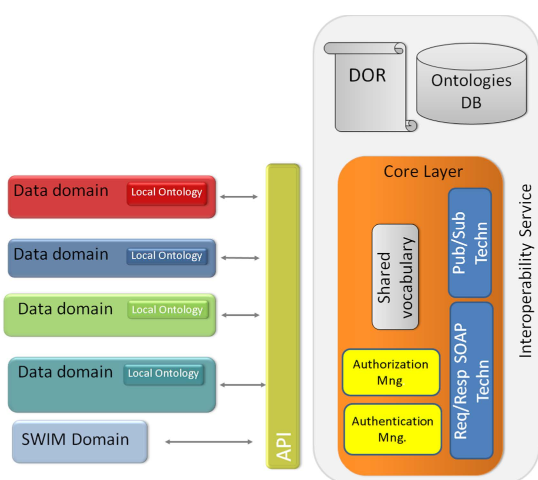


Figure 4. Interoperability service architecture

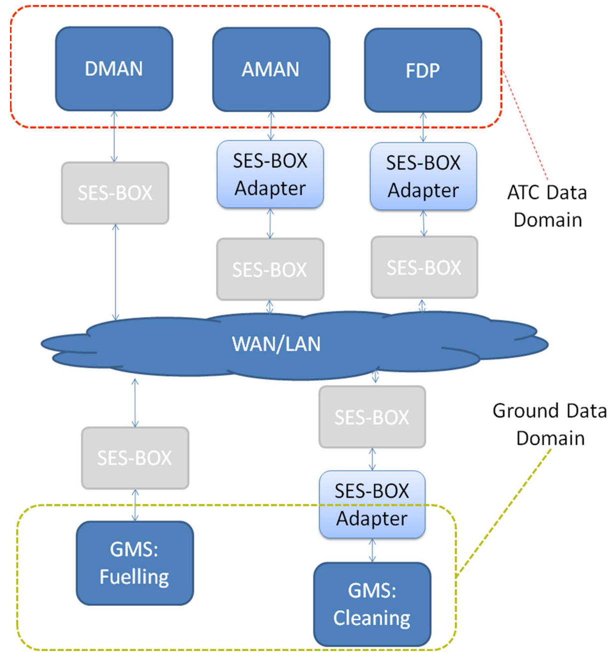


Figure 5. Interoperability service with legacy system and future IS native systems

vocabulary within a specific data domain.

The functionalities for sharing the vocabulary and the ontologies are provided by the core layer of IS. Moreover, the core layer issues some interfaces to the Data Domain for performing common services, such as *register_service*", "*invoke_service*", and "*deregister_service*". Through these services, each system of *Data Domain* can register its operations into a Distributed Operations Register, describing their signature via xml language, in order to be invoked from other subsystems to transmit and receive data. By the Distributed Operations Register (i.e. DOR) all systems can find out (through white and yellow pages services provided by the core layer) the operations for interoperating with other systems belonging to the same Data Domain.

The physical transmission and reception of data among the systems is instead performed through two complementary mechanisms. The former, namely, publish/subscribe pattern is mainly exploited for exchanging huge amount of data in synchronous and asynchronous communications among several actors (usually, in a one-to-many or many-to-many fashion), whereas the latter stands for the request/replay pattern suitable for synchronous request/response. Furthermore, the core layer of IS implements services for managing all functionalities related to the security policy and requirements that the infrastructure has to assure and perform, i.e. Authorization, Authentication, and Integrity.

Figure 4 shows the high level architecture of the interoperability service, in which several Data Domains are connected through the same core layer. The provided architecture assures high level of modularity and scalability adapting itself to different context. In fact, according to the system location, each system that would plug in the

network will have to import the core layer functionalities and interface its business application through the provided API. For the legacy system, this operation needs to be performed through a further layer, namely SES-BOX Adapter, which implements all the adaptation policy to/from the system and the SES-BOX.
Figure 5 shows a possible deployment schema for different kinds of systems that either rely directly on the SES-BOX (i.e. new system already compliant with SES-BOX specifications) or make use of an adapter to interoperate with the other systems through SES-BOX.

3.2 Decision orchestration service

Decision orchestration service (DSO) takes advantage from the interoperability service for gathering useful information from the actual stakeholder's systems, and tries to forecast the arising of inefficiencies and drawbacks in the airport management in presence of unexpected events. Examples of such events are lag in the turnaround processes, delay of flights during a cruise, passenger caught in a long queue at check-in affecting the boarding procedure, break down of security screening, reduction of check-in counters. To do that, the DSO service relies on the capacity of each stakeholder's system to publish a description, in XML format, of the provided service level through a SLA (Service Level Agreement) format. SLA will describe the metrics needed to measure the service level, along with the expected values. Thanks to that, DSO is capable of characterizing in a quantitative way the stakeholder's services and tries to forecast their future performance in the context of the other systems involved in the airport. In particular, DSO cooperates with all systems either passively or actively collecting information from them through the IS service, which provides the input for running fast-time simulation within a user-defined time window of the total airport. The fast-time simulation depends on the agent-based model [x] characterizing all stakeholders' processes involved into the airport such as turnaround process, baggage process, check-in, boarding, security, etc. Throughout the simulation each one of the simulated processes is monitored and the metrics defined in the SLA assessed for foreseeing any quality level reduction. These events give rise to a *tactical phase* in which the DSO service advises all systems that will suffer some inefficiencies and drawbacks, in order to put in place proactive and suitable countermeasures. The loop among forecasting, assessment and tactical phase is continuously performed for preventing airport performance reduction or, at least, make aware the stakeholders of critical situations that they will have to face in the next time window.

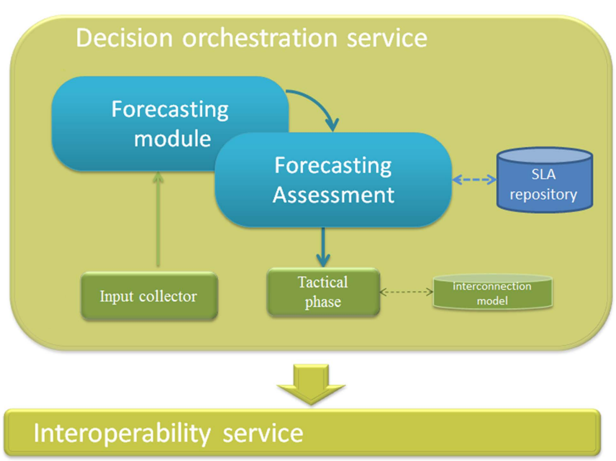


Figure 6. High level architecture of decision orchestration service

DSO relies on the architecture sketched in Figure 6, whose main modules are as follows:

- Forecasting module (i.e. FM) hosts an *agent-based model* of the airport, which is able to simulate all the process/behaviors involved in the airport. The inputs to initialize the executions of the simulation are achieved by the Input Collector Module. This module gathers all the needed inputs from the actual system relying on the interoperability service.
- Forecasting Assessment allows assessing the performance related to each process/behavior simulated by the FM. This module takes advantage form the SLA repository for pinpointing the metrics to be monitored and the thresholds for detecting the service quality decreasing;
- Tactical phase: is the module deputed to interact with the system that in the forecasts will face inefficiencies or drawbacks. It has in charge to prepare the messages to be exchange according to the type of system.

4 Experimental scenario

A suitable testbed has been built-on for assessing the capability of the framework to enable the interoperability among systems operating in Air Traffic Management, Ground/Apron area, and auxiliary (as meteorological systems). The scenario emulates the presence of critical meteorological conditions, such as snow storm or Vulcan eruption, on an airport. In this context different stakeholder's have to share information among them for managing as best as possible the inconvenient trying to reduce the inefficiencies to the passengers. In particular, the testbed involved into the scenario the following systems (Figure 7):

- Improved weather information systems (*IWIS*) is a system in charge of dispatching weather nowcasts and computing a prediction of airport capacity in terms of both arrivals and departures, taking into account weather conditions and runway layout;
- *Departure MANagement* (*DMAN*) stands for a system able to manage the aerodrome departure flow, confirming the start-up approval time to the aircrafts and the sequence of take-off of each aircraft. This system is located in the airport tower (TWR) and is under air-traffic controller responsibility.

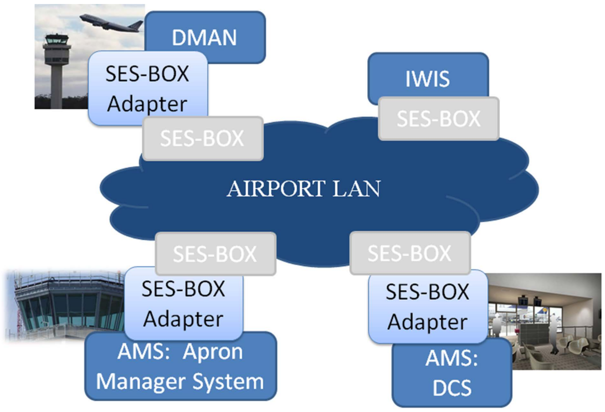


Figure 7 Experimental testbed

- *Apron Management System* (*AMS*) are systems capable of transmitting/receiving information regarding the state of the operation performing during the turnaround process. Turnaround process involves all the operation to be performed from the parking into the stand area (such as fuelling, catering, cleaning, luggage's loading, and boarding) up to the block-off (pushback). Anyway, within this process many stakeholder's are involved, each one for making one or more specific operations (e.g. one stakeholder performs catering and another both fuelling and cleaning) and they can make use of different systems. As an instance the Departure Control System (DCS) is one of the most utilized systems for managing the passenger's departure procedures (e.g. boarding, check-in, etc.). Apron manager system, indeed, is in charge of supervising all the operations performing surrounding all aircrafts in the aerodrome (it is located in the apron tower), and it communicates with the TWR for exchanging information about delays inbound from the air and outgoing from the apron area.

All enlisted systems are legacy systems, expected for the IWIS, which has been already developed compliant with the SES-BOX specifications. The scenario aims to assess the capability of the proposed framework to enable the interoperability among different system operating in different domain. Orchestration decision service is not involved into the scenario, because it's still under development. The scenario is activated from the IWIS system, which forecasts in the short time a snowstorm in the aerodrome. A proper message is published by IWIS with the nowcast of snowstorm and the calculated airport capacity. The message, formatted according to the data model defined in ATC ontology, is published by the SES-BOX to the Airport LAN on a proper *Topic*. DMAN, by the SES-BOX adapter has subscribed itself on the same *Topic* for receiving the IWIS messages. Once the message is received and decoded through the ATC ontology by SES-BOX adapter, the information related to the reduced capacity estimated from IWIS, are utilized to reschedule the take-off slots for the outgoing aircrafts. Of course, the DMAN tries to reduce as best as possible the waiting time in queue on the taxi. Once the aircrafts departure sequence has been rescheduled, it publishes a new message to the Apron Manager System with the new estimated Target Start Up Approval Time (TSAT) for each one of flights. Thanks to this information the Apron Manager System promptly advises, by means the message, the DCS system of the air company associated with the delayed flights. This information allows to the air company stakeholder of blocking the boarding process, if it is performing, or lagging it for avoiding to the passengers a long and bother waiting time on the aircraft. The same information is utilized from Apron Manager System for managing the turnaround operations still in progress on the apron area.

5 Conclusions

A novel framework has been proposed in the paper, which aims on the one hand to provide a means for enabling the interoperability among different systems operating in separate domains and on the other hand to propose a decision orchestration service capable of prevent inefficiencies in the airport services. The framework relies on a SOA architecture, which assures scalability and modularity properties. So far only the Interoperability service has been implemented and it ability in allowing the interoperability has been shown through the experimental scenario description.

References

[1] Long-Term Forecast Flight movements 2010-2030, EUROCONTROL, 2010.

[2] E. Gilbo, "AirportCapacity: Representation, Estimation, Optimisation", IEEE Trans. on Control Systems Technology, vol. 1, pp. 144-145, Sep. 1993.

[3] L. Brunetta, L. Righi and G. Andretta, "An Operations Research Model For The Evaluation Of An Airport Terminal: SLAM (Simple Landside Aggregate Model)", Journal of ATM, 1999.

[4] A. Pick, T. Rawlik, "Optimization of Airport Process," Proc. of ACC, May 15-18, Taiwan, pp. 1054-1059, 2011.

[5] P. Eriksen, "Collaborative decision making information management in airports," Proc. of Digital Avionics Systems, pp. 1-8, vol. 2, 2002

[6] EUROCONTROL and IATA, "Manual of Airport CDM Implementation", 2012.

[7] http://www.swim-suit.aero/swimsuit/

[8] TOTAL AIRPORT MANAGEMENT (Operational Concept & Logical Architecture) Version 1.0

[9] http://www.sesarju.eu/

[10] A.M. Khattak, Z. Pervez, A.M.J. Sarkar, L. Young-Koo, "Service Level Semantic Interoperability," Proc. of Applications and the Internet (SAINT), pp. 387-390, 2010.

[11] http://www.cfmu.eurocontrol.int/

The Impact of Multi-Institutional Semi-Structured Learning Environments (MISSLE)

Raymond R. Buettner Jr.
Information Sciences Department
Naval Postgraduate School
Monterey, CA, USA
buettner@nps.edu

Abstract - *A specific instance of the multi-institutional semi-structured learning environment (MISSLE) informing system construct is introduced and described. The description places an emphasis on aspects that have proven successful in accelerating new systems to the customer and in the enhancement of systems for those participating in these collaborative learning events. Examples are provided that include unmanned/robotic systems, command and control systems, and surveillance/reconnaissance systems.*

A review is provided of some of the major challenges associated with developing systems of systems (SoS)and the challenges of applying traditional systems engineering (SE) approaches as identified (independently) by Blanchard and Jamshidi. Challenges that may be addressed, in whole or part, using the MISSLE construct are identified and discussed.

Evidence collected from quarterly MISSLE events is offered that the informing system perspective that guides the design and execution of the MISSLE construct has successfully bridged some of these challenges and has the potential for broader application across both the government and commercial sectors for enabling more effective SoS development.

Keywords: systems engineering, systems of systems engineering, collaborative learning environments, informing systems.

1 Introduction

Earlier work [1] has described a novel informing system, the multi- institutional semi-structured learning environment (MISSLE) and provided two examples of these entities. MISSLEs created by the Naval Postgraduate School have evolved into ongoing series of complex events involving hundreds of experiments, thousands of participants and dozens of products annually as part of a unique business process [2]. These environments are established in multiple locations with regular events held at the Naval Postgraduate School's Tactical Network Testbed (TNT) north of Paso Robles, California and The United States Air Force Auxiliary Field in Avon Park, Florida. In Section 1.1 a detailed description of the two primary MISSLE constructs created and executed for the United States government is provided.

Section 1.2 reviews some of the challenges faced by today's engineers and program managers. Derived from the Jamshidi's [3] excellent paper on the subject and Blanchard's [4] superb textbook on systems engineering management these challenges serve as the framework by which the impact of the MISSLE construct are identified and compared.

In section 2.0 specific examples of SoSs developed using the MISSLE construct are described and used to illustrate the utility of the construct.

1.1 MISSLE Details

This research is focused on lessons learned and data gathered during the conduct of more than 40 quarterly field events. Currently two MISSLE environments are routinely executed that can be described using essentially the same general characteristics. Each learning environment has participants from government, industry and non-governmental/non-profit organizations (NGO/NPO) to include academia – this provides the multi-institutional form. Each environment has multiple individual experiments occurring simultaneously and for each experiment the specific organization that conducts the experiment has specific objectives generally only loosely aligned with any other activities taking place.

The first MISSLE is focused on military operations environment and is more institutionally aligned in that the military community that it serves is generally centered on the United States Department of Defense (DoD). The second environment, focused on humanitarian assistance and disaster relief is less aligned, with government participation, for example, ranging from cities and counties within the United States up to the United Nations. Similarly, the nature of industry and NGO/NPO participants is less homogenous for the second environment than it is for the first and has more participation by citizens of countries other than the United States.

In both cases, the activities are semi-structured. There is just enough structure to be safe, secure and legal while optimizing the potential for collaboration. These events feature dozens of technical activities (experiments) that are conducted by different organizations from different institutions that may not collaborate with one another routinely. The semi-structured approach seeks to balance the benefits of both structured and unstructured collaboration [3]. The design and construction of each environment follows exactly the same processes. The community stakeholders/sponsors determine the domain focus and this is shared with the community via a "request for information" or RFI. Organizations that have potential solutions to challenges related to the RFI focus domain submit white papers that are reviewed by the stakeholders and organizations that have submitted white papers describing potentially useful capabilities are invited to attend. The NPS, as the host, coordinates the many details associated with bringing hundreds of people together to conduct dozens of experiments. Some of these details include, scheduling, air space and frequency de-confliction, risk management planning, badging and briefings.

An average event in 2011 featured over 500 individual participants representing more than 80 companies. These companies range from defense giants to small firms with less than a dozen employees. Each organization proposed a capability of interest to at least one stakeholder with the stakeholders representing all 4 military services and several higher level headquarters. More than 50 government agencies, from the military services, and other defense related entities, to police and fire organizations typically participate. Government laboratories often send representatives both as observers and as participants in experimentation. Universities and non-governmental agencies (such as the Red Cross) participate as well. These participants spend an average of 4 days each conducting or observing experiments. Usually more than 50 planned experiments are conducted broadly categorized as Sensors, Weapons, Transport, Communications and Miscellaneous. Each event includes formal and informal evaluations of each experiment, regular morning and afternoon briefings, educational presentations over the lunch periods, and the opportunity to conduct additional "adhoc" experiments if desired.

During each event it is emphasized that the MISSLE is first and foremost a learning environment. The invitations to participate emphasize that these events are intended to explore applications of technology within the subject domains and that these are not sales events. Systems engineers and scientists are the intended audience on the technology side. On the stakeholder side the audience is focused on those who set requirements for new systems. The morning safety and coordination brief each day emphasizes that these are learning events with failure not only accepted but actually encouraged due to the learning associated with boundary exploration.

1.2 Systems of Systems Engineering Challenges

This section reviews some of the challenges faced by today's systems engineers and program managers. Derived from the Jamshidi's excellent paper on the subject and Blanchard's popular textbook on systems engineering management these challenges serve as the framework by which the impact of the MISSLE construct are identified and compared.

The cited works of Jamshidi and Blanchard each offer a variety of characteristics, as opposed to concise definitions, of "systems of systems". Blanchard's [3] perspective essentially approaches the systems of systems environment as part of the evolutionary development of systems engineering and views the increased emphasis on systems, greater utilization of commercial off the shelf products, rapidly changing requirements, increasing system complexities and other challenges as an extension of the challenges facing systems engineering in the current environment.

Jamshidi's [5] offers an excellent description of the challenges facing the systems of system engineering community and essentially concurs with Blanchard's approach by stating that "systems engineering (SE) needs to undergo a number of innovative changes to accommodate and encompass SoS."

In both instances the authors have identified a need to deal with emergence, integration, simulation and architectures in an environment that often denies the system engineer or program manager the ability to control systems that he or she must rely on as part of the system of systems that will deliver the capability the must be provided.

Between these two authors there are many examples of systems of systems implementation challenges ranging from the design, construction, and operation of a modern airliner to global command and control systems to robotic systems to renewable energy systems. NPS MISSLEs have featured participants working on projects that would address (in part) 15 of the 18 implementation challenges that Jamshidi identifies.

2 Systems of Systems Exemplars

Two specific examples of SoSs developed using the MISSLE construct are described with regards to the challenges identified in section 1.2 and used to illustrate the utility of the construct. The utility of other aspects, primarily educational and research, of the informing system construct to the challenges is also discussed.

Section 2.1 describes the development of an austere force protection system. Section 2.2 describes the development of an organic surveillance and reconnaissance system. Due to sensitivities related to military capabilities the performance aspects of these systems are not discussed rather the relationship between the system design and integration process and the SoSE challenges identified in Section 1.2 are emphasized.

Section 2.3 explores the potential of the MISSLE construct to enhance what Jamshidi terms the “theoretical problems” as opposed to the application challenges already addressed. The potential utility of the approach in enhancing SoSE and SE education is also explored.

2.1 SoSE Example: Force Protection

This section will describe role of the MISSLE in the development of a small team force protection system. It will address the challenges associated with the SoSE and how they were addressed for this example.

The nature of current warfare includes an increased emphasis on small combat teams that had not been anticipated in the pre-war environment. The result was that sometimes a remote outpost might have the same force protection capability, a sentry or look out, that has existed since wars were first fought. While it might be true that the sentry might be equipped with high quality optics and infrared or night vision devices he was essentially just a look out. Contrast this to any of the major outpost in the theater of operations that has coordinated aerial observation, 360 degree cameras and a variety of acoustic and other non-visual sensors all tied to a command post where the lookouts essentially monitor computer screens.

Military leadership determined that this was unacceptable and requested a small unit kit that would allow the single sentry to take advantage of modern technologies and to restore the asymmetrical advantage that high technology is supposed to provide to the western soldier.

Neither time (the request was urgent) nor resources (budget) would permit the creation of a single purpose built system to restore the asymmetry so instead the program manager (PM) sought to create a system of (existing) systems that could be integrated to permit a lone sentry to effectively watch over his small team in a potentially hostile region.

The PM relied heavily on a team that had been participating regularly in the military MISSLE events. This had exposed the team to more than 50 potential capabilities (systems) that could be integrated into the kit. This allowed the team to coordinate with end users (soldiers) to rapidly identify the key performance attributes and parameters for an effective kit.

Unusual in modern acquisition efforts, though critical for systems of systems design efforts, this meant that existing capabilities were as important as the formal requirement in creating the design. Over the course of 3 quarterly MISSLEs an initial integration menu was selected and explored, the initial design and integration conducted and the procedures and tactics created so that the initial system of systems, none of which had been originally intended to be integrated with the other, could be deployed for a successful combat evaluation.

The end result was that a final design was approved and less than one year from the release of requirement dozens of kits were on their way to the battlefield. The continuous exposure created by regular MISSLE participation served as an enabler for the engineering team. One could speculate that the participation of the industry and lab entities, where they are regularly encouraged to try new things (unplanned or ad-hoc collaborative experiments usually represent ~30% of the total experiments conducted) and to integrate with others made it easier for the individual system’s teams to work with the PM’s team to integrate their technologies into the new system of systems.

The MISSLE provided a mechanism to at least partially compensate for emergence, lack of PM control, non-concurrent systems development and other challenges.

2.2 SoSE Example: Organic ISR

This section will describe role of the MISSLE construct in the development of an organic intelligence, surveillance and reconnaissance (ISR) system design. More specifically it will address the challenges associated with the SoSE and how they were addressed.

Western nations in general and the United States in particular, have a substantial ability to collect and distribute imagery from a variety of systems from satellites to unmanned aerial systems. There have been reports that U.S. systems produce more video than could ever be reviewed by human eyes. However in spite of this amazing collection, storage and distribution effort the individual soldier executing in the current warfare environment is often at the end of a long line of higher priorities and does not get the near real time imagery that he needs to more effectively accomplish his mission. There remains a strong desire for small units to have organic ISR systems

The learning emphasis and failure tolerance of the MISSLE construct led to a participating soldier suggest to a UAS industry participant that it would be beneficial if they could attach an inexpensive commercial off the shelf (COTS) camera to their vehicle and make the pictures available via a wireless mesh network that another participant was developing.

After coordinating with the network, UAS and event hosts an experiment was conducted that demonstrated that sufficiently high resolution photographs could indeed be obtained and shared over the network back to the soldier operating the system. The system flown was literally held together by duct tape and tie wraps with a camera purchased at a nearby department store and removed from its casing.

At the following quarterly event the same camera system and network was properly integrated (sans tape and wraps) into the UAS. Again results were exceptional and a military sponsor directed the design and construction of a small set of prototype systems for combat evaluation.

The next MISSLE, the third for this effort, featured the participation of soldier operators to validate tactics, techniques and procedures for the prototypes so that they could be deployed for combat evaluation. This design is now part of the inventory of US forces in the war zone.

Two years following this deployment a UAS manufacturer involved in traditional acquisition with the US government was still struggling to integrate digital wireless mesh networking into its aircraft.

As in the first example, the existence of the MISSLE community provided a place, a community and a set of norms that helped to overcome some of the challenges associated with the current SOS environment. It could be argued that these environments, by creating intersection points across wide varieties of systems in relatively close temporal and physical proximity actually take advantage of those aspects of the SOS environment to create emergent solutions that would not be identified by the current requirements process.

2.3 Other Utility

Also to be explored is the potential of the MISSLE construct to address (in part) what Jamshidi terms the “theoretical problems” as opposed to the application challenges addressed previously. Additionally there is a potential for the MISSLE approach to enhance SoSE and SE education and research.

Essentially Jamshidi’s paper provides a list of theoretical problems that the SoSE community needs to tackle. These include standards, architecting, simulation, integration, emergence and management. In fact the planned upgrade of the test bed and its facilities provides an opportunity for research that includes all of these challenges.

Education can be included as well, systems engineering is still a relatively young discipline on the campus and the implications of SoSE have not been fully integrated in to teaching programs.

The military MISSLE has been integrated into an educational pilot degree program at the Naval Postgraduate School. A master’s degree in Technology Integration was created that included participation in 4 MISSLE events and the design of an integrated system as part of the degree requirements. The MISSLEs are regularly attended by NPS students and faculty. The Information Systems and the Systems Engineering students are by far the most numerous student participants. However it is worth noting that there has been participation by students and faculty from all four schools (Engineering, Business, Strategy and Operations), the three Institutes (Modeling, Networks and Systems Engineering) and literally every academic department in events related to one or both of the two MISSLEs.

MISSLEs may provide an opportunity to use both traditional and participatory research methods to explore process improvement for the SE and SoSE communities. As repeatable, regular and low cost venues these informing systems can be used to explore theory and practice as well as providing data sets for modeling and simulation. It should be remembered that the host entity is a university and works well with other academic organizations to incorporate research and educational opportunities. These environments should be used to explore the range of theoretical and implementation problems.

In a sense, the MISSLE may be the equivalent of the Hubble space telescope allowing the systems engineering community to more closely examine the emerging systems of systems world just as astronomers use the Hubble to examine the emerging universe. Where the Johns Hopkins Applied Physics Laboratory operates Hubble, the Naval Postgraduate School operates the MISSLEs.

3 Conclusions

The MISSLE informing system construct was described with an emphasis on the potential utility of the construct to address the challenges associated with systems of systems engineering. The ability of the MISSLE informing system, already claimed to support the development of education, research and practice, to bridge the gap between traditional SE and the SoSE approach was discussed through the comparison of two successful projects with significant MISSLE application to the SE/SoSE challenges derived from Blanchard and Jamashidi.

Additionally evidence was offered that these collaborative learning environments may offer some utility in addressing education, theory and research issues associated with the SoSE domain.

The primary intent of this work is to explore the utility of the MISSLE construct to enhance the understanding the systems of systems environment and to inform the practice and scholarship associated with systems engineering in the modern environment. The initiation of a robust dialogue regarding the potential of these multi-institutional semi-structured learning environments to advance such understanding (and performance) for the systems engineering community would be an exceptionally rewarding outcome.

References

[1] R. Buettner, "Multi-Institutional Semi-Structured Learning Environments", Proceedings of the 3rd International Conference on Society and Information Technologies, 2012.

[2] A. Bordetsky and D. Netzer, "Testbed for Tactical Networking and Collaboration", **International C2 Journal**, Vol. 4, No. 3, 2010, pp. 19-20.

[3] M. Swink, "Building Collaborative Innovation Capability", **Research Technology Management**, Vol. 49, No. 2, 2006, pp. 37-47.

[4] B. Blanchard, Systems Engineering Management, John Wiley & Sons, New York, 2008.

[5] M. Jamashidi, "Systems of Systems Engineering - New Challenges for the 21st Century", **IEEE A&E Systems Magazine**, Vol. 23, No. 5, 2008, pp. 4-19.

Proc. of the 2012 7th International Conference on System of Systems Engineering, Genoa, Italy - 16-19 July 2012

System of Systems to Provide Quality of Service Monitoring, Management and Response in Cloud Computing Environments

Paul Hershey, Shrisha Rao, Charles B. Silio Jr., and Akshay Narayan
Paul_C_Hershey@raytheon.com, shrao@ieee.org, silio@umd.edu and akshaynarayan@ieee.org

***Abstract*—As military, academic, and commercial computing systems evolve from autonomous entities that deliver computing products into network centric enterprise systems that deliver computing as a service, opportunities emerge to consolidate computing resources, software and information through cloud computing. Along with these opportunities come challenges, especially to service providers and operations centers that struggle to monitor and manage Quality of Service (QoS) for these services in order to meet customer service commitments. Traditional approaches fall short in addressing these challenges because they examine QoS from a limited perspective rather than from a System of Systems (SoS) perspective applicable to a net-centric enterprise system in which any user from any location can share computing resources at any time. This paper presents a SoS approach to provide QoS monitoring, management, and response for enterprise systems that delivers computing as a service through a cloud computing environment. Concrete examples are provided that identify the key components of this SoS and their application to real-world scenarios. Simulated results demonstrate the effectiveness of the approach for a representative scenario.**

***Index Terms*—Systems of systems, Cloud Computing, Enterprise Systems, Net-centric, Service Oriented Architecture, Quality of Service(QoS), Security, Information Assurance, DDoS**

I. Introduction

As economic pressure intensifies for network and enterprise operations centers, those responsible for these centers seek a method to lower costs in the presence of data overload. This dramatic increase in the quantity of data is a product of the evolution of complex, net-centric enterprise systems over which multiple disparate users in dispersed locations share gigabytes, terabytes, or even petabytes of data at high speeds over production networks. Cloud computing provides one possible method to this challenge by reducing cost through shared computing resources while distributing data to multiple end users efficiently [1], [2]. However, present approaches to cloud computing are weak with respect to ensuring Quality of Service (QoS) to end-users [3].

The approach presented in this paper introduces a System of Systems (SoS) to provide a clear and concise view of QoS events within cloud computing environments that pro-actively informs enterprise operators of the state of the enterprise and, thereby, enables timely operator response to QoS problems. Section II provides a step-by-step description of the SoS approach; Section III provides the mathematical model for the QoS metrics considered in our work. Section IV identifies real-world applications scenarios and presents concrete examples. Section V details the experimental results obtained in the prototype system created to verify the system of systems approach. Section VI presents benefits and conclusions.

II. Approach

Structural	A SoS has a structure that comprises interdependent systems that integrate to form a higher order system, usually resulting in a hierarchy. This hierarchy can include monitoring and response at the highest-level system down to the smallest sub-component system (i.e., bit-level).
Coupling	The systems that comprise a SoS include coupling with respect to such areas as data, information, functions, state, and algorithm. A loss of any portion of the SoS will degrade the overall performance or capabilities of the higher order system; therefore, the systems are interdependent.
Behavioral	Integration of decisions and actions of systems occurs in the higher order system through governance in contrast to non-SoS where the sharing of information is the basis for collaboration.
Inter-operable	Systems that comprise a SoS interface with one another and interoperate by design in contrast to non-SoS where systems are not designed to do so.

TABLE I
SoS Characteristics

Step 1: *Define a SoS for monitoring, management, and response*. A SoS [4]–[7] possesses the characteristics shown in Table I.

Figure 1 presents a SoS comprising a system of multiple administrative domains operating within a Service Oriented Architecture (SOA)-based cloud computing system [8]. For the system depicted in Figure 1, a single authority provides governance services to multiple heterogeneous administrative domains in which SOA-based applications enable business and collaboration services that support end-users who are pro-

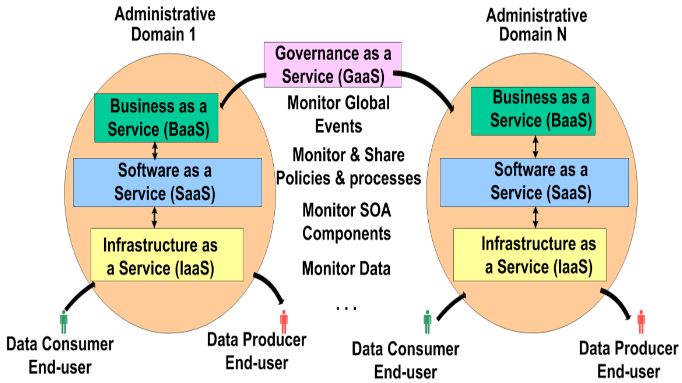


Fig. 1. Net-centric SOA-based System of Systems (SoS)

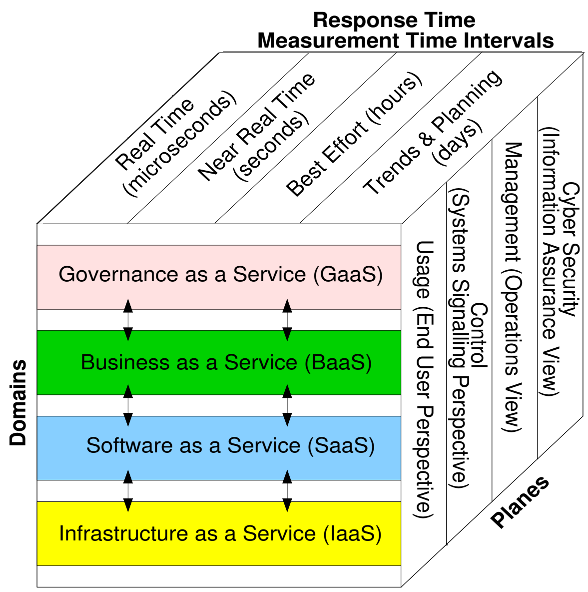


Fig. 2. Enterprise Monitoring, Management, and Response Architecture (EMMRA) for Cloud Computing Environments - EMMRA Cloud

ducing and consuming data using software and infrastructure services.

Step 2: *Derive Framework for Quality of Service Monitoring, Management and Response in Cloud Computing Environments*. The Enterprise Monitoring, Management, and Response Architecture (EMMRA) framework for Cloud Computing Environments (EMMRA Cloud), shown in Figure 2, extends previous work [8], to provide structure from which to identify points within the administrative domains of Figure 1 where key QoS metrics may be monitored and managed.

This framework is multi-dimensional to enable an end-to end view of the system where metric viewpoints may be located within traditional Open Systems Interconnection (OSI) layers (infrastructure through applications/software) and SOA-based layers (business and governance), as well as, across these layers.

The X dimension (Response Time) defines time-based services based on measurement time intervals (MTIs) ranging from microseconds to days or beyond, depending on the time criticality of the information to the mission. In each domain, the term service refers to any program, algorithm, function, analysis technique or monitoring activity that uses or interprets enterprise system information. A service can be a single individual or group of individuals (e.g., individuals doing problem determination or analysis), a program providing statistics, a piece of code performing network functions (such as load balancing), or another entity using network data. Different services in which network managers are interested require different time periods (i.e., MTIs) for data collection. For example, a critical mission in which lives may be at stake, could require real-time (i.e., micro-second) monitoring, analysis, and response for QoS event problem determination. On the other end of the spectrum, a department store studying customer trends may be content to collect and analyze data on a daily basis as they plan for an upcoming sales event. Traditional performance measurement and analysis approaches do not have the flexibility to collect data and assess QoS performance in real-time for the wide range of MTIs required by enterprise-level services. A need exists for a flexible data collection device that can collect information within complex enterprise systems for varying MTIs. EMMRA Cloud provides this flexibility and thereby meets critical QoS mission requirements for diverse missions with varying MTIs.

The Y dimension (Domains) detects and responds to enterprise events using similar techniques and instrumentation. A key contribution from this step is the extension of the domain dimension from Infrastructure as a Service (IaaS) and Software as a Service (SaaS), where traditional techniques attempt to enforce QoS in cloud computing environments, to include Business as a Service (BaaS) and Governance as a Service (GaaS) to meet the challenges of SOA-based net-centric enterprise systems. For BaaS, the EMMRA Cloud focuses on monitoring business processes and managing these to ensure their uniform implementation among end-users. For GaaS, EMMRA cloud identifies ownership of governance services and enables the institution and enforcement of policies that influence enterprise-wide behavior.

The Z dimension (Planes) introduces structures that monitor and manage particular end-to-end events. Planes (namely, usage, control, management, and cyber-security) encompass all domains; and thereby, provide a cross-domain solution that enables enterprise-wide monitoring, management, and response. Planes also span multiple MTIs enabling them to address multiple and diverse services. Implementing a plane-based approach within an SOA methodology is highly effective compared with existing domain-only QoS parameter evaluation techniques because the EMMRA Cloud approach fills the gaps between the domain monitoring layers. For example, the Usage Monitoring Plane encompasses those activities that filter, collect, analyze and disseminate information about the user data. The monitoring and management of this information could originate with an activity in the GaaS domain (e.g., dissemination, implementation, and assessment of operational polices), but this activity can then influence similar activities within the BaaS, SaaS, and IaaS domains. The Usage Monitoring Plane provides this enterprise-wide view.

Step 3: *Identify Cloud Computing Metrics*. To enable effective QoS monitoring, management, and response across and among multiple and diverse enterprise operations centers, we define relevant service-based QoS metric categories and metrics within those categories. The term "service-based metric" represents an object that has a name (i.e., metric name), a definition, a value (i.e., measure), an observation time period, a computing or collection method (i.e., measurement), and one or more threshold values for setting alarms (e.g., minor, major, critical). In this paper, whenever the word "metric" is used, it refers to the "metric name" of the metric object. The term measure is a quantitative value of the metric object derived for the observation time interval. The term "measurement" is the computing or collection process or method of determining the value of the metric object.

To clarify the meaning of these definitions, consider the example of a toll bridge with two parameters of interest: the length of the bridge and the number of cars that have passed over the bridge. Let the service provider be the local department of transportation whose mission is to determine whether the capacity of the bridge is adequate or if a second bridge span will be required to handle the traffic at some time in the future. A common characteristic of metrics is that they change over time. The length of the bridge is static over time and does not change regardless of how often it is measured. Thus, this parameter is not a metric. By contrast, the number of cars that have passed over the bridge requires counting cars over an observation time interval. This observation time interval could be the last hour, the last day, or from the time that the bridge first opened. The number of cars counted can be reset to zero at the beginning of the observation time interval as required by the end user in order to best meet the purpose of collecting the traffic flow information, e.g., throughput analysis or capacity planning. For this example, the terminology number of cars that have passed over the bridge is a service-based metric. The value representing the counted number of cars per observation time interval provides the measure for the metric. The measurement method for the metric could be to use a sensor placed across the lanes of the bridge that would advance a counter each time a car passed over it.

This paper focuses on the QoS metrics categories of performance and security and their respective metrics, shown in Table II. In Section V, we present results for some of these categories that we now define.

Delay is the elapsed time observed for a completed or on-going task and is caused by processing, queuing and transmission of data. Examples of other terminology used interchangeably with delay include response time, round-trip time, and latency. Delay metrics identify when the cumulative effects of processing and transport hinder the end user's ability to accomplish the mission. It is obvious to the end user and, therefore, influences customer satisfaction.

Category	Metric
Performance	Delay
	Delay Variation
	Throughput
	Information Overhead
Security	Authentication
	Authorization
	Non-repudiation
	Integrity
	Information Availability
	Certification & Accreditation
	Physical Security

TABLE II
METRICS CATEGORIES

The comparison of delay for different observation time intervals is **Delay Variation**. Examples of delay variation include: the variation in application response time between peak and non-peak hours, the edge-to-edge delay variation in the pattern of packet arrival events, and jitter. Delay variation identifies system instability that either presently prevents the end users from successfully executing their missions or that is a forewarning of upcoming problems that will do so.

Throughput metrics describe the amount of work completed over a time period [9]. Throughput metrics identify the level of work accomplished by the team, application, computer and network. Proper interpretation of these metrics enhances productivity, efficiency, and resource allocation.

The **Authentication** metrics category includes metrics that confirm the identity of users, systems, or data sources. Authentication employs one or more mechanisms such as passwords, key exchange, digital certificates, and biometrics. Confirmation of entity identity is a fundamental requirement in establishing trust and confidence in the service and enabling other security functions.

User **Authorization** metrics report the success and failure of access to resources based on policy and permission levels. Authorization extends authentication - confirming an entitys identity - to define the entitys privileges (i.e., those functions that the entity can be trusted to perform). Authorization metrics enforce the principle of least privilege. By ensuring that entities are assigned the fewest privileges consistent with their mission, the overall service integrity is maintained and mission effectiveness is enhanced.

Certification & Accreditation is the comprehensive evaluation of the technical and non-technical security features of an IT system and other safeguards, made in support of the accreditation process, to establish the extent to which a particular design and implementation meets a set of specified security requirements [10]. DoD requires C & A security through DITSCAP. Metrics within the C & A metrics category identify security risks and deficiencies, provide information to help ensure that steps are taken to correct these deficiencies and vulnerabilities, and provide information to help ensure the safeguarding of applications, networks, systems, data and information.

Step 4: *Identify suitable locations within the cloud computing environment for metric detection*. This step identifies suitable locations to observe and collect the metrics in Table II. Figure 3 shows these locations for an expanded set of metrics for security (green boxes), delay (yellow boxes) and throughput (blue boxes) for a representative cloud computing environment realized through a SOA-based net-centric enterprise system. This system includes four user communities: (i) end users shown with the client workstation; (ii) help desk shown with the trouble management system database; (iii) operations shown with the configuration management database; and (iv) engineering shown with the project control and development tracking database. With respect to the system components, the end user client machines that run applications over the base/-camp/post/station network appear at the left side of the figure. The network edge appears as a Customer Edge (CE) router attached to a High Assurance IP Encryptor (HAIPE) device and a Provider Edge (PE) router, along with a cache. The core network includes information transported via MPLS, Dense Wave-Division Multiplexing (DWDM) and Synchronous Optical Network (SONET) services. The network terminates at a Defense Enterprise Computing Center (DECC) with the connection from a second PE router to a HAIPE device and then to a DECC Edge (DE) router and a DECC LAN. To the right of the DE, are the computing services to be shared through a cloud computing environment, including those for dynamic host configuration protocol (DHCP), Distributed Names System (DNS), Discovery, App Server, Portal, Message Queue, Web Server, Service Node, Security, and Database. We explain the reasons for selecting these locations for metric monitoring in Section IV.

Step 5: *Identify potential implementation schemes from which to collect and analyze the cloud computing QoS metrics*. One possible scheme presented in this paper embeds EMMRA Cloud Computing (CC) agents within multiple diverse cloud computing components where they can continuously monitor the enterprise system for QoS metrics associated with cloud computing environments (e.g., delay, and throughput, and security metrics). These agents communicate over an out-of-band (OOB) monitoring network to EMMRA Cloud Collection and Analysis (CA) nodes that are located at local, regional, enterprise and global operations centers as shown in Figure 4.

III. System Model

In this section, we describe the mathematical model for the QoS metrics considered in our work. Further we describe the possible response actions to be considered on a QoS breach. We describe the individual metric classes considered in Table II.

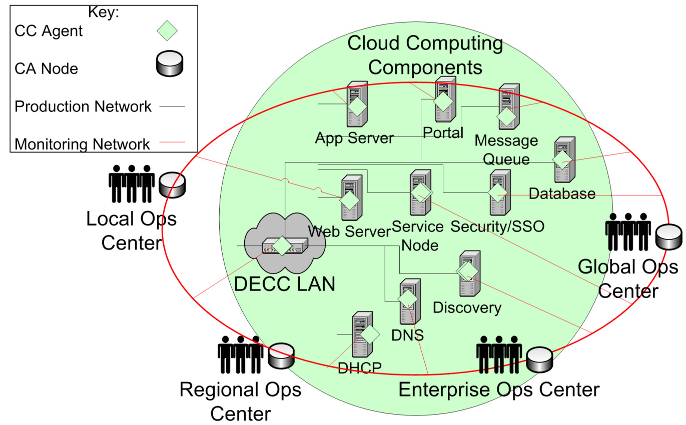


Fig. 4. Locations for IA/Security, Delay, and Throughput Metric Observation

A. Performance

The *performance QoS metrics* are additive in the numerical sense. The system of systems view from the top level domain in Figure 2 (i.e., GaaS) perceives *delay* as a sum of the delays experienced in the other lower domain levels of the cloud. This is also dependent on the infrastructure components used to provide the service. Hence we must include the component induced performance degradation. The delay metric can be represented as shown in (1).

$$D_{SoS} = p_1\, D_G + p_2\, D_B + p_3\, D_S + p_4\, D_I, \qquad (1)$$

where each p_i is a parameter that is dependent on the infrastructure component used. D_j is the delay experienced in each layer j in EMMRA, where the specific letter for j is the domain (i.e., Governance, Business, Software, Infrastructure).

We define *throughput* at the system level as the number of transactions that are completed per unit time. Throughput can be visualized at different levels. Throughput at the GaaS level of EMMRA is of the order of few days. This must be captured in a different scale. However, the throughput at the lower levels of EMMRA is multiplicative in nature. Throughput at every level is a function of the throughput at a lower level in EMMRA. Hence, we have:

$$T_I = n \times TransactionThroughput$$
$$T_S = m \times T_I$$
$$T_B = q \times T_S \qquad (2)$$

Here m, n and q are the numbers of transactions at the lower domain needed to complete the transaction at the higher domain.

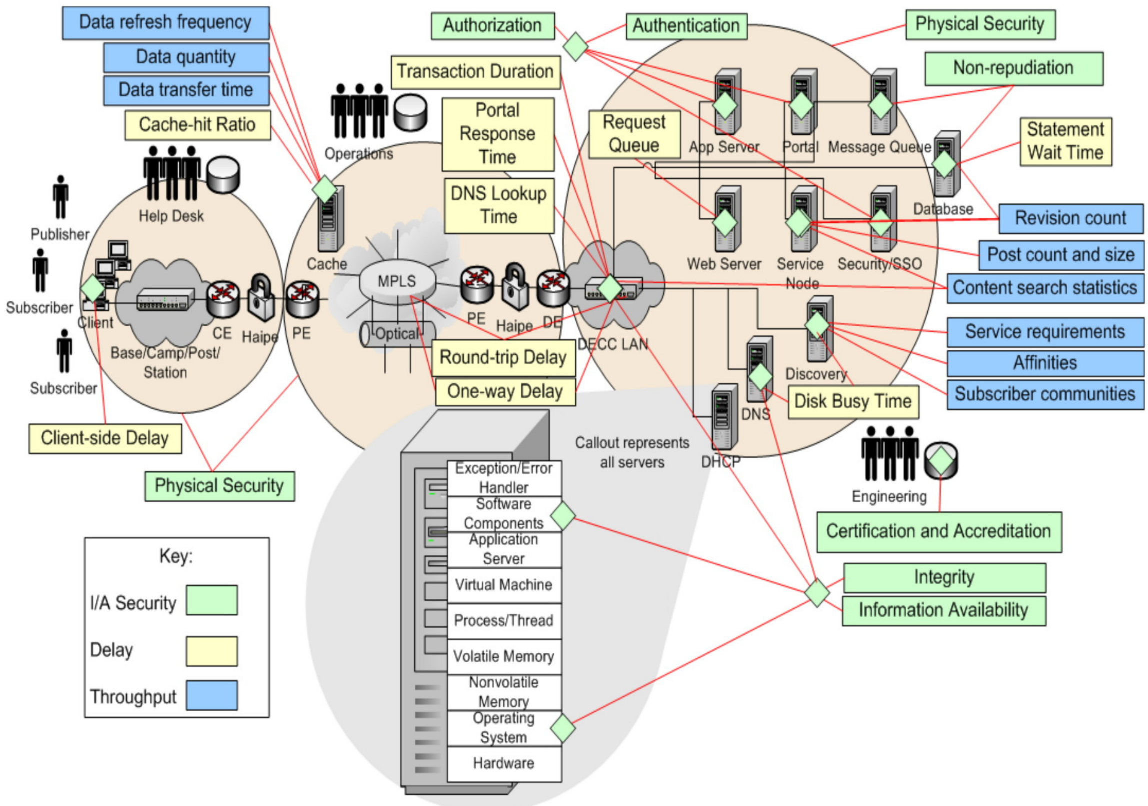


Fig. 3. Locations for IA/Security, Delay, and Throughput Metric Observation

B. Security

Security can be thought of as a functional requirement of the system. It comprises authentication and authorization using certificates and accreditation.

The *authentication* QoS metric is the logical conjunction at each level in EMMRA. The user's access to the system ceases at the level authentication fails. Hence the system of systems view of authentication is a logical AND of the authentications at various levels in EMMRA. Security can be viewed as a top-down metric.

$$A_{SoS} = A_G \wedge A_B \wedge A_S \wedge A_I \qquad (3)$$

The lower level EMMRA components have to be kept secure from the end user. A user at the top level can obtain service from the bottom levels, but, is not authorized to access the components directly. Only specific personnel are allowed access to the lower level components (*viz*, administrators). Hence in order to obtain access to lower level components the user needs to be authenticated at the top level. This is modeled in (3).

Authorization however is a bottom-up metric and is applicable at each level. User access to the service at any layer of EMMRA is subject to authorization. The authorization is such that the least privilege is granted sufficient to accomplish the operation. Authorization is applicable at each level in EMMRA Cloud. For example, in a banking application, an administrator is not authorized to access account details of the customer of the bank. Authorization at the IaaS level can be represented as shown in (4).

$$Auth_I = \min \left\{ \bigcap_{i \,\in\, \text{Set of actions}} p_i \right\} \qquad (4)$$

where p_i is the permission to perform action i at the IaaS level. The $\min$ operator is used here as an indication of the least privilege level that is granted to the user.

Similarly, authorization is defined for the rest of the levels in the EMMRA Cloud. The system of systems view of authorization can be obtained using methods such as linear logic [11]. However, we do not present the details here owing to page length constraints.

IV. Real World Application Scenarios

This section describes a potential real-world application scenario (i.e., use case) for the SoS approach to QoS monitoring, management, and response in cloud computing environments. This use case focuses on using the new approach presented in Section II to monitor, manage, and respond to QoS in the presence of Distributed Denial of Service (DDoS) attacks in cloud computing environments [12]. Assume that we use the SoS, framework, and metrics defined in Section II, Steps 1, 2 an 3, respectively. Step 4 is used to identify the locations at which to observe those metrics, and Step 5 is used to deploy EMMRE CC agents at those locations.

Figure 4 identifies the desired locations from which to monitor and respond to the DDoS-relevant security metrics of Table II. The rationale used to select these locations follows.

Authentication: cyber security can be monitored at the Apps, Portal, and Security/SSO servers. For example, EMMRA CC agents can monitor Security Assertion Markup Language (SAML) authentication assertions at the Security/SSO server. These assertions facilitate the secure exchange of authentication information between systems regardless of their underlying security mechanisms.

Likewise, EMMRA CC agents can effectively monitor and respond to *Authorization* events from the Apps, Portal, and Security/SSO servers where they can monitor and respond to information such as need-to-know determination required to grant authorization to a resource. This is typically implemented in the DD Form 2875, System Authorization Access Request (SAAR). For this example, EMMRA CC agents embedded within the Apps server would report the number of roles supported and the number of users assigned to each role, including the number added and the number removed over an observation period. The EMMRA CC agent would also monitor system resources access attempts to determine authorization failures.

Non-repudiation confirms that a transaction between two parties took place. Thus, a good location for EMMRA CC agents would be at the receiving end or a processing intermediary, such as the Message Queue in Figure 4. Here, the EMMRA CC agents can observe confirmation of messages both from the sender to the agent and from the agent to the receiver.

A prime location to monitor and respond to *Integrity* CC events is the DNS server. For example, web applications use DNS name resolution to resolve IP addresses. An attack called DNS poisoning corrupts the DNS domain-to-IP address database such that requests are redirected from the legitimate server to the attackers server for the purpose of presenting fraudulent content or collecting sensitive information. To prevent DNS integrity problems, EMMRA CC agents should audit all administrative updates to the master database and data transfer activity and provide a response so that operators configure DNS such that the end user never directly contacts the master database. EMMRA CC agents should search the DNS logs for particular events such as spikes in DNS traffic that could indicate a redirection. In addition, EMMRA CC agents should work with Intrusion Detection Systems, Intrusion Prevention Systems, anti-virus and configuration tracking software at the DNS servers to deter unauthorized changes to the server.

For *Information Availability* monitoring and response, EMMRA CC agents should be installed at the DECC LAN router and DHCP server to observe malicious traffic that could reduce information availability below its required threshold, thereby indicating spurious threat activity at individual servers. The commander Joint Task Force (JTF) for Computer Network Defense (CND) can use this metric to set and track the Information Operations Condition (INFOCON) [13]. Based on the INFOCON status level, a CA node in the operations center can invoke countermeasures to uniformly heighten or reduce defensive posture, defend against computer network attacks (CNA), and mitigate sustained damage to the DoD information infrastructure.

Certification and Accreditation monitoring and response supports the DoD Information Assurance Certification and Accreditation Process (DIACAP) Instruction [14]. EMMRA CC agents distributed within the engineering project control and development-tracking database can provide the relevant information to support ongoing certification and accreditation.

The DoD requires *Physical Security* for every enclave of a net-centric enterprise system. Thus, each user community could benefit from embedded EMMRA CC agents located at perimeter devices that indicate destruction, theft, or sabotage from unauthorized access to facilities, equipment, material, data, information, or documents.

A typical real-world scenario to which this use case applies is that of security monitoring and response for a financial/banking application. The SoS and framework apply because, in a banking enterprise environment, to complete one transaction at the business domain, policies established and enforced at the governance domain require that multiple sub-transactions occur at the applications/software domain that are distributed to end-users through the infrastructure domain. The MTI for this scenario could be microseconds to seconds, depending on the significance of the transaction. The Cyber Security Plane would be used to monitor across all domains to detect and enable proactive response to a DDoS security breach within any of these domains that could compromise the transactions and cause potentially devastating consequences to the end-user customer.

Other potential real-world applications that can use the SoS approach include the business processes of corporate enterprise systems and e-governance portals of governments.

V. Results

The performance metrics were verified using a prototype online transaction processing application. Delay, throughput and variation in delay were recorded at various time granularities. The graph in Figure 5 indicates that the system of systems perspective of delay metric is additive in nature as discussed in Section III. Variation in delay over time (Figure 6) and throughput (Figure 7) are indicators of the overall system performance. As a part of this work, we determine the QoS parameters that provide a SoS view of the proposed cloud environment. We fix QoS thresholds (e.g., throughput per

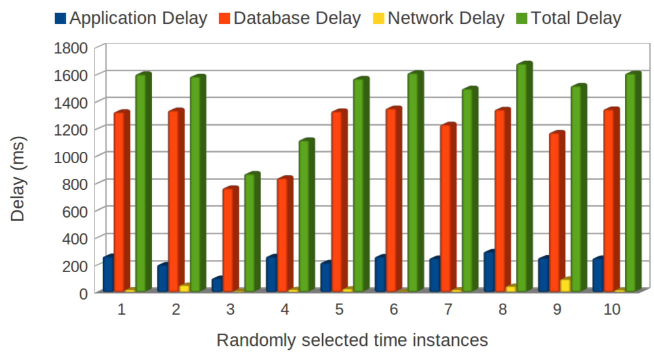


Fig. 5. Delay recorded in 10 sample transactions

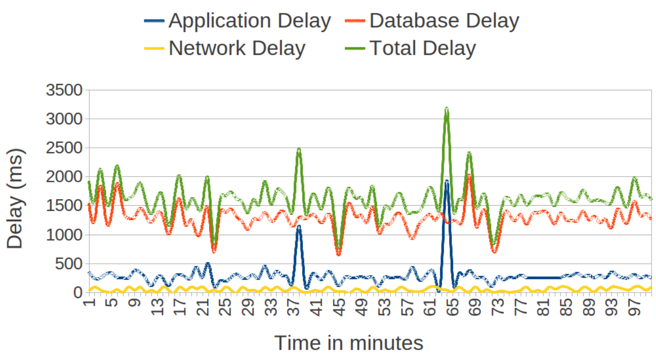


Fig. 6. Variation in delay recorded over time

second and delay per millisecond) for the system and monitor the service itself or the service components for QoS breach. The QoS thresholds can be specific to the application scenario and need not be fixed a priori for all applications that can be deployed on a cloud. For a QoS breach, a response action (RA) (i.e., an automated action to rectify the breach) is initiated. Well-established QoS monitoring guidelines and frameworks exist for IaaS and SaaS cloud deployments. We establish a method to correlate the IaaS/SaaS QoS breach events to the business process and governance. This correlation provides a SoS view of the QoS monitoring and management in a cloud environment.

VI. Conclusion

The new approach presented in this paper enables cloud computing service providers and operations centers to meet committed customer QoS levels using a trusted QoS metric collection and analysis implementation scheme that extends traditional monitoring, management and response for IaaS and SaaS to a complete SOA-stack that includes business logic (BaaS) and governance (GaaS). The paper includes real-world scenarios that describe the applications of this approach to voice and data systems for performance metrics and to DDoS for security metrics. Next steps include simulating these scenarios to quantify the effectiveness of this approach with respect to operations center response time to restore QoS in the presence of anomalous enterprise events.

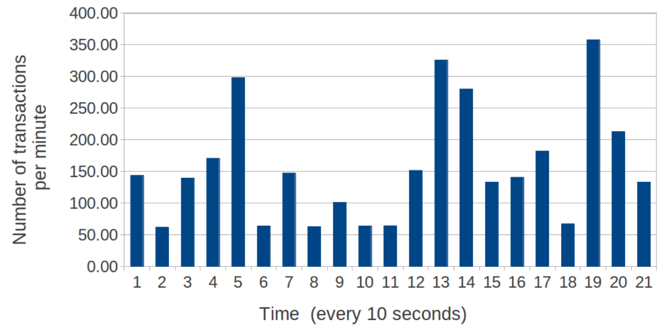


Fig. 7. Throughput: Number of transactions per minute

References

[1] P. Mell and T. Grance, *The NIST Definition of Cloud Computing*. National Institute of Standards and Technology (NIST), US Dept. of Commerce, Sep. 2011, NIST Special Publication 800-145, http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf.

[2] B. P. Rimal, E. Choi, and I. Lumb, "A taxonomy and survey of cloud computing systems," in *NCM '09: Proceedings of the 2009 Fifth International Joint Conference on INC, IMS and IDC*, Aug. 2009, pp. 44–51, doi:10.1109/NCM.2009.218.

[3] M. F. Mithani, M. Salsburg, and S. Rao, "A decision support system for moving workloads to public clouds," in *Annu. Intl. Conf. on Cloud Comput. and Virtualization (CCV 2010)*, Singapore, May 2010, doi:10.5176/978-981-08-5837-7_135.

[4] G. A. Lewis, E. Morris, P. Place, S. Simanta, and D. B. Smith, "Requirements engineering for systems of systems," in *3d Annu. IEEE Intl. Systems Conf. (IEEE SysCon 2009)*, Mar. 2009, pp. 247–252, doi: 10.1109/SYSTEMS.2009.4815806.

[5] S. M. White, "Modeling a system of systems to analyze requirements," in *3d Annu. IEEE Intl. Systems Conf. (IEEE SysCon 2009)*, Mar. 2009, pp. 83–89, doi:10.1109/SYSTEMS.2009.4815777.

[6] "Defense Acquisition Guidebook (DAG)," Jan. 2012, [Online], Available: https://dag.dau.mil/Pages/Default.aspx, US Dept. of Defense.

[7] *Systems Engineering Guide for Systems of Systems, Version 1.0*, ser. OUSD (A & T) SSE. Office of the Deputy Undersecretary of Defense Acquisition and Technology, Systems and Software Engineering, Aug. 2008.

[8] P. Hershey and D. Runyon, "SOA monitoring for enterprise computing systems," in *11th Intl. IEEE Enterprise Distributed Object Comput. Conf. (IEEE EDOC 2007)*, Oct. 2007, pp. 443–450, doi: 10.1109/EDOC.2007.33.

[9] P. J. Denning and J. P. Buzen, "The operational analysis of queueing network models," *ACM Comput. Surv.*, vol. 10, no. 3, pp. 225–261, Sep. 1978, doi: 10.1145/356733.356735.

[10] "DoD instruction 5200.40: DoD information technology security certification and accreditation process," [Online], Available: http://csrc.nist.gov/groups/SMA/fasp/documents/c&a/DLABSP/i520040p.pdf.

[11] J.-Y. Girard, "Linear logic," *Theoretical Computer Science*, vol. 50, no. 1, pp. 1–101, 1987, doi: 10.1016/0304-3975(87)90045-4.

[12] P. Hershey and C. Silio, "Procedure for detection of and response to distributed denial of service cyber attacks on complex enterprise systems," in *6th Annu. IEEE Intl. Systems Conf. (IEEE SysCon 2012)*, Mar. 2012, pp. 85–90, doi:10.1109/SysCon.2012.6189438.

[13] "DoD instruction 5200.80: Security of DoD installations and resources," 2005, [Online], Available: http://www.dtic.mil/whs/directives/corres/pdf/520008p.pdf.

[14] "A one-way delay metric for ippm," IETF: RFC2679, [Online], Available: http://www.ietf.org/rfc/rfc2679.txt.

Design and Realization of the Simulation Component-based Parallel Framework for HLA Federate

Zhang Jing, Zhang Yingchao, Yu Qin zhang, Guan Chuan fang, Li Wei
Science and Technology on Complex Systems Simulation Laboratory, Beijing Institute of System Engineering, Beijing, China
zhangjing.benben@gmail.com

Peng Yong
National University of Defense Technology
Changsha, China

Abstract - *While the difficulty and complexity of simulation system increasing, the development efficiency and execution performance of the large-scale distributed simulation system (LSDSS) must be improved effectively. Furthermore, while the multi-core has become popular hardware, yet the traditional federate mechanism couldn't utilize and schedule the CPU resource, which would limit the simulation performance. Then PFHF, Parallel Framework for HLA Federate, was designed and realized to solve the problem above. PFHF could combine and construct federate from simulation model components, perform simulation engine management, data distribution management, object management, component management and load balancing function, etc., then support the integrated execution environment based on multi-core and parallel technique. Several experiments proved that PFHF could utilize multi-core CPU adequately, and reduce the LSDSS execution time efficiently.*

Keywords: Modeling & Simulation; HLA; Parallel Federate; Simulation Component; Multi-core CPU; Dynamic Load Balancing.

1 Introduction

Modeling and Simulation (MS) is one kind of effective methods to analyze the complex systems (CS). HLA has been applied widely in the LSDSS field for the reason that it could solve the simulation reuse problem and save system development consuming [1]. Additionally, Component Based Software Engineering (CBSE) , is another kind of effective methods to implement software reuse. Yet, compared to federate, component is provided with much more reusability, for its special characteristics, i.e. function cohesion, execution agility, combination speediness, etc. To take advantage of component technique, the functions of Object Class and Interaction Class could be encapsulated by component, and the HLA federate could be constructed accordingly. Then supported by the Component Management Framework (CMF), the federate could be combined and assembled to construct a new federation or system as required with the same function just as based on HLA mechanism. So, Federate Component Method (FCM) is propitious to model combination and reuse, and could provide solutions to realize federate parallelization.

Furthermore, as the computer hardware developing rapidly and the simulation system execution speed meeting higher requirements, the research on HLA's performance keeps deepened. Under the traditional manner, federate would execute sequently, so it's necessary to ameliorate the federate development method to utilize the multi-core CPU's computing capability, which is a bottleneck to improve simulation performance. So federate parallelized is the other efficient method to improve simulation speed.

Based on the two points above, Parallel Framework for HLA Federate (PFHF) was advanced, researched, and realized to solve model reuse and federate parallelization. PFHF could support Simulation Component Model (SCM) which should be compatible with HLA to allow federate to be constructed by Simulation Component (SC). PFHF could support series of services to schedule federate synchronously, including simulation engine management, data distribution management, object management, component management and dynamic federate load balance function. PFHF could communicate with RTI by the special adapter.

2 Background and Challenge

2.1 Theory Basis

HLA, High Level Architecture, is a software architecture standard in distributed simulation field to solve the mutual-operation and reuse problem of different simulation systems. Then HLA prescribed the interface requirement of federate and the interaction procedure between federates in a simulation system to ensure the validity. Yet it didn't prescribe how to realize the federate itself, which caused large numbers of various different federates formats with different execution mechanisms and operation performance. To improve the federate speed became one of the most significant works. Furthermore, as the development of hardware, federates being operated by the multi-core CPU became normally. A new method to improve simulation system's operation speed aroused researcher's great interests, i.e. how to parallelize federate and realize simulation entities (in a federate) operation synchronously by different CPU's cores. Usually, the methods could be classified to 3 types [3], [4], [5].

Yet each of the methods above has its advantages and disadvantages respectively. A new inspiration emerged during our research, i.e. if component could support some kind of API similar with RTI function, it wouldn't cause great difference between developing SC and federate. It would be benefit to transplant the existed federates to new SCs. Therefore, improvement and amelioration to HLA based on component and parallelization techniques was researched roundly in the paper.

2.2 Correlative Research

Modeling and Simulation based on component technique has been studied by many researchers and achieved many achievements. The relative works about HLA simulation system includes extensible HLA Federate Architecture[6], Component-Based Federate Development [7] and Simulation Component Model based on HLA [8] etc. Research on Parallelized Distributed Discrete Event Simulation (PDDES) is very wide, for example, event ordering, memory management, time synchronization, message transmission, and so on. And the Time Synchronization Algorithm (TSA) is the kernel factor to affect the performance of PDDES and researched much more deeply and widely, in which the Time Warp Algorithm was the most popular [9] [10] [11] and gets the most application. Jeffrey [6] advanced a Standard Simulation Architecture (SSA) to solve the inter-connection problem between simulation systems based on SPEEDES and HLA.

Yet all the work above only dealt with one aspect between SC and parallelization. Therefore, PFHF, Parallel Framework for HLA Federate, was researched and developed in our work to improve the performance of HLA simulation system when being executed on multi-core CPU. Furthermore, PFHF could correspond with RTI very well to fulfill data transform and transmission, support RTI component and non-RTI component, and permit multi-entities operation parallelization among component.

3 Parallel Framework for HLA Federate

3.1 Summary

PFHF is composed of three parts, i.e. Simulation Component Model (SCM), Parallel multi-core emulatOr for coMponent (PCOM), and RTI Adapter, shown in Fig. 1. Simulation Component (SC) is the User Model that realize SCM standard interface. SCM standard interface includes the Management Interface (MI) and Operation Support Interface (OSI). MI would be called by PFHF, and OSI would be called by Simulation Engine to deal with the events sent and received by simulation model. Each simulation component maybe includes several entities. For instance, the black circles on the corresponding bar of each SC represent the Simulation Entities (SE). So, component1 includes 4 SEs, and component2 includes 2 SEs.

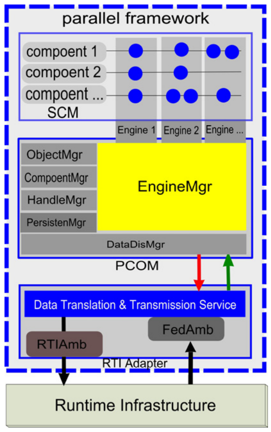


Figure 1. Parallel Component Framework

PCOM is the kernel module of PFHF and composed of several services, including the Object Management (ObjectMgr), Component Management (CompoentMgr), Handle Management (HandleMgr), Persistent Management (PersistenMgr), Data Distribution Management (DataDisMgr) and Engine Management (EngineMgr).

RTIAdapter comprises of RTIAmb, FedAmb, and Data Translation and Transmission Service (DTTS). RTIAmb provides the interface for federate to call RTI service. And FedAmb is used to realize RTI callback function interface.

3.2 Principles of PFHF

There are many popular standard component models, such as the EJB of Sun, the COM/.NET of Microsoft, the CORBA of OMG, and so on. Enlightened by these techniques, PFHF introduced some concepts and methods used by them. Meanwhile, to keep consistent with the federate development technique, which would benefit to reuse the existed federate resource, SC design adopted some glossaries and concepts of HLA. Therefore, the requirements of PFHF must be obeyed when construct SC.

- SC should be passive and driven by simulation engine with the discrete events scheduling manner.
- Interaction between SCs should be fulfilled by event manner completely.
- Model combination needn't change component's code to finish the dynamic combination.
- SC could base on both RTI mechanism and non-RTI mechanism to keep the reusability and flexibility of component development mode.

- Data requirements between SCs should adopt the method similar with HLA Declaration Management.

3.3 Definitions and Design

Def1: Definition of SCM

SCM = <MPort, Attri, BaseModel, UserModel, ObjectBaseClass, InteractionBaseClass>

- MPort—InputPort and OutputPort, the management interfaces.
- Attri—the attributes of SCM.
- BaseModel—the base model of SCM
- UserModel—the user model of SCM
- ObjectBaseClass—the object base model of SCM
- InteractionBaseClass—the interaction base model of SCM

BaseModel was designed to simplify model reconstruction and transplant process. Each SC should contain one BaseModel, which would be the root class of all the UserModel, and would play the most important role in the SC. BaseModel would provide several functions to UserModel, such as simulation logic time management, event programming and execution, data sending and receiving, and so on. Furthermore, BaseModel provided the interface functions similar with HLA Object Management Service (OMS). UserModel must realize these functions. Meanwhile, BaseModel could manage event effectively, and fulfill the functions that entities attributes updating, transfer, and exchange.

BaseModel could support combination and management of sub-model to ensure that the simulation entity could own and schedule the sub-entities, and could increase and decrease sub-entities dynamically. The coupling relationship between In/OutPort of sub-entities would be managed by their father-entity. According to changing the coupling relationship of entities dynamically, the structure of simulation system could be adjusted dynamically. Meanwhile, SCM would support the object base class and interaction base class in accordance with HLA to fulfill data filtering function for simulation entity.

3.4 Realization of PCOM

PCOM parallelized the computing process of Simulation Entities (SE) among the federate, and provided the thread-level parallelization computing capability for SEs. Furthermore, to parallelize SEs based on different simulation engines, it's necessary to encapsulate the data access and functions between different SEs to events and transfer these events by simulation engine. Fig. 3 described the relationships among each service of PCOM. Limited by the paper length, only the simulation engine management service, i.e. EngineMgr, would be discussed emphatically.

Simulation EnGine (SEG) adopted the thread technique based on multi-core CPU platform. So each SEG could be called as a Logical Process (LP). LP, a basal concept in the Parallelized Discrete Event Simulation (PDES) field, represents one serial process in the real world. So PDES could be looked as composed of series of LP. The communication among LPs would be implemented by sending events with timestamp. Furthermore, the communication cost would be decreased as possible by the high efficient event ordering and scheduling functions.

Time Synchronization Algorithms (TSA) should ensure events of LPs to satisfy the restriction of local causality. So PCOM improved the centralized barrier conservative time synchronization mechanism [13, 14, 15].

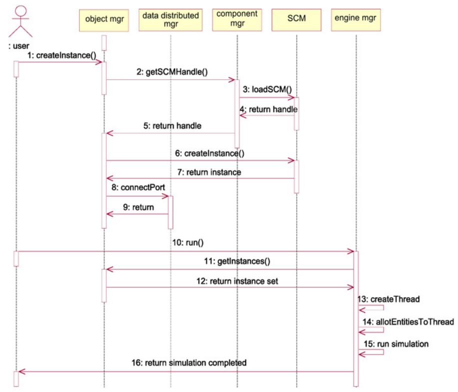


Figure 2. Interactions Between Services of PCOM

During event processing, once received the Time Synchronization Accomplish Signal (TSAS), LP would exit event processing procedure and check whether there was any event's timestamp in the event queue earlier than those dealt ones. If existed, LP would perform withdrawal operation and recover the status of event queue and model to ensure that event queue ordered by timestamp. At last LP would deal with the safe events returned from synchronization algorithm. Computing the LBTS and sending GVT of LP and TSAS were responded by the LP which sent Synchronization Request last. According to the Conservative Time Synchronization algorithm, each LP had a lookahead, which should be the lowest lookahead of its all simulation entities. Similarly, the definition and computing algorithm of LBTS could cite Reference 1.

3.5 Realization of RTI Adapter

Although SCs could be executed parallelizedly based on the multi-core CPU hardware supported by PCOM, to connect with other federate and combine federation, PFHF need to interact with RTI, which supported by RTIAdapter. On one hand, RTIAdapter would call RTI service by RTIAmb, realize RTI callback function, and interact with RTI. On the other hand, RTIAdaper would interact with PCOM by DTTS. So, the RTIAdapter should support the ordinary federate function, i.e. to public / order declaration, register object instance, update attribute, send interaction, reflect attribute, receive interaction, and synchronize time.

RTI would only respond to transfer data, and not for data package / dis-package services, which should be completed by federate. When executing simulation, the data distribution management service would send the UPDATE_ATT_MSG and SEND_INTER_MSG events to the DTTS of RTIAdpater. DTTS would extract the value attributes of event as it was received, which represented the object instance attribute value or the interaction parameters. If received a UPDATE_ATT_MSG event, DTTS would put the value to a AttributeHandleValueMap variable, and call the function of updateAttributeValues of RTI to update attribute. And if received a SEND_INTER_MSG event, DTTS would put the value to a ParameterHandleValueMap variable, and called the sendInteraction function of RTI to send interaction. If it's the first time to update the attribute of object instance, DTTS would register instance to RTI firstly, and update the value. Furthermore, RTIAdapter also responded to harmonize the time management between RTI and PCOM.

4 Performance Monitoring and Dynamic Load Balancing

In the real simulation procedure, the computing resource of each SE wouldn't be steadfast and should take the Dynamic Load Balancing Technique (DLBT) to reduce the execution time and communication response time so as to optimize the computing resource. Furthermore, DLBT was the base of parallelization.

Usually DLBTs include three types. The first, search and determine the imbalance status of simulation system through testing LP's occupied CPU time, SE execution speed [13], CPU load or communication load on each node [17], and so on. The Second, take the centralized, distributed or level load data gathering architecture [17]. The Last, control the LP's CPU slice [18] [19], thread transfer [20] [21], or federate transfer [17] [22] to balance the load. By reference those techniques above, PCOM did some improvement. For the reason that all the LPs of PCOM would be executed in one physical thread space all together, the load data of each LP could only be accessed by its Occupied CPU Time (OCT). To keep load balancing more exactly, except the OCT, the CPU time of each SE to deal with the corresponding simulation event should be tested either. Furthermore, the cost of thread communication among one physical process is too low to regard. So, to PCOM, the Centralized Load Balancing Schedule (CLBS) [13] would be the best choice for the reason that all the SEs and LPs were in the same process space.

To a simulation mission, suppose that T was the execution time without load balancing, and T_b was the execution time with load balancing. Then the change of simulation system performance β was defined as follows:

$$\beta = \left(T - T_b\right) / T$$

if $\beta > 0$, means the performance improved.

if $\beta < 0$, means the performance reduced.

5 Experiments and Analysis

5.1 Experiment Environment

The experiment platform with 8 nodes was composed of series of computers connected by LAN. Each node included Quad-core 3.00 GHz Intel Core CPU and 4GB RAM, and was connected through 100M Ethernet. The operation system was Windows XP, and the compiling environment was Visual Studio 2008. RTI adopted the KD-RTI 1.3NG [13], and the RTI server was deployed on one node. The number of SEs of test data varied from 800 to 1200, and each SE possessed some special computing load. The difference between the minimum load and the maximum was 50 times. And the load of SEs changed obeyed the uniform distribution randomly between the minimum and maximum.

Each federate was controlled both by time and event. The federate adopted time-step to request time advance. The step of ordinary federates were equal, i.e. 1, and the parallel federates hadn't fixed step and would request time advance by event. The lookahead of each federate was 1. Each federate would declare / order the attributes of SEs, i.e. federates would send and receive the SEs' value mutually.

5.2 Analysis on Federate Parallelized Effect

To compare the performance between the ordinary federate and parallel federate, the same simulation system with the same number SEs was executed by the ordinary and parallel federate respectively. The execution time was compared.

To the ordinary federate, we tested whether adding the number of federate would improve the performance of simulation system. And to the parallel federate, we evaluated whether the simulation performance would be affected by adjusting the number of LPs. To compared conveniently, in allusion to the same number of SEs, we constructed the federate with 8, 16, 32, 64 ordinary SEs and tested respectively. During each test, the SEs were distributed to federates equably, and the number of federates on each computing node were same either. Executing several federates on one node meant that distributed SEs to different federates so as to decrease the number of SEs for each federate. According to RTI synchronization, all of the federates could utilize the multi-core CPU's resource to reduce the computing cost of federates.

To the parallel federate, only one federate would be deployed on one node, i.e. no matter how many SEs, there would be 8 parallel federates executing. Yet the number of LPs of each federate could be changed to test whether the degree of parallelization would affect the simulation performance. To compare conveniently, the number of LPs assumed as 1, 2, 4, 8, and the SEs were deployed to each LP equably during initialization procedure without dynamic load balancing.

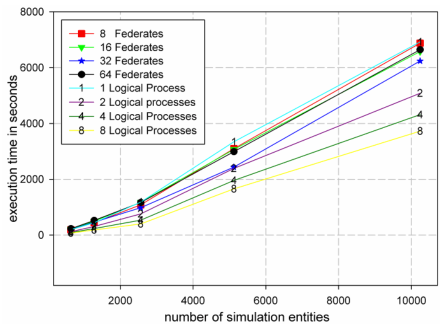


Figure 4. Execution Time for Different Number of Simulation Entities

The result was shown as Fig.4. The x-axis meant the number of SEs, and the y-axis meant the simulation execution time. According to Fig.4, the Simulation Execution Time (SET) increased as the complexity of simulation increasing. To the ordinary federates, lines with different figure symbols, the conclusion could be drawn that no matter the number of SEs was, while the number of SEs was same, the SET wouldn't be decreased evidently by adding federate number to reduce the SEs on each federate because the time advance cost for communication and request among federates increased meanwhile. The four lines in Fig. 4 were close to each other.

To the parallel federates, lines with different number symbols, the conclusion could be drawn as follows:

- To the same number of SEs, SET would decrease as the number of LPs increasing.
- The number of LPs increasing, the more SEs could be disposed and the parallelization degree increased.
- The simulation performance of the parallel federates with 1 LP was close to the ordinary federates, which proved that the additional cost caused by PFHF was very little.
- While the simulation scale was large, the performance would be improved 25% with 2 LPs, 35% with 4 LPs, and 45% with 8 LPs by PFHF.

5.3 Analysis on Load Balanced Effect

The Federate Execution Time (FET) would be compared in these experiments between with/without Dynamic Load Balancing (DLB). During initialization procedure, the SEs were deployed to each LP equably. While simulating, the Load Balancing Algorithm would monitor the Load Difference (LD) among LPs. If the LD was higher than a threshold, , the entities transfer among LPs would be executed to balance the load. How to set was depended on the practical applications and determined as 8 in this test. The effects of DLB with 4 LPs and 8 LPs for each parallel federate were researched respectively. And the number of simulation entities of federates increased from 80 to 1280. The result of 8 LPs was shown as Fig.5. The x-axis meant the number of SEs, the y-axis on left meant the simulation execution time, and the y-axis on right meant the Number of Migrations (NM).

According to Fig.5, when the number of SEs was not large, the LD among LPs was less than , so the DLB procedure was not carried out. Furthermore, we could find that the cost of DLB algorithm was very little by judging from the two lines being near extremely before the inflexion. While the number of SEs increasing, the LD among LPs enhanced either, the DLB carried out and the number of migrations increasing, as shown by the right y-axis, and the SET of federate with DLB was much less than the federate without DLB distinctly, and the NM was about 1 to 300 as shown by the broken line. Then the conclusion could be drawn that the DLB algorithm advanced by our research could improve the performance of parallel federates about 17%.

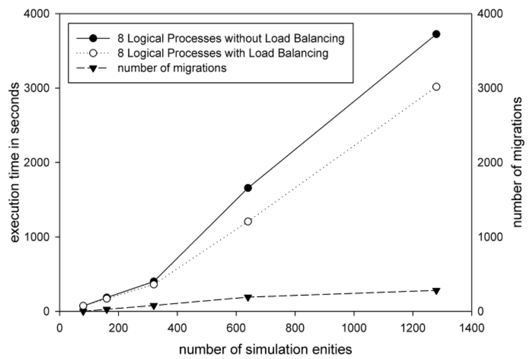


Figure 5. Load Balancing of Parallel Federate with 8LPs

Furthermore, the comprehensive test was designed and executed with large amount of SEs and DLB algorithm. The parallel federates with 8 LPs could improve the simulation system performance about 56%.

6 Conclusions

PFHF, designed and realized by our efforts, would develop simulation federate by component technique and utilizing the multi-core CPU's computing resource to improve federate execution speed. Comprehensive experiments proved that PFHF could increase the performance of simulation systems based on HLA observably and steadily. In our future work, we would analyze the exact time cost of DLB algorithm, keep on optimize the performance of simulation system, and research the smart and effective method to adjust the LD threshold, , adaptively and elaborately.

References

[1] Simulation Interoperability Standards Committee (SISC) of the IEEE Computer Society. IEEE Standard for Modeling and Simulation (M&S) High Level Architecture (HLA)-IEEE Std 1516-2000, 1516.1-2000, 1516.2-2000. New York: Institute of Electrical and Electronics Engineers Inc., 2000.

[2] OpenMP. OpenMP specification. http://openmp.org/wp/openmp-specifications/.

[3] MPI. MPI document. http://www.mpi-forum.org/docs/docs.html.

[4] Microsoft. Task Parallel Library. http://msdn.microsoft.com/en-us/library/dd537609.aspx.

[5] Intel. Intel Threading Building Blocks. http://threadingbuildingblocks.org/documentation.php

[6] Gong JX, Hang C, Qiu XG. Constructing the Extensible HLA Federate Architecture. Journal of System simulation, 2006, 18(11):3126-3130.

[7] Alex Radeski, Shawn Parr, Russell Keith-Magee. Component-Based Development Extensions to HLA.Proc. of the 2002 Spring Simulation Interoperability Workshop. 2002. Paper ID: 02S-SIW-046. Orlando, FL,USA.

[8] Alex R , Shawn P .Towards a Simulation Component Model for HLA. Proc. of the 2002 Fall Simulation Interoperability Workshop. 2002:PaperID:02F-SIW-079. Orlando, Florida,USA.

[9] Jefferson DR. Virtual Time. ACM Transactions on Programming Languages and Systems,1985, 7(3):404-425.

[10] Fujimoto R. 2000. "Parallel and Distributed Simulation Systems." John Wiley& Sons, Inc., 605 Third Avenue, New York, NY 10158-0012.

[11] SPEEDES. SPEEDES Document. http://www.speedes.com/documentation.html. Steinman Jeff, 1991. "SPEEDES: Synchronous Parallel Environment for Emulation and Discrete Event Simulation." In proceedings of Advances in Parallel and Distributed Simulation Pages 95-103. January 1991, Anaheim, California. Society for Computer Simulation Press.

[12] Jeffrey S. S and Douglas R. H. Evolution of the Standard Simulation Architecture. Proc. of the Command and Control Research Technology Symposium. San Deigo, CA, USA: IEEE Press, 2004: papered: 067.

[13] Huang J, Hao JG, Huang KD. The Research and Application of HLA-Based Distributed Simulation Environment KD-HLA. Journal of System Simulation, 2004,16(2):214-220.

[14] Ayani, R. A parallel simulation scheme based on the distance between objects. In Proceedings of the SCS Multiconference on Distributed Simulation 21, 2 (March 1989), pp. 113-118. 18. SCS Press.Tampa, Florida,USA..

[15] Chandy, K.M., and Sherman, R. The conditional event approach to distributed simulation. In Proceedings of the SCS Multiconference on Distributed Simulation 21, 2 (March 1989), pp. 93-99. SCS Press.Tampa, Florida,USA..

[16] Peter L. Reiher , David Jefferson. "Virtual time based dynamic load management in the Time Warp operation system. Transactions of the Society for Computer Simulation International 22,1,1989.

[17] Robson E. De Grande, Azzedine Boukerche. Dynamic balancing of communication and computation load for HLA-based simulations on large-scale distributed systems. J. Parallel Distrib. Comput. 71 (2011) 40–52.

[18] D. Glazer and C. Tropper. On process migration and load balancing in Time Warp. IEEE Trans. On Parallel and Distributed Systems. Vol.4, No.3, March,1993,pp318-327.

[19] Ming-Ru Jiang, Shiuh-Pyng Shieh, Chang-Long Liu. Dynamic Load Balancing in Parallel Simulation Using Time Warp Mechanism. IEEE 1994. pp.222-247.

[20] Glazer, D.W., Tropper, C.: On process migration and load balancing in time warp.IEEE Transactions on Parallel and Distributed Systems 4(3), 318–327 (1993).

[21] Wilson, L.F., Shen,W.: Experiments in load migration and dynamic load balancing in speedes. In: Proc. of the Winter Simulation Conference, pp. 483–490. IEEE Comp. Society, Los Alamitos (1998)

[22] Gan, B.P., Low, Y.H., Jain, S., Turner, S.J., Hsu, W.C.W.J., Huang, S.Y.: Load balancing for conservative simulation on shared memory multiprocessor systems. In: Proc. of the Workshop on Parallel and Distributed Simulation, pp. 139–146.IEEE Comp. Society, Los Alamitos (2000)

Reconstitution of Electromyographic Signals from Pen-Tip Velocity

Inès Chihi
Laboratoire de Recherche en Automatique "LA.R.A",
Ecole Nationale d'Ingénieurs de Tunis,
BP 37, Le Belvédère 1002 Tunis,
Tunisie
E-mail: chihi4ines@hotmail.fr

Afef Abdelkrim
Laboratoire de Recherche en Automatique "LA.R.A",
Ecole Nationale d'Ingénieurs de Tunis,
BP 37, Le Belvédère 1002 Tunis,
Tunisie
E-mail: afef.abdelkrim@esti.rnu.tn

Mohamed Benrejeb
Laboratoire de Recherche en Automatique "LA.R.A",
Ecole Nationale d'Ingénieurs de Tunis,
BP 37, Le Belvédère 1002 Tunis,
Tunisie
E-mail: mohamed.benrejeb@enit.rnu.tn

***Abstract* - This study deals with a new identification approach, based on Recursive Least Squares algorithm (RLS) to reconstruct the electromyographic signals (EMG) of the forearm muscle. The present study uses the relationship between EMG signals and the velocities profiles of the pen-tip moving on (x, y) plane during the human handwriting motion.**

An experimental approach has been carried out to measure the forearm EMG signals and the pen-tip displacements on a digital writing tablet. These measurements are used to predict the electrotromyographic signals of the most active forearm signals during the human handwriting motion. In this research, a new third order, linear model is proposed to identify these muscular activities. Good qualitative and quantitative agreement was found between the proposed model response and the recorded experimental data.

Quantitative agreement was found between traces and trajectories calculated with identified system.

Index Terms - Recursive Least Squares algorithm, EMG signals, velocities profiles, handwriting motion.

I. INTRODUCTION

The detection of EMG signals with powerful and advance methodologies is becoming an important requirement in biomedical engineering. The main reason for the interest in EMG signal analysis is in clinical diagnosis and biomedical applications. These signals provide an important source of information for the diagnosis of neuromuscular disorders and controlling assistive devices like prosthetic/orthotic devices.

Plamondon shows in [1] that the writing velocity is a sequence of bell shapes of different amplitudes and during. Each bell corresponds to a witting primitive and depends on the dimensions and during of this primitive. The important role of the handwriting velocity to control the human handwriting process was proved in [2]. Moreover, this information is taken into account to reconstruct the pen-tip trace.

[3] shows that handwriting motion according to (x, y) plan is mainly based on the interconnection of the electromyographic activity of two forearm muscles and that EMG signals contain the adequate information to present the pen-tip movement in (x, y) plan.

This paper considers structural and parametric identification to model the electomyographic signals from the experimental approach that allows to record the movements of the pen-tip and the EMG signals. This approach is described in the second section. In order to characterize these biological signals, the third section focus on the structural and parametric identification technique. In the last section, a validation step of the proposed model is presented. We finish this research by a conclusion.

II. EXPERIMENTAL APPROACH

Several studies showed that the natural component of the graphic trace corresponds to the spatial movements of the pen tip during the handwriting process, [3], [4] and [5]. These trajectories can be described as movements in tow-dimensional space of the writing plane.

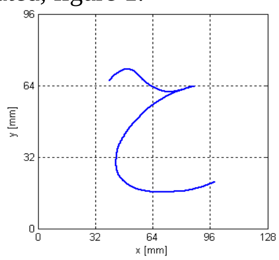
The study presented in [3] has permitted to characterize a handwriting model allowing to generate the written graphic traces from the muscular activities of two major forearm muscles, responsible for controlling the hand during the act of writing. This result allows finding the relationship between the generation of handwriting traces and recorded EMG signals of these muscles.

The measurements of these signals during the writing process require proposing the experimental approach already taken in [4]. This experimental study requires different writers who have written different Arabic letters and some basic geometric forms, [4].

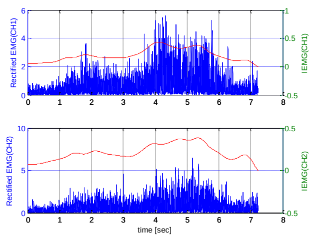
In this experimental approach a measuring system allows to write some Arabic letters and geometric forms on a digital tablet and to memorize the positions of the pen-tip and allows obtaining two electromyographic signals, EMG_1 and EMG_2.

An example of experimental recordings of Arabic letter "HA" is illustrated by figure 2(a). Pen-tip movements according to x and y directions and EMG_1 and EMG_2 recorded during the handwriting process by surface electrodes are illustrated by figure 2(b).

EMG waveforms present transient phenomena and other disruptive signals resulting from various sources, such as electromagnetic phenomenon sector and noise associated with electrodes and uncertainties of measures.
The fluctuation of EMG's magnitudes that corresponding to pen-tip movements can be obtained by using rectifying technique.
Hence, the full wave rectified EMG, called Integrated EMG (IEMG), is calculated, figure 1.



(a)



(b)

Fig.1 Recorded data of the Arabic letter "HA" , [4]
(a) Form, (b) Full wave EMG signals and IEMG signals

III. PROPOSED EMG MODEL

In this section, we propose a new mathematical structure model based on the real data measurements recorded from the experimental approach already presented.
In the literature several recursive and not-recursive algorithms are presented, [6], [7] and [8]. The first type of algorithm has the advantage to be executed on real time and requires less memory resources in terms of calculations.

In this section, the structure of the model is initially proposed. Then, the Recursive Least Square algorithm is used to reconstruct the integrated electromyographic signals recorded during the handwriting process, using velocity's profile V_x and V_y, equation (1).

$$\mathrm{V(k)} = \sqrt{\mathrm{V_x(k)^2 + V_y(k)^2}} \tag{1}$$

with:
V : Pen-tip speed,
V_x,V_y : Pen-tip speeds according to x and y axis respectively.

A. *The proposed mathematical structure*

The new proposed structure allows to reconstruct IEMG signals from the pen-tip velocity on (x, y) plan. This new third order model generates the electromyographic signals IEMG at k instance from the velocities, V_x and V_y at k, k-1, k-2 and k-3 instances and the electromyographic signals at k-1, k-2 and k-3 instances, equations (4) and (5) . The proposed structure is presented as follows:

$$e_1(k) = \sum_{i=1}^{4} -\left[\hat{a}_{i1}(k)\, e_1(k-i) + \hat{b}_{i1}(k) e_2 V_{xe}(k-i)\right] + \left[\hat{c}_{i1}(k) V_x(k-i+1) + \hat{d}_{i1}(k) V_y(k-i+1)\right] \tag{2}$$

$$e_2(k) = \sum_{i=1}^{4} -\left[\hat{a}_{i2}(k)\, e_2(k-i) + \hat{b}_{i2}(k) e_1 V_{xe}(k-i)\right] + \left[\hat{c}_{i2}(k) V_x(k-i+1) + \hat{d}_{i2}(k) V_y(k-i+1)\right] \tag{3}$$

with :

e_1	: EMG signal of the first forearm muscle,
e_2	: EMG signal of the second forearm,
V_x, V_y	: Velocities according to x and y movements respectively,
a_{i1}, b_{i1}, c_{i1}, d_{i1}	: Estimated velocitie V_{xe} parameters,
a_{i2}, b_{i2}, c_{i2}, d_{i2}	: Estimated velocitie V_{ye} parameters,
k	: Discrete time.

B. *Parametric identification*

The parametric identification is an experimental approach of determining the parameters of a mathematical model of a given process to generate output response as close as possible to the real system. This technique is based on the prediction error, which is the error between the output of the process and the output predicted by the model. In order to minimize the prediction error, the model parameters are modified at each sampling time by using an algorithm estimating the parameters.

In order to identify the IEMG signals recorded during the handwriting process, the classical Recursive Least Squares (RLS) algorithm with constant forgetting factor in the particular case of linear systems is used. RLS is one of the fastest converging adaptive filtering algorithms. It's is used to identify and model dynamic systems from experimental data. This algorithm helps to get a sequential processing of experimental data available at each time.
The structure of such identification technique is presented by the equations (4) to (6), [6].

$$\hat{\theta}(k) = \hat{\theta}(k-1) + P(k) \sum_{i=n+1}^{k} y(i)\, \Psi(i) \tag{4}$$

$$P(k) = P(k-1) - \frac{P(k-1)\,\Psi(k)\,\Psi^T(k)\,P(k-1)}{1 + \Psi^T(k)\,P(k-1)\,\Psi(k)} \tag{5}$$

$$\varepsilon(k) = y(k) - \hat{\theta}(k-1)\,\Psi(k) \qquad (6)$$

where:

$\hat{\theta}(k)$: Vector of estimated parameters,

$P(k)$: Adaptation matrix,

$\varepsilon(k)$: Estimated error,

$y(k)$: Actual output of the system to identify,

$\psi(k)$: Observation matrix,

k : Discrete time.

The estimated electromyographic signals can be resumed by the equations (7) and (10) .

$$e_1 = \psi_1^T \theta_1 + \varepsilon_1 \qquad (7)$$

$$e_2 = \psi_2^T \theta_2 + \varepsilon_2 \qquad (8)$$

ε_1 and ε_2 : Error vectors, relative to e_1 and e_2 signals respectively,

ψ_1^T and ψ_2^T : Matrices which elements are the delayed inputs and outputs components, relative to e_1 and e_2 signals respectively.

$$\hat{\psi}_1^T(k) = \begin{bmatrix} e_2(k) & e_2(k-1) & e_2(k-2) & e_2(k-3) \\ e_1(k) & e_1(k-1) & e_1(k-2) & e_1(k-3) \\ -V_{xe}(k-1) & -V_{xe}(k-2) & -V_{xe}(k-3) & -V_{xe}(k-4) \\ -V_{ye}(k-1) & -V_{ye}(k-2) & -V_{ye}(k-3) & -V_{ye}(k-4) \end{bmatrix} \qquad (9)$$

$$\hat{\psi}_2^T(k) = \begin{bmatrix} e_1(k) & e_1(k-1) & e_1(k-2) & e_1(k-3) \\ e_2(k) & e_2(k-1) & e_2(k-2) & e_2(k-3) \\ -V_{xe}(k-1) & -V_{xe}(k-2) & -V_{xe}(k-3) & -V_{xe}(k-4) \\ -V_{ye}(k-1) & -V_{ye}(k-2) & -V_{ye}(k-3) & -V_{ye}(k-4) \end{bmatrix} \qquad (10)$$

Variations of parameters in time show that they converge to constant values, figures 3 and 4.
Using the estimated parameter's values, the new proposed model is presented by equations (11) and (12).

$$e_1(k) = \sum_{i=1}^{4} -\left[\hat{a}_{i1}\, e_1(k-i) + \hat{b}_{i1} e_2\, V_{xe}(k-i)\right] + \left[\hat{c}_{i1}\, V_x(k-i+1) + \hat{d}_{i1}\, V_y(k-i+1)\right] \qquad (11)$$

$$e_2(k) = \sum_{i=1}^{4} -\left[\hat{a}_{i2}\, e_2(k-i) + \hat{b}_{i2} e_1\, V_{xe}(k-i)\right] + \left[\hat{c}_{i2}\, V_x(k-i+1) + \hat{d}_{i2}\, V_y(k-i+1)\right] \qquad (12)$$

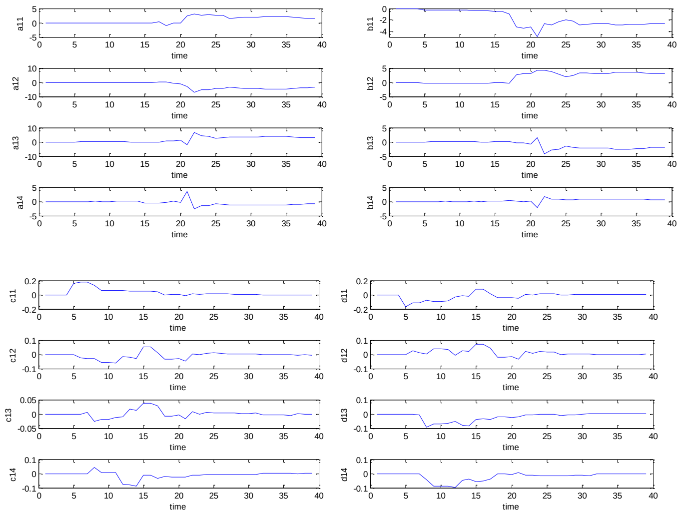


Fig.2 Parameter's evolutions relative to the estimated IEMG$_1$ signals

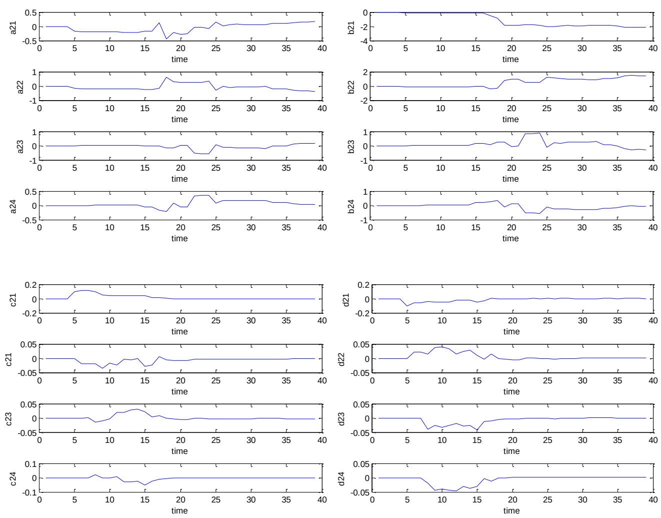


Fig.3 Parameter's evolutions relative to the estimated IEMG$_1$ signals

The estimated parameters of velocities vectors $\hat{\theta}_1$ and $\hat{\theta}_2$ accoding to e_1 and e_2 signals, respectively, will take the following forms:

$$\hat{\theta}_1 = \left[\hat{a}_{11}\ \hat{a}_{12}\ \hat{a}_{12}\ \hat{a}_{14}\ \ \hat{b}_{11}\ \hat{b}_{12}\ \hat{b}_{13}\ \hat{b}_{14}\ \ \hat{c}_{11}\ \hat{c}_{12}\ \hat{c}_{13}\ \hat{c}_{14}\ \hat{c}_{15}\ \ \hat{d}_{11}\ \hat{d}_{12}\ \hat{d}_{13}\ \hat{d}_{14}\ \hat{d}_{15}\right] \qquad (13)$$

$$\hat{\theta}_2 = \left[\hat{a}_{21}\ \hat{a}_{22}\ \hat{a}_{23}\ \hat{a}_{24}\ \ \hat{b}_{21}\ \hat{b}_{22}\ \hat{b}_{23}\ \hat{b}_{24}\ \ \hat{c}_{21}\ \hat{c}_{22}\ \hat{c}_{23}\ \hat{c}_{24}\ \hat{c}_{25}\ \ \hat{d}_{21}\ \hat{d}_{22}\ \hat{d}_{23}\ \hat{d}_{24}\ \hat{d}_{25}\right] \qquad (14)$$

with:

$\hat{a}_{1i},\ \hat{b}_{1i},\ \hat{c}_{1i},\ \hat{d}_{1i},$: Estimated parameters relative to the estimated e_1 signal,

$\hat{a}_{2i,}\ \hat{b}_{2i,}\ \hat{c}_{2i,}\ \hat{d}_{2i,}$: Estimated parameters relative to the estimated e_2 signal,

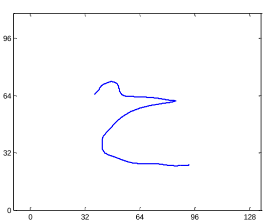
i : 1, 2, 3 or 4.

RESULTS

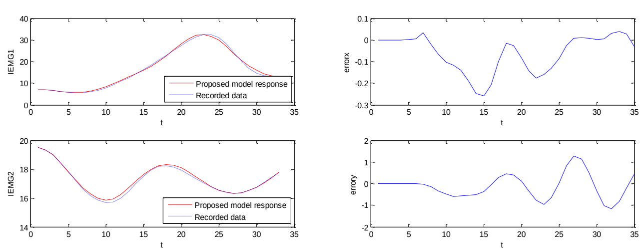
A. *Identification's results*

Comparisons between the recorded IEMG and the reconstructed integrated electromyographic signals with the same starting point are shown in figures 4 and 5. Indeed, this figure illustrates some identification's results for the Arabic letters "HA" and "SIN", written by different writers. For the first shape, an important correspondence is shown between the proposed model's response and the recorded data. A refinement should be considered for the second letter "SIN" which represents more curves.

The solid line presents the recorded experimental data and the dotted line is relative to the answer given by the estimated parameters of the obtained model.



(a)



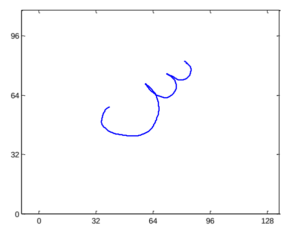
(b) (c)

Fig.4 Results of the identification of the proposed model (a) Form (Arabic letter "HA"), (b) IEMG signals according to x and y directions, (c) Errors

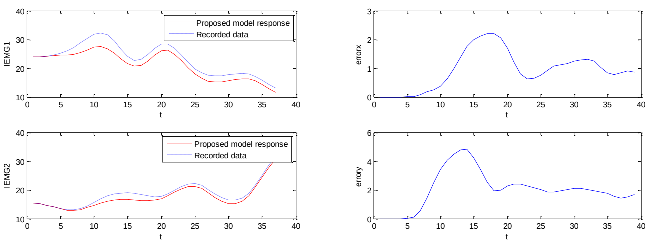
A. *Validation's results*

The parametric identification of handwriting process is completed by a validation step of the proposed model. Different tests are carried out to retain or reject the proposed structure. We propose two types of validation, the first in the case of monowriter and the second in the case of multiwriter.

The first one is a validation of the proposed model in the case the integrated in monowriter case. This means, integrating parameters of a model characterizing IEMG signals with data saved from another example of this signals characterizing another example of the same type of drawing trace and the same writer.



(a)



(b) (c)

Fig.5 Results of the identification of the proposed model (b) Form (Arabic letter "SIN"), (b) IEMG signals according to x and y directions (c) Errors

The second is a prediction of the integrated electromyographic signals in the case of in multiwriter case. We applied the experimental recorded input data of IEMG signals for a first writer with parameters of another model characterizing other IEMG signals of the same kind of shape but of a second writer. The

Monowriter case: Recorded data relative to one model of a graphic trace are applied to different models for the same writer and the same type of the trace.

Figure 6 shows the predicted results $IEMG_1$ and $IEMG_2$, and the real recorded data in monowriter validation. This figure shows compatibility between the predicted data and the experimental curves. Dotted lines denote recorded results and solid lines represent the proposed model response. Consequently, for many types of Arabic letters and geometric drawing by the same writer, our proposed model confirms the possibility of direct prediction of the forearm muscle activities, IEMG signals.

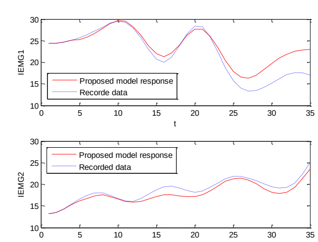


Fig.6 Monowriter validation:
Predicted IEMG signals of the Arabic letter "SIN"

Multiwriter case: In order to propose one model for signals allowing to the same types of trace but for two different writers. This is obtained by applying velocities data, V_x and V_y, recorded for a trace written by writer1 into an IEMG model belonging to the same type of drawing relative to a writer2. Simulation's result of the multiwriter validation is illustrated in figure 7 that shows an important error between the proposed IEMG model and the real recorded signals.

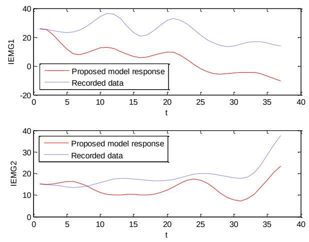


Fig.7 Multiwriter validation:
Predicted IEMG signals of the Arabic letter "SIN"

CONCLUSION

In this study, the IEMG signals recorded during the human handwriting process is proposed by using structural and parametric identification technique based on the Recursive Least Square algorithm. The obtained model uses mainly the relationship between the electromyographic signals of the forearm muscles and the velocity profile of the pen-tip moving in (x, y) plane.

According to the results shown in this study, good concordnce is observed in the case of the IEMG signals'prediction belonging to th same writer and the same shape. In order to reconstruct the electromyographic signals representing a specific type of shape, an important error is observed in multiwriter case.

The proposed model can reconstruct the biological waves for several Arabic letters and shapes written by the same writer.

An improvement of the predicted model results must be made not only in the monowriter case but also in the multiwriter case.

REFERENCES

[1] R. Plamondon, A kinematics theory of rapid human movements. Part II: Movement time and control, Biological Cybernetics, 1995.

[2] R. Plamondon, Alimi M.A., Yergeau P., Leclerc F., Modelling velocity profiles of rapid movements: a comparative study, Biological Cybernetics, vol. 69, 1993, pp. 119-128.

[3] I.Chihi, C. Ghorbel, A. Abdelkrim and M. Benrejeb, Parametric identification of handwriting system based on RLS algorithm, ICCAS, Seoul, Corea, Oct, 2011.

[4] M. Sano, T. Kosaku and Y. Murata, Modelling of Human Handwriting Motion by Electromyographic Signals on Forearm Muscles, CCCT'03, Orlando-Florida, 2003.

[5] M. Benrejeb, A. El Abed-Abdelkrim and M. Sano, 2006. Sur l'étude du processus d'écriture à la main. Approches classiques et non conventionelles, Sciences et technologies de l'Automatique (e-sta) , 2006, Vol. 3, N°1.

[6] T. Tutunji, M. Molhim, E. Turki, Mechatronic systems identification, using an impulse response recursive algorithm, ScienceDirect, Simulation Modeling Practice and Theory, 2007, pp 970-988.

[7] W. Apley, and J. Shi. An order downdating algorithm for tracking system order and parameters in Recursive Least Squares Identification. IEEE Trans. Signal Proces, 1999

[8] T. Tutunji, , M. Molhim and E. Turki. Mechatronic systems identificatio using an impulse response recursive algorithm. Simulation Modeling Practice Theory, 2007.

Validation of Swarms of Robots: Theory and Experimental Results

Luis Mendez¶, Sidney N. Givigi*, Howard M. Schwartz¶, Alain Beaulieu*, Gerard Pieris†, Giovanni Fusina‡
¶Department of Systems and Computer Engineering
Carleton University
Ottawa, ON. Canada
Email: lmendez@sce.carleton.ca, schwartz@sce.carleton.ca
*Department of Electrical and Computer Engineering
Royal Military College of Canada
Kingston, ON. Canada
Email: sidney.givigi@rmc.ca, alain.beaulieu@rmc.ca
†Defence Research and Development Canada-Ottawa
Ottawa, ON. Canada
Email: Gerard.Pieris@drdc-rddc.gc.ca
‡Defence Research and Development Canada-Toronto
Toronto, ON. Canada
Email: Giovanni.Fusina@drdc-rddc.gc.ca

***Abstract*—In this paper we present a formulation of a swarm algorithm for some predefined formations and we show that using the control laws derived, the whole system is stable. Furthermore, the algorithm for the formation is implemented in microcontrollers, which effectively makes the experiment a system of systems. The only input for the algorithm is the position of the robot and the relative positions of neighbours within two metres of the robot. The results of experiments presented show that the algorithm presents the same general behaviour as predicted in simulations. However, due to noise in measurements and commands, the behaviour is slightly different.**

***Index Terms*—swarm, flocking, formation, distributed algorithm, experimental validation, UAV, autonomous systems.**

I. Introduction

Swarming and flocking are collective behaviours that can be observed in several different species of animals, from insects to birds and from fish to mammals. Biologists have analyzed these behaviours for several decades already [1], [2]. These behaviours have also generated analogies in physics [3], [4], mathematics [5] and computer science [6].

In engineering, there are several applications that mimic the swarming or flocking behaviours. One of the most important nowadays is the use of swarming and flocking algorithms for controlling the movement of autonomous Unmanned Aerial Vehicles (UAVs) [7] in diverse missions such as, for example, search and rescue. The vehicles are supposed to be independent and their control to be decentralized. Also, the formation is assumed to be due to self-organization, therefore ruling out any type of centralized control.

Some works have dealt with the problem of stability and convergence of flocking algorithms [8]. Most of these approaches focus on the analysis based on Lyapunov functions. However, few deal with the linearization of the system [9]. In any case, experiments may be necessary to assess the stability of the system with the control law derived for the swarm.

Relatively few works have dealt with the implementation of a decentralized self-organizing algorithm in actual robots/vehicles [10]. In these cases, one needs to consider problems common in other applications in engineering, such as filtering, estimation and low-level feedback control.

This paper presents a very important formation for linear flocking. In several applications, such as surveillance, robots need to move in a linear formation in order to sweep an area. It is also desirable that the robots count only on local information and decentralized processing, avoiding possible problems with dropped communications. The algorithm with the control law for flocking is then shown to be stable. Finally, the system is implemented in actual physical robots. The algorithm was initially presented in [11], but was only shown to work in simulation and no proof of stability was provided. This paper also extended the results presented in [12].

The paper is divided as follows. Section II presents the flocking algorithm as introduced in [11]. Section III contains the mathematical analysis of the control algorithm and its proof of stability. In section IV we present the model and constraints of the robots used in the reported experiment. In section V we introduce the experimental setup for the experiment described in section VI. Finally, we discuss our conclusions in section VII.

II. Flocking Algorithm

The flocking formation we are interested in this paper is called in [11] "*Straight-Line Formation*". In this formation, the flock needs to move in a line and the relative distance between neighbours is to be kept approximately constant. In other words, the robots self-organize in a straight line, at the same distance from each other, at a specified altitude, at a specified angle with respect to their common direction of flight. Figure 1 shows this formation. In this figure, the robots are represented by small triangles and the vector $\overrightarrow{v_0}$

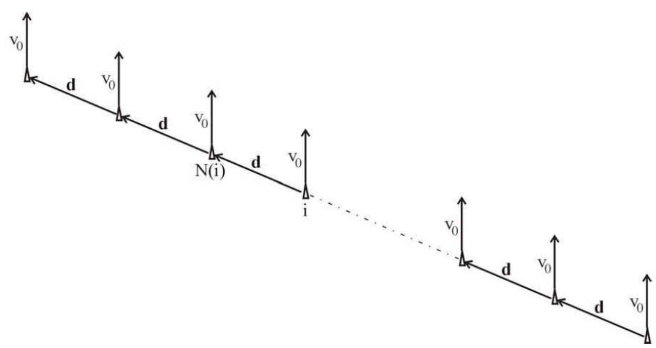


Fig. 1. Depiction of robots in straight line formation

corresponds to their velocity. Each robot, except the first one on the line, has a neighbour situated at position $\overrightarrow{d}$ with respect to itself [11, p.41].

Each robot in the formation has the following dynamics:

$$\dot{\bar{x}}_i = \bar{v}_i \quad (1)$$

$$\dot{\bar{v}}_i = \bar{u}_i \quad (2)$$

where x_i represents the position, v_i represents the velocity and u_i represents the acceleration of the i^{th} robot. As discussed in [8], this model is sufficient for the study of most swarming problems.

As described in [12], the control signal produced by the algorithm can be divided in two terms as:

$$\overrightarrow{u_i} = \overrightarrow{u_i}^F + \overrightarrow{u_i}^N \quad (3)$$

where $\overrightarrow{u_i}^F$ is related to the positioning of the i^{th} robot with respect to other robots (α-lattice forming term) and $\overrightarrow{u_i}^N$ is the navigational feedback term. We will discuss each one of the forming control signals in the next subsections.

A. α-lattice term

The α-lattice forming term of Equation (3) can be broken down into two different parts

$$\overrightarrow{u_i}^F = \overrightarrow{u_i}^F(r) + \overrightarrow{u_i}^F_d \quad (4)$$

The first part is an attraction or repulsion force exerted by all robots locally perceived by the i^{th} robot and can be written as

$$\overrightarrow{u_i}^F(r) = \sum_j \overrightarrow{u_j}^F(r)(\overrightarrow{x_i x_j}) \quad (5)$$

where j represents all the robots except robot i; r is the distance between robots i and j; and $\overrightarrow{x_i x_j}$ is the unit vector connecting robots i and j. Also, each element of the sum in equation (5) is given by

$$\overrightarrow{u_j}^F(r) = \begin{cases} A[1 - r/a] & \text{when } 0 \leq r \leq b \\ -B & \text{when } b < r \end{cases} \quad (6)$$

where, again, r is the distance from robot i to robot j; the constants A and B represent the magnitude of the repulsive and attractive forces respectively; a is the radius at which the force changes from repulsive to attractive; and b is the point of transition from the quadratic to the linear form of the potential.

In order to define the term $\overrightarrow{u_i}^F_d$ in equation (4) it is necessary to define $\mathbf{N}(i)$, the neighborhood of robot i. By definition, $\mathbf{N}(i)$ contains the closest robot that is in the closed half space $\mathbf{E}(i)$ defined as:

$$\mathbf{E}(i) = \{\bar{x} | \overrightarrow{d} \cdot (\bar{x} - \bar{x}_i) \geq 0\} \quad (7)$$

Given definition (7), $\overrightarrow{u_i}^F_d$ is given by:

$$\overrightarrow{u_i}^F_d = c_d \begin{cases} (\overrightarrow{x} - \overrightarrow{x_i} - \overrightarrow{d}) & \text{for N(i) not empty} \\ [0,0,0]^T & \text{for N(i) empty} \end{cases} \quad (8)$$

Note that c_d can be seen as a proportional gain related to the error in position for the i^{th} robot.

B. Navigational term

The navigational feedback term of the control signal is responsible for keeping the robots moving in a prescribed direction. The signal is given by

$$\overrightarrow{u_i}^N = c_v(\overrightarrow{v_0} - \overrightarrow{v_i}) + c_h(h_0 - z_i)\hat{k} \quad (9)$$

where c_v is a constant; $\overrightarrow{v_0}$ is the desired velocity vector; $\overrightarrow{v_i}$ is the current velocity vector of the i^{th} agent; c_h is a constant; h_0 is the desired altitude of the robot; z_i is the current position of the i^{th} robot in the Z-plane; and $\hat{k}$ is the unit vector in the Z-direction (the altitude).

Note that c_v can be seen as a derivative gain and c_h as a proportional gain for the navigational error of the i^{th} robot.

With all the signals defined, we can proceed to the analysis of the stability of the control algorithm.

III. Theoretical convergence

In order to show that the algorithm converges to a stable formation, we will linearize the system for each one of the robots around the equilibrium points and show that all the eigenvalues are negative, meaning that the states will converge to the desired behaviour. As such, we will be able to conclude that once around the stable behaviour, the system will not diverge.

This section will be broken in two subsections. Subsection III-A will deal with the linearization of the states when the robot is close to its equilibrium. Subsection III-B deals with the convergence of the algorithm to the desired behaviour.

A. Linear representation for states close to equilibrium

Let us start the analysis by considering first the navigational force $\overrightarrow{u_i}^N$. Close to equilibrium $\overrightarrow{v_0} \approx \overrightarrow{v_i}$; therefore, we can write the navigational control signal (9) as

$$\overrightarrow{u_i}^N = c_v \begin{bmatrix} v_{x0} - (v_{x0} + \Delta v_{xi}) \\ v_{y0} - (v_{y0} + \Delta v_{y_i}) \\ v_{z0} - (v_{z0} + \Delta v_{zi}) \end{bmatrix} + c_h \begin{bmatrix} 0 \\ 0 \\ z_0 - (z_0 + \Delta z_i) \end{bmatrix}$$
$$\overrightarrow{u_i}^N = \begin{bmatrix} -c_v \Delta v_{xi} \\ -c_v \Delta v_{y_i} \\ -c_v \Delta v_{zi} - c_h \Delta z_i \end{bmatrix} \quad (10)$$

We can write equation (10) in matrix form as

$$\overrightarrow{u_i}^N = \begin{bmatrix} 0 & 0 & 0 & -c_v & 0 & 0 \\ 0 & 0 & 0 & 0 & -c_v & 0 \\ 0 & 0 & -c_h & 0 & 0 & -c_v \end{bmatrix} \begin{bmatrix} \Delta x_i \\ \Delta y_i \\ \Delta z_i \\ \Delta v_{xi} \\ \Delta v_{y_i} \\ \Delta v_{zi} \end{bmatrix} \quad (11)$$

Now let us consider the force acting on the i^{th} robot due to the robot in its neighborhood (as defined by equation (8)) when close to equilibrium. Recall that this force is given by

$$\overrightarrow{u_{i}}_d^F = c_d \begin{cases} (\overrightarrow{x} - \overrightarrow{x_i} - \overrightarrow{d}) & \text{for N(i) not empty} \\ [0, 0, 0]^T & \text{for N(i) empty} \end{cases}$$

If a robot is located very close to a desired equilibrium state, its position and the position of its neighbour robot is depicted by figure 2.

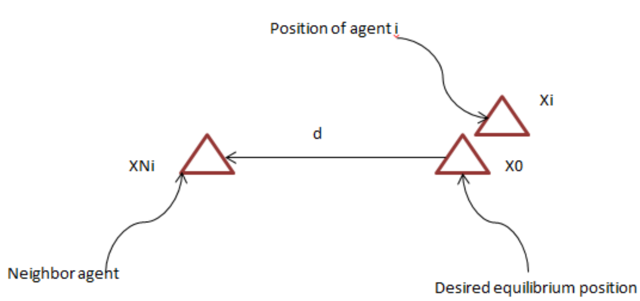


Fig. 2. Robot located close to equilibrium and its neighbour robot

At equilibrium the position of the neighbour robot is $\overrightarrow{x_{Ni}} = \overrightarrow{x_i} + \overrightarrow{d}$; therefore, equation (8) can be written as

$$\overrightarrow{u_{i}}_d^F = c_d [\begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} + \begin{bmatrix} d_x \\ d_y \\ d_z \end{bmatrix} - \begin{bmatrix} x_0 + \Delta x \\ y_0 + \Delta y \\ z_0 + \Delta z \end{bmatrix} - \begin{bmatrix} d_x \\ d_y \\ d_z \end{bmatrix}]$$
$$\overrightarrow{u_{i}}_d^F = c_d \begin{bmatrix} -\Delta x \\ -\Delta y \\ -\Delta z \end{bmatrix} \quad (12)$$

We can write equation (12) in matrix form as

$$\overrightarrow{u_{i}}_d^F = \begin{bmatrix} -c_d & 0 & 0 & 0 & 0 & 0 \\ 0 & -c_d & 0 & 0 & 0 & 0 \\ 0 & 0 & -c_d & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \\ \Delta v_x \\ \Delta v_y \\ \Delta v_z \end{bmatrix} \quad (13)$$

The last term of the algorithm that needs to be considered is equation (6)

$$\overrightarrow{u_j}^F(r) = \begin{cases} A[1 - r/a] & \text{when } 0 \leq r \leq b \\ -B & \text{when b} < r \end{cases}$$

Since we are assuming the i^{th} robot is close to the equilibrium, the magnitude of the displacement vector $\overrightarrow{d}$ is greater than the value of parameter b (the robots cannot be at the repulsion region), therefore, the force that acts on robot i due to all other robots in the formation will be given by the second case of equation (6), that is, $-B$.

Without loss of generality, we may choose the vector $\overrightarrow{d}$ to have only one nonzero component such that

$$\overrightarrow{d} = \begin{bmatrix} d_x \\ 0 \\ 0 \end{bmatrix} \quad (14)$$

Then, close to equilibrium the distance between two robots is $r_{ij} \approx d_x$. Therefore, the attractive force can be written as:

$$\overrightarrow{u_i}^F(r) = -B \begin{bmatrix} ((x_0 + d_x) - (x_0 + \Delta x))/r_{ij} \\ ((y_0) - (y_0 + \Delta y))/r_{ij} \\ ((z_0) - (z_0 + \Delta z))/r_{ij} \end{bmatrix}$$
$$\overrightarrow{u_i}^F(r) = -B \begin{bmatrix} 1 - \Delta x/r_{ij} \\ -\Delta y/r_{ij} \\ -\Delta z/r_{ij} \end{bmatrix}$$
$$\overrightarrow{u_i}^F(r) = \begin{bmatrix} -B + B\Delta x/d_x \\ B\Delta y/d_x \\ B\Delta z/d_x \end{bmatrix} \quad (15)$$

Writing equation (15) in matrix form we obtain

$$\overrightarrow{u_i}^F(r) = \begin{bmatrix} B/d_x & 0 & 0 & 0 & 0 & 0 \\ 0 & B/d_x & 0 & 0 & 0 & 0 \\ 0 & 0 & B/d_x & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \\ \Delta v_x \\ \Delta v_y \\ \Delta v_z \end{bmatrix} \quad (16)$$

We can now proceed to the analysis of convergence and stability for the flocking system described in section II.

B. Convergence

Around the equilibrium, in general, any robot system may be approximated by the linear equation of the form

$$\dot{\bar{x}}_i = A\bar{x}_i + B\bar{u}_i \tag{17}$$

In our case, since equations (1) and (2) are already linear, we have

$$A = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \; and \; B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{18}$$

The control signals in equations (11), (13) and (16) can be combined, resulting in

$$\overrightarrow{u_i} = K \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \\ \Delta v_x \\ \Delta v_y \\ \Delta v_z \end{bmatrix} \tag{19}$$

where

$$K = \begin{bmatrix} \frac{B}{r} - c_d & 0 & 0 & -c_v & 0 & 0 \\ 0 & \frac{B}{r} - c_d & 0 & 0 & -c_v & 0 \\ 0 & 0 & \frac{B}{r} - c_d - c_h & 0 & 0 & -c_v \end{bmatrix} \tag{20}$$

Applying the control signal (19) to the system in (17) close to the equilibrium, one gets that the system is then described by the matrix

$$A' = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ \frac{B}{r} - c_d & 0 & 0 & -c_v & 0 & 0 \\ 0 & \frac{B}{r} - c_d & 0 & 0 & -c_v & 0 \\ 0 & 0 & \frac{B}{r} - c_d - c_h & 0 & 0 & -c_v \end{bmatrix} \tag{21}$$

Calculating the eigenvalues of (21), we get

$$\lambda(A') = \begin{bmatrix} \frac{1}{2}\frac{-rc_v + \sqrt{r^2c_v^2 + 4Br - 4r^2c_d - 4c_hr^2}}{r} \\ -\frac{1}{2}\frac{rc_v + \sqrt{r^2c_v^2 + 4Br - 4r^2c_d - 4c_hr^2}}{r} \\ \frac{1}{2}\frac{-rc_v + \sqrt{r^2c_v^2 + 4Br - 4r^2c_d}}{r} \\ -\frac{1}{2}\frac{rc_v + \sqrt{r^2c_v^2 + 4Br - 4r^2c_d}}{r} \\ \frac{1}{2}\frac{-rc_v + \sqrt{r^2c_v^2 + 4Br - 4r^2c_d}}{r} \\ -\frac{1}{2}\frac{rc_v + \sqrt{r^2c_v^2 + 4Br - 4r^2c_d}}{r} \end{bmatrix} \tag{22}$$

From equation (22) we note that only two eigenvalues $\lambda(A')$ are dependent on c_h. Also, since r and B are constants, we may easily find the range of c_d and c_v for stability to be $c_d > 0$ and $c_v > 0$. Therefore, by using the control laws defined in equations (3), (4) and (9), the system will converge to the desired behaviour. Lastly, notice that even if the system is nonlinear, the same rational of this section can be applied to the linearized system and the system may be shown to be stable close to equilibrium. However, we do not guarantee that the system will converge to equilibrium if the initial conditions are far from the desired equilibrium.

In the next sections we will report an experiment that shows that the system with the control laws presented is indeed stable for ground robots.

IV. Experimental Platform

The algorithm presented in section II was ported to a ground robotic platform. These physical robots follow the differential drive robot model, i.e., the turning radius is more responsive and tighter than other drive models [13]. The kinematic difference equations that represent the robot's model are [14].

$$\begin{aligned} x(k+1) &= x(k) + D(k) * cos(\Theta(k) + \frac{\Delta\Theta(k)}{2}) \\ y(k+1) &= y(k) + D(k) sin(\Theta(k) + \frac{\Delta\Theta(k)}{2}) \\ \Theta(k+1) &= \Theta(k) + \Delta\Theta(k) \end{aligned} \tag{23}$$

where $x(k)$ and $y(k)$ are the cartesian coordinates of the robot's position; $D(k)$ is the distance traveled between time steps k and $k+1$; $\Theta(k)$ is the angle between the vehicle and the x axis; $\Delta\Theta(k)$ is the rotation angle between time step k and $k+1$.

The wheel speeds that describe the robot's movement will be under the influence of errors that are not completely known. Critical unknown errors are the drive wheel circumference speed measurement for each wheel ($\varepsilon_L, \varepsilon_R$) and the axle length ($\varepsilon_b$).

$$\begin{aligned} v_t(k) &= \frac{(\omega_L(k)R + \varepsilon_L) + (\omega_R(k)R + \varepsilon_R)}{2} \\ \omega(k) &= \frac{(\omega_R(k)R + \varepsilon_R) - (\omega_L(k)R + \varepsilon_L)}{L + \varepsilon_b} \end{aligned} \tag{24}$$

where $v_t(k)$ is the robot's translational speed; $\omega(k)$ is the robot's rotational speed; R is the radius of the two drive wheels; and L is the axle length.

The longitudinal velocity $v_t(k)$ and angular speed $\omega(k)$ are then used to calculate the inputs $D(k)$ and $\Delta\Theta(k)$ of equation (23) such that

$$\begin{aligned} D(k) &= v_t(k)\Delta t \\ \Delta\Theta(k) &= \omega(k)\Delta t \end{aligned} \tag{25}$$

where Δt is the time step used in the control loop of the system.

Notice that even though this system is different from equations (1) and (2), in a small area around the equilibrium, the system can be easily linearized.

V. Experimental Environment

In order for the robots to perform the flocking algorithm described in section II, some information is required, namely, the robot needs to know its distance to its neighbours.

The approach we chose to use is based on a localization system. In our approach, the localization system is able to acquire the robot's own position as well as the positions of neighbours, robots within a two-metre radius. This mimics the use of a sensor such as a camera with limited sensing radius. This forces the robots to work with only local information and the conditions of the algorithm in section II are satisfied.

Also, the position measurements returned from the localization system are noisy. This is not a problem when we are calculating the inter-robots acceleration signal $\overrightarrow{u}_i^F$ of equation (3), but becomes a major issue when calculating the navigational acceleration $\overrightarrow{u}_i^N$. In order to minimize the adverse behaviours, we made use of an Extended Kalman Filter to calculate the orientation of the robot. The controller structure is shown in Figure (3).

The *Kalman Filter* block, as already mentioned, is used to get rid of the noise in the measurements provided by the *Localization System* block and it runs at a rate of $5\ Hz$.

The only information provided by the *Localization System* block is the position (x_i, y_i) of the i^{th} robot as well as the distances and angles from the i^{th} robot to all its neighbours. Let us now consider the neighbouring set $\mathbf{N}_i$ as described in section II. For all $j \in \mathbf{N}_i$, the *Localization System* block sends the i^{th} robot the data (d_{ij}, α_{ij}), where d_{ij} is the distance from robot i to robot j and α_{ij} is the angle between the robots. With this information, the robot can calculate right away the inter-robots acceleration signal $\overrightarrow{u}_i^F$.

As for the navigational acceleration $\overrightarrow{u}_i^N$, the robot needs to estimate what is its orientation angle $\Theta(t)$. This information is not sent to the robot, but, using the *Kalman Filter*, the robot can estimate the heading angle based on previous measurements.

All this observed information is then passed on to the algorithm described in section II and the acceleration signals $\overrightarrow{u}_i^F$ and $\overrightarrow{u}_i^N$ are generated. These signals are then passed on to a PID controller that runs at a much higher rate of $25\ Hz$ to keep the robot running at the desired speed and orientation. At the next measurement step a new position is passed on to the *Kalman Filter* and the process starts again.

VI. Experiment

The conducted experiments were devised in order to show that the algorithm of Section II could be ported to a physical platform such as the one introduced in Section IV and with the sensorial capabilities of the experimental facility described in Section V would lead to the same behavior already shown to work in simulations [11].

The robots started at known positions and the only restriction was that each robot had at least one other robot in its neighbourhood. This condition is necessary so the algorithm can work from the start. Otherwise, given the limited space and time for the experiment, some robots could never find a neighbour and the flocking algorithm would not produce the desired response. An alternative would be for a robot to wander around until it finds another one and then start moving towards the flocking zone. However, this would be impractical in our experimental setup, as it would lead to possible collisions.

The robots start at the left of the environment and they feel a navigational acceleration in the direction of the right of the play area. The desired speed of the "flow" (equation (9)) is set to $\overrightarrow{v}_0 = 25\,cm/s$ and the navigational gain is set to $c_v = 1.0$.

When robots are in the neighborhood, a "swarming" signal is generated as described in (5). In order to avoid overshoot in the response, we chose $c_d = 0.1$. This will give us only real eigenvalues. Moreover, the separation distance $\overrightarrow{d}$ is chosen so the robots do not hit each other or the walls. Therefore, we select $\| \overrightarrow{d} \| = 100\,cm$.

With all these parameters defined, we can now describe the experiment. As already said, the robots start at known poses (x_i, y_i, θ_i). This is done so the Extended Kalman filter can accurately track the future positions of the robots. We chose the robots to start at the following poses:

$$\begin{aligned}(x_1, y_1, \theta_1) &= (2.68, 1.71, \frac{\pi}{2}\,rad)\\(x_2, y_2, \theta_2) &= (2.32, 3.17, 0\,rad)\\(x_3, y_3, \theta_3) &= (2.77, 4.70, -\frac{\pi}{2}\,rad)\end{aligned}$$

With these positions, initially, it may be observed that the neighbouring sets are as follows:

$$\begin{aligned}\mathbf{N}_1 &= \{2\}\\\mathbf{N}_2 &= \{1, 3\}\\\mathbf{N}_3 &= \{2\}\end{aligned}$$

Therefore, the condition that $\mathbf{N}_i \neq \emptyset,\ \forall i$ is satisfied. Also, all the robots start at rest and $v_i = 0,\ \forall i$.

The paths described by the robots are described in Figure 4. As it may be seen in the figure, the robots start and after a transient response, they stabilize around a distance, but never actually reach the desired distance. This has already been observed in simulation. It takes a long time for the robots to reach a steady state. Since the robots cannot run for long, the steady state is never reached. Also, notice that none of the robots actually overshoot. This, again, is due to the choice of gains c_d and c_v as discussed before.

The collaborative behaviour of the robots is very similar to what has been observed in simulation [11]. The results shown in Figure 4 suggest that the algorithm generates the same type of control inputs as in the simulations. But some differences must be emphasized.

First, noise that has not been introduced in the simulations is dealt with in the experiments. This shows that the algorithm is robust. But with the data collected, it is not possible to conclude what is the level of robustness of the algorithm. More

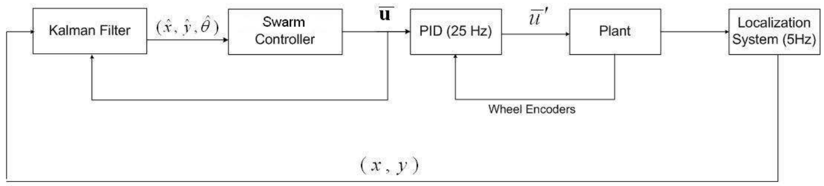


Fig. 3. Robot System Block Diagram

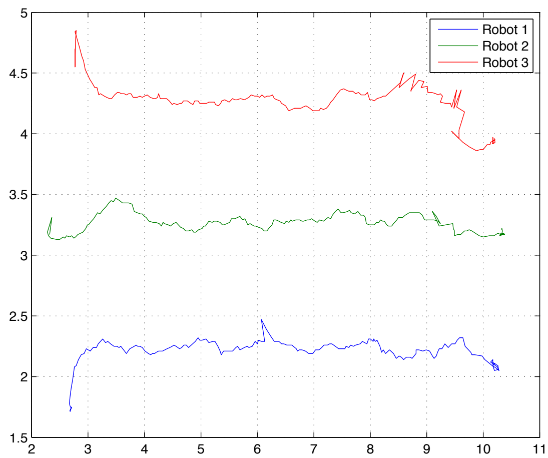


Fig. 4. Flocking paths for all robots (units in m)

tests with more robots are necessary to show the robustness of the system.

Second, the gains c_d and c_v chosen are not general. In different cases, with different parameters, these gains could change. However, with the analysis done in section III, we have very good guidelines for the selection of the gains.

Third, the robots move on a plane, therefore, gain c_h is not used. As observed in Section III, this parameter does not influence the stability of the system (at least as described). However, experimentation with flying robots is necessary in order to assess the impact of the third dimension on the stability of an actual system.

VII. Conclusion

In this paper we have shown how the flocking algorithm presented in Section II is stable and that it can be implemented in physical robots. The necessary information and parameters for the algorithm were described and how local information can be used to drive the robots into flocking discussed.

Simulations have already shown that the algorithm works in a simulation environment [11]. However, no stability analysis had been provided. This is necessary to show that the algorithm can be used in actual platforms. Even though limited, the analysis provided in this paper points to the possibility of using the algorithm in actual flight.

Finally, experimental results were reported. This is also a critical step in order to show that such solutions can be ported to more general platforms and scenarios. The results presented show that the algorithm produced the expected behaviour. However, more extensive tests are required in order to show that the algorithm would be resilient to failures.

As next steps, we intend to implement the algorithm in a larger group of ground robots. This would show that the algorithm is indeed portable to higher scales. Also, we intend to port the algorithm into flying platforms in order to prove that it is in fact able to run in real time in one extra dimension.

References

[1] E. Shaw, "Fish in schools," *Natural History*, vol. 84, no. 8, pp. 40–45, 1975.

[2] B. L. Partridge, "The structure and function of fish schools," *Scientific American*, vol. 246, no. 6, pp. 114–123, 1982.

[3] T. Vicsek, A. Czirook, E. Ben-Jacob, I. Cohen, and O. Shochet, "Novel type of phase transition in a system of self-driven particles," *Physical Review Letters*, vol. 75, no. 6, pp. 1226–1229, 1995.

[4] J. Toner and Y. Tu, "Flocks, herds, and schools: A quantitative theory of flocking," *Physical Review E*, vol. 58, no. 4, pp. 4828–4858, 1998.

[5] A. Mogilner and L. Edelstein-Keshet, "A non-local model for a swarm," *J. Math. Biology*, vol. 38, pp. 534–570, 1999.

[6] C. W. Reynolds, "Flocks, herds, and schools: a distributed behavioral model," in *Computer Graphics (ACM SIGGRAPH 87 Conference Proceedings)*, vol. 21, no. 4, 1987, pp. 25–34.

[7] P. Vincent and I. Rubin, "A framework and analysis for cooperative search using UAV swarms," in *Proceedings of the 2004 ACM symposium on Applied computing*, ser. SAC '04. New York, NY, USA: ACM, 2004, pp. 79–86. [Online]. Available: http://doi.acm.org/10.1145/967900.967919

[8] R. Olfati-Saber, "Flocking for Multi-Agent Dynamic Systems: Algorithms and Theory," *IEEE Transactions on Automatic Control*, vol. 51, no. 3, pp. 401–420, 2006.

[9] H. Schwartz and S. Givigi, "A practical approach to robotic swarms," in *Tenth IASTED International Conference on Control and Applications*, 2008, pp. 8–13.

[10] C. Churavy, M. Baker, S. Mahta, I. Pradhan, N. Scheidegger, S. Shanfelt, R. Rarick, and D. Simon, "Effective implementation of a mapping swarm of robots," *IEEE Potentials*, vol. 27, no. 4, pp. 28–33, 2008.

[11] G. Labonte, "Self organization of formations and swarms," DRDC - Ottawa, Tech. Rep. CR 2009-244, 2010.

[12] J. Lindsay, S. Givigi, G. Pieris, G. Fusina, and G. Labonte, "Experimental validation of swarms of robots," in *Systems Conference (SysCon), 2012 IEEE International*, march 2012, pp. 1–5.

[13] R. Siegwart and I. Nourbakhsh, *Introduction to Autonomous Mobile Robots*. Cambridge, Massachusetts: The MIT Press, 2004.

[14] E. Ivanjko, A. Kitanov, and I. Petrović, "Model based Kalman Filter Mobile Robot Self-Localization," in *"Robot Localization and Map Building"*, H. Yussof, Ed. Vukovar, Croatia: In-Tech, 2010, pp. 59–89.

An Autonomous Image-guided Robotic System Simulating Industrial Applications

Raza Ul Islam, Jamshed Iqbal,
Sarah Manzoor, Aayman Khalid, Sana Khan

Robotics and Control Research (RCR) Group,
COMSATS Institute of Information Technology,
Islamabad, Pakistan
jamshed.iqbal@comsats.edu.pk

Abstract–*This paper presents a robotic system based on a serial manipulator. The robot is a vertical articulated arm with 5 revolute joints having 6 Degree Of Freedom. Actuated with six precise servo motors, the system offers positional accuracy of ±0.5mm with a movement speed of 100mm/s. Forward and Inverse Kinematic model of the robot has been developed and its workspace has been analyzed to facilitate the use of robotic arm as a simulated industrial manipulator. Image processing has been done to make system more autonomous. Followed by a user's commands, the system acquires image of the environment using on-board camera. This image is processed to extract information about object's coordinates. Based on these coordinates, Inverse Kinematic model computes the required joint angles for the end-effector to reach at desired position and orientation thus enabling it to manipulate the object. The proposed system can be used in wide range of industrial applications involving pick and place, sorting and other object manipulation tasks. The system can also be potentially useful for heavy and 'giant' industrial applications after scaling up i.e. using huge robotic arm, employing multiple and better cameras and optimizing algorithms.*

Keywords: Robotic system, Manipulator arm, Industrial mechatronics

1 Introduction

The advancement in robotics has greatly influenced industrial automation processes. An industrial robot. according to Mikell Groover is "a general-purpose, programmable machine possessing certain anthropomorphic characteristics"[1]. Most of the industrial robots look like mechanical arms and many of these actually function like a human arm and are termed as articulated in scientific literature. Robot base turns around in a similar fashion as that of twisting torso of a human. On most articulated robotic arms, the shoulder and elbow pivots on the axis perpendicular to the arm axis and parallel to the plane containing the robot base. The wrist of articulated robotic arms offers motion in pitch while roll and yaw motions may or may not be there.

Industrial robots find their enormous applications in assembling, welding, spray painting, machine loading, manufacturing, construction, fabrication and so on. Considering an assembly task, following sequences are often executed [2]: (a) Pick up a complete part (or its piece) vertically from a horizontal work bench (b) Move the part in a horizontal plane to a point just above another place on the bench (c) Lower the part to the bench at the proper point thereby accomplishing the task. Not only in assembly, many other industrial applications require pick and place. Pick and place is probably the most common task in industry. Several industrial robots have the required capability to automate picking a part up from one location and placing it onto the other. They not only offer more accurate and fatigueless operations but they also speed up the process thereby increasing production rates. Most of the movements performed by pick and place robots are extremely cumbersome for the humans to do. Achieving repeatability and quality is essentially the outcome of consistent output of a robotic system. Moreover, in industrial robots, tooling can be interchanged as per application requirements.

A key word in the definition of an industrial robot is its programmability. These robots can be reprogrammed as per application scenario, operational environment and requirements of object to be manipulated. Although industrial robots are usually intended and trained for a specific task, some level of autonomy is certainly welcomed. Non-autonomous robotic systems need to be taught about the location of objects. This can be achieved by employing a teaching loader. Letting a robot know about the object location in this fashion is cumbersome, tedious and time consuming. Moreover, a robot will be needed to re-teach if the object location varies. This paper presents an autonomous robotic system having capability to detect objects through image guidance. Various sub-systems have been integrated to realize this concept. These sub-systems include mechanical, kinematics, electronics, computer vision and image processing .

The paper first briefly introduces overall system in Section 2. Developed kinematic model of the robotic arm is presented in Section 3. This includes forward kinematics as well as Inverse Kinematics (IK). Workspace analysis of the robot is also discussed in this section. Details of image guidance for object detection is explained in Section 4. Finally Section 5 comments on conclusions and future work.

2 Overall System

The overall system consists mainly of a Control Station (Standard PC), an on-board camera for image acquisition and a robotic arm with its controller interfaced with a PC. Block diagram of the overall system is illustrated in Figure 1.

The system waits for user command. As soon as it encounters a command, an image of the robot's surrounding is captured. The image is then processed using developed algorithm (Section 4) to extract the object centre coordinates. These coordinates are then used to calculate the necessary joint angles of the robot using IK model approach (Section 3). The values of the computed joint angles make sure that the gripper of the robot will reach at the required point as dictated by the object coordinates. If the object to be manipulated lies outside the robot's workspace, an error 'out of bound' is prompted to the user. Otherwise a low-level mapping routine converts the required joint angles into encoder ticks followed by the execution of the command. The system then iterates in the same loop unless user wants to exit. Detailed functional flowchart of the system is presented in Figure 2.

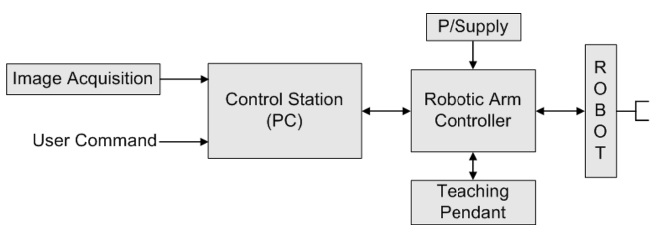


Figure 1. Block diagram of overall system

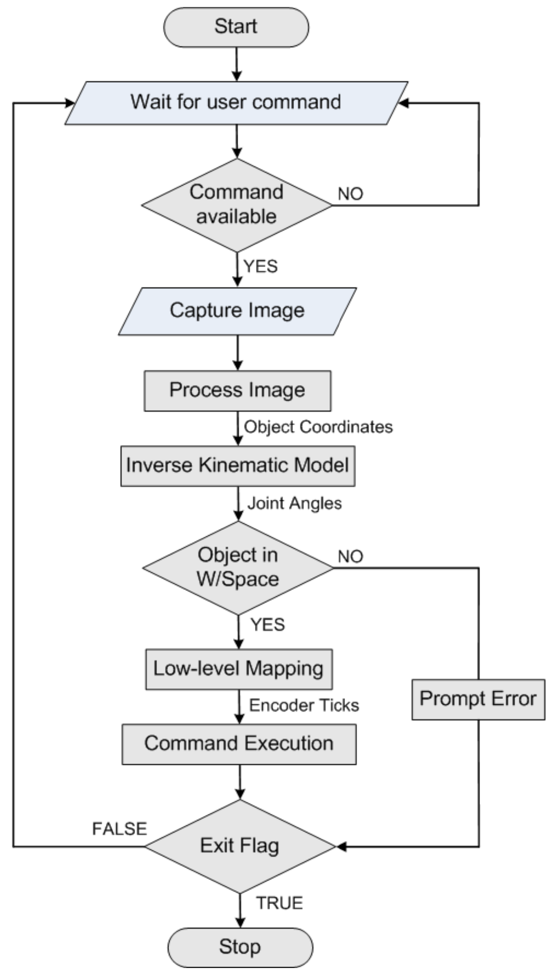


Figure 2. Functional flow chart

3 Kinematic Model of Robotic Arm

The system is primarily based on a 6 Degree Of Freedom (DOF) vertically articulated robotic arm, ED7220C. All its joints are revolute. Driven by six servomotors, the robot is equipped with optical encoders to close the feedback control loop. A gripper having capability to manipulate small sized objects (order of 0.5mm) is attached for object manipulation. The robot can mobilize itself provided the appropriate options are incorporated. This feature makes the robot a good candidate for simulating industrial tasks e.g. an automated assembly line or pick and place etc. Figure 3 shows the robotic arm while Table 1 depicts its joint and link specifications [3].

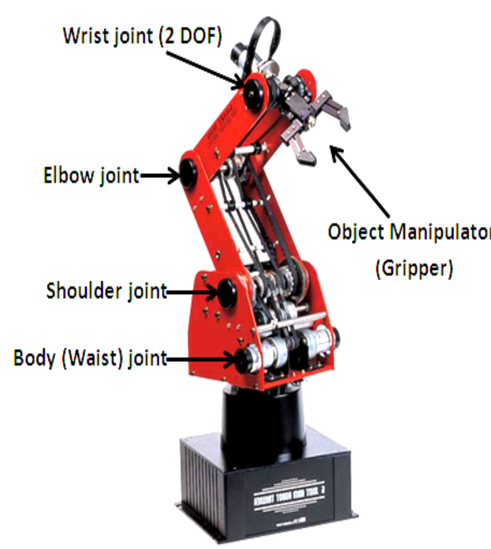


Figure 3. ED7220C–A 6 DOF articulated arm

Table 1. Arm joints and links specifications

Joint no.	Name	DOF	Link length [mm]
1	Waist	1	385
2	Shoulder	1	220
3	Elbow	1	220
4	Wrist	2	155

Kinematics is the study of motion without considering the forces involved in that motion. Kinematic analysis of a robotic mechanism essentially include: (a) Forward or direct kinematics (b) Inverse Kinematics.

3.1 Forward Kinematics

Given the link lengths and joint angles of a robotic manipulator, forward or direct kinematics computes the end-effector position and orientation. This always leads to a unique solution. Figure 4 illustrates the kinematic model.

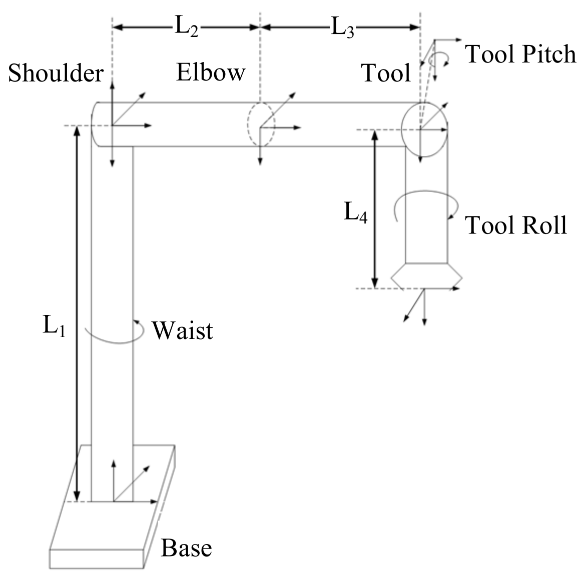


Figure 4. Kinematic model of ED7220C

The well known convention using Denavit-Hartenberg (DH) parameters has been followed. The frame assignment of various joints is elaborated in Figure 5. Wrist having 2 DOF is represented as tool roll and tool pitch.

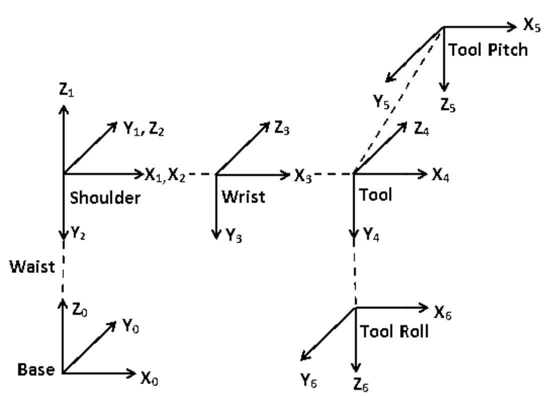


Figure 5.Frame assignment

After frame assignment, DH parameters of the robotic arm have been determined (Table 2). The nomenclature is as follows:

α_{i-1} = Angle from Z_{i-1} to Z_i measured about X_{i-1}
a_{i-1} = Distance from Z_{i-1} to Z_i measured along X_{i-1}
d_i = Distance from X_{i-1} to X_i measured along Z_i
θ_i = Angle from X_{i-1} to X_i measured about Z_i

Table 2. DH parameters of the robot

i	α_{i-1}	a_{i-1}	d_i	θ_i
1	0	0	L_1	θ_1
2	-90°	0	0	θ_2
3	0	L_2	0	θ_3
4	0	L_3	0	θ_4
5	-90°	0	0	θ_5
6	0	0	L_4	0

Based on these DH parameters of the robot, each joint frame has been expressed in its preceding frame using general transformation matrix (1) [4]

$$ {}^{i-1}_{i}T = \begin{bmatrix} C\theta_i & -S\theta_i & 0 & a_{i-1} \\ S\theta_i\, C\alpha_{i-1} & C\theta_i\, C\alpha_{i-1} & -S\alpha_{i-1} & -d_i\, S\alpha_{i-1} \\ S\theta_i\, S\alpha_{i-1} & C\theta_i\, S\alpha_{i-1} & C\alpha_{i-1} & d_i\, C\alpha_{i-1} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1)$$

Having obtained the transformation matrix for each joint, exploiting the compound transformation property yields the overall transformation matrix from end-effector to base of the robot (2). The symbols A and B in (2) are given by (3) and (4) respectively.

$$ {}^{0}_{6}T = \begin{bmatrix} C_1C_5C_{234} + S_1S_5 & -C_1C_{234}S_5 + S_1C_5 & -C_1S_{234} & C_1A \\ S_1C_5S_{234} - C_1S_5 & -S_1S_{234}C_5 + C_1C_5 & -S_1S_{234} & S_1A \\ -S_{234}C_5 & S_{234}S_5 & -C_{234} & B \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2)$$

$$A = L_2C_2 + L_3C_{23} - L_4S_{234} \quad (3)$$
$$B = L_1 - L_2S_2 - L_3S_{23} - L_4C_{234} \quad (4)$$

Where sin and cosine terms involve summation of the corresponding angles e.g.

$$C_{234} = \cos(\theta_2 + \theta_3 + \theta_4)$$

The first part (3X3) of the given transformation matrix determines the orientation of the end-factor with reference to base while the last column expresses its position ((5)-(7)).

$$X = C_1A \quad (5)$$
$$Y = S_1A \quad (6)$$
$$Z = B \quad (7)$$

3.2 Inverse Kinematics

Inverse kinematics (IK) being more useful in industrial robotic applications than forward kinematics, determines how much joints of a robot should be manipulated to reach a specific position and orientation. IK can be computed using analytical, geometrical or numerical iterative approach. Modeling a robotic arm for its IK is a complex challenge as it does not converge to a unique solution. At least two solutions (termed as 'Elbow-up and Elbow-down in literature) are always obtained. The IK algorithm in real industrial robots addresses which solution a robot should opt for in such circumstances of multiple solutions.

Both analytical and geometrical techniques have been used to develop IK model of the robot arm (ED7220C). Analytical approach computes equations for the first three joints including waist, shoulder and elbow ($\theta_1,\theta_2,\theta_3$) while for joint angle θ_4 (tool pitch), geometrical method is utilized. The last angle θ_5 (tool roll) has been computed using the image processing algorithm (Section 4).

Considering first three joint angles, the general form of the transformation matrix from elbow to base is given by

$$^{Base}_{Elbow}T = \begin{bmatrix} n_x & o_x & a_x & p_x \\ n_y & o_y & a_y & p_y \\ n_z & o_z & a_z & p_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (8)$$

Now computing the inverse of homogeneous transformation matrix of frame {1} represented in base, (9) is obtained.

$$(^0_1T)^{-1} = \begin{bmatrix} C_1 & S_1 & 0 & 0 \\ -S_1 & C_1 & 0 & 0 \\ 0 & 0 & 1 & -l_1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (9)$$

Multiplying (8) and (9) and then comparing the product with the result of$(^0_1T)^{-1}(^0_4T)$, equations (10), (13) and (16) can be written for required joint angles ($\theta_1,\theta_3,\theta_2$).

$$\theta_1 = Atan2(p_x, p_y) \quad (10)$$

$$c_3 = \frac{(c_1p_x + s_1p_y)^2 + (p_z - l_1)^2 - {l_2}^2 - {l_3}^2}{2l_2l_3} \quad (11)$$

$$s_3 = \pm\sqrt{1 - {c_3}^2} \quad (12)$$

$$\theta_3 = Atan2(s_3, c_3) \quad (13)$$

$$c_2 = \frac{(c_1p_x + s_1p_y)(c_3l_3 + l_2) - (p_z - l_1)s_3l_3}{(c_3l_3 + l_2)^2 + {s_3}^2{l_3}^2} \quad (14)$$

$$s_2 = -\frac{(c_1p_x + s_1p_y)s_3l_3 + (p_z - l_1)(c_3l_3 + l_2)}{(c_3l_3 + l_2)^2 + {s_3}^2{l_3}^2} \quad (15)$$

$$\theta_2 = Atan2(s_2, c_2) \quad (16)$$

Having calculated values of θ_2 and θ_3, θ_4 has been computed as

$$\theta_4 = \theta - (90\,° + \theta_2 + \theta_3) \quad (17)$$

Where θ is an element of the user-defined matrix (18). The matrix lists priority wise values of desired angles with which end-effector should approach the object for manipulation.

$$\theta = [\theta_a \quad \theta_b \quad \theta_c \quad \theta_d \ldots] \quad (18)$$

For the present work, the most priority angle value (θ_a) has been set to 90°. This employs that the best possible scenario for the end-effector to manipulate the object is exactly at its top vertically.

3.3 Workspace Analysis

An essential parameter to design a robotic arm based system is the analysis of robot's workspace. Workspace of a robotic arm is primarily a function of Range Of Motion (ROM) of its joints and associated link lengths. Table 3 mentions ROM of ED7220C joints while link lengths have been tabulated in Table 1.

Table 3. ROM of ED7220C joints

Joint no.	Name	ROM [Degrees]
1	Waist	-244→66
2	Shoulder	-120→-30
3	Elbow	-106→66
4	Wrist	Pitch: -220→40 Roll: 0→360

Based on the mentioned link lengths and ROM of ED7220C, (2)-(4) have been used to determine overall robotic workspace. Figure 6 illustrates the top view of the workspace. Because of the constraints on joint ROM, the robot can manipulate in the radius of 580 mm in a defined region. The 'V' shaped region in Figure is because of ROM constraint of the waist joint.

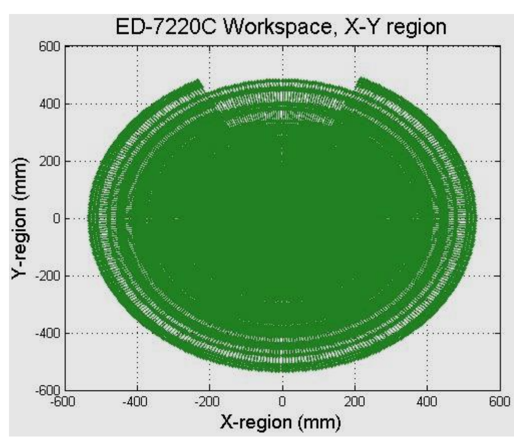
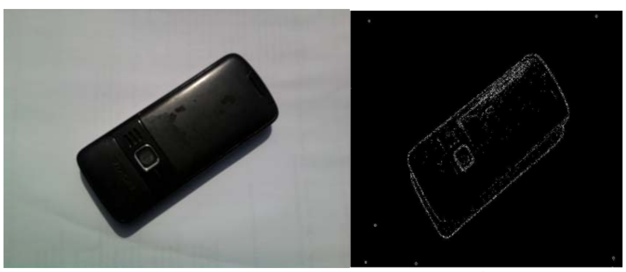


Figure 6. Workspace in XY Region

4 Image Guidance for Target Detection

The proposed system is autonomous in the sense that it does not have to rely on a manual teaching to know the object's coordinates. Employing computer vision enhanced robot autonomy. User just has to input the object of interest. Together with the developed robot model and image processing algorithm, the robot navigates to the required point for object manipulation thus accomplishing its task.

The main aim of the developed image processing algorithm is to detect the target object's whereabouts by extracting its center point and the orientation at which it has been placed. Figure 7a shows a sample original captured image while result of edge detection without improving the picture attributes is illustrated in Figure b. To cope with non-linearities and shading problem, the image has been pre-processed before edge detection. The image has been first smoothened using averaging filter (Figure c) and then sharpened-up to enhance edges of the original object using Laplacian filter (Figure d) followed by edge detection (Figure e) [5]. Finally based on the extracted corner point indices (Figure f), centre coordinates of the object and thus orientation θ_5(tool roll) have been computed. The results from image processing algorithm are then fed to IK model as discussed in Section 3.



(a) (b)

(c) (d)

(e) (f)

Figure 7.Image processing on sample image: (a) Original image (b) Simple edges (c) Averaging filter (d) Laplacian filter (e) Edge detection (f) Center and corner points

5 Conclusions

This work combines image processing with robot modeling to simulate industrial application. Image processing provides autonomy to the system. Prior to extraction of necessary features from image, various filtering techniques have been applied on captured image. Object's coordinates form input of the developed robot model. A complete solution to the IK of a widely used 6 DOF robotic arm, ED7220C has been derived. The derived IK model always provides accurate joint angles to make sure that the end-effector points to the required position and orientation, provided the point of interest is in robot's workspace. In case the object is out of workspace, the model prompts the user. Overall system integrated with on-board web cam is shown in Figure 8. Teaching pendant is also shown here. However use of computer vision in the proposed system has eliminated the role of pendant. To demonstrate the concept, a common task in industrial automation i.e. pick and place has been considered. After testing and verifying individual subsystems, the complete system has been already integrated and currently undergoing in-lab trials. Multiple irregular shaped objects will be considered in the future.

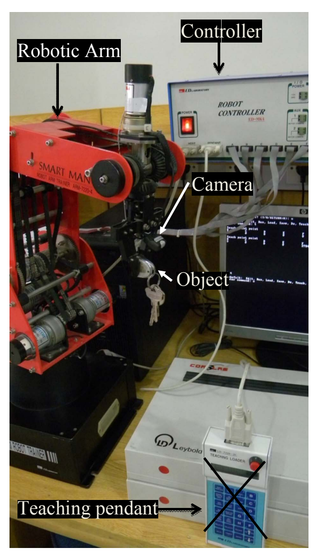


Figure 8. Overall System

References

[1] M. P. Groover, M. Weiss, R.N. Nagel, N. G. Odrey, “Industrial robotics – Technology, programming and applications”, Tata McGraw-Hill publishing company, 2008

[2] C. R. Asfahl, “Robots and manufacturing automation”, John Wiley & Sons Inc., 2nd edition

[3] 7220C robotic arm data sheet, ED Corporation, Korea, p. 21

[4] J. J. Craig, “Introduction to robotics: Mechanics and Control”, Prentice Hall, 3rd edition

[5] R. C. Gonzalez and R.E. Woods, “Digital Image Processing”, Prentice Hall, 3rd edition

A System Architecture for Heterogeneous Moving Objects Trajectory Models using Different Sensors

Azedine Boulmakoul, Lamia Karim, Adil Elbouziri
University of Hassan II-Mohammedia,
Faculty of Science and Technology, ILIS department
BP 146 Mohammedia 20650, Mohammedia, Morocco
azedine.boulmakoul@gmail.com
lkarim.lkarim@gmail.com
a_elbouziri@yahoo.fr

Ahmed Lbath
University of Grenoble 1 - Joseph Fourier
LIG Lab. BP 53, 38041 Grenoble
Cedex 9, France
ahmed.Lbath@ujf-grenoble.fr

Abstract - *This paper proposes a general trajectory's meta-model and its architecture. In order to provide a unified meta-model and a powerful framework for trajectory's services and queries, the proposed trajectory's data-model has benefited from advantages of both conceptual and ontological space-time. However, it extends the basic data-model of trajectory with new patterns as Space Time Path to describe activities of the moving object, and Composite Region of Interest. Additionally, the proposed system is distinguished by providing framework for dealing moving objects trajectory, in an interoperable way, using heterogonous sensors that traditional data model alone are incapable for this purpose. The proposed system architecture for moving object trajectory's data-model is focused on service composability and data interoperability combining OGC standards, Service-Oriented Architecture and streaming technology, to allow applications constructed using the framework have better performance.*

Keywords: Trajectory data modeling, Trajectory framework, moving object database, space time path, space time ontology, trajectory meta-model, spatial data engineering, trajectory meta-model instantiation.

1 Introduction

Technological evolution of GPS devices has allowed capturing easily, with low cost and high quality, spatio-temporal coordinates (x, y, t) of a moving object (human beings, vehicles, etc.). The massive information, generated by GPS devices, is needed in different fields of applications and studies e.g. a system for destination and future route prediction based on trajectory mining [1], real-time monitoring of water quality using temporal trajectory of live fish [2], analyzing bird migrations trajectory[3], etc. By operating ontology, our proposed meta-model focuses on practical problems providing a shared understanding and common data model for different presentation of trajectories (raw, structured, semantic region of interest and space time path). Moreover, in our general meta-model, we use event approach to allow integration of data from heterogeneous spatio-temporal data source with diverse spatial and temporal sampling protocols, facilitates the recovery and requests of different types of trajectories. Once our proposed moving objects trajectories meta-model has been instantiated, trajectories could be organized in spatio-temporal database, to support representing and querying of moving objects and their trajectories [4], or into trajectory data warehouse to analyze and make decision. This research's goal is to provide a system architecture, focused on service composability and data interoperability, for the proposed general moving objects trajectory's meta-model that is based space-time ontology and analytical geo-semantic. Using object approach, the Oriented Object Trajectory Meta model integrates previous models of geometric, structured and semantic of trajectory. Our model includes also the hybrid spatio-semantic model given in [5] and models the following, by using space-time event approach, analytical geo-semantic, and space time ontology: (i) Spatial Model according to OGC Spatial Data Model. (ii) Observation domain of trajectory according to OGC Sensor Meta Model and OGC Feature Type. (iii) All activities between the begin and the end of Space Time Path [6]. (iv) Mechanism of detection used to collect generated positioning data. (v) Movement patterns using composite Region of Interest.

The remaining of the paper is organized as follow: Section 2 presents basic concepts relating to trajectories of moving objects whereas in Section 3 we discuss the issues related to the representation of trajectories. Section 4 focuses on presenting the proposed moving object's trajectory meta-model. Section 5 describes the system's architecture. Section 6 presents expressiveness query types for unified trajectory database. Finally, in Section 7 we conclude the paper and present some directions of future work.

2 Basic concepts of trajectory

A trajectory is a description of physical movement of moving objects changing over time, in the following basic presentation of trajectories, shown in figure 1: (i) Raw trajectory is the recording of the positions of an object at specific space-time domain, for a given moving object and a given time interval, it is presented as a sequence of geometric location in 2D spatial system (x_i, y_i, t_i). (ii)

Structured trajectory [3] defined as a raw trajectories structured into segments corresponding to meaningful steps in the trajectory trace (e.g. travel). (iii) Semantic trajectory [3] express the application oriented meaning using four component (stop, move, begin and end). Stop, move, begin and end are no more spatio-temporal position, but semantic objects linked to general geographic knowledge and application geographic data (figure 3). (iv) Other recent approach describes movement patterns in both spatial and temporal contexts based on Region of Interest [7] by defining spatial neighborhood and temporal tolerance.

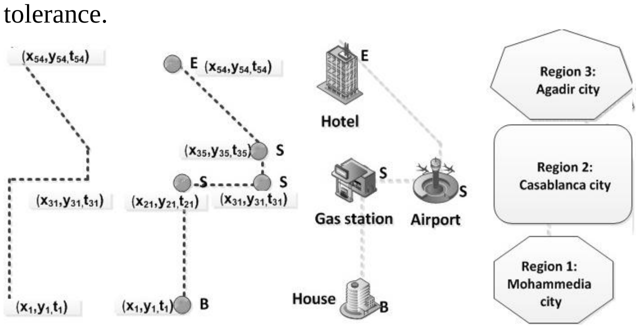


Figure 1. Raw trajectory, structured trajectory, semantic trajectory and Region of interest

3 Related Work

Event ontology has already been proven useful in a wide range of context, due to its simplicity and usability. Several ontology describing “events” and related concepts have been investigated. The SHOE General Ontology [8] defines an event as something that happens at a given place and time. Kate Beard in [9] describes an event based approach to model space and time, instead of presenting it using geographic and identities characteristics which could be repeated, where dynamic aspect dominates instead of geographical features. In [3], Spaccapietra stated that there are two facets of a trajectory: geometric facet which only considers the point geometry and a semantic facet which gives a meaning or semantic interpretation of application objects (Move, Stop, Begin and End). Patterns in [3] provide a trajectory structure as a sequence of moves and stops in between, with begin and end events to represent a trajectory in a relational model. However, there are many inconveniences when using a relational trajectory model, such as complex maintenance when upgrading application. Furthermore, the current trajectory design pattern models trajectory just from the semantic point of view. Giannotti [7] describes movement patterns in both spatial and temporal contexts, based on static and dynamic RoI (Region of Interest). However, it is still very hard to explain and understand movement behaviors based on these patterns, because most of these researches methods define trajectory and neglect having observation and description of trajectory/moving object (e.g. when taking a taxi or a bus, it seems to be very interesting to have information on what it takes for a number of passengers). Whereas, marketing studies main interest for moving object’s physical and virtual activities in a specific space-time. Furthermore, queries using current models database can not give detection’s mechanism type used to collect data, e.g. which mechanism of detection was provided to collect data of a person at 8 a.m?

Adding activities to trajectory presentation, like transportation mode, allow users to have an extended query type, like asking how a person X went to work. Or how much time this person spends in walking every week? Activities studies are conducting information and knowledge discovering. Also, it allows understanding people behaviors, and other phenomena. Shaw [10] extended concept of space-time path, shown in figure 2, to represent both physical (e.g. walking) and virtual activities (sending email). As each activity has a geographical location and time interval, space-time path has been profiled as a container of all activities occurring by a moving object.

Figure 2. Space-time path presentation

The proposed data-model introduce the composite Region of Interest concept and deal Mobile Objects trajectories with diverse spatio-temporal sampling protocols and different sensors available that traditional data model alone are incapable for this purpose. Figure 3 present an example of composite region of Interest, where region Hypermarket is composed from other regions prepaid parking, bank, supermarket, department store and restaurant, respectively spatio-temporal data have been captured using different sensors GPS, RFID, data base transactions, and camera.

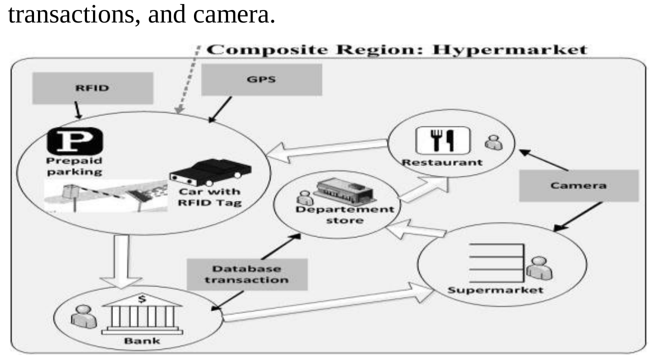


Figure 3. Trajectory with Composite Region of Interest and different sensors

Below in table 1, we provide a brief tabular comparison of previous moving object's trajectories models used in literature with our proposed meta-model on the following criteria's basis: supporting presentation of Structured trajectory, Semantic trajectory, Region of Interest, Composite Region of Interest, Activities/ space time path, Model type, Mechanism of detection provided, Based ontology and Event approach.

Table 1. Comparison of previous models of trajectory

	Structured trajectory	Semantic trajectory	Region of Interest	Composite Region of Interest	Activities/ space time path	Model type	Based ontology and Event approach	Mechanism of detection provided (other than GPS)	OGC Sensor meta-model	OGC Observation feature-type
[9]	√	×	×	×	×	Relational	×	×	×	×
[11]	×	×	×	×	×	Function of time and bound deviation	×	×	×	×
[3]	√	√	√	×	×	Relational	×	×	×	×
[6]	√	√	√	×	×	Relational	×	×	×	×
[8]	√	×	√	×	×	T-Patterns mining	×	×	×	×
Our meta-model	√	√	√	√	√	Object oriented	√	√	√	√

4 The proposed Unified Trajectories Pattern

We use object approach to increase developer productivity and higher quality applications. A description of class diagram will be presented in future work to avoid exceeding number of pages allowed. The most important packages used, shown in figure 4, in our trajectory patterns are described as follow.

4.1 Space Time Path Domain package

Space Time Path Domain, the central package, combines classes of Raw, Structured, Semantic Trajectory and Space Time Path to provide the different models of trajectory that can be used.

4.2 Activity Domain package

As monitoring activity of moving objects is a crucial task in many urban, economic and social systems applications, we added a UML package Activity Domain including classes of physical activities (e.g. drive to work, have lunch, drive to school) and virtual activities (e.g. send email, receive a call). This package allows us to represent, manage and analyze activities that occur in physical and virtual spaces, events that involve multiple individual's activities and projects that consist of multiple events, also objects of this package are used to create space-time path.

4.3 Observation Domain package

In the context of Open Geospatial Consortium Sensor (OGC) meta-model and OGC Feature type, we added, to our trajectory meta-model, Observation Domain package to describe action with a result describing some phenomenon. For example, when tracking a terrorist's trajectory, it may be important to record observations and activities?

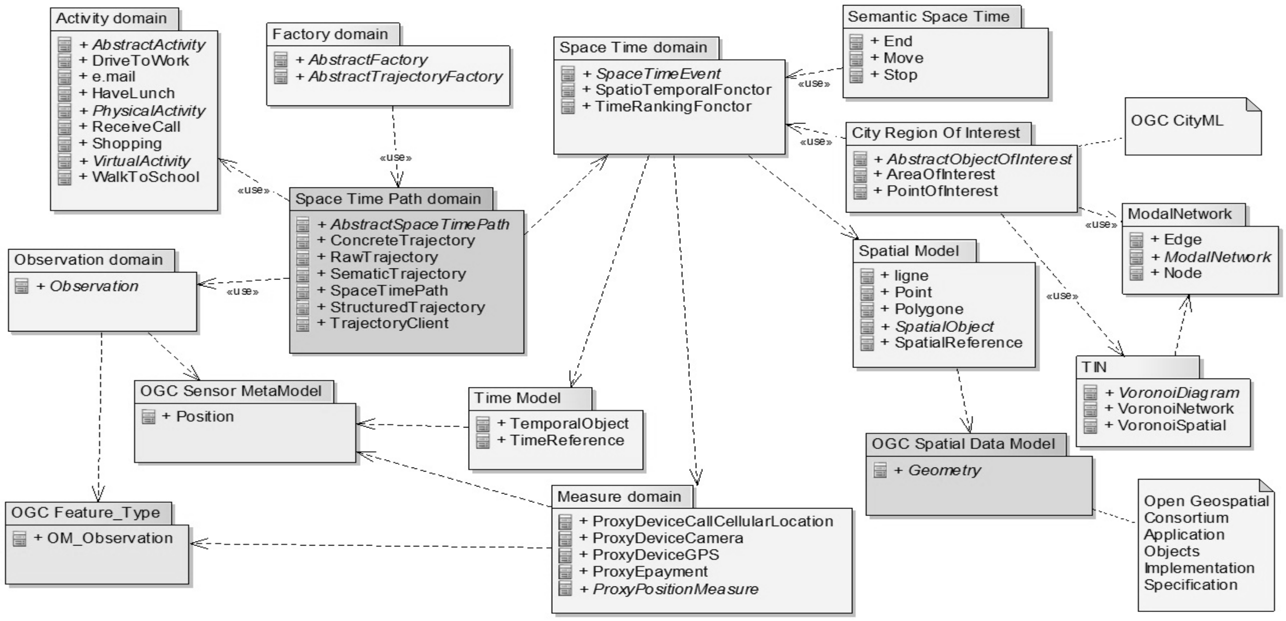
4.4 Measure Domain package

GPS equipped mobile phones, device camera, vehicles with navigational equipment or location based services, give digital traces (sequences: $<(x_1,y_1,t_1), ..., (x_n,y_n,t_n)>$) to describe mobility behavior (figure 1). In order to define the basic models for how geospatial information is to be characterized and encoded according to the OGC, these devices are arranged in Measure domain package which use OGC Sensor Meta-Model package and OGC Feature_Type package.

4.5 Region of Interest packages

Because in many applications, it is expensive to work with trajectory as a sequences of Spatio-temporal event: $<(x_1,y_1,t_1), ..., (x_n,y_n,t_n)>$. Giannotti [7] describes movement patterns in both spatial and temporal contexts, based on static and dynamic Region of Interest (RoI), by allowing approximation in both spatial neighborhood and temporal tolerance.

In this package, we model RoI and introduce a composite RoI class, e.g. in a tourist guide application, we can consider Morocco as a RoI, however in other user's request; the RoI could be the city of Casablanca, monument or beach. Hence in our meta-model, a RoI could be composite of one or several RoI. Also, we modeled RoI as a Voronoi polygon in which a site can represent a point of interest.



patterns

5 Unified Moving object System Architecture

This paragraph describes the proposed system architecture for our unified moving object data model framework that combines the streaming technology and Service-Oriented Architecture to improve the application performance and interoperability. By architecting trajectory's components using Web Services based on public OGC standards, and be placing these services within a SOA messaging substrate, we may integrate trajectories services (e.g. tracking, visualizing and querying moving object's trajectory) with other applications and locations based services. Also, it provides a uniform means to offer, discover, interact with and use capabilities to produce desired effects consistent with measurable preconditions and expectations. However, we use streaming technology to upload and deal videos from cameras in real time. Target system architecture, shown in figure 5, consists of the five main layers: Data collector layer, trajectory pre-processing layer, trajectory's generator layer, trajectory database layer and trajectory applications layer.

(A) Data collector layer contains mobile devices, E-payment and video streaming proxy servers in Demilitarized zone, respectively to collect data from Mobile devices, self services terminal and cameras using HTTP, FTP and SOAP protocols. All depends on criticality of data to collect, data collector services are used to collect online or offline data.

(B) Trajectory pre-processing layer employs: (a) Data reducer service: mobile devices with location positioning capability generate a huge of redundant spatio-temporal locations of (x_i, y_i, t_i), the aim of data reduce service is to reduce number of discrete spatio-temporal points to record to save storage space and increase quality and trajectories' queries speed. In literature, there are two reduction techniques[12]: (i) Batched compression techniques: is an off-line compression uploading strategies where the full positioning data of tracked moving object is taken into consideration by the compression algorithms, the results aim to approaching the global optimal better than the techniques in the other category. There are various algorithms used to replace the original trajectory by an approximate line segment [13]. (i) On-line data reduction techniques: batched compression techniques cannot deal with the second category of application that requires an instantaneous update of position, e.g. traffic monitoring service must display traffic situation constantly. Thus, depending on precision requirement and criticality of application, a selective update is applied. (b)Errors measures service: data reduction, especially using online technology, can cause errors. Hence, we added error measures service to control the efficiency and performance of the technique of reduction used in a previous step. Evaluation criteria are: (i) the execution time spent to run a trajectory data reduction algorithm, (ii) size of an approximate trajectory vs. the size of its original trajectories, (iii) the deviation of reduced trajectory from its original trajectory [14]. (c) Activity recognition service: activities could be captured using sensors, e.g. GPS, WIFI, cameras and Self Terminal SSL. [15] Deal the problem of activities recognition using mobile phones. (d) Reverse geo-coding service used to convert a coordinate recorded to a readable street address which is easier to understand by the end user.

(C) Through Trajectory's generator layer, a set of spatio-temporal points are transformed from a cleaned raw

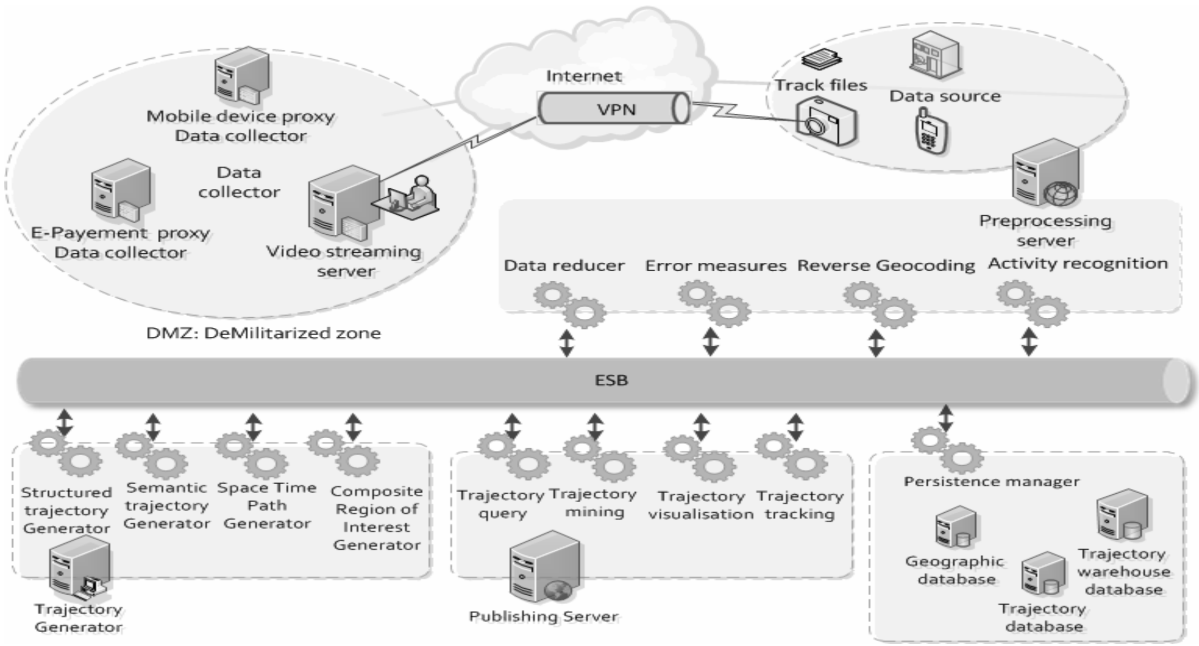


Figure 5. Unified moving objects system architecture

trajectory presentation to structured, semantic, region of interest or space time path using respectively structured trajectory, semantic trajectory, space time path, and composite region of interest generator.

(D) Trajectory database layer contains Geographic, and generic trajectory databases. We use POSTGIS as generic trajectory databases that support geographic objects of open source object-relational database PostgreSQL. In effect, PostGIS follows the OpenGIS Simple Features Specification for SQL and has been certified as compliant with the "Types and Functions" profile.

(E) Trajectory applications layer contains specialized Web services that are self-contained, self-describing, modular applications of trajectory mining, tracking, visualizing and querying which could be published, located and invoked across the web, using wide spectrum of Web-enabled stationary (desktops, workstations, Web TV) and mobile devices (PDAs, mobile phones, laptops, etc.) our framework use MapServer as an open source platform for publishing spatial data and interactive mapping applications to the web.

6 Unified Moving Object Trajectory Query

As mentioned above, our work differs in that we provide a mean of trajectories data's exploitation to answer a wide range of complex trajectory queries that whether traditional trajectory database are capable for this purpose or not. For example, query like "cars license plate that park in front of my house noticing it all, when I am on the way to school" can not be answered using traditional models. This paragraph is interested in evaluating the efficiency, performance and utility of models that instantiated through a set of spatio-temporal queries. Using our proposed unified meta-model, trajectory queries can be classified into six types according to their instantiated spatiotemporal data-model:

- Raw trajectory queries: ask for spatio-temporal coordinates of a specified moving object (MO) at a given time t to specified trajectory segment(s), but as raw trajectory store no semantic information, just sequence of (x,y,t), asking for semantic information need using of spatial join, e.g. find all places (restaurant, supermarket and administrations) visited by a moving object (MO):

select r.name from rawtrajectory t , restaurant r
where t.id='MO' and ST_Contains(r.ps , t.spatialpoint)
Union
select s.name from rawtrajectory t , supermarket s
where t.id='MO' and ST_Contains (s.ps , t.spatialpoint)
Union
select a.name from rawtrajectory t , administrations a
where t.id='MO' and ST_Contains (a.ps , t.spatialpoint)

- Structured trajectory query: asks for spatio-temporal coordinates where the moving object's stop, move, beginning and the end, in specified trajectory segment(s). E.g. finding all roads with moving object (MO) when it took him/her over 10min.

select r.name from structuredstop t , roads r
where t.id='MO' and intersects (t.spatialpoint , r.geom) and (t.timeEndStop - t.timeBeginStop)>'10min'

- Semantic trajectory query: ask for trajectories where moving object (MO) stayed in a given semantic place (restaurant, cinema, stadium…) for a while (e.g., 1 hour). E.g. find all trajectories where moving object MO got on the road and took him/her over 10min.

select t.name , t.timeBeginStop , t.timeEndStop
from SemanticStop t
where t.id='MO' and t.cat='Road' and (t.timeEndStop - t.timeBeginStop)> '10min'

- Trajectories and regions of interests query: ask for trajectories crossed a point of interest, area of interest, modal network or voronoi diagram in a given time interval, e.g. Figuring out the number of trajectories that visited each commercial region.

select t.nameRegion, count(*) nb_visits
from RoITrajectory t where t.cat='commercial'
group by t. nameRegion

- Space time path query: ask for activities or process of a moving object in a spatio temporal location, e.g. finding the space time path of a specific person.

select t.name, t.timebeginactivity, t.physicalActivity, t.virtualActivity
from SpaceTimePath t
where t.id='MO'

- Trajectories and mechanism of detection query: ask for devices and their reliability degree used to capture information in a specific spatio-temporal location, e.g. which mechanism of detection used to capture information when the moving object was at the airport.

select d.name, d.reliability from RoItrajectory t , devices d where t.name like '%airport%' and t.id= d.id and t.timebeginstop =d.timed

7 Conclusions

Our framework has many benefits. Firstly, the object oriented data-model provides a higher level of interoperability and information sharing, where we happen to have traditional trajectories models. Secondly, we have used UML which is easy to learn and can model all real world objects. Thirdly, clearly answering a wide range of complex queries, as it shows us new interesting patterns (i) space–time path to describe physical and virtual activities of a moving object, (ii) composite region of interest. In the future, we are in the process of testing the proposed framework on a case study.

References

[1] Chen, L., L. Mingqi, and G. Chen, "A system for destination and future route prediction based on trajectory mining," Pervasive and Mobile Computing, Elsevier Science Publishers, pp. 657-676, 2010.

[2] Heng, M., T. F. Tsai, and C. Cheng, "Real-time monitoring of water quality using temporal trajectory of live fish," Expert Systems with Applications, pp. 5158–5171, 2010.

[3] Spaccapietra, S., C. Parent, M. D. Damiani, J. A. Macedo, F. Porto, and C. Vangenot, "A Conceptual view on trajectories," Data and Knowledge Engineering, pp. 26–146, 2008.

[4] E. K. Frentzos "Trajectory Data Management in Moving Object Databases," PhD Thesis, University Of Piraeus, Department of Informatics, 2008.

[5] Yan, Z., C. Parent, S. Spaccapietra, and D. Chakraborty "Hybrid Model and Computing Platform for Spatio-Semantic Trajectories", 7th Extended Semantic Web Conference (ESWC), Heraklion, Greece, 2010

[6] Shaw, S. L. "A Space-Time GIS for Analyzing Human Activities and Interactions in Physical and Virtual Spaces," Center for Intelligent Systems and Machine Learning, UTK, 2011.

[7] Giannotti, F., M. Nanni, D. Pedreschi, and F. Pinellin "Trajectory Pattern Mining," International Conference on Knowledge Discovery and Data Mining, pp. 330-339, 2007.

[8] http://www.cs.umd.edu/projects/plus/SHOE/onts/general1.0.html

[9] Beard, K. "Modeling Change in Space and Time: An Event Based Approach, in Innovations in GIS," Edited by Taylor & Francis, pp. 55-76, 2007.

[10] Hongbo Yu, Shih-lung Shaw "Revisiting Hägerstrand's time-geographic framework for individual activities in the age of instant access", Springer Science, pp. 103-118, 2007.

[11] Meng, X., and Z. Ding "DSTTMOD: A Discrete Spatio-Temporal Trajectory Based Moving Object Databases System," DEXA, LNCS 2736, Springer verlag, pp. 444-453, 2003.

[12] Wang. L. and J. Krumm. Trajectory Preprocessing, Springer, chapter 1, 2011.

[13] Hershberger, J., and J. Snoeyink. Speeding up the Douglas-Peucker Line simplification Algorithm. International Symposium on Spatial Data Handling, pp. 134–143, 1992.

[14] Potamias, M., K., Patroumpas, and T. Sellis. Sampling Trajectory Streams with Spatio-Temporal Criteria. International Conference on Scientific and Statistical Database Management (SSDBM), pp. 275–284, 2006.

[15] Reddy, S., M. Mun, J. Burke, D. Estrin, M.H. Hansen, M.B. Srivastava. Using mobile phones to determine transportation modes. ACM Transactions on Sensor Networks (TOSN) 6(2), 2010.

Exploiting Cloud Computing for enabling distributed testing of complex systems: the SELEX-SI roadmap

Gabriella Carrozza, Ph.D.
SESM s.c.a.r.l
Via Circumvallazione esterna di Napoli, Giugliano in Campania, 80014, Naples, Italy
gcarrozza@sesm.it

Massimo Loffreda
SESM s.c.a.r.l
Via Circumvallazione esterna di Napoli, Giugliano in Campania, 80014, Naples, Italy
mloffreda@sesm.it

Vittorio Manetti, Ph.D.
SESM s.c.a.r.l
Via Circumvallazione esterna di Napoli, Giugliano in Campania, 80014, Naples, Italy
vmanetti@sesm.it

Abstract - *Complex systems are usually made up of several heterogeneous components glued together to get a System of System (SoS) demanding more and more effort in terms of integration, testing and maintenance due to the number of components, as well as to the several sources of failures that rise from heterogeneity. On the other hand, the strict reliability requirements of these systems ask for massive testing campaigns since they mostly fail due to software defects that can be either triggered systematically during system execution or manifest in a transient way during its operational phase. In the SELEX-SI scenario, where systems get developed and tested across different premises distributed all over Europe, performing traditional, manual, and on-site testing becomes dramatically expensive in terms of time and human resources. Cloud computing represents the most promising way for allowing the seamless access to distributed testbed from any site and for allowing remote testing activities, either at system and integration level. A cloud based infrastructure in charge of connecting all the company premises would allow to run testing experiments from anywhere and, more important, the possibility of reproducing distributed systems deployment scenarios to run integration testing in a pre-installation phase thus dramatically reducing company costs. This paper aims to describe the i) cloud research roadmap that SELEX-SI has been designing, ii) the architectural design of the cloud infrastructure and iii) the real ROI that the company expect from introducing such an innovation into the traditional software production process.*

Keywords: cloud computing, Open Source, testing.

1 Introduction

The need for massive testing campaigns in safety and mission critical systems comes from the strict dependability requirements and the catastrophic effects that any failures may have both on company business and human lives. However, performing exhaustive testing experiments, intended to test and assess the dependability of very complex SoSs, is not a trivial task due to the residual faults that usually manifest during system operational life and never before. These failures, indeed, are due to operational conditions that are very hard to reproduce in pre-operational testing environment and sometimes manifest in a transient way despite the deterministic root cause, which is always a software defect, i.e., a bug into the source code. It can be definitely taken for granted that early testing is the best way to i) reduce the probability of operational failures, ii) improve system dependability and iii) reduce maintenance costs. Actually, discovering and fixing bugs before system installation represent a powerful mean for reducing costs, minimizing the number of people trips on site, and improving company's credibility towards customers that would be otherwise impacted by the greater number of system failures that can manifest during acceptance tests on the field. To actually get these benefits, companies are required to invest into the testing phases from a twofold perspective. First, testbed that mimic the real operational scenarios must be set up in order to reproduce system working condition as much as possible. Second, massive testing campaigns, aiming at checking both functional and nonfunctional requirements, must be performed despite of the great effort (in terms of infrastructures and human resources) that this may require.

This work aims at illustrating the SELEX-SI strategic roadmap that has been developing to get these challenges and that is mainly based on cloud computing investments. Actually, setting up an extended enterprise private cloud computing scenario, would allow the company to i) reproduce the real world scenarios, that usually encompass systems distributed over more operational premises, e.g., several ATC centers belonging to the same system and deployed over different cities in a given country, and ii) set up testbed platforms to perform distributed testing campaigns from different premises thus reducing people mobility costs. The paper discusses the technical and technological investigations that have been started to sustain and argument this intuition, providing preliminary results and illustrating the next future steps.

2 Rationale and motivations

Getting the great challenge of reducing costs and improving the dependability and quality of delivered software systems, requires both methodological and technological investigations with the ultimate aim of developing a service oriented infrastructure to provide people working over different premises with remote facilities to perform testing and integration activities from anywhere and at any time. Different teams would be allowed working on the *same* platform, reducing lots of inefficiencies that actually hold mainly in terms of:

- Different platforms management overhead;
- Effort replication for installation and configuration in different premises;
- Need to move people from one site to another to work on different *physical* testbed;
- Static allocation of plenty of resources against the same system/task;
- Need to purchase as many hardware and software infrastructures as the testbed platforms;
- Maintenance costs increase due to the number of platforms;
- Need to train employers, with very specific skills, on any site to assure productive work and prompt support in case of problems.

2.1 EXPECTED BENEFITS

In order to minimize the process lacks discussed above, and the impact they have on productivity and costs, SELEX-SI has been investigating on both technological and architectural solutions, able to provide the following tangible benefits:

- Get the most of resources usage, to avoid they to get under loaded, aka “Too many servers for too little works!”;
- Maximize resource sharing in an heterogeneous environment, even enabling remote facilities (always keeping security and performance requirements in mind);
- Increase hardware and software support tools usability;
- Minimize inconsistencies among different testbeds;
- Minimize configuration and installation effort;

Through the activities and the investigation described in this work, the company does expect to identify the best way for getting all these benefits at a reasonable cost and in a medium term time horizon. Cloud computing paradigm looks the most promising solution to keep this goal. Furthermore, the availability of a plethora of Open Source platforms, would contribute to the overall costs reduction by keeping infrastructure setup, development and implementation costs much lower than any commercial solution.

The following section is devoted to provide basics on cloud computing, as well as to illustrate the technological roadmap followed by SELEX-SI in this direction.

3 Towards the cloud

SELEX-SI has been working and investigating in this field for years, passing through several technological options and methodological approaches that never revealed to be winning. As for example, dedicated automated testing environments have been setup over years for specific systems, to reduce costs or agile development and testing attempts have been made sometimes. We approached at cloud computing to define a one-for-all solution that can be used in every domain and applied to all the systems and solutions provided by SELEX-SI worldwide.

3.1 CLOUD COMPUTING BASICS

Cloud Computing (CC) is a paradigm having recent and growing popularity. Aim of this paradigm is to provide IT resources (such as computational resources, software components and storage resources) as services delivered trough the network, hiding in such a way the sophistication of the underlying infrastructures.

CC allows to manage resources in a *pay per use* way, and it guarantees the dynamic allocation of such resources against the current load requirements of the overall infrastructure (both in terms of users and operative load). As support for CC, the system-level virtualization techniques realize an abstraction of physical resources by multiplexing them in several virtual resources. The virtualization Hypervisor is a software layer having the goal to manage a number of Operating Systems on a single physical node, allowing in this way to optimize the use of the available resources. This is the basic principle to implement the so called *cluster consolidation* technique, which has its natural evolution in the CC concept .

3.2 SOLUTIONS FOR SYSTEM-LEVEL VIRTUALIZATION

First step in the our technological roadmap has been the selection of a proper Open Source solution for system-level virtualization, being this mechanism the main brick to realize a private CC infrastructure. The results we obtained from the analysis and the comparison among solutions coming from a scouting campaign, led us to choose KVM (Kernel-based Virtual Machine), a full virtualization solution for Linux on x86 hardware containing virtualization extensions.

3.3 BUILDING THE CLOUD

Exploiting the aforementioned technologies, our final goal is to configure the corporate sites in Rome, Fusaro, Genova and Giugliano (both SELEX-SI and SESM lab) in order to build the cloud shown in Figure 1.

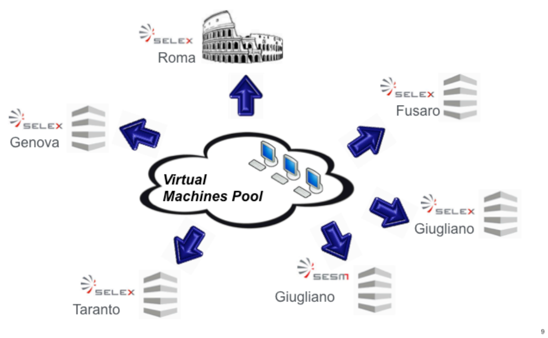


Figure 1 - SELEX-SI Cloud, a logical view

The remote availability of hardware and software resources physically installed into the sites, is guaranteed by a dedicated network infrastructure configured as single backbone among the sites, and by Web interfaces providing entry points for the available services. The management of the overall CC infrastructure is in charge of the technical staff working at the Rome site, while the management of each single site platform is in charge of the technical staff working at such site (see Figure 2).

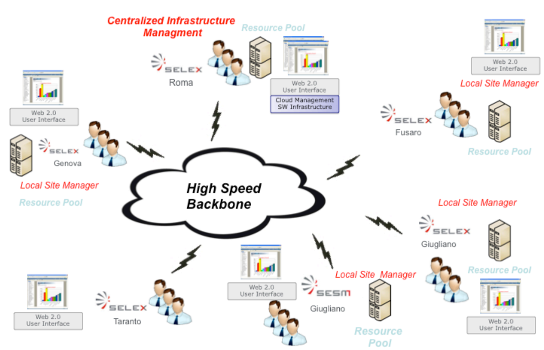


Figure 2 – SELEX-SI Cloud, physical view and access channels to the resources

In order to observe cost and performance requirements, combinations of enterprise and Open Source solutions have to be taken into account, both concerning the site level and the infrastructure level. More in detail, the following alternatives may be considered:

- Site Level (SL)

SL1. Proprietary SW: proprietary solutions for both the virtualization system and the CC platform

SL2. Hybrid : Open Source virtualization sysem and proprietary CC platform, or vice versa

SL3. Open Source: Open Source solutions for both the virtualization system and the CC platform

- Corporate Level (CL)

CL1. Proprietary SW: the overall infrastructure is based on proprietary solutions

CL2. Hybrid: the SELEX-SI sites are managed by proprietary solutions, while the SESM Lab site is managed by Open Source solutions

CL3. Open Source: the whole corporate infrastructure is based on Open Source solutions

The final decision, will be driven by the system application requirements as well as from company management indications in this direction. However, (SL3, CL3) is the final point we wish to reach, even if intermediate "mixed" alternatives will be applied for a while to allow the complete transition from legacy enivronments to the cloud.

4 Experimental campaigns

The experimental session we performed has the goal to verify the robustness of the virtualization hypervisor we selected. In order to reach this goal we chose a reference class of applications, namely applications for automatic testing of software components in the Air Traffic Control and the Surveillance of Battlefields domains. Later in this section we refer such kind of applications as V&V (Verification and Validation) applications.

4.1 AIMS AND ORGANIZATION

The approach we devised and implemented consists in using virtual machines as execution environments for V&V applications, and in generating virtual machines to increase the computational load for physical machines on which the KVM Hypervisor is running. In the following we refer the virtual machines belonging to the first class as V&V VMs, and those belonging to the second class as Load VMs.

The virtual cluster we realized is then composed by both kind of VMs, while the whole testbed is accordingly made-up of a number of physical hosts and a number of virtual hosts.

We aim at determining and evaluating the behaviour of each element of such testbed. More in detail, we need to assess the following conditions:

1. No problems reported in the behaviour of physical machines running the KVM Hypervisor while increasing number and

computational load of the hosted virtual machines;
2. No differences found in the behaviour of Virtual Machines running different Operating Systems;
3. No hangs and no crashes manifested on both virtual and real machines;
4. No problems reported during the execution of V&V applications on Virtual Machines.

The following section describes the used tools and the overall organization of the testbed.

4.2 THE TESTBED

The testbed is made-up of physical machines hosting the KVM Hypervisor, and a virtual cluster composed by VMs having different purposes. These VMs present also different configurations in terms of Operating System (we used Windows XP, Windows Server 2003, and several Linux distributions), RAM, CPU slot, Disk.
The physical hardware we exploited consists in 8 Dell Power Edge M600 blade with the following configuration:

- CPU: 2 * QUAD-CORE Intel Xeon E5420, 2.5 GHz
- RAM: 32GB
- Network: 4* Gb Ethernet NIC
- Operating System : CentOS 6 64-bit

In the following we refer to such nodes with the label *CentOS-KVM*.

As stated above, the reference class of software we selected, consists of applications for automatic testing of software components belonging to the Air Traffic Control field and the Surveillance of Battlefields field. In particular, we chose as V&V applications for our experiments two frameworks for the automatic testing of the two systems respectively[1]:

- YYY
- XXX

Both frameworks present several components; while the former requires a single VM (Win2003_YYY), each component of the latter runs on a single VM; in the following the structure of the XXX automatic testing framework:

- XXX-ClientTST02 (Windows XP)
- XXX-GIS-TST02 (Windows Server 2003)
- XXX-CS-TST02 (Windows Server 2003)

Regarding the Load VMs, instead, in order to stress the virtual devices they are equipped with, we selected and exploited ad hoc applications; these last are classified per Operating System and listed in the following :

Windows platform:

- MemAlloc to stress the Memory
- Core Damage to stress the CPU
- DITG to stress the Network

Linux platform:

- The *stress* Linux command to stress both the memory and the CPU
- The *iperf* traffic generator to stress the network

Since we aim to determine and evaluate the behaviour of both the physical nodes on which the KVM Hypervisor is running, and the VMs composing our virtual cluster, one important task is the monitoring of the overall testbed.
In order to reach this goal we chose the Nagios infrastructure monitoring system. We installed and properly configured the server side on a VM, and we created a number of plugins by modifying some scripts available on the Nagios Excange database, to monitor the remote Linux hosts. Concerning the Windows VMs, we used the NSClient application and the Windows performance counters.
A list of the activities we monitor with Nagios is listed below :

- CPU Load
- Disk Status
- Memory Usage
- Network Transfers

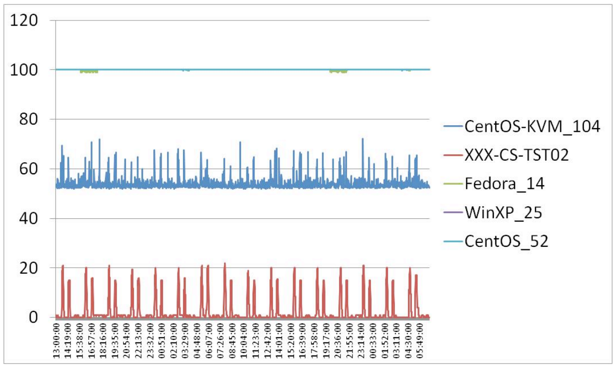
The schema we implemented is then enriched by the Convirture Enterprise-class management system, and an OpenFiler iSCSI SAN. The Convirture management system is equipped with a very friendly web interface accomplishing the organization and the management of VMs composing the virtual cluster. Convirture is itself running on a dedicated VM. Concerning the SAN, OpenFiler is an Open Source storage management appliance that we use to manage a 12 Terabyte storage space RAID-10 configured.

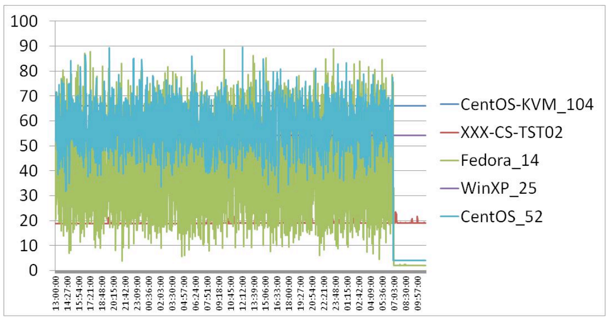
5 Preliminary results

The experimental campaign has been splitted in two phases. The former aims to verify the first three points listed in section 5.1; the latter to demonstrate the last one. In the first phase, the resources composing the overall testbed have been used, both virtual and real. More in details:

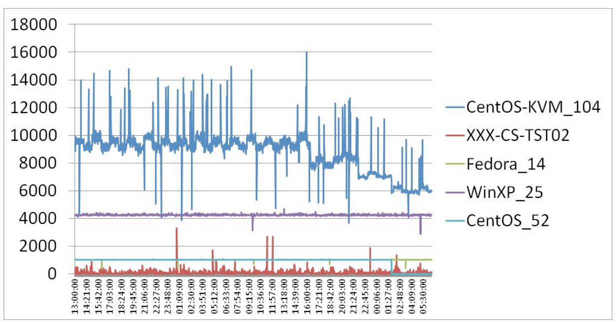
- 8 physical nodes (one of them hosts management and monitoring applications, the other host VMs);
- 3 V&V VMs composing the XXX testbed;
- 1 V&V VM composing the YYY testbed;
- 18 Load VMs; each of them is characterized by a particular configuration.

[1] System names are omitted for the sake of information confidentiality





b) Memory Usage



c) Network Transfer

Figure 3 – First phase

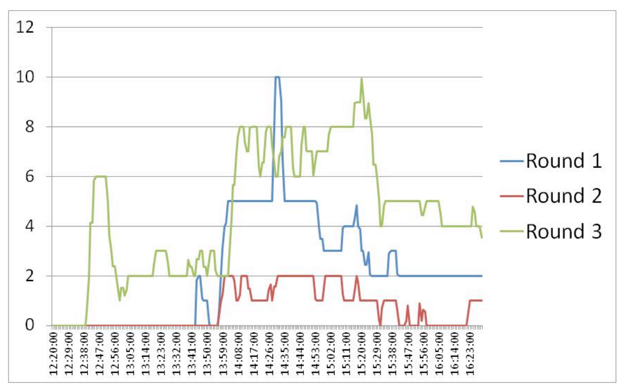
The duration of such experiment has been fixed in 48 hours.
The data we collected during the experiment is related to the main metrics we took into account, namely CPU Load, Memory Usage, Network Transfer. In Figure 3 we present for each metric the data related to one physical node and the relative hosted VMs. The first two metrics express a percentage, while the last metric is in Kbps. Time on the x axe.
The graphs in the figure show that the behaviour of the physical nodes meets our needs in terms of scalability, robustness, and support of different Operating Systems. Values related to such nodes never overcome properly selected thresholds, and this demonstrates that the nodes work fine also when the computational load on the hosted VMs (and then on the physical node itself) increases.

Concerning the VMs behaviour, no hangs and no crashes occurred on both Linux and Windows platforms, even though a heavy computational load stressed the core components of such machines.

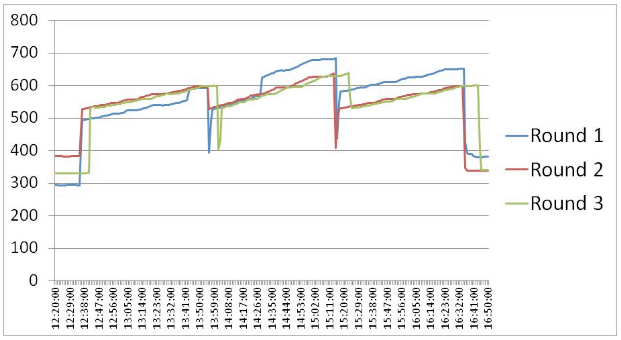
The second phase is in turn splitted on three rounds having each one a duration of 4 hours. Only three physical nodes have been used to perform such second phase, except for the node hosting the management and monitoring applications.
What changes among the three rounds is the number of VMs hosted on each physical node. The main goal of this experiment is to verify that the V&V applications (hosted themselves on VMs) do not suffer the increasing computational load (consisting of VMs) on the physical nodes.

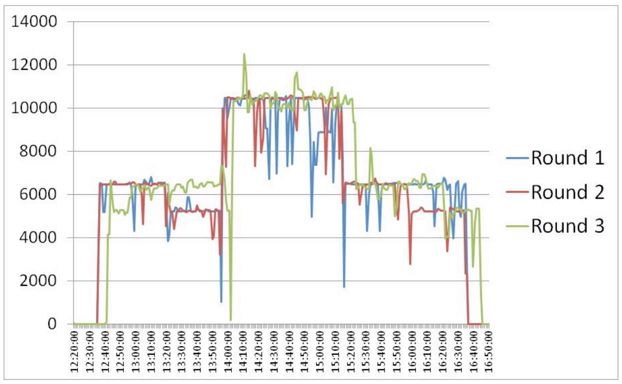
In Figure 4 we present for each metric the data related to one V&V VM (Win2003_YYY); each graph presents a comparison among the lines related to each round of the experiment. The first metric (CPU Load) express a percentage, the second one (Memory usage) is in MB, while the last metric (Network Transfer) is in Kbps. Time on the x axe in this case too.

The graphs show that the behaviour of the V&V VM under test (Win2003_YYY) is the same in each of the implemented rounds. This is a demonstration of what we need to verify: the increasing load on the physical nodes does not impact the performance of the applications running on the hosted VMs.



a) CPU Load





c) Network Transfer

Figure 4 – Second phase

6 Lessons learnt and future work

The need for improving the efficiency of software V&V process, in terms of costs and product quality, has been driving the investigation and the experimental campaigns described in this paper. Although the work is at a very early stage, some interesting considerations emerged from this first bundle of experiments. First, they confirmed the intuition that virtualization is the best way to get the most of your hardware infrastructure if you set it up properly. However, what made we happy with results is the good behavior we get from an Open Source hypervisor, namely KVM. Second, the platforms we are going to select are ready to support the kind of load that is typical of SELEX-SI application scenarios, without any impact on performances and improving the efficiency of real world testing campaigns. This was the *condicio sine qua non* we wanted to assess before going towards CC platform setup that, actually, represents the second part of the story. What we plan to do actually in the next future is to:

1. Enrich experimental campaigns on different applicative workload, even more demanding in terms of number of test procedures and computational load.
2. Run the same experiments described here, and the ones we are going to do more, on top of a commercial hypervisor to compare the performance results;
3. Start a CC platform scouting, on top of KVM, to select the one in charge of meeting the requirements of both the hardware we have and the applications we must address. Actually this is a task already running and interesting feedback are coming from OS CC environment like OpenNebula, OpenStack and Cloud Stack;
4. Start performing distributed campaigns on at least to different SELEX-SI premises to demonstrate the real effectiveness of the approach.

To conclude, the results gained and illustrated in this work show that KVM is a valuable Open Source alternative to build up a CC environment against both the performance and application requirements exposed by SELEX-SI scenarios. This paves the way to the actual development of a CC platform aimed at supporting the company V&V processes and optimizing costs.

References

[1] P.M. Chen, B.D. Noble. When Virtual is Better Than Real. Proceedings of the 8th IEEE Workshop on Hot Topics in Operating Systems. Schloss Elmau. Germany. May 2001.

[2] M. Armbrust, A. Fox, et. al. Above the Clouds: A Berkeley View of Cloud Computing. Technical Report No. UCB/EECS-2009-28. University of California at Berkley, USA. Feb. 10, 2009.

[3] Q. Zhang, L. Cheng and R. Boutaba. Cloud Computing: state-of-the-art and research challenges. Journal of Internet Services and Applications, Volume 1. Number 1, 7-18. Springer 2010.

[4] R. Canonico, P. Di Gennaro, V. Manetti, and G. Ventre. Virtualization techniques in network emulation systems, volume 4854/2008, Euro-Par 2007 Workshops: Parallel Processing of Lecture Notes in Computer Science, pages 144-153. Springer Berlin / Heidelberg, 2008. ISSN:0302-9743, ISBN:978-3-540-78472-2

[5] A. Shah. Kernel-based virtualization with kvm. Linux Magazine, vol. 86, pp. 37{39, 2008.

[6] A. Botta, A. Dainotti, A. Pescapè. A tool for the generation of realistic network workload for emerging networking scenarios. Computer Networks (Elsevier), 2012.

[7] W. Barth. Nagios, System and Network Monitoring. No Starch Press, u.s (2006)

[8] B. Childers. Openfiler: an Open- Source Network Storage Appliance. Linux Journal 2009. ACM Digital Library. Retrieved on June 22, 2009 from http://www.linuxjournal.com/ article/10414. appliance. Linux J., 2009(181):6, 2009.

[9] http://www.soft.tahionic.com/download-memalloc/index.html

[10] http://damage.vigilantesoftware.com/

http://www.nsclient.org/

Automated context aware composition for convergent services

Armando Ordonez, Juan C. Corrales
Telematics Engineering Group
University Of Cauca
Popayán, Cauca, Colombia
{jaordonez,jcoral}@unicauca.edu.co

Paolo Falcarin
ACE School
University of East London
London, UK
falcarin@uel.uk

Abstract – *Automated Services composition is an active research area nowadays, most of the approaches are based in Artificial intelligence techniques. However, most of these approaches focus on specific steps of Web services composition and lacks of details for general application in broader fields such as Convergent Services. In the present work, an architecture based on Natural Language analysis for AI planning processing and automatic deployment in JSLEE is presented. The preliminary experiments show promising results.*

Keywords: NLP Processing, Automated composition, automated deployment.

1 Introduction

Combining services and technologies of Web with traditional Telecommunication services is known as convergent composition. As the number and dynamics of services grow, it becomes unfeasible for human capacity to perform this process manually. Automated composition in Web domain has been attached widely [1],[2]. However, some technical details remain unsolved for application in convergent environments.

Composition of convergent services is fundamentally different from Web services composition [3]. The highlighting differences are: i) End-users prefer to personalize the composed services. ii) Mobility provided for different devices require special treatment of services provisioning. ii) Including telecommunication services require high reliability and fault tolerance in the composed service execution.

Previous works deal with this problem and present some techniques and architectures; some of these works comes from European projects like SPICE [4], OPUCE [5] and OMELETTE [6]. These approaches do not automate the whole service composition process; besides, they lack of ways for expressing user request through voice, finally, they do not include user context information in plan generation.

On the other hand, some approaches from Web domain propose to include preferences in the automated Web Services composition. The inclusion is done through extension of planning languages such as PDDL (Planning domain definition language) and the use of particular planners [7]. The planning domain specification is a complex process that can become prohibitive. The above is due to the fact that planning domain creation implies reality modelling of the whole domain. In the Convergent composition it implies that all the services (Web and Telco) must be described in semantic format to be translated to planning domains.

In this context, the main contributions of our research work are: (1) Presents an architecture for automated services composition considering issues associated with convergent composition for the whole process, (2) Describes a metamodel for user context including device, preferences and situation profiles. (3) Depicts a mechanism for knowledge representation of user request from Natural language to PDDL including control flow information (4) proposes a technique for planning domain feeding based on expert made mashups (5) Propose a novel mechanism for automatic deployment in JSLEE environments.

This paper is organized as follows: section 2 describes the whole architecture of the system. Section 3 presents the Natural Language module for translating user requests to PDDL. Section 4 describes the Automatic Deployment mechanism for JSLEE environments from planning outputs. Section 5 presents the related work and Section 6 draws the conclusion and future work.

2 Global view of the architecture

Figure 1 depicts a global view of the architecture. In order to ease the interaction with the system, the user can introduce the request through their voice from his mobile. In the literature other alternatives have been found for user request treatment, such as Mashups [8] and Services creation environments [9]. However the natural language offers a better mechanism for end users without expertise to express their requests [10].

The *Natural Language Analysis* decomposes the request in constitutive parts. Equally, infer semantically which words are verbs (possible actions), nouns (possible parameters), control flow or context information. For context analysis, the present approach uses three dimensions: user preferences, device capabilities and situational context.

Knowledge Representation module makes a translation from the processed request into a problem domain in PDDL. PDDL is the standard language for the encoding of the planning domains and problems.

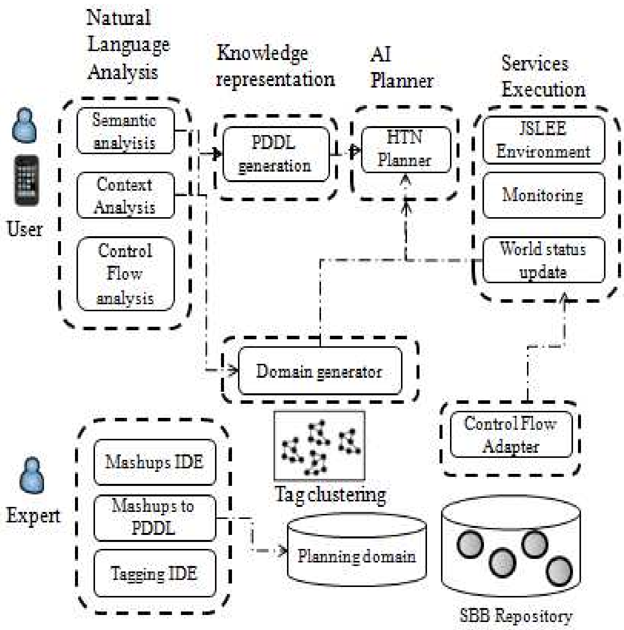


Figure 1. Global view of the architecture

The ranking is created using a *Planning Domain Repository*. In AI Planning based Web services composition, the planning domain is a set of actions (representing services) that could possibly participate in the composition plans. The further planning process will select the suitable services creating thus the composed service

HTN planners have shown a good performance in Web services composition. However one of the biggest drawbacks for HTN planning is the difficulty for create planning domains. In the present architecture, the planning repository is fed by a set of mashups translated into PDDL domains. Those mashups are created by experts. These experts create functional compositions and tag them according to their expertise. Besides, the experts can include non functional properties to the services description.

The *Domain Generator* module creates dynamically a ranking of planning domains. In order to do so, the *Domain Generator* analyses the tag clustering according to the user request and creates a subset of plans that are used in the planning process.

This dynamic generation creates a subset of the whole planning domain based in tagging process. By creating a subset of the planning domain, the search space is pruned for the further planning generation. This represents a reduction of the planning computational time, and an easy way to include new services and preferences.

The *AI planner* receives as input the problem and planning domain and creates a set of ordered task representing the real composition plan. This plan is sent to the *Control flow Adapter*. The *Control Flow Adapter* takes as input the words from the control flow subset identified in the *Natural Language Analysis.* Next, the *Control flow Adapter* matches the words subset with a set of control flow patterns: Sequence, Parallelism, Exclusive Decision, Multiple Decision, and Looping. This module select precompiled SBBs (JSLEE software components) and perform a code injection to perform in execution time the deployment of the new composed service. The latter is possible since planning process is performed with descriptions of implementation services deployed in the JSLEE environment. Given the above, the *Control Flow adapter* just executes a linkage of the running services.

The *Services Execution* module is continually monitoring the services and updating the world status (the set of variables representing the reality in the planner). If something goes wrong, the planning process is performed again starting from the current world status.

The above architecture can be modified including interaction with the user in different steps. This interaction drives to a better control from the user and a lower automation of the process.

3 Natural Language to PDDL

This transformation is performed by three modules: the *Natural Language Analysis*, The *knowledge Representation* and the *Domain Generator*.

The *Natural language analysis* module receives as input the user request, and generates a set of terms in first order logic described in PDDL. The first element of the process is the *Semantic Analysis*. This analysis is performed through the following steps: Initially, a user makes a request from his/her mobile device in NL, which is received by the *Tokenizer*, where tokenization operation starts, it obtains simple lexical units from complex sentences, by removing existing spaces. Additionally, this module corrects simple lexical errors that may arise in the request, i.e., misspelled word errors that are easily identifiable. Afterward, the sentences are processed through *Filter*

Words Module and later they pass through *Words Tagging*, with which, it is pretended to classify (tag) words of the sentence according to their grammatical category. The module also aims to undertake an analysis based on linguistic rules, trying to identify and compensate syntactic and structural errors. Some techniques used to implement the modules above are GateNLP, OpenNLP, Apache UIMA, among others. One of the most important is GateNLP [11], which offers an architecture that contains functionality for plugging in all kinds of NLP software: (POS taggers, sentence splitters, named entity recognizers) and all are java based.

Once completed these operations, the request is more consistent, but remains complex. Therefore, the *Semantic Analysis* module performs a classification between "*Control*", "*Functional*" and "*Situational*" words according to its meaning. In order to do so, this module identifies the correct sense of words according to their context, i.e., identifies the correct one from a word within multiple meanings that can occur in a sentence. This allows an easy identification of keywords with their respective grammatical category (e.g., noun, verb, adjective, conjunction) that define the user's request and with which domain selection will be made. Thus, this stage offers the user greater flexibility in the use of language and allows the establishment of a wider range of possibilities. On the other hand, this phase also identifies conditional words identification (e.g., *if*, *then*, *later*) and words of order (sequence) (e.g., *first*, *second*) important for the *Generic Flow Generator* in the inference phase.

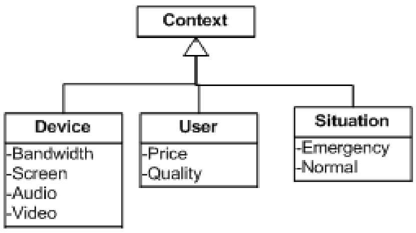


Figure 2. Dimensions of user context

The *Context Analyzer* classifies the information according to three dimensions: device, user and situation. Each one of them gets the information from different sources and defines the selection of domains.

The *Device* dimension gathers the information from the device and the network. Device references are checked in capabilities repositories like the Wireless Universal Resource File (WURFL) and the Composite Capability/Preference Profiles (CC/PP) in order to analyze devices capabilities and features.

The *User Profile* gathers the information from the user ID and the preferences repository in the system. Finally, the *Situation* D*imension* takes information from Natural Language request. The latter situational dimension is activated when the user request contains specific words such as: *"urgently"*, *"emergency"*.

The relation between the dimensions is calculated using a preferences function, assigning weight to each one of them. Consider a user connected through a Smartphone with video capabilities. This user has registered low cost preference; therefore the cheaper service using SMS or voice is selected because. The user price preference has a higher weight in preferences calculation. On the other hand, if the System detects an emergency situation, the most reliable services are selected no matters the price.

The planning domains selection is performed as follows. The words tagged as nouns in the *Semantic Analysis* are compared with the high level task in the planning domain repository. To do so, the semantic matchmaking is performed using the *Tags Clustering*. Tags clustering define a classification of words assigned by experts. Figure 3 shows a simplified scenario where three services (high level tasks in HTN) represented by ovals, are tagged with related words. In this example, if the user expresses "contact" in his request, two services are selected: "Call" and "Send SMS"

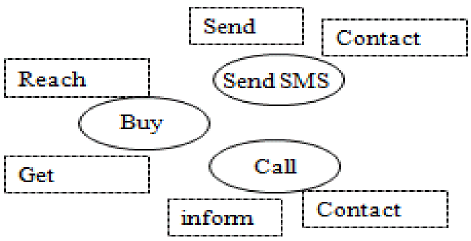


Figure 3. Global view of the architecture

The second filter is done used the context information. To do so, the request is translated into a first order logic (see table 1).

Table 1. NL to PDDL transformation (Simplified)

NLP output	PDDL
noun (parameter)	precondition
verb (service)	task name
noun (preference)	precondition

For example, if the request contains "contact Mark urgently", the systems translate the request:

```
(define (problem contact)
    (:objects person quality )
    (:init (person Mark)
```

(quality high))
(:goal (call Mark)))

The problem definition and the dynamically generated planning domain are the input for the AI planner.

4 Automated deployment

In Telecommunication domains, high performance environments are mandatory; in order to warranty high availability and Fault tolerance. Most of the proposals for Next Generation Networks include application servers as core element of the architecture [12]. Traditionally most of the related works in the area of automated planning for Web Services composition lack of details for implementation in application servers. On the other hand, high reliability servers require that most of the components are compiled for optimal performance. The process of automated composition is not complete unless a smart mechanism for automatic deployment is included [13].

JSLEE stands for Jain SLEE [14] and is a robust server for event driven execution. The main components of JSLEE architecture are the SBBs. The standard defines that the components are composed of Java compiled classes with additional descriptors. Fortunately, JSLEE defines a hierarchy of classes allowing interaction between SBBs with the purpose of reach SBB composition.

In the present architecture, the planning is done using implementation services descriptions instead of abstract descriptions. This approach leads to a high number of possibilities. However, the planning domain is based on mashups previously tested and tagged. The latter avoids the needing for parameters matchmaking and heavy semantic descriptions that could be time consuming. This way the planning is focused in adaptation and composition of previous processes templates.

Each service in the Mashup, and consequently in the composed plan has a representation in the JSLEE environment. Thus, the control flow is on charge of articulate all the functionalities according to the user request. To do so, specific words from the natural language analysis are matched with the patterns of behaviour between member services in a composition (see table 2). As the exact word is commonly not found in the request, the words tagged as control flow words are analyzed using wordnet in order to find related terms.

Each one of these patterns is coded in a special controller SBB. The controller SBB performs a runtime Java class loading to create the composed service. Preliminary tests of this module are performed using Mobicents distribution. This idea was previously tested by Lehmman et a.[13]

Table 2. NL to PDDL transformation (Simplified)

Pattern	Example word
Sequence	And
Parallelism	Both
exclusive decision	If
multiple decision	Both
Looping	While

5 Related work

Our approach is focused on automation of user centred service composition, and we intend to apply our framework to convergent environments. Previous works have proposed frameworks for automated services composition, like Kim et.al [15] . The authors present phases for automatic composition, focusing only on web domain without concern on execution phase.

Shia et al. [16] and da Silva et al. [17] present frameworks for automatic composition. These frameworks exploit natural language processing and semantic annotations for services matchmaking based on SPATEL language. They do not address the validation of the non-functional properties and are focused only on the request analysis and plan generation. Our approach deal with all the phases including execution and reconfiguration based on JSLEE environments.

Sirin et al. [18], provide an algorithm to translate OWL-S service descriptions to a SHOP2 domain and makes planning based on services. However, this works lacks of details of implementation in real environments and focus only on Web services. Other authors present approaches for Web service composition based on AI planning with preferences [19] [20]. However, they propose extensions to standard planning language and adaptation of planners, adding new levels of complexity to the automated composition process.

Zhu et al. [21] present Hybrid Service Creation and Execution Environment (HSCEE), a Template-based service creation platform with low latency service execution. HSCEE is based in on BPEL templates but most of the design tasks are manually. Equally natural language processing is not considered.

Table 2 outlines a comparison of related work, based in the following criteria: first, if the work deals with all the phases in the service composition process including reconfiguration. Second, if the approach for service composition is user-centred. Third, if the approach takes in account convergent considerations.

TABLE I. COMPARISON OF RELATED WORKS

works	Include all phases	User centred	Consider Convergence
[15]	No, just the request analysis and the service creation	No	No, they focus on web domain
[16] [17]	No, just the request analysis and the service creation	No	Yes
[18] [19][20]	No, just the OWL-S based planning	No	No, focused on Web domain
[21]	Yes, but not all of them are automatic or Natural language based.	Yes	Yes

6 Conclusion and future work

Automated convergent composition is a very intensive research area; previous works have worked on some aspects of this process. However, there is not a complete framework to develop this process. None of the above proposals presents details to apply proposed frameworks to Telco environments in a real environment, the works presented by Shia y da Silva et al. [16] [17] are the most relevant for us in the literature, many elements are similar to our work but our approach has a different direction. They deal with complex treatment of SPATEL language and ontologies in order to reach automated User centred composition. We deal with user profile information to customize AI planning. Equally, our approach is tending to reach low execution times and include mechanisms for automated reconfiguration. This paper describes the general components of architecture for automated services composition applied to environmental management. Specific elements of each module are detailed previously [22][23]. The present architecture performs an analysis of the request in Natural Language and translates it to PDDL. Likewise, the architecture includes mechanism for context consideration and automated deployment and mashup based planning domain creation.

In the near future, we want to research and propose on developing of mechanisms for automation of planning domain creation, and experimentation with other planners in order to consider reconfiguration in different phases of the process and better execution times. Equally we are extending the preferences criteria in order to get a better personalized experience for the user.

References

[1] K. Sycara, M. Paolucci, A. Ankolekar and N. Srinivasan, “Automated discovery, interaction and composition of semantic web services, “ Web Semantics: Science, Services and Agents on the World Wide Web, Vol. 1, No. 1, pp. 27 -46, 2011.

[2] P. Bertoli and M. Pistore and P. Traverso, “Automated composition of Web services via planning in asynchronous domains, “Artificial Intelligence, Vol. 174, No 3-4, pp. 316 – 341, 2010.

[3] G. Bond, E. Cheung, I. Fikouras and R. Levenshteyn, “Unified telecom and Web Services Composition: problem definition and future directions,” Proc. 3rd International Conference on Principles, Systems and Applications of IP Telecommunications, pp. 1-12, Atlanta, 2009.

[4] P. Falcarin, “Service Composition Quality Evaluation in SPICE Platform,” High Assurance Services Computing, Springer US, pp. 89-102, 2009.

[5] J.C. Yelmo, J.M. del Alamo, R. Trapero, P. Falcarin, Y. Jian, B. Cairo and C. Baladron, “A User-Centric service creation approach for Next Generation Networks,” Proc. Innovations in NGN: Future Network and Services (K-INGN), May 2008, pp. 211-218, doi: 10.1109/KINGN.2008.4542268.

[6] Universidad politecnica de Madrid, “Initial discovery and composition,” Technical Report D5.2, Omelette project, 2011.

[7] J. Baier and S. McIlraith, “Planning with Preferences,” AI Magazine, vol. 29, pp. 25-36, 2008.

[8] Z. Zhao, S. Bhattarai, J. Liu, Ji and N. Crespi, “Mashup services to daily activities: end-user perspective in designing a consumer mashups,” Proc. 13th International Conference on Information Integration and Web-based Applications and Services, pp.222-229, Ho Chi Mihn, December 2011.

[9] N. Laga, “Widgets and composition mechanism for service creation by ordinary users”, IEEE Communications magazine, Vol. 50, Issue 3, pp. 52-60, March 2012.

[10] K-H. LEE, J. Lim, “Constructing Composite Web Services from Natural Language Requests,” Journal of Web Semantics, Vol. 8, No. 1, 2010.

[11] H. Cunningham, V. Tablan, I. Roberts, M. A. Greenwood, and N. Aswani, “ Information Extraction and Semantic Annotation for Multi-Paradigm Information Management” M. Lupu, K. Mayer, J. Tait, and A. J. Trippe, editors, Current Challenges in Patent Information Retrieval, Vol. 29. Springer, 2011.

[12] C. Chrighton, “JAIN SLEE vs SIP Servlet Which is the best choice for an IMS application server?,” Proc. Australasian Telecommunication Networks and Applications Conference, 2007. ATNAC 2007. pp.448-453, Christchurch, Dec. 2007.

[13] A. Lehmann, T. Eichelmann, U. Trick, U, R. Lasch, B. Ricks and R. Toenjes, “TeamCom: A Service Creation Platform for Next Generation Networks,” Proc. Fourth International Conference on Internet and Web Applications and Services. ICIW 2009, p.p. 12-17, Venice 2009.

[14] JSR 240, JAIN SLEE vl.1 Web Site, http://jcp.org/en/jsr/detail?id=240

[15] A. Kim, M. Kang, C. Meadows and J. Sample, “A Framework for Automatic Web Service Composition,” Technical Report, Naval research Lab USA, 2009.

[16] M. Shiaa, P. Falcarin, A. Pastor, F. Lécué, E. Silva and L. Ferreira, “Towards the Automation of the Service Composition Process: Case Study and Prototype Implementations,” Proc. ICT Mobile Summit, pp. 1-8, Stockholm, June 2008.

[17] E. da Silva¸ L. Ferreira, M. van Sinderen, “Towards runtime discovery, selection and composition of semantic services,” Computer Communications, vol. 34, pp.159-168, Feb. 2011.

[18] E. Sirin, B. Parsia , D. Wu, J. Hendler, D. Nau, D, “HTN Planning for Web Service Composition using SHOP2,” Proc. Web Semantics: Science, Services and Agents on the World Wide Web In International Semantic Web Conference 2003, Oct. 2004.

[19] N. Lin, U. Kuter, E. Sirin, “Web Service Composition with User Preferences,” Proc. European Semantic Web Conference (ESWC2008), Springer-Verlag, May 2008.

[20] S. Sohrabi, N. Prokoshyna and S. McIlraith, “Web Service Composition via the Customization of Golog Programs with User Preferences,” Conceptual Modeling: Foundations and Applications, Springer-Verlag, pp. 319-334, 2009.

[21] D. Zhu, Y. Zhang, B. Cheng, B. Wu, and J. Chen, “ HSCEE: A Highly Flexible Environment for Hybrid Service Creation and Execution in Converged Networks,” Journal of Convergence Information Technology, Volume 6, Number 3. March 2011.

[22] E. Pedraza, J. Zúñiga, L. Suarez, J.C. Corrales, “Automatic Service Retrieval in Converged Environments Based on Natural Language Request,” Proc. The Third International Conferences on Advanced Service Computing, pp. 52-60, Rome, Sep. 2011.

[23] A. Ordonez, J.C. Corrales, P. Falcarín, J. Guzman, “User Centred Automated Composition in Telco 2.0,” The Fourth International Conferences on Advanced Service Computing. Nice, July 2012 (in revision).

Proc. of the 2012 7th International Conference on System of Systems Engineering, Genoa, Italy - 16-19 July 2012

Executable System-of-Systems architecting based on DoDAF Meta-model

Long Li[1] eric.longlee@gmail.com
Yajie Dou[1] wdxhxcbdyj@163.com
Bingfeng Ge[1,2] bingfengge@nudt.edu.cn
Kewei Yang[1] kayyang27@hotmail.com
Yingwu Chen[1] ywchen@nudt.edu.cn

[1]Department of Management, National University of Defense Technology, Changsha, China

[2]Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada

Abstract - *Methods used to gain executable architecture are all based on transforming view models to executable models, which can not take the advantage of "data-centric" approach adopted by DoDAF2.0. A novel executable architecting approach transforming DoDAF Meta-model (DM2) to executable model directly is proposed, which can overcame the limitation of traditional approaches that information consistency and data completeness may be damaged. An example to analyze Command and Control (C2) capability in Air Defense System-of-Systems (ADSoS) is provided in detail to illustrate the process to transform DM2 to Coloured Petri Net (CPN). Besides, the input and output data is also given in detail along with the process illustration.*

Keywords: *system-of-systems; executable architecting; DM2; CPN*

1 Introduction

SoS is composed by multiple component systems, which are usually geographic distribution, operational and managerial independent. Besides, SoS usually have emergent behavior. Those characteristics make it very different from traditional systems [1][2]. Multi-view method is verified as an effective way to facilitate capability planning and developing of SoS. Multi-view method can be used to create conceptual models describing specific roles and interfaces of component systems, which ensure the possibility that disparate component systems can collaborate to achieve general capability. In military, as defense communities continue to face a critical challenge to integrate multiple capabilities across heterogeneous developing and legacy weapon systems [3]. Department of Defense Architecture Framework (DoDAF) which was issued by the U.S. Department of Defense provides a standard modeling framework of SoS architecture for not only military but also other application areas [4]. The latest version of DoDAF (DoDAF 2.0) [5] is published in 2009. Compared to previous versions, DoDAF 2.0 incorporates a "data-centric" approach, which focuses on architectural data, rather than on developing individual products as described in previous versions. Data collection, store and maintenance turn into the major work in SoS architecture development. Data in view models must be in accordance with DM2 concepts, associations, and attributes. The conformance ensures the reuse of information, architecture artifacts, models and viewpoints.

DoDAF is only a conceptual modeling framework of SoS, which can not facilitate behavior validation and performance evaluation. However, SoS can be treated as networked information system. Information systems are dynamic in nature. Events occur that trigger the execution of functions. An executable model of SoS architecture enables the architect to analyze the dynamic behavior of the architecture, identify behavioral errors not easily seen in the static descriptions, and demonstrate to the customer or user the capabilities that the architecture enables [6]. Completed executable SoS architectures are usually models that can be executed by computers automatically, some existing approach include Petri Net and ExtendSim.

There are already some studies on how to gain and operate executable architectures. Levis and his colleagues are the first proposing executable models can be gained by transforming three DoDAF view models to

Coloured Petri Net or to other executable formalisms. The three view models are activity model (OV-5), rule model (OV-6a), data model (OV-7) [6][7][8][9]. Besides the studies within DoDAF or C4ISR (previous version of DoDAF), Huynh and Osmundson [10] studied gaining executable models of a coalition U.S.-Singapore maritime SoS architecture with conversion from UML and SysML conceptual models, respectively. Wang [11] studied modeling the Global Earth Observation SoS with SysML and transforming SysML models to CPN to do architecture validation. MDA (Model Driven Architecture) method was introduced to guide the transformation process in Wang's study.

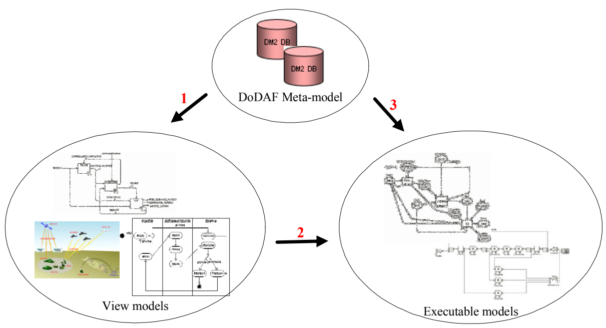


Figure 1. Comparison of two executable architecting methods

However, previous studies still gain executable models based on converting architecture view models. As shown in Figure 1, traditional executable modeling method is conversion 2 (convert view models to executable models). Executable modeling based on DM2 is conversion 3 (convert DM2 data models to executable models). With the guide of DoDAF 2.0, view models should be built by extracting data from DM2 (conversion 1, convert DM2 to view models). If we still extract data from view models to build executable model, the completeness and conformance of architectural data could be damaged due to the semi-formal characteristic of view models. A better choice is converting DM2 data models to executable models directly, which not only ensures the completeness and conformance of architecture data, but also reduces the work to build view models as mediate process. Therefore, this study is desired to verify the feasibility to convert DM2 data models to executable models directly.

2 Framework

DM2 provides a new data meta-model to organize semantically related concepts or elements into a common set of data types and define their associations and attributes based on formal ontology. With elements in DM2, user can model nearly all necessary aspects of an SoS architecture.

A framework of executable architecting and analyzing based on DM2 is presented below. There are three necessary steps to accomplish this work.

Step1. Requirements driven development of DM2 models for SoS architecture.

Step2. Extracting data from DM2 models and converting them into executable models.

Step3. Analyze architecture based on executable models.

Table 1. Mapping matrix between partial DM2 and CPN elements

DM2 elements		CPN elements
Capability		transition collection
Activity	ConsumingPartOfActivity	transition, input arc
	ProducingPartOfActivity	transition, output arc
Capability, Activity, activityPartOfCapability, wholePartType		substitution transition subpage
Resource, Association between Resource and Activity	Performer, activityPerformedByPerformer	place, arc, colour set
	Information Flow, activityResourceOverlap, activityChangesResource	place, arc, colour set
Rule, ruleConstrainsActivity	if	Conditional arc inscriptions, monitor function
	then	output arc inscriptions
MeasureType		colour set
Measure		value of colour set, arc inscriptions

This study concentrates on the case study of Step2. The major jobs are extracting data from architectural data organized following DM2 formalism and transforming them into desired executable data models. The executable formalism can be selected according to specific application. This study selects CPN as the executable formalism. CPN possess adequate modeling ability and abundant analysis methods. Besides, CPN has been demonstrated as an effective way to validate architecture [6][8][11]. DM2 can be treated as a language to model architecture and its logical behavior. CPN is another language that can be used to model system activity and state. Therefore, a prior work of model transformation is to create a mapping matrix between modeling elements of DM2 and CPN. The result is as shown in Table 1.

3 Case study: Conversion from DM2 to CPN

A case study to analyze C2 capability in ADSoS is presented in this section to illustrate the detailed process of model transformation. A typical scenario can be described as follows: When the radar detects intruders, it sends the intruder information to C2 center immediately. Then the C2 center assesses the situation of intruders. Afterwards, other intelligence resource, superior instructions, geographic and weather information are fused to generate a comprehensive situation assessment. Then the combat command is generated based on the comprehensive situation assessment. Then the general combat task is divided and allocated to combat units. The combat units will intercept intruders according to its combat task.

Suppose that the DM2 data models for ADSoS have been created. The scope and granularity of DM2 data models are determined by the purpose to analysis the C2 capability. A five steps process conversing from DM2 to CPN is illustrated as follows.

Step 1. Decide which activities should be included to create the transitions structure in CPN model according to the purpose to analyze C2 capability.

Suppose that the DM2 data models are stored in XML format, search in the database. Four activities have association called "activityPartOfCapability" with C2 capability, they are Situation Assessment, Information Fusion, Command Interpretation and Target Distribution. Three activities have association with Situation Assessment, they are Target Identification, Target Optimization and Situation Production. A simplified search result is in Figure 2.

```
<?Capability?>
<Capability ideas:FoundationCategory="IndividualType" id="ca1">
        <ideas:Name exemplarText="C2" id="n153"/>
</Capability>
<?Activity?>
<Activity ideas:FoundationCategory="IndividualType" id="a1">
        <ideas:Name exemplarText="Situation Assessment" id="n10"/>
</Activity>
<Activity ideas:FoundationCategory="IndividualType" id="a2">
        <ideas:Name exemplarText="Information Fusion" id="n161"/>
</Activity>
<Activity ideas:FoundationCategory="IndividualType" id="a3">
        <ideas:Name exemplarText="Command Interpretation" id="n162"/>
</Activity>
<Activity ideas:FoundationCategory="IndividualType" id="a4">
        <ideas:Name exemplarText="Target Distribution" id="n13"/>
</Activity>
<Activity ideas:FoundationCategory="IndividualType" id="a5">
        <ideas:Name exemplarText="Target Identification" id="n161"/>
</Activity>
<Activity ideas:FoundationCategory="IndividualType" id="a6">
        <ideas:Name exemplarText="Target Optimization" id="n162"/>
</Activity>
<Activity ideas:FoundationCategory="IndividualType" id="a7">
        <ideas:Name exemplarText="Situation Production" id="n13"/>
</Activity>
<?activityPartOfCapability?>
<activityPartOfCapability ideas:FoundationCategory="WholePartType"
id="apoc1" place1Type="ca1" place2Type="a1"/>
<activityPartOfCapability ideas:FoundationCategory="WholePartType"
id="apoc2" place1Type="ca1" place2Type="a2"/>
<activityPartOfCapability ideas:FoundationCategory="WholePartType"
id="apoc3" place1Type="ca1" place2Type="a3"/>
<activityPartOfCapability ideas:FoundationCategory="WholePartType"
id="apoc4" place1Type="ca1" place2Type="a4"/>
<?WholePartType?>
<WholePartType ideas:FoundationCategory="WholePartType" id="wpt1"
place1Type="a1" place2Type="a5"/>
<WholePartType ideas:FoundationCategory="WholePartType" id="wpt2"
place1Type="a1" place2Type="a6"/>
<WholePartType          ideas:FoundationCategory="WholePartType"
id="wpt3" place1Type="a1" place2Type="a7"/>
```

Figure 2. C2 capability, activities and associations between them in DM2 database

With the mapping matrix in Table 1, the DM2 data in Figure 2 can be transformed into a CPN transition structure. Each activity maps to a transition. Convert the first level activities composing C2 capability into first level transitions in CPN page, convert activity with sub-activities into substitution transition and subpage. The conversion result is shown in Figure 3.

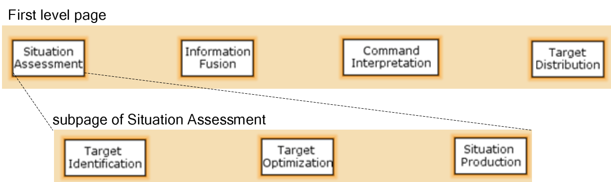


Figure 3. CPN model with only transitions

Step 2. Add places and arcs to CPN model according to resource flow between activities, including information flow and material flow.

The major information flow to achieve C2 capability in ADSoS is as follow: sensed information — situation information — comprehensive situation information — command interpretation result—targets distribution result, they are named as X, Z, Z', V and Y, respectively, as shown in Figure 4. Besides, situation information may be acquired from or send to other organizations.

Two material resources relate to C2 capability, they are commander and C2 system, as shown in Figure 4. Commander can only handle limited tasks in unit time. C2 system also is limited in data processing, different modules of C2 system (Situation Assessment module, Information Fusion module, Command Interpreter module and Target Distribution module) can operate in parallel, but each module can only process one task each time.


```
<Information ideas:FoundationCategory="IndividualType" id="di1"
target_id="t001" target_status="sensed" target_dsp="a black one with big
bomb">
        <ideas:Name exemplarText="X" id="n19"/>
</Information>
<Information ideas:FoundationCategory="IndividualType" id="di2"
target_id="t002" target_status="incoming" threat_level="3">
        <ideas:Name exemplarText="Z" id="n20"/>
</Information>
<Information ideas:FoundationCategory="IndividualType" id="di3">
        <ideas:Name exemplarText="Z'" id="n19"/>
</Information>
<Information ideas:FoundationCategory="IndividualType" id="di4">
        <ideas:Name exemplarText="V" namingScheme="ns1" id="n20"/>
</Information>
<Information ideas:FoundationCategory="IndividualType" id="di5">
        <ideas:Name exemplarText="Y" id="n20"/>
</Information>
<?System?>
<System ideas:FoundationCategory="IndividualType" id="s1"
SA_module_id="sm1" module_status="busy">
        <ideas:Name exemplarText="SA module" id="n34"/>
</System>
<System ideas:FoundationCategory="IndividualType" id="s2">
        <ideas:Name exemplarText="IF module" id="n34"/>
</System>
<System ideas:FoundationCategory="IndividualType" id="s3">
        <ideas:Name exemplarText="CI module" id="n34"/>
</System>
<System ideas:FoundationCategory="IndividualType" id="s4">
        <ideas:Name exemplarText="TD module" id="n34"/>
</System>
<?PersonType?>
<PersonType ideas:FoundationCategory="IndividualType" id="pt1"
commander_id="cmd1" commander_status="free">
        <ideas:Name exemplarText="Commander" id="n34"/>
</PersonType>
```


Figure 4. Resources in DM2 database


```
<?activityResourceOverlap?>
<activityResourceOverlap ideas:FoundationCategory="TripleType" id="aro1"
place1Type="" place3Type="a1" place2Type="di1"/>
<activityResourceOverlap ideas:FoundationCategory="TripleType" id="aro2"
place1Type="a1" place3Type="a2" place2Type="di2"/>
<activityResourceOverlap ideas:FoundationCategory="TripleType" id="aro3"
place1Type="a2" place3Type="a3" place2Type="di3"/>
<activityResourceOverlap ideas:FoundationCategory="TripleType" id="aro4"
place1Type="a3" place3Type="a4" place2Type="di4"/>
<activityResourceOverlap ideas:FoundationCategory="TripleType" id="aro5"
place1Type="a4" place3Type="" place2Type="di5"/>
<activityResourceOverlap ideas:FoundationCategory="TripleType" id="aro6"
place1Type="a4" place3Type="a1" place2Type="pt1"/>
<activityResourceOverlap ideas:FoundationCategory="TripleType" id="aro7"
place1Type="a1" place3Type="a1" place2Type="s1"/>
<activityResourceOverlap ideas:FoundationCategory="TripleType" id="aro8"
place1Type="a2" place3Type="a2" place2Type="s2"/>
<activityResourceOverlap ideas:FoundationCategory="TripleType" id="aro9"
place1Type="a3" place3Type="a3" place2Type="s3"/>
<activityResourceOverlap          ideas:FoundationCategory="TripleType"
id="aro10" place1Type="a4" place3Type="a4" place2Type="s4"/>
```


Figure 5. activityResourceOverlap association in DM2 database

As shown in Figure 5, activityResourceOverlap is a triple, place1Type and place3Type represent two activities. place2Type represents resource. activityResourceOverlap represents that activity place1Type produces resource place2Type and activity place3Type consumes resource place2Type. According to mapping matrix in Table 1, convert place2Typ into place, link transition place1Type to place place2Typ with input arc, link place place2Typ to transition with place3Type output arc, as shown Figure 6.

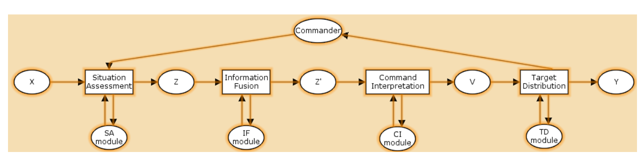


Figure 6 Add places and arcs to CPN model

As shown in Figure 6, place was added for each module of C2 system, which can model the limitation of data processing ability for each module. Due to the same reason, a place is also added for commander.

Step 3. Accomplish colour sets and variables declaration according to resource and measure data in DM2 database.

Resource data is shown in Figure 4. For each resource, declare a compound colour sets.The name of compound colour sets is named as the resource name, simple colour sets in compound colour sets are named as the name of attribute of the resource. For example, “sensed information” is named as X in DM2 database, which has three attributes, including target_id, target_status and target_dsp. Then a compound colour sets can be declared as: colset X =record target_id:int*target_dsp:string*target_status: Target_status.

Notice that unit of each colour set should be declared according to Measure and application situation, default choice is string. Partial list of colour sets is shown in. Define a variable for each colour set, as shown in Table 3.

Table 2. Partial colour sets declarations

<table>
<tr><th>Resource name</th><th>Attributes</th><th>Colour sets declaration</th></tr>
<tr><td rowspan="3">X</td><td>target_id</td><td rowspan="3">colset Target_status = with incoming|sensed|destroyed|escaped|damaged;
colset X = record target_id:int*target_dsp:string*target_status: Target_status timed;</td></tr>
<tr><td>target_dsp</td></tr>
<tr><td>target_status</td></tr>
<tr><td rowspan="3">Z</td><td>target_id</td><td rowspan="3">colset Z = record target_id:int* target_status: Target_status*threat_level:int timed;</td></tr>
<tr><td>target_status</td></tr>
<tr><td>threat_level</td></tr>
<tr><td rowspan="2">Commander</td><td>commander_id</td><td rowspan="2">colset Command_status = with busy|free|unavailable;
colset Commander = record commander_id:int*commander_status:Command_status;</td></tr>
<tr><td>commander_status</td></tr>
<tr><td rowspan="2">SA_module</td><td>SA_module_id</td><td rowspan="2">colset Module_status = with busy|available|hanged
colset SA_module = record SA_module_id:int*module_status:Module_status;</td></tr>
<tr><td>module_status</td></tr>
</table>

Table 3. Partial variables declaration

var x:X;	var c:Commander;
var z:Z;	var sm:SA_module;

Step 4. Define place types and arc inscriptions according to colour set declarations, variables and measures.

Define each place type with its colour set. Define arcs links to each place with its variable, and a weight may be added to arc inscription according to measure. The result is shown in Figure 7.

Step 5. Define arc inscriptions and monitor functions according to rules.

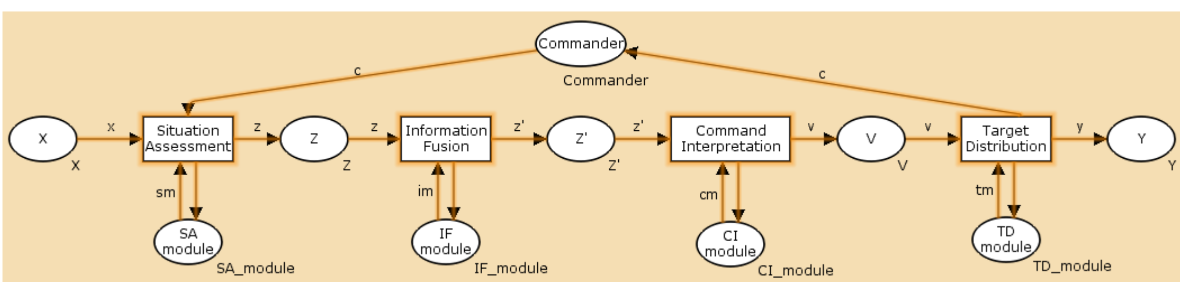


Figure 7. Place type and arc inscription definition

As shown in Figure 8, rule is organized in a fixed if-then formalism. Contents of “if” attribute are converted into conditional arc inscriptions or monitor functions. Contents of “then” attribute are converted into monitor functions. Rule_SA defines the conditions for Situation Assessment, and Rule_IF defines conditions for Information Fusion. The conversion result is shown in Figure 9.

When the five steps were finished, a CPN based executable architecture would be completed. According to the framework, next works are validating the logic correctness and evaluating the effectiveness based on executable architecture.

```
<?RuleDM2?>
<Rule ideas:FoundationCategory="IndividualType" id="ru1" if="
x.target_status = 'incoming' And x.target_dsp = ‘bomber’"
then="Z.target_status = 'sensed' and Z.threat_level = '4'">
        <ideas:Name exemplarText="rule_SA" id="n187"/>
</Rule>
```

<Rule ideas:FoundationCategory="IndividualType" id="ru2" if=" Z.target_status = 'sensed' and Z.threat_level = '4'" then=" Z'.situation = 'The target need to be head off'">
<ideas:Name exemplarText="rule_IF" id="n187"/>
</Rule>

Figure 8. Two rules in DM2 database

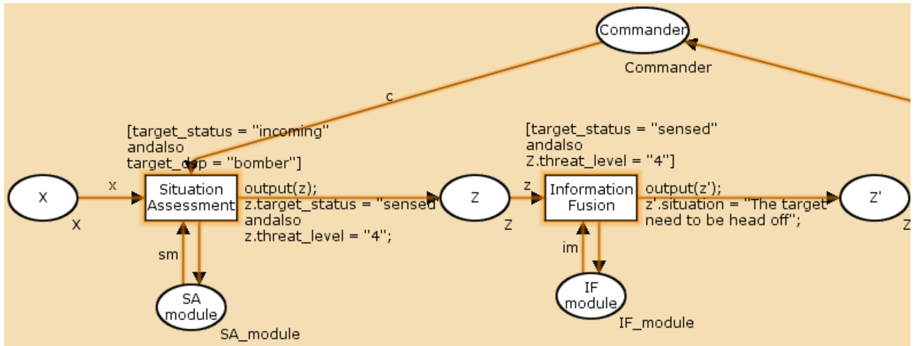


Figure 9. Rules conversion result

4 Conclusions

Executable SoS architecting based on DM2 is a “data-centric” approach. Compared to existed approaches, it has several benefits. First, DM2 provides a unified and standard formalism to store architecture data, which ensures data consistency among different view models. Second, DM2 is defined based on formal ontology with rigorous semantics. Rigorous semantics can be converted to get rigorous executable architecture. Third, various view presentations can be ignored due to using DM2 data models as initial model, which greatly reduces the complexity and time cost of model transformation.

The study illustrates an application example of ADSoS to introduce the five steps converting DM2 data models to executable architecture. However, the novel approach also has one limitation. DM2 is a data storage formalism, which can not be viewed directly. So the model transformation should be processed automatically by machines, which is a challenge that more works need to be done to achieve.

References

[1] Maier M. Architecting Principles for Systems of Systems [C]. Proceedings of the Sixth Annual International Symposium, International Council on Systems Engineering. Boston: 1996: 567-574.

[2] Stevens R. Engineering Mega-Systems: The Challenge of Systems Engineering in the Information Age [M]. New York: CRC Press, 2011.

[3] Jamshidi M. System of Systems Engineering: Principles and Applications [M]. New York: CRC Press, 2008.

[4] Ni F，Wang M Z，Zhou F, et al. Approach to executable architecture modeling from DoDAF to HCPN [J]. Systems Engineering and Electronics, 2010, 32(5): 959-965.

[5] DoD Architecture Framework Working Group. DoD architecture framework version 2.0 [R]. USA Department of Defense, 2009.

[6] Levis A H，Wagenhals L W．C4ISR Architectures: I. Developing a Process for C4ISR Architecture Design [J]. Systems Engineering, 2000，3(4)：225-247．

[7] Wagenhals L W, Shin I, Kim D, et al. C4ISR Architectures: II. A Structured Analysis Approach for Architecture Design [J]. Systems Engineering, 2000，3(4)：248-287．

[8] Wagenhals L W, Haider S，Levis A H．Synthesizing executable models of object oriented architectures[J] ． Systems Engineering．2003，4(6)：266-300．

[9] Wagenhals L W, Levis A H. Service Oriented Architectures, the DoD Architecture Framework 1.5, and Executable Architectures [J]. Systems Engineering, 2009, 12(4): 312-343.

[10] Huynh T V ， Osmundson J S. A System Engineering Methodology for Analyzing System of Systems Using the Systems Modeling Language (SysML) [R]．Naval Postgraduate School，2007．

[11] Wang R, Dagli C H. Executable System Architecting using System Modeling Language in Conjunction with Coloured Petri Net in a Model-Driven Systems Development Process [J]. Systems Engineering, 2011，14(4)：383-409．

A Data-Centric Executable Modeling Approach for System-of-Systems Architecture

Bingfeng Ge [1, 2] bingfengge@nudt.edu.cn

Keith W. Hipel [2] kwhipel@uwaterloo.ca

Long Li [1] eric.longlee@gmail.com

Yingwu Chen [1] ywchen@nudt.edu.cn

[1] College of Information System and Management, National University of Defense Technology, China

[2] Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada

Abstract - *A data-centric executable modeling approach is proposed for system-of-Systems (SoS) architecture by taking full advantage of the Department of Defense Architecture Framework (DoDAF) Meta-model (DM2), which provides more flexibility and adaptability to the automated construction of executable models directly from the architectural data. Firstly, the architectural data meta-model is established to guide architectural data modeling of core data elements and associations in DM2 as the common and consistent data dictionary for architecture modeling, and the executable formalism meta-model is designed to formally define executable models. Then, the mapping rules between both meta-models are defined as the common transformation specification regardless of what modeling language or methodology is employed in developing architectural descriptions. Finally, XML (eXtensible Markup Language) technologies are discussed to facilitate the automated transformation of executable models from architectural instance data. Colored Petri Net (CPN) is used as an illustrative executable formalism in the discussion of the proposed approach.*

Keywords: System-of-systems architecture, data-centric, meta-models, architectural data elements, executable formalisms.

1 Introduction

Recently, the paradigm of system-of-systems (SoS) has emerged as a popular choice for being an economic and strategic approach for enhancing existing system capabilities to address challenging systems engineering and management problems in the military, academia, industry, and elsewhere [1]-[3]. However, the development of an SoS is evolutionary over time and challenged by its characteristics of increasing complexity, emergent behavior, and uncertainty in requirements and context; thus, there is increasing integration of its evolutionary components and functions with demands for more interoperability to fulfill desired effects as a whole [4], [5]. Model Based Systems Engineering, which focuses on the system models as the leading artifacts of the systems engineering process, can help address these significant challenges [6]; while systems architecture is the conceptual model that characterizes the structure of the component systems, their relationships and behavior, and provides the principles and guidelines governing their design and evolution over time [4], [7].

Architecture based capability engineering in particular addresses the complexity and uncertainty early in the SoS design process, and therefore conceptualizes the capabilities expected to be achieved by the entire SoS via the development and continuous evolution of its systems architecture to accommodate more possibilities and unpredictable operating environments [1], [4]. Over the past few decades, many studies on architecture modeling and analysis, such as architecture frameworks and design methodologies, have served well in the development and analysis of architectural descriptions for traditional systems that meet a set of fixed requirements. However, the products of an architectural description, being static models of architectural elements in nature, fail to support dynamic analysis, validation and verification of whether all elements combined together behave as expected and the overall architecture as modeled achieves the desired capabilities [4], [7], [8]. Consequently, executable modeling has become increasingly important as executable models are still strongly needed to enable the time-dependent dynamic simulations to allow a more complete examination and early exploration of the logical and behavioral characteristics, and to provide cost-benefit analysis of the capabilities as modeled in the architecture against the capability requirements.

Currently, the most popular approach for executable modeling has been concerned with the construction of executable models from static architectural models, which in essence is the model transformations using the principles of model driven architecture (MDA) at the model level [6], [8], [9]. The latest Department of Defense Architecture Framework (DoDAF 2.0) [7] establishes the DoDAF Meta-model (DM2) along with a new “data-centric” approach to developing architectural descriptions in a semantically consistent and interoperable fashion. However, current studies of executable modeling usually rely heavily on static models specified with different modeling languages, without taking full advantage of DM2 and the “data-centric” approach. Since the inconsistent representation for the same semantic content in various modeling languages, the target executable models transformed from static models are very diverse even for the same architecture. Furthermore, any changes detected or any errors found in the creation of the executable model and subsequent detailed analysis must be reflected back to guide revisions in the static models of source architectural description [9], while the iterative architecture refinements (e.g., modifying an architecture design or correcting errors in the design) are needed to better support the SoS evolution and accommodate its changing requirements and context. As most of current executable modeling studies are still a manual model transformation process with weak model consistency checking due to semi-formal semantics in modeling languages [8], [10], it connotes that they may require more significant effort to preserve the complete bi-directional traceability and consistency between all elements of static and executable models.

A data-centric executable modeling approach is proposed for SoS architecture by taking full advantage of DM2 and the "data-centric" approach while addressing the collective weaknesses of current studies to provide more flexibility and adaptability to the automated construction of executable models directly from architectural data. The remainder of this paper is organized as follows. Related work leading to executable modeling is reviewed in Section 2 with the outline of weaknesses. The proposed data-centric executable modeling approach is discussed in Section 3, including the meta-models of architectural data and executable formalism, mapping rules between them, and model transformation to generate executable models. Appropriate conclusions are finally provided in Section 4.

2 Related Work

Since Unified Modeling Language (UML) dominates the modeling language for developing object-oriented architectural descriptions, several variants of UML have been created to make it executable, including executable UML ($_X$UML) and executable and translatable UML (^X_TUML). Systems Modeling Language (SysML) has recently evolved as an extension of UML for systems engineering applications. However, being weak in formal execution semantics, both these variants of UML and SysML still do not support the formal specification, validation and verification of executable models [8], [10].

Currently, an alternative approach for executable modeling is the conversion of static architectural models specified by different modeling languages to executable models based on various executable formalisms. Levis et al. [4], [9] described a framework for architecture design and evaluation, in which the DoDAF compliant complete architectural description, produced by either Structured Analysis or Object-Oriented Methodology based on UML, can be transformed into Colored Petri Net (CPN) executable models. Liles [9] has created mapping rules to enable the automated transformation to a CPN executable model from the UML-like architectural description, where a UML activity diagram with swim lanes is used to capture the complete static behavioral description. Wang et al. [8] proposed an executable system architecting paradigm with a SysML-based MDA design process for discrete-event system modeling and analysis, in which a new conversion procedure is developed for converting SysML models into CPN models. Huynh et al. [11] proposed a systems engineering methodology for performing SoS architecture analysis, involving process modeling with SysML, and the conversion of the resulting SysML models into an executable model via Extend (renamed ExtendSim in the latest version). Ring et al. [12] employed the Activity-Based Methodology to generate integrated static DoDAF models, and then to facilitate the transition from these models to executable process models in Bonapart, which is an object-oriented business process modeling tool based on a CPN simulation engine. Other target executable formalisms include Agent-Based Simulation [13] and DEVS (Discrete Event System Specification) [14].

Overall, the aforementioned current executable modeling studies in essence are the model transformations using the principles of MDA [6], [8], [9]. They firstly describe the system context and requirements with computation independent models (CIMs); CIMs are then refined to static architectural models (platform independent models (PIMs)) specified by different modeling languages to define system functionality and behavior; and PIMs can be further transformed into executable models (platform specific models (PSMs)) with more implementation details based on various executable formalisms of any desired dynamic simulation platforms.

Clearly, the primary advantage of current studies is the ability to permit the same concept to be realized with different modeling tools (e.g., modeling languages and executable formalisms) to draw on the strengths of each tool. There also exist, however, the following collective weaknesses due to the model transformation process that relies heavily on the static models of architectural descriptions, without taking advantage of DM2 and a "data-centric" approach in the executable modeling.

- Current studies are grounded on the premise that the information collectively contained in static models must be sufficient and consistent to fully specify the executable models [4], [8]. Either design methodology can produce all the information needed, however, semi-formal semantics in modeling languages make them weak at model consistency checking [10], and maintaining concordance among static models (still manually in many cases) is difficult, especially for a "product-centric" architectural description.
- Different modeling languages have different symbols (vocabulary), semantics, and syntax [4], which lead to inconsistent representation for the same semantic content in static models. Therefore, the model

transformation process based on the mappings between elements of static and executable models is methodology-dependent and diverse even for the same architecture and same target executable formalism, and needs someone to be familiar with both the original modeling languages and target executable formalisms. It is difficult to be commonly understood and compared across multiple instances.

- The model transformation, usually a manual process, should keep the complete bi-directional traceability between both static and executable models [9]. Due to the lack of a common foundation and specification, current studies offer weak flexibility and adaptability to guide traceable and consistent revisions in the architectural description and executable modeling for an SoS, which needs iterative architecture refinements to better support its evolution and accommodate any changing requirements and context.

3 Proposed Approach

In response to the collective weaknesses of current studies, a data-centric executable modeling approach is proposed for SoS architecture by taking full advantage of DM2 and the "data-centric" approach to provide more flexibility and adaptability to the construction of executable models, as shown in Figure 1. More specifically, several processes are needed to be addressed: a process for establishing the meta-models of architectural data and executable formalism to guide the architectural data modeling and define the executable models, respectively; a process for defining the mapping rules between both meta-models as a common specification at a higher level for the model transformation regardless of the modeling languages; and a process for performing an automated model transformation conforming to established mapping rules and with the support of XML (eXtensible Markup Language) technologies.

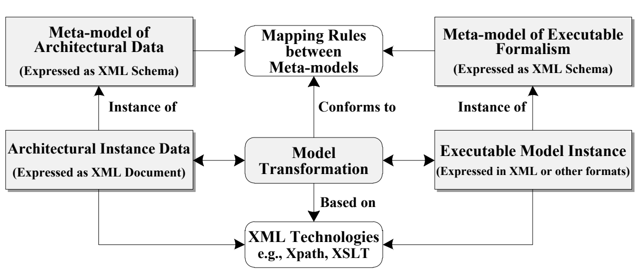


Figure 1. Executable modeling overview

3.1 Meta-models for Executable Modeling

According to the MDA with MOF (Meta Object Facility) designed by OMG (Object Management Group) [15], architectural models can also be organized within a four-layered architecture. At the top layer, the level M3 is the self-defined meta-meta-model and allows defining meta-models at level M2. A model conforming to its meta-model is defined at level M1, and represents a real system at the bottom level M0. A meta-model (model of model, or model of metadata) typically defines the abstract syntax of models and the interrelationships between model elements [15]. That is, it is usually defined as a set of concepts or model elements of a language (or within a certain domain), as well as the constraints and rules of how they may be arranged and related to build models without necessarily providing the concrete syntax of the language. Accordingly, the creation of a model populated with instance data can be equivalent to define the model elements in its meta-model with specific attributes of an instance.

DoDAF 2.0 [7] presents DM2, which is an entirely new data meta-model used to organize semantically related data concepts or elements into common taxonomies (terminology having common definitions) of data types and define their associations and attributes based on several important properties of a formal, higher-order, four-dimensionalism ontology called IDEAS (International Defence Enterprise Architecture Specification). The new "data-centric" approach along with DM2 places greater emphasis on architectural data as the necessary ingredient for architecture development, and then enables static models or views (DoDAF-described Models populated with instance data) and other user-defined views of a subset of architectural data to be built "Fit-for-Purpose" and in a semantically consistent and interoperable fashion.

By providing the standard terminology and formal/ well-defined semantics, which allow machines to interpret in an automated manner, DM2 can also be used to provide more flexibility and adaptability for executable modeling. It is a fundamental architecture principle that architectural data elements should be grouped into the six semantically complete interrogatives (i.e., WHO, WHERE, WHAT, WHEN, WHY, and HOW (5W1H)) as data taxonomies at the highest level to ensure consistency in the meaning of each data element [7], [12]. Moreover, DM2 is still undergoing a major evolution, and some of its broad and diverse data types and associations probably would be redundant, while an executable model can be defined as an integrated dynamic model of sequenced Activities (HOW) performed by Performers (WHO) to produce and consume Resources (WHAT) in Locations (WHERE) under specified Rules and Conditions (WHY) [12]. Thus, it is necessary and advantageous for architecture modeling to capture core elements around 5W1H in DM2 for the collection of architectural instance data, which is only needed to be sufficient enough to fully specify the executable models, while not necessarily providing all of the required details demanded by the specific scope of an architecture. In this respect, a simplified meta-model mainly based on the Capability and Activity meta-models of DM2 is established to formally describe core data elements around 5W1H and their associations at the high-level, as depicted in Figure 2. This architectural data meta-model is also the basis for model concordance; thus, architectural instance data, which conforms to the meta-model with standard terminology and formal semantics, can be collected, maintained, and shared among various stakeholders as a common, consistent and comprehensive data dictionary to increase the potential for application interoperability and reuse, and provide more flexibility and adaptability for executable modeling regardless of what modeling language or methodology is employed to develop architectural descriptions.

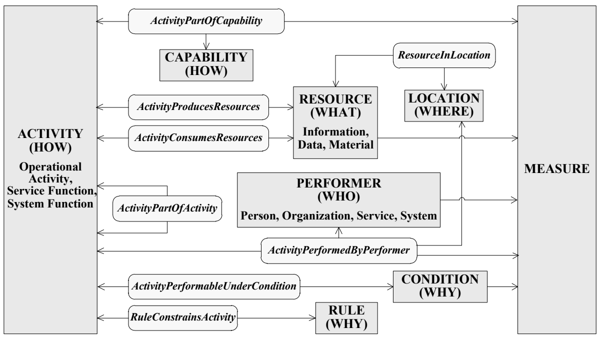


Figure 2. Meta-model of high-level data elements

The capability based architectural data modeling process conforming to the aforementioned meta-model can be executed to collect architectural instance data regarding these core elements and associations for executable modeling. Supposing the capability requirements of SoS have been derived via the requirements analysis process, the sequences of *Activities* (*ActivityPartOfActivity*) with required resource flows can be identified to achieve the desired *Capabilities* (*ActivityPartOfCapability)*. Then, system or service based solutions which satisfy the resource flows and support the capabilities are finally obtained by answering a set of questions consistent with the meta-model, as to whether these *Activities* are performed by *Performers* (*ActivityPerformedByPerformer*) to produce and consume *Resources* (*ActivityProducesResource*, *ActivityConsumesResource*) in *Locations* (*ResourceInLocation*) under specified *Rules* and *Conditions* (*RuleConstrainsActivity*, *ActivityPerformedUnderCondition)* from both operational and solution-related perspectives. The reader is referred to the literature of DoDAF 2.0 [7] for a detailed exposition on these core elements and associations of DM2 (denoted by italics in this paragraph).

Executable models, which are built based on different executable formalisms, have different interests and involve various abstraction levels. CPN models are well-suited for information systems that consist of a number of communication and synchronous processes, while the others based on ExtendSim are better for performing key performance indicator oriented, data and rule driven simulations for complex systems. Similarly, a meta-model can also be established to formally define the model elements and their relationships of the executable formalism selected to create “Fit-for-Purpose” executable models. Wagenhals et al. [9] have established such a meta-model for CPN, as illustrated in Figure 3. CPN will hereafter be used as an illustrative executable formalism in the discussion of the proposed approach.

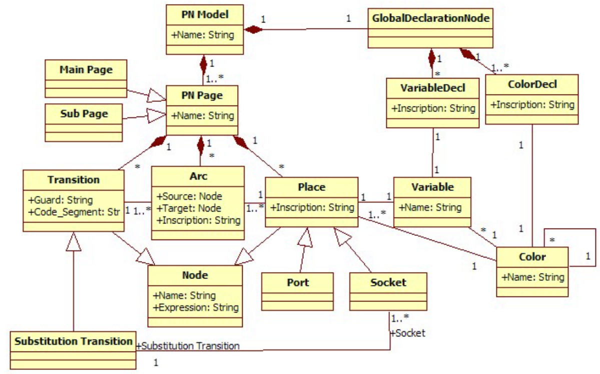


Figure 3. CPN Meta-model [9]

3.2 Mapping Rules between Meta-models

Each static model focuses on particular aspects of the architecture; on the contrary, an executable model defines the time-dependent dynamic behavior around the core elements and associations from holistic 5W1H. Therefore, one can take full advantage of DM2 and the “data-centric” approach in the model transformation process to construct an executable model automatically by directly pulling related instance data of core data elements and associations collected in the previous data modeling process. This in essence extracts all the architectural instance data needed to populate the executable formalism meta-model.

In order to provide a common transformation specification for executable modeling to be effective on a family of model transformations using same target executable formalism, the mapping rules should be defined at the higher meta-model level, rather than between the elements of static and executable models as is done in current studies. Accordingly, the proposed approach establishes the mapping rules between the meta-models of architectural data and executable formalism as the common transformation specification regardless of what modeling language or methodology is employed in developing architectural descriptions. That is, it requires an understanding of concepts or model elements, semantics, and syntax of the executable formalism, and then assigns its formal execution semantics to the core data elements and their associations of DM2. One needs to compare the meta-models of architectural data and executable formalism, and establish the mapping rules between the elements of both meta-models with semantic consistency and complete bi-directional traceability. That is, the concepts captured in the architectural data meta-model and conveyed by the target executable formalism must be consistent; those expressing the same concept should have the same semantics; and all of the elements in the target executable formalism must have a mapping from a source architectural data element. Table 1 gives some key mappings from the core architectural data elements (see Figure 2) to the model elements of CPN.

Table 1. Mappings between core data elements and CPN

Core Data Elements	CPN
CAPABILITY	Transition
ACTIVITY	
PERFORMER	Place
RESOURCE	
MEASURE	Color Sets, Intital Markings
ActivityPartOfCapability	Substitution Transition & Subpage
ActivityPartOfActivity	
ActivityPerformedByPerformer	Place-Arc-Transition
ActivityProducesResource	Transition-Arc-Place
ActivityConsumesResource	Place-Arc-Transition
RuleConstrainsActivity	Arc Inscriptions, Guard Functions, Code Segments
ActivityPerformedUnderCondition	

3.3 Model Transformation

After the mapping rules between both meta-models are defined, model transformation is a process of converting the architectural instance data conforming to architectural data meta-model to the executable model conforming to the executable formalism meta-model. That is, the mapping rules can be seen as a series of operations to construct an executable model by extracting and transforming all the related information from the architectural data meta-model populated with architectural instance data to populate the target executable formalism meta-model. According to the key mappings given in Table 1, the procedure for the model transformation to construct a CPN executable model from architectural data can be defined, as shown in Table 2.

XML is "a markup language produced by the W3C that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable", and nowadays a dominant representation format of arbitrary data structures used in many areas [16]. It has become the default format for many office-productivity tools and the common use for the interchange and sharing of data that can occur in a toolset-agnostic, methodology-agnostic environment [7], [16]. Many established XML technologies or specifications can also provide new possibilities for expressing the key mapping rules and the procedure for model transformation in code to enable the automated creation of executable models.

Table 2. Procedure for model transformation

1. Construct the *declarations* and define the *color sets* in the *index* using the attributes of the *Measures*. 2. Create a *substitution transition* for each *Activity* (or *Capability*) having *Sub-Activities,* or create a *transition* for each leaf *Activity* according to the *ActivityPartOfCapability* or *ActivityPartOfActivity*. 3. Create a *place* for each *Performer* or *Resource*, and assign the appropriate *color set*. 4. Create *arcs* to connect *transitions* and *places* based on the associations specified in *ActivityPerformedByPerformer*, *ActivityProducesResource*, and *ActivityConsumesResource*. 5. Use *Rules* and *Conditions* associated with each *Activity* based on *RuleConstrainsActivity* and *ActivityPerformedUnderCondition* to add *arc inscriptions*, *guard functions*, or *code segments*. 6. Create a *subpage* for each *substitution transition*. 6.1. Follow steps 2 to 5 to create all the related *transitions*, *places*, and *arcs*. 6.2. Assign *places* to the appropriate *Input*, *Output*, and *I/O ports*. 7. Specify *initial markings* for each related *place* that represents the *Performers* or *Resources*.

Thus, the architectural instance data and executable models can be expressed as an XML document conforming to an XML schema, such as XSD (XML Schema Definition) or DTD (Document Type Definition). Since such XML schema languages typically check and formally describe XML documents for validity by constraining the set of elements, their attributes, and the logical structure of an XML document [16], the meta-models of architectural data and executable formalism can also be expressed as XML schemas by mapping every element in each meta-model to appropriate constructs available in the corresponding XML schema. Actually, the meta-models of architectural data in DoDAF 2.0 are expressed in the XSD conforming to the DM2 Physical Exchange Specification; and the model files of executable formalism, such as the CPN in CPN Tools, are XML documents with a DTD/XSD as the model template. Since architectural instance data, as well as some executable models, can be expressed in an XML document with an XML schema, the mapping rules can be expressed in code using the specifications of XSL (eXtensible Stylesheet Language) to automatically produce an executable model. XPath (XML Path Language) defines XPath expressions to select nodes from the XML document of architectural instance data to extract all the information needed for the construction of executable models; and XSLT (XSL Transformation) can be employed for defining the transformation from the XML document of architectural instance data to the document of executable models in XML or other formats according to the procedure for model transformation (see Table 2). Consequently, during each of the iterative architecture refinements for an SoS architecture, one does not need to develop the architectural models from scratch, but only to execute actions such as creation of new elements, update of existing elements and deletion of elements in appropriate XML documents.

4 Conclusions

In this paper, a data-centric executable modeling approach for SoS architecture is proposed by taking full advantage of DM2 and the "data-centric" approach to provide more flexibility and adaptability to the construction of executable models. The meta-model of architectural data is firstly defined to guide the architectural data modeling to collect architectural instance data of core data elements and their associations in DM2, which can act as a common, consistent and comprehensive data dictionary for architecture modeling; and the executable formalism meta-model is designed to formally define "Fit-for-Purpose" executable models. Then, the mapping rules are established at the higher meta-model level as the common transformation specification for executable modeling to be effective on a family of model transformations, no matter what modeling language or methodology is employed in developing architectural descriptions. Finally, XML technologies are discussed to facilitate the automated transformation of executable models from architectural instance data.

Acknowledgements

This research was supported in part by the State Scholarship Fund of China under Grant No. 2011611075, and the National Natural Science Foundation of China under Grant Nos. 70971131, 71001104, and 70901074.

References

[1] M. Jamshidi, *System of Systems Engineering: Innovations for the 21st Century*. New York: Wiley, 2009.

[2] K. W. Hipel, M. M. Jamshidi, J. M. Tien, and C. C. White III, "The future of systems, man, and cybernetics: application domains and research methods," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 37, no. 5, pp. 726–743, Aug. 2007.

[3] C. B. Keating and P. F. Katina, "Systems of systems engineering: Prospects and challenges for the emerging field," *Int. J. System of Systems Engineering*, vol. 2, nos. 2/3, pp. 234-256, 2011.

[4] A. H. Levis, "System architectures," in *Handbook of Systems Engineering and Management*, A. P. Sage and W. B. Rouse, Eds., 2nd ed. Hoboken: John Wiley & Sons, Inc., 2009, ch. 12, pp. 479–506.

[5] J. F. Andary and A. P. Sage, "The role of service oriented architectures in systems engineering," *Inform. Knowledge Syst. Management*, vol. 9, no. 1, pp. 47-74, Jan. 2010.

[6] C. Piaszczyk, "Model based systems engineering with Department of Defense architectural framework", *Syst. Eng.*, vol. 14, no. 3, pp. 305–326, Jul. 2011.

[7] DoD Architecture Framework Working Group, *DoD Architecture Framework ver. 2.02*. US Department of Defense, Aug. 2010.

[8] R. Wang and C. H. Dagli, "Executable system architecting using systems modeling language in conjunction with colored petri nets in a model-driven systems development process," *Syst. Eng.*, vol. 14, no. 4, pp. 383–409, Oct. 2011.

[9] L. W. Wagenhals, S. W. Liles, and A. H. Levis, "Toward executable architectures to support evaluation," in *Proc. 2009 Int. Symp. on Collaborative Technologies and Systems*, pp.502-511, May 2009.

[10] I. Khan, "Methodology for the development of executable system architecture," in *Proc. 8th Int. Conf. on Frontiers of Information Technology*, Islamabad, Pakistan, Dec. 2010.

[11] T. V. Huynh and J. S. Osmundson, "A systems engineering methodology for analyzing systems of systems using the systems modeling language (SysML)," in *Proc. 2nd Annu. Conf. on System of Systems Engineering*, Fort Belvoir, VA, Jul. 2006.

[12] S. J. Ring and D. Nicholson, "Activity-based methodology for development and analysis of integrated DoD architectures," in *Handbook of Enterprise Systems Architecture in Practice*, P. Saha, Ed. Hershey, PA: Information Science Reference, 2007, ch. 5, pp. 85–113.

[13] A. W. Zinn, "The use of integrated architetures to support agent based simulation: An initial investigation," M.S. thesis, Air Force Inst. of Technol., Wright Ptrsn Afb, OH, 2004.

[14] S. Mittal, "DEVS unified process for integrated development and testing of service oriented architectures," Ph.D. dissertation, Univ. of Arizona, Tucson, AZ, 2007.

[15] Object Management Group, *Meta Object Facility (MOF) Core Specification ver. 2.0*. Jan. 2006.

[16] "XML," http://en.wikipedia.org/wiki/XML, May 2012.

Governance Mechanism Pillars for Systems of Systems

Darabi H. R. Author
School of Systems and Enterprises
Stevens Institute of Technology
Hoboken, NJ, USA
hdarabi@stevens.edu

Gorod A. Co-author
The University of Adelaide
Adelaide, SA, Australia
alex.gorod@adelaide.edu.au

Mansouri M. Co-author
School of Systems and Enterprises
Stevens Institute of Technology
Hoboken, NJ, USA
mo.mansouri@stevens.edu

Abstract – *Traditional management frameworks that have been applied to governance of systems are no longer effective in the context of systems of systems. To provide an effective governance framework for a system of systems, it is necessary to understand the governance mechanism in its entirety, including its fundamental components, or pillars.*

The aim of this paper is to provide a framework to analyze the pillars of governance in systems of systems. This framework enables the researchers to study the complexity in the dynamics of interactions. It also provides a tool for modelers to simulate the interactions. Moreover, the framework is intended to facilitate governance of systems of systems for practitioners. Five proposed pillars of the governance mechanism are: (1) purpose integration, (2) belonging regulation, (3) incentivizing device, (4) interactions protocol, and (5) principles dissemination & perception distortion. A simulated model of supply chain demonstrates these five pillars in action within the system of systems context.

Keywords: governance of systems of systems, SoS modeling and simulation, socio-technical systems governance, supply chain management.

1 Introduction

Over the last several decades, the evolution of novel technologies resulted in the emergence of inter-connected and interdependent complex adaptive systems. Creation of a system of systems (SoS) concept is an attempt by engineers and scientists to cope with the problems in engineering, design, and governance of SoS. Although there is no accepted definition of a SoS, there have been many attempts to define it [2]. A working definition of a SoS is proposed by the United States Department of Defense as "a set or arrangement of systems that results when independent and useful systems are integrated into a larger system that delivers unique capabilities" [3]. These SoSs possess five different characteristics: (1) autonomy, (2) belonging, (3) connectivity, (4) diversity, and (5) emergence, which are defined by Boardman and Sauser [4].

In the early stages of the evolution of a SoS, practitioners and scientists used traditional management paradigms in the SoS governance. The fact that each SoS is in general a system itself justified the use of established knowledge of managing systems. This relationship between systems theory and management science has a long-standing history. In 1955, Koontz mentioned the idea of management as a system [5]. According to Koontz, it performs a set of functionalities to direct the organizations toward expectable outcome. These functionalities are planning, organizing, staffing, controlling, and leading.

In a similar fashion, the general systems theory in management accommodated a systems perspective to organizations and attempted to manage human organizations as a system [6]. For example, Mintzberg used a systems perspective to analyze the functions of organizations, and proposed to use organizational structure to produce and deliver the outcome of the organizations [7, 8].

Likewise, Jay Forrester proposed that mathematical control theory, which has been applied in the exploration of mechanical and electrical systems, can be used to understand the complex dynamics of human systems [9]. He created the discipline of System Dynamics (SD) and applied it as a mathematical formulation of organizational interactions [10]. SD grew into a powerful tool to study the feedback loops and causal relationships in organizations [11]. In the 1990s, systems thinking in organizations became a dominant dialogue, which was reflected in the widespread popularity of "The Fifth Discipline," a book by Peter Senge [12].

SoS researchers and practitioners commonly indicated the shortcomings of conventional methods in managing of SoSs. For example, Rinaldi et al. emphasized the importance of considering interconnectivity in infrastructure engineering [13, 14], Dahmann showed the SoS management challenges in the Department of Defense context [15], and Darabi et al. presented this challenge in extended enterprise systems domain [16].

This gap is demonstrated by Mansouri and Mostashari, who highlighted the significance of governance in a systems' domain, which they define as "a set of architectural guidelines, strategic allocations, directing policies, inter-organizational communicating regulations, transactional structures, and network protocols that allow Extended Enterprise Systems to achieve their objectives in a systemic way within given contextual and resource constraints" [17].

Several attempts have been made to develop a framework for governance for SoSs. Some researchers focused on offering solutions for specific problems, such as coordinating multiple autonomous systems [18], optimizing robotic sensor system of systems [19], and participatory water infrastructure decisions [20]. A few examples of the frameworks are network management of Gorod et al. [21-23], Binder's enterprise management [24, 25], Mansouri and Mostashari's enterprise governance systems [17], and the adaptive network governance system of Nooteboom [26].

The current frameworks do not allow to study the internal complexity of the SoS governance mechanism in its entirety. To understand the internal dynamics of the governance mechanism and its interplay with the SoS, its fundamental pillars need to be examined. The objective of this paper is to propose a framework to analyze the SoS governance mechanism through the lens of its pillars. This framework enables scholars to extend the knowledge about the SoS governance and its application [27].

The next section presents the context for applying governance and classification of different management styles. Section 3 describes the proposed pillars of governance mechanism in a SoS. Section 4 provides a modeling example of the governance mechanism's pillars in a supply chain SoS case study. The final section includes conclusion and further research suggestions.

2 Governance Mechanism

A typology of systems is a useful tool to understand the requirements of a governance mechanism in a system. Gorod et al., [1] proposed a typology to classify systems into four categories, namely: (1) assembly system, (2) traditional system, (3) flexible SoS, and (4) chaotic form of a system. The choice of an effective management style varies based on a system's category, as depicted in Figure 1 [27].

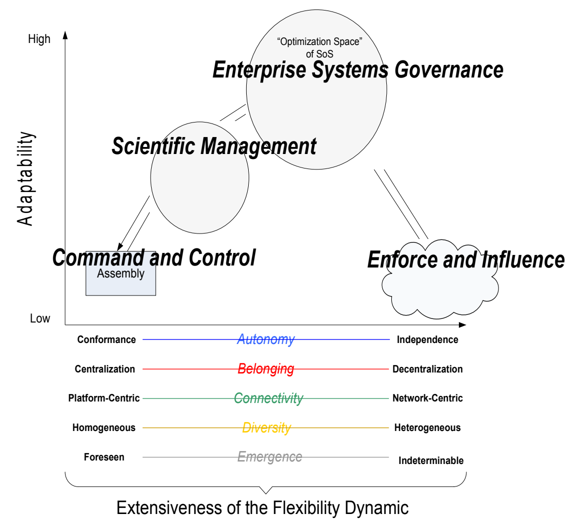


Figure. 1. A typology of systems based on adaptability and flexibility [1]

For assembly systems, command and control is an effective mean to produce the output of the system. The traditional management style is the most suitable methodology to direct the traditional system and veer it toward the suitable outcome. In a chaotic form of a system, the boundaries are not clearly defined. Therefore, enforce and influence is the most applicable style. In a SoS, however, governance is the most appropriate style of directing, coordinating and leading the SoS. As we proposed, the governance mechanism includes a set of principle guidelines, or pillars, that form the mechanism and influence the SoS.

3 Five Pillars of SoS Governance

The governance mechanism in a SoS exhibits the characteristics of a complex system and can be seen as a complex system itself. Therefore, it is necessary to view it as a whole, including the behavior and interplay of its individual components. The proposed framework

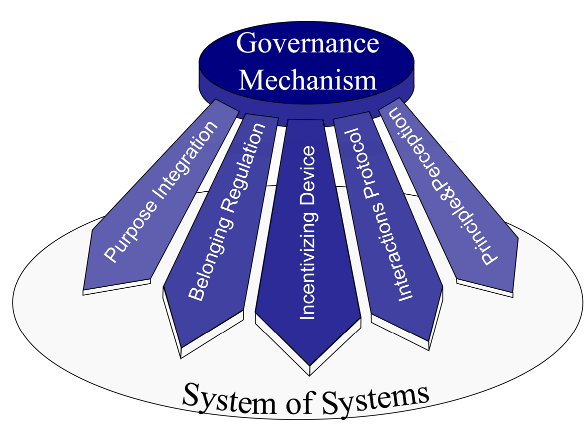


Figure. 2. Five pillars of governance in SoSs

consists of five pillars to governance, as illustrated in Figure 2. The governance body uses a combination of these five different pillars to influence the behavior of a SoS. The five pillars of governance, subsequently, can be understood as triggers that could be utilized integratively to influence or govern the SoS.

3.1 Purpose Integration

Any SoS is designed or emerges to serve a purpose, or a set of objectives. According to the Department of Defense's System Engineering Guide for Systems of Systems, there are four types of a SoS based on their purpose [28]:

- virtual SoS, in which there is no central authority and the interactions emerges,
- collaborative SoS that is voluntarily and the authority use collaborative means to decides,
- acknowledged system of systems, which have recognized objectives and a designated manager, and;
- directed SoS that is built and managed to fulfill an specific purpose.

In a SoS, the constituent elements are capable of negotiating and collaborating on a common SoS goal. The purpose could be also defined depending on a perspective. It is then critical for a SoS governing body to utilize negotiation techniques as a way to establish, sustain, and influence the purpose of the SoS.

3.2 Belonging Regulation

A SoS's boundary is often not clearly defined [29]. Thus, it is the responsibility of a governing body to define and constraint the boundaries of the SoS. The governing body uses regulation of belonging as a tool to negotiate with existing and potential constituent elements of the SoS. For example, it can set new regulations to exclude the constituent elements that do not meet a certain criteria or performance measures. Moreover, these regulations can be used to persuade outsiders to become part of the SoS and bring in new resources.

3.3 Incentivizing Device

A condition of shared resources is a central issue in a SoS. The collection and redistribution of these shared resources is a common responsibility of the SoS's governing body. Therefore, the governing body can utilize the collection and redistribution mechanism, or incentivizing device, to influence the behavior of the individual constituents inside the SoS.

Incentives are widely used in all types of systems, including SoSs, to govern constituent elements. Incentive device has an instantaneous effect on the behavior of constituent elements. For example in air transportation SoS, the increase in incentives for green energy sources can stimulate the usage of alternative energy sources by airlines.

3.4 Interactions Protocol

In a SoS, the interactions between constituents are determining factors in the eventual outcome of the SoS. Consequently, it is the responsibility of the governing body to define, maintain and update the protocols of interactions within the SoS. For example, a standard protocol can determine the outcome of a tradeoff between two constituent elements. It can further regulate the work flow and the information flow within the SoS.

The protocol of interactions is in a close relationship with the incentivizing device. It can be used as a mean to change the cost of each interaction between constituent elements.

3.5 Principles Dissemination and Perception Distortion

Unlike the other pillars of governance in a SoS, the final pillar does not use direct force of the governing body and can be perceived as a "soft governance method." The enforcement tool for this pillar is more of an informal social mechanism rather than a formal mechanism of the SoS governance.

Principles are unwritten norms that guide the choice of constituent elements of a SoS without enforcing a formal rule. These principles can be best practices (according to the incentive structure), or moral guidelines. The social mechanisms, such as collective sanctions and reputation [30], can be used to make these principles compulsory.

Altering the perception of the constituent elements within a SoS is another way through which the governing body can influence the SoS. Because of structural embeddedness of constituent elements and their bounded cognitive capability, the governing body can distort their perception of the environment to affect their dynamics.

4 Governance Mechanism Model

A supply chain case study is used to demonstrate the applicability of the proposed framework through agent-based modeling. The ultimate common goal of a supply chain SoS is to minimize the overall costs of the supply chain while meeting the demand. Although the SoS's multiple constituents share this goal, they also strive toward fulfilling their individual aims to maximize their own profitability. Various agent-based models of supply chains have been developed for such uses as risk management [31], revenue sharing [32], and planning and optimization [33, 34]. The study is used in the paper to illustrate the governance mechanism and its pillars in action. Therefore, we have developed a simple supply chain model with four echelons to show the application of the five pillars of governance.

Figure 3 depicts the modeling representation of the chosen supply chain case study. There are four echelons in the case study, including suppliers, manufacturers, a single distributor, and retailers. The number of corporations in each layer (except distribution), is assumed to be n=3. While the flow of the materials is forward, the information moves backward – from retailers to the distributor. The arrows show the possible flow of the materials. It is not necessary for any company to have a material flow to a specific company, which means the number of real arrows can be fewer than the fully connected network in Figure 3. The flow of the information is only possible between the companies with material flow in the reverse direction.

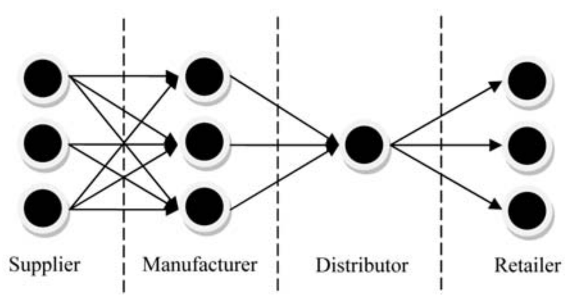


Figure. 3. Modeling of the Supply Chain Case-Study

The role of a supply chain is to deliver the products and meet the demand of end users. Since there is a tradeoff between the cost of available inventory and the loss of demand, there is an optimal level of inventory in a supply chain. This level for the supply chain is not necessarily the optimal level of inventory for all of the individual firms within it. As a result there are various systemic failures due to autonomy of the players within the SoS.

One of the interesting effects in supply chain is bullwhip effect. The bullwhip effect occurs when a series of companies in a supply chain use orders from a downstream member as the input for their production decisions. The variance of the orders in upstream will be much higher than the variance of the real demand because of distortion in information [35].

The modeling example reflects a multi-step decision-making framework for the optimal level of inventory in each node. The assumption is that all the players, or constituents, use the periodic review system to manage their inventory. On each step (R), for example once a week, each player checks its inventory level and decides its ordering level (Q). Q is equal to the difference between inventory and a desirable level of stock (S). This inventory management system can be also referred to as (R, S). Since the desirable level of inventory for each one of the nodes is but suboptimal to the whole SoS, the governing body, which is the distributor in this case, requires influencing the choice of players. The distributor uses five pillars of the governance mechanism as described in the following sections.

4.1 Purpose Integration

The first pillar of governance is to integrate the purpose of a SoS. The purpose of each player within the supply chain is to minimize its costs over a time period is formulated as follows:

$$C_t = \sum_i C_{i_t} \quad (1)$$

where t is the decision-making timeframe, i is the cost index, and C_{i_t} is the *i-th* type of supply chain cost in the selected timeframe. As a result the distributor can persuade the other players to change their purpose from minimizing the short-term costs to minimizing the long-term costs. This modification in decision-making timeframe can provide an effective method to improve the collaborative inventory design.

4.2 Belonging Regulation

Belonging to the supply chain is determined by a dyadic contract between two constituents. The distributor can influence the behavior of the players by setting specific measures for regulating contracts. For example, the mean and variance for the lead time can be set. If a contractor does not meet the requirements of the set lead times, it will be excluded from the SoS. In addition, the substitute companies can join the SoS if they can meet this requirement, among others

4.3 Incentivizing Device

The constituents determine the optimal level of the inventory based on minimizing the costs. If the governing body can calculate the optimal level of inventory for the whole supply chain, it can enforce it by setting subsidies or taxes. For example, if the optimal level of inventory for player i is equal to 500 units in each period, and the systemic optimal level is 700 units, the distributor can subsidize this player to increase its inventory.

For effective use of this governance pillar, the distributor requires information about the demand, cost, and optimal level of inventory in the entire SoS. However, this, in turn, can also limit the efficacy of the incentivizing device.

4.4 Interactions Protocol

Earlier mentioned bullwhip effect can occur in a supply chain due to determining the demand solely based on the information about orders. As a result, a modest increment in the demand of the end user can cause a dramatic shift in the order level of suppliers.

This adverse effect can be governed by setting an appropriate interactions protocol. For example, the distribution of information within the supply chain can be standardized in a way that the supplier can obtain updated information about the changes in demand. The distributor can use this pillar of governance to influence the behavior of higher layers of the supply chain.

4.5 Principles Dissemination and Perception Distortion

Principles are informal rules that guide the behavior of constituent elements of a SoS. The process of decision-making in a supply chain always includes a number of choices. In the presented example, the selection of optimal level of inventory, order size, and preferred contractor are part of the decion-making. These guiding principles for each choice are enforced by the collective decision of the SoS constituents.

In the case study example, the reputation of suppliers to meet the lead time requirement is a general principle used to guide their behavior. The manufacturers are able to share the information about trustworthiness of suppliers to meet deadlines. As a result, the suppliers are expected to maintain a good reputation and deliver the orders on time. This principle (minimizing the backlog), is enforced by the collective ability of manufacturers.

5 Conclusion

The authors proposed a framework based on five pillars for governance in a SoS. The governing body can utilize these pillars to influence and veer the SoS toward a desirable outcome.

The five introduced pillars of the governance mechanism are (1) purpose integration, (2) belonging regulation, (3) incentivizing device, (4) interactions protocol, and (5) principles dissemination and perception distortion.

In the paper, the applicability of the framework exhibited by an agent-based model of a SoS supply chain case study. Since authority in a SoS is distributed among different constituent elements, utilizing the five proposed pillars of the governance mechanism, the governing body can influence the behavior of the SoS.

The proposed framework enables researchers to model the complexity of a SoS governance and can serve as an important step towards developing a SoS governance standard..

References

[1] A. Gorod, J. Gandhi, B. Sauser, and J. Boardman, "Flexibility of System of Systems," Global Journal of Flexible Systems Management, vol. 9, pp. 21-31, 2008.

[2] A. Gorod, et al., "Sytem of Systems engineering management: A review of modern study and a path forward," IEEE systems journal, vol. 2, pp. 484-499, 2008.

[3] DoD, "System of Systems Engineering," in Defense Acquisition Guidebook, ed Washington, DC: Department of Defense, 2004.

[4] J. Boardman and B. Sauser, "System of Systems - the meaning of "of"," presented at the IEEE Conference on System of Systems Engineering, Los Angeles, CA, 2006.

[5] H. Koontz and C. O'Donnell, Principles of management, An analysis of managerial functions, 4th ed. New York: McGraw-Hill book company, 1955.

[6] H. Koontz, "The managment theory junble revisited," Academy of Management Review, vol. 5, pp. 175-187, 1980.

[7] H. Mintzberg, "The manager's job: folklore and fact," Harward business review, vol. 53, pp. 49-61, 1975.

[8] H. Mintzberg, The structuring of organizations. Englewood Cliffs: Prentice-Hall, 1979.

[9] J. W. Forrester, Industrial dynamics. Cabridge: MIT Press, 1972.

[10]J. W. Forrester, "System dynamics, systems thinking, and soft OR," System Dynamics Review, vol. 10, pp. 245-259, 1994.

[11]J. D. Sterman, Business Dynamics: Systems thinking and modeling for a complex world. Boston: McGrawHill, 2000.

[12]P. M. Senge, The fifth discipline. New York: Doubleday, 1990.

[13]S. M. Rinaldi, et al., "Identifying, understanding, and analyzing critical infrastructure interdependencies," IEEE Control Systems Magazine, vol. December, pp. 11-25, 2001.

[14]I. Eusgeld, et al., "“System-of-systems” approach for interdependent critical infrastructures," Reliability Engineering & System Safety, vol. 96, pp. 679-686, 2011.

[15]J. S. Dahmann, "Systems Engineering for Department of Defense Systems of Systems," in System of Systems Engineering, ed: John Wiley & Sons, Inc., 2008, pp. 218-231.

[16]H. R. Darabi, et al., "A framework for decision making in extended enterprises: the FAA NextGen case," in International Congress on Ultra Modern Telecommuncation and Control Systems (ICUMT), Moscow, Russia, 2010.

[17]M. Mansouri and A. Mostashari, "A Systemic Approach to Governance in Extended Enterprise Systems," in 4th Annual IEEE Systems Conference, San Diego, CA, USA, 2010, pp. 311 - 316.

[18]A. Abel and S. Sukkarieh, "The coordination of multiple autonomous systems using information theoretic political science voting models," in IEEE/SMC International Conference on System of Systems Engineering, 2006, pp. 1 - 8.

[19]H. Azarnoush, et al., "Towards optimization of a real-world robotic-sensor system of systems," in World Automation Congress (WAC '06), Budapest, 2006, pp. 1-8.

[20]K. M. Faust, "A system-of-systems approach to participatory water infrastructure decisions," M.S.C.E. 1490646, Purdue University, Indiana - United States, 2010.

[21]A. Gorod, "System of Systems engineering management framework: a 'satisficing' network approach," 2010.

[22]A. Gorod, et al., "System of Systems management: a network management approach," presented at the IEEE International Conference on System of Systems engineering, San Antonio, TX, 2007.

[23]B. Sauser, et al., "System of Systems Management," in System of Systems Engineering: Innovations for the 21th century, M. Jamshidi, Ed., ed Hoboken: John Wiley & Sons Inc., 2009.

[24]M. Binder and B. T. Clegg, "A ceonceptual framework for enterprise management," International Journal of Production Reserach, vol. 44, pp. 3813-3829, 2006.

[25]M. Binder and B. Clegg, "Enterprise management: A new frontier for organisations," International Journal Production Economics, vol. 106, pp. 409-430, 2007.

[26]S. Nooteboom and P. Marks, "Adaptive networks as second order governance systems," System research and behavioral science, vol. 27, pp. 61-69, 2010.

[27]M. Mansouri, A. Gorod, and B. Sauser, "A typology-based approach to adopting effective management styles for enterprise systems," in 4th Annual IEEE Systems Conference, San Diego, CA, 2010, pp. 300 - 305.

[28]DoD, "Systems engineering guide for systems of systems," Deputy Under Secretary of Defence (Acquisition and Technology)2008.

[29]A. Gorod, et al., "A quantitative approach to analysis of a system of systems operational boundaries," in International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT 2010), Moscow, 2010, pp. 655-661.

[30]C. Jones, et al., "A general theory of network governance: exchange conditoins and social mechanism," Academy of Management Review, vol. 22, pp. 911-945, 1997.

[31]M. Giannakis and M. Louis, "A multi-agent based framework for supply chain risk management," Journal of Purchasing and Supply Management, vol. 17, pp. 23-31, 2011.

[32]I. Giannoccaro and P. Pontrandolfo, "Negotiation of the revenue sharing contract: An agent-based systems approach," International Journal of Production Economics, vol. 122, pp. 558-566, 2009.

[33]D. Ivanov, et al., "A multi-structural framework for adaptive supply chain planning and operations control with structure dynamics considerations, "European Journal of Operational Research, vol. 200, pp. 409-420, 2010.

[34]O. M. Akanle and D. Z. Zhang, "Agent-based model for optimising supply-chain configurations," International Journal of Production Economics, vol. 115, pp. 444-460, 2008.

[35]H. L. Lee and V. Padmanabhan, "Information distortion in a supply chain: The bullwhip effect,"Management Science, vol. 43, p. 546, 1997.

Approaches in addressing System of Systems

Vernon Ireland, Antonella Cavallo, Amina Omarova, Yasmin Ooi-Sanches and Barbara Rapaport
Entrepreneurship, Commercialisation and Innovation Centre (ECIC)
The University of Adelaide
Adelaide, SA, Australia
vernon.ireland@adelaide.edu.au

Abstract - *The concept of waste through lack of coordination of independent organisations into a larger system is examined and an estimate of losses reported. An approach for single organisational structural analysis is extended to multi-organisational analysis to provide examples of benefits. However it is recognised that system of systems fall into two types: those which require only traditional integration tools and others which require systems thinking tools such as Checkland's Soft System Methodology and SAST. Techniques to use SSM are explored.*

Keywords: system integration, multi-organisational, Soft System Methods, Strategic Assumptions Surface Testing..

1 Introduction

We recognize that a key aspect of system of systems is the inclusion of autonomous and independent systems [1]. An extreme case of this is the Air Operations Centre of the US DoD in which over eighty independent and autonomous systems operate. Norman and Kuras [2] make the points that they:

- Don't share a common conceptual basis;
- Aren't built for the same purpose, or used within specific AOC workflows;
- Share an acquisition environment which pushes them to be stand-alone;
- Have integration enabling technologies (glue ware) drafted onto these elements and integration developments are undertaken after delivery of the component systems to their foreign customers.

The importance of addressing system of systems issues has been emphasised by the relatively recent IBM report which identifies losses of $15 trillion pa involved in the lack of integration of organisations of which, they estimate, $4 trillion can be eliminated [3]. This will occur through viewing functions at a higher systems level than currently done.

Various examples of such losses can be illustrated:

- Significant amounts of food is wasted throughout the world by grocery stores and restaurants not sending unused and fresh food to needy people;
- Rivers pass through state boundaries and the national benefits are reduced by selfish states;
- Roads, each controlled by different states, can be integrated in terms of speed and safety standards and shared maintenance equipment.

2 Structural analysis

There are some obvious tools and techniques for integration. Leung and Brockstedt [4] outline an approach for a single organisation which can be developed to address multi-organisational integration. They develop an ontology for structural analysis of an enterprise. They categorise business entities and relationships into four groups of activities and processes, organisation, strategy and marketing. They then recognise that the key aspects at the Enterprise Level are Organisational view, Activity People View and Product View, Activities and Processes, the organisation, the people and the processes.

Obvious examples of the benefits of integration include the highway system and rivers as they pass through multiple states, food supplies in which waste occurs in restaurants and supermarkets, electric power supplies in which the boundaries of the system are limited and generating capacity is wasted through the generators being supported by coal or oil, however they are just spinning as the load reduces in a region. Integration into a larger system reduces waste. Korsten and Seider [3] offer many other examples as shown in Figure 1.

One example is provided of using Leung and Brockstedt's [4] approach for a highway system passing through multi-jurisdictions as shown in Table 1.

Axelsson and Axelsson [5] explore inter-organisational collaboration in health which requires multi-disciplinary teams spanning boundaries of different organisations and

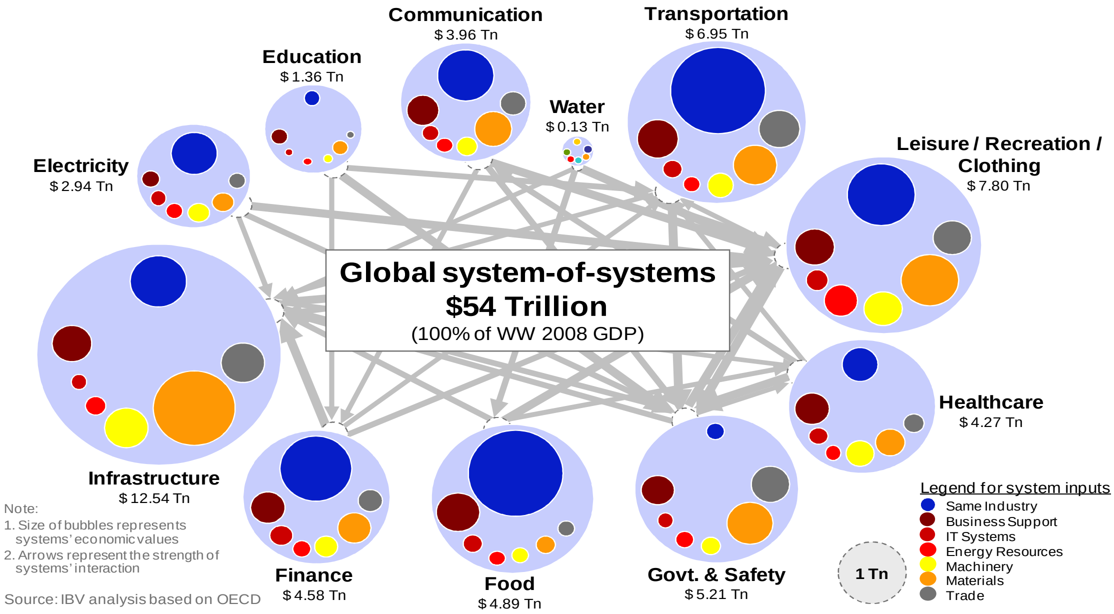


Figure 1. Total world complex, dynamic and interconnected system of systems [3]

Table 1. Examples of loss reduction by seeing multiple roads as a system.

Key Aspects At The Higher Level System	Roadway Integration
Organisational View – additional organisations integrated	Integrate adjacent highway control authorities
Higher system capabilities	Integrates services and capabilities
New processes	Integrated traffic management, Integrated multi-agency traffic lights management, Energy
Benefits	Increased average speed, Reduced maintenance costs, Reduced accidents, Reduced times to attend crises
Activities or focuses	Integrated CCTV, Shared maintenance, Shared emergency services
KPIs	Average speed, Maintenance costs, Time to attend crisis
Potential emergent properties	Introduced toll to support traffic separation intersections

sectors. They recognise Lawrence and Lorsch's paper which outlined the benefits of differentiation and integration within organisations [6]. Management needs to drive integration and remove structural barriers such as different administrative systems, different laws, rules and regulations, different budget and financial streams, different information systems and databases (p83). However such also requires addressing different organisational cultures, different values and interests and differences in commitment of individuals in the organisations. Axelsson and Axelsson found the process includes forming, storming, forming and performing applied (p84).

Reid et al [7] propose that a four level approach be used to address the integration of systems in health care, the levels being the Patient, the Care Team, the Organisation and the Political and Economic Environment. They assert that real time monitoring of patients would save costs and lives. However, individual physicians would need to both have different training and change their methods of operation into a team based mode. Reid et al also believe this will require changes to the guild structure under which individual physicians operate. The organisation of hospitals will require substantial changes in their decision making systems, operating systems and human resources practices. At the political and economic environmental level, many regulatory policies do not support the proposed method of operation and changes are required. One optimistic note is that there are some clinics which operate in this systems centred manner.

Water management has potential to provide a model for cooperation as rivers pass through state and national boundaries. Ferreya and Beard [8] outline eight practical insights for protecting groundwater in rural areas of Ontario. The practical insights are:

a. Knowledge and expertise: scientific knowledge should not dominate collaboration efforts; acceptance of goals can be achieved but not necessarily ownership;

b. Collaborative management goals and targets: building relationship is as important as improving water quality;

c. Collaborative advantage: the benefits of collaboration and minimising drawbacks must be emphasised;

d. Foundational assumptions: as collaborative partnerships are dynamic and can change very quickly, it is important to regularly review foundation assumptions;

e. Partnership evaluation: the partnership should be assessed not only for improvements in water quality but the quality of the alliance and relationships among stakeholders;

f. Inter-organisational leadership: the ability to effectively and simultaneously guide and facilitate stakeholder interaction should not be taken for granted: developing inter-organisational skills should be a top priority;

g. Ownership of actions: clear boundaries between individual organisations and those of the partnership need to be established; this builds trust;

h. Communication strategy: designing and implementing a multi-level communication strategy with common messages for target audiences when communicating with the broader watershed community outside the limits of the inter-organisational partnership.

Governments, in recognising both the challenges and benefits of multi-organisational integration, can provide both legislation and taxation benefits to force and encourage enterprise integration Li [9].

Finally, such integration requires amending practices which have become entrenched and have become the traditional and accepted approach. Changing this requires both a fresh look at project boundaries and new approaches to project solutions. An approach to addressing project stakeholders and boundaries, and for developing solutions, is found in the systems thinking approaches [10]. Strategic Assumptions Surface Testing (SAST) recognises the benefits of various stances in addressing a problem of both stakeholders who are participative and adversarial. SAST develops an integrative approach in which different options must eventually be brought together in a higher order synthesis. It focuses on developing the benefits of a higher order solution.

Rosenhead [11] points out that the developed solution should include:

- A satisfying rather than optimising rationale;
- Acceptance of conflict with over goals;
- Different objectives measured in their own terms;
- The employment of transparent methods that clarify conflict and facilitate negotiation;
- The treatment of human elements as active subjects.

Elfitri [12] found the mechanisms of Collaborative Decision Making were:

- Face to face interactions;
- Meet and communicate regularly;
- Develop solutions cooperatively;
- Decide goals collectively;
- Conduct joint problem solving sessions;
- Significant information sharing and exchange;
- Work together as a team;
- Identify new stakeholders along the way; voluntary and open partnership;
- Organised discussion;
- Recognise interdependence.

There are a number of problems in the integration of businesses. An enterprise has the first responsibility to optimise its own performance for the benefit of itself and not the larger system [9].

3 Types of complex systems

Addressing SoSs is assisted by developing granularity in describing complexity. Snowden and Boone [13] take up the classification of systems into categories of simple, complicated, complex and chaotic. This is used by Glouberman and Zimmerman [14] in the classification of health care systems. Tools for distinguishing complicated from complex are provided by Cotsaftis [15]. The test to identify whether it is complicated or complex is: Identify whether the system can be explained by reduction (i.e. are there equations or obvious hierarchic relationships between the system and its components):

1. Identify the degrees of freedom in the system (the number of variables or aspects free to vary)

2. Decide if it is simple or complicated – how many degrees of freedom

3. Check the number of control tools.

If the number of control tools is less than the number of degrees of freedom, the system is complex.

Furthermore, there are advantages in distinguishing between types of complex problems and projects. The most common form of addressing complex systems is that completed by defense projects, an example of which has been provided by Norman and Kuras [2]. In these projects the overall and reasonably detailed objectives have been defined by requirements. The goal in integrating the systems is to integrate the legacy system into the system of system. Such approach is labeled Complex System Type A.

4 Soft systems use in complex systems

Van Haperen has developed a methodology that enables coherent development and definition of user requirements. Traditional system development and engineering methods no longer suffice and more qualitative methods and techniques need to be embraced. An evolutionary relationship exists between the methodologies and techniques used to define requirements, to design and develop the system and to assess its effectiveness [15]. Change is a product of our era. If an organisation is to absorb the technological, organisational and social changes successfully, it should consciously conceptualise these elements and their interactions as a 'whole'. Customers change their preference over shorter time spans. Competition can be global and is often fuelled by the onward march of technological innovation. Governments impose new regulations. Transformation in society and in ways of thinking imposes fresh responsibilities on organisational decision-makers. Using merely a 'hard' approach would ignore these and it would remain questionable that such a solution would ever be effective. Wilson [17] highlights that organisations, rather than dealing with 'how' to solve a problem, firstly should concern themselves with determining 'what the problem is'.

Worm [18] highlights that 'adequate performance in complex, high risk, tactical operations requires support by highly capable management'. Measuring performance, developing systems and conducting operational testing that cope with such complex conditions are a challenge.

Hence, Complex Type B projects, dealing with issues such as terrorism, managing climate change, addressing illegal drugs, disputes between countries which are traditional enemies, and others, require very different methods, primarily including the use of systems thinking methods, especially Checkland's Soft Systems Methods (SSM), to identify a potential solution [19].

Checkland provides some guidelines as to what should be included. These are:

- Structures
- Processes
- Climate
- People
- Issues expressed by people
- Conflicts

The first step is to understand the concept of different perspectives that are possible to draw out of the rich picture. The SSM process of using CATWOE standing for Customers, Actors, Transformation process, Weltanschauung or World View, the Owner to whom the "system" is answerable and the Environment that influences but does not control the system, all provide a tight process necessary for the breadth of vision required to see integration of systems possibilities.

Bob Will [20] recommends the process follow the order starting from transformation. One way of ensuring this is to construct the CATWOE in the following order:

1. Transformation
2. Weltanschauung
3. Customer
4. Actors
5. Owners
6. Environment.

Will also comments that the CATWOE approach can be amended to replaced C with two concepts; B for Beneficiaries, and V for Victims producing BATWOVE.

SAST can be used as already discussed but Checkland's approach of developing multiple CATWOEs (possibly 10-20), and comparing them for additional perceptions, contributes to development of a solution.

SSM is a methodology for resolving problems and assists in understanding the many simultaneous views which may exists on what an organisation is trying to achieve. SSM is particularly useful where business requirements are unclear, conflicting interests exist, or the proposed system is contentious. It is noted that SSM may be used to complement other approaches e.g. SSADM and System Dynamics, rather than replacing them.

Bergvall-Kareborn [21] focuses on adopting different perspectives, which are suitable for contributing to a solution of national disputes, such as perspectives from: the world's leaders, leaders of each country indirectly involved, the view of common people, religious leaders both extreme and moderate, the military in each country, mothers and fathers who are just trying to bring up their children, etc.

Bergvall-Kareborn [21] suggests the perspectives of ethical, judicial, aesthetic, economic, social, lingual, historic, logical, physical, faith, love, harmony, frugality, social intercourse, symbolic representation, energy, vitality, and motion among others.

Will [20] points out that the roles in the CATWOE or BATWOVE will differ depending on the perspective taken. Exploring each perspective suggested by Bergvall-Kareborn [21] may not be appropriate – other perspectives may be more relevant to the systems being integrated. However, it is the recognition of the results from each and the comparison of these which provides the power of the method.

Integrating organisations to provide a larger system, as described in section 2, can be seen as Complex Type C.

5 Conclusion

It has been found that there are benefits in:

- Addressing the losses due to inadequate integration of systems which are optimised for their own benefit but which induce losses and inefficiencies when considered at a higher system level;

- Recognising the difference between complicated systems which obey reductionist approaches and complex systems which are independent and autonomous;

- Separating complex systems into two levels in which Level 1 is the more usual defense example, and similar in other industries, in which the objectives are reasonably clear by comparison with Complex Level 2 in which the precise problem needs to be defined in parallel with the solution;

- For Complex Level 2 projects, soft systems methods, such as SAST and Checkland's Soft Systems Methodology are very useful approaches for exploring how systems can be integrated leading to perceptions by the comparison of perspectives.

References

[1] J. Boardman, and Sauser, B. System of Systems -the meaning of of, IEEE International System of Systems Conference. Los Angeles, CA. 2006.

[2] D. Norman, and M. Kuras, Engineering Complex Systems, Complex Systems edited by Dan Braha, Ali Minai and Yaneer Bar-Yam, Springer 2006.

[3] P. Korsten, and C. Seider. The worlds 4 trillion dollar challenge, IBM Global Business Service, 2010.

[4] Y. T. Leung, and J. Brockstedt. Structural analysis of a Business enterprise, Service Science, 2009, pp 169-188.

[5] R. Axelsson, and B. Axelsson Integration and collaboration in public health - a conceptual framework, International Journal of Health Planning and Management, 21, 2006, 75-88.

[6] P. Lawrence and J. Lorsch, "Differentiation and Integration in Complex Organizations" Administrative Science Quarterly 12, (1967), 1-30.

[7] P. Reid,, W. Compton and J. Grossmamn, Editors, A Framework for a systems approach to health care delivery, Building a better delivery system: a new engineering/health care partnership. Washington (DC) National Academies Press (US) 2005.

[8] C. Ferreya and P. Beard. P. Assessing collaborative and integrated water management in the Maitland River watershed, University of Guelph, 2005.

[9] D. Li, The objectives of the corporation under the entity concept, The Accounting Review, Vol 39, No 4, October, 1964, pp 946-950.

[10] M. Jackson, Systems Thinking - Creative Holism for Managers, John Wiley and Sons, Chichester, UK, 2003.

[11] J. Rosenhead, J. (1998). Complexity Theory and Management Practice (LSE Working Paper 98.25). London School of Economics, London, 1998.

[12] R. Elfithre, Collaborative Decision Making within the context of Integrated Water Resources Management in Langat Basin Malaysia, PhD Thesis Universiti Kebangsaan Malaysia (UKM), 2006.

[13] Snowden, D. and Boone, M. A Leaders framework for decision making, Harvard Business Review, 2007.

[14] S. Glouberman and B. Zimmerman, B. Complicated and Complex Systems: What Would Successful Reform of Medicare Look Like?, Discussion Paper 8, Commission on Future Health in Canada, 2002.

[15] M. Cotsaftis, M. What makes a system complex? An approach to self-organisation and emergence, mcot@ece.fr, 2007.

[16] K. Van Haperen, K. Working Towards Information Superiority: Application Coherence for Digitisation Programmes – A Method for Coherently Defining Requirements for Future Command and Control Information Systems, Koios Group Ltd, Hampshire, United Kingdom, 2002.

[17] B. Wilson, “Systems: Concepts, Methodologies and Applications”, 2nd Edition, John Wiley and Sons: Chichester, New York, Brisbane, Toronto, Singapore, 1990.

[18] A. Worm, Integrated Modelling and Analysis of Cognitive Mechanisms in Distributed Command Teams and Command and Control Systems, European Symposium on People in Digitized Command and Control, Royal Military College of Science, Shrivenham (UK), 12-13 December 2001.

[19] P.B. Checkland, Systems Thinking, Systems Practice, John Wiley, 1981.

[20] B. Will, Soft Systems Methodology, http://users.actrix.co.nz/bobwill/Resources/ssm.doc, 2012.

[21] B. Bergvall-Kareborn [13], Qualifying Function in SSM Modelling—A Case Study, Systemic Practice and Action Research, Vol. 15, No. 4, August 2002.

Understanding the Dynamics of System-of-Systems in Complex International Negotiations

Barbara Rapaport
Vernon Ireland
Alex Gorod
Entrepreneurship, Commercialisation
and Innovation Centre (ECIC)
The University of Adelaide
Adelaide SA 5005, Australia
barbara.rapaport@adelaide.edu.au
vernon.ireland@adelaide.edu.au
alex.gorod@adelaide.edu.au

Abstract - *This paper explores the intersection between the complex international negotiation processes and the more traditional negotiation modes, emphasizing their strong connection to the issue of the behavioural leadership. The purpose of the study is to demonstrate why the traditional negotiation approaches to complex conflicts have frequently failed in the past, and how the recognition of system-of-systems (SoS) can be helpful to political leaders during the negotiations, from the standpoint of achieving a more sustainable outcome. The concept of leadership style serves as a linking point to determining the dynamics of negotiations and recognition of multiple systems in the study. Specifically, the leaders' personal characteristics -- for instance, their personal views with regards to the conflict, level of comfort with ambiguity, physical agility and adaptability in challenging environments -- can be crucial to the achievement of successful settlements.*

Keywords: Complex International Negotiation Processes, Complex Conflicts, System of Systems (SoS), Behavioral Leadership, SAST, CSH

1 Introduction

The purpose of this paper is to examine the negotiation process between nations which have enmity towards each other in terms of complexity theory and to identify if there are any lessons that could be learned by including processes used by complexity theory within the negotiation process.

The term *complex conflicts* frequently refers to seemingly intractable problems, characterized by being highly resistant to any kind of resolution including socially, economically, and politically sophisticated issues [1]. Brown et al's main characteristics are associated with the high levels of uncertainty concerning the agreeable outcomes, due to the broad diversity of stakeholders' views, which often leads to political gridlock rather than the political resolution [2]. It is widely recognized that the complexity of conflicts is related to the implicit non-linearity of the issues involved, whereby the multiple disparate elements of the whole are interconnected and interdependent of the general environment. From the foregoing, some studies put forth the argument that adopting a reductionist approaches of 'substracting', 'splitting', and 'dividing' the complex conflicts into modules, would serve to decrease the overall levels of complexity [3]. This is contrary to complexity approaches.

2 Complexity in negotiation processes

It is vital to emphasize that the field of complexity theory more than often refers to multilateral negotiations, where many parties and sides (more than three) are involved in the negotiations process. However although the agenda of bilateral negotiations mainly relates to two parties, it still involves levels of complexity. Despite the inherent complexity in negotiations, an enhancement of structural clarity classifies negotiations into 'multilateral', 'bilateral', 'multiparty' and 'biparty' [4].

In terms of multilateral negotiations, the following presents a number of various theories as applied to the complex negotiation processes during the Uruguay Round of the General Agreement on Tariffs and Trade (GATT) and the establishment of the Single European Act in the European Community [3]:

- Coalition theory
- Decision theory
- Game theory
- Leadership theory
- Organizational theory
- Small group theory

It is noted that in the past Coalition and Mediation theory to some extent applied complexity theory to negotiations when dealing with multilateral European (regional) conflicts [3]. This is surprising given Jacksons definitions of complexity. Jackson [5] describes complex systems as those which are 'interconnected and complicated further by lack of clarity about purposes, conflict, and uncertainty about the environment and social constraints. Such have been called *wicked problems*. In tackling wicked problems, problem structuring assumes greater importance and problem solving using conventional techniques. If problem formulation is ignored or badly handled, managers may end up solving, very thoroughly and precisely, the wrong problem'. Mason and Mitoff [6] added that such ill structured problem situation is made up of highly interdependent problems. 'Complexity builds on and enriches systems theory by articulating additional characteristics of complex systems and by emphasizing their inter-relationship and interdependence'. 'The theories of complexity provide a conceptual framework, a way of thinking, and a way of seeing the world' [7].

Focusing on the international negotiations, the core complexity issue found within both multilateral and bilateral negotiation processes is the inability of various parties to first agree among themselves as to their negotiating strategy. A complementary feature of complexity is the monolithic nature of the negotiating party, which means the capacity to reduce internal conflict and represent a single voice. A single voice means a high level of cohesion, whereas too many voices produce the opposite [4].

A low degree of cohesion has been found to disrupt unitary behaviour of entities, often arising from internally conflicting opinions, and thereby affecting 'one voice' communication to the other party. The high degree of non-monolithic and non-unitary internal behaviour of entities of a party has been referred to with regard negotiations, however systems thinking techniques such as Strategic Assumption Surfacing and Testing (SAST) does not appear to have been used. In order to achieve high levels of cohesion, entities/parties should overcome any divergence of opinion within their own internal environment [4, 16].

Hence the conceptualization of a relationship between internal to external entities and their degrees of interaction would be a key factor to understanding the inherent complexity within negotiations [4].

Further, it has been acknowledged that the agenda for failure of multilateral negotiations is due to inherent complexity related to 'parties', 'preferences' and 'proposals'. Of central concern, therefore, is multidimensionality, which mainly refers to constantly shifting coalitions with respect to proposals which are being considered and overall decisional aspects of negotiations. As a consequence the agenda for negotiating parties is often unpredictable [8].

Furthermore, parties can be at conflict over the format of the agenda, therefore they tend to put substantial efforts and resources to pre-negotiate the terms and conditions of the agenda which later need to be debated during negotiations. Another aspect is the multidimensionality of international politics, which cannot be reduced to one dimension. [8].

In this event, system thinking techniques such as Critical System Heuristics (CSH), with its recognition of victims and its technique of focusing on a series of 'ought' questions, which include the boundaries of a problem, the strategy chosen, the measures of success, what expertise is required, and a number of other issues, does not appear to have been used.

Although the innate nature of the complexity in the negotiation processes has been widely acknowledged, these theories do not directly embrace the system of systems (SoS) methodologies, which address the dynamics of the inter-connectedness in complex international negotiations. This is the core focus of the paper. Furthermore, they do not address the issue of top-down processes being a hindrance in complex systems. This issue is taken up in section 6.

3 Behavioural leadership in negotiation processes

A starting point to SoS methodology is to understand the nature of behavioural leadership in negotiations. This is one perspective ; another is to question the concept of top-down leadership and to question whether this is counter productive (see section 6 which takes up a slightly contrary view).

Negotiation processes are aimed at facilitating and resolving conflicts through the parties to the conflict, which have been designated to come to an agreeable solution on the divergent positions and views on certain crucial issues. In the initial phases of negotiation processes, parties can still be weak and uncertain with regards to the appropriate representation of their rights, as they still may be struggling to form the structure of their authority and/or formulate a clear vision of their demands toward the other party. Under such complex circumstanes, the identification of a solid partner or a leader in future dialogues is the key and most pressing objective [9].

Therefore, the conceptualization of leadership style serves as a linking point to determining the dynamics of negotiations and the recognition of multiple systems involved. Specifically, the leaders' personal characteristics -- such as their *personal views* ; *levels of comfort with ambiguity ; physical agility & adaptability* to complex environment -- can be crucial to the achievement of successful settlements [10].

These characteristics refer to the adaptive agent concept found in complex systems. It points out to individuals who are highly responsive to the environment, and who are through their extraordinary sensitivity are able to perceive interdependencies and changes within the environment. At the same time they can adapt their own behaviour with respect to processes, whether it be decision making or leading negotiation processes [12].
Elaborating further, as an ideal-type, adaptive agents should be leaders who are intellectually agile and highly adaptable to the changing environment, with an inner awareness of complex worlds and its inherent connectivity [11]. The interaction of agents occurs in a non-linear way with observation, communication, physical interaction, and communication to create a social

network to achieve common goals and aims. It is suggested by Malcolm Gladwell that there are three main types of adaptive agents who can have profound influence and play a vital role in creating a change. These are so-called *mavens* - people whose focus is on gathering information, *salespeople* - people who possess skills of convincing others as to their point of view, and *networkers* - those who are able to make the right connections with various stakeholders [12]. Interestingly, the level of connectedness and interactions of agents is measured through Social Network Analysis (SNA), which focuses on measurement of agents' ability to access information or measure the level of cohesion from the established relationships [12].

In view of this, the adaptive agent concept redefines the leadership style which serves as a linking point to determining the dynamics of negotiations. It is believed that leaders aim to stabilize systems through their directive actions of plan and control. However, if to account for complexity in the system, then there is a strong necessity for a different leadership style [12]. Actions of leaders in complex dynamic systems are reflected in their ability to disrupt traditional leadership behaviours, by thought-provoking those, whose approaches are certain, however, who are unable to openly communicate difficult or uncomfortable issues. Further, such leaders encourage novelty where innovation is a standpoint to this tactic. This refers to the ability of leaders to establish interaction and strengthen connection with various groups of people, thus increasing levels of 'unpredictable dialogue' and eventually promoting non-traditional models which differ to plan and control models [12]. Moreover, leaders of this category are considered to be sensemakers, they are the ones who behave in a coherent and articulate manner, with clarity as to the image of the organization and co-jointly influence actions [12].

Finally some behaviour modification as a prelude to negotiation is worth considering. Systems dynamics, popularised by Peter Senge [13], employs the science of feedback and recognizes both favourable or positive and unfavourable or negative responses from a system. An example is provision of funds to a developing country may induce a negative response of discouraging growth as people become reliant on the aid and less reliant on their own initiatives. Another example of a negative feedback loop is when advertising encourages greater sales however the increased productivity required reduces the quality of the product or service thus discouraging additional sales. These two examples are unintended consequences. A positive response occurs when an individual takes of physical exercise as more exercise allows one to address even more difficult physical tasks.

A further example of the use of systems dynamics is bullying behaviours by children, in which the bully usually takes on someone with low self-confidence. A systems dynamic approach is to build a self-confidence of the bullied person rather than directly attempting to stop the bullying. This is interfering by use of a positive process. Systems dynamics may offer possibilities of working with national groups who only hark back to injuries they have received at the hands of the other nation, other than moving forward.

4 Adoption of complexity methodologies to negotiations

The significance of complexity theory has emerged in various fields and professions as a concept called upon to assist, comprehend and deal with the growing phenomena of the increased environment-determined complexity. Mainly, the SoS definition refers to the fields of sociology, biology, engineering, and the military [14], where the SoS methodologies have been created as a response to the highly complex issues, such as the ones related to the environment and dealt with by the United Nations. SoS definition varies upon the system in which professionals operate. For instance, Sage and Cuppan emphasize the operational and managerial independence, geographic distribution, emergent behaviour and evolutionary development of the complex adaptive systems; while Lukasik outlines the integration of systems into SoS, ultimately contributing to the social infrastructure's evolution. Finally Kotov puts emphasis on the complexity found within the systems [14].

5 Use of Soft Systems Methodology in the negotiation processes

One of the key issues to consider is whether use of Checkland's soft system methodology (SSM) may be of value in both assisting the parties to understand each other and in identifying a solution. The core of SSM is comparing alternative solution approaches in order to identify common meaning. This is done through addressing a number of approaches and going through the development of a rich picture, identifying CATWOE or BATWOVE, and comparing potential solutions with each other. Given that the parties to negotiation often do not have shared meaning over a number of incidents in the past, and the future, gives this methodology credibility.
An example would be looking at a dispute from each of the perspectives of [21]:

- Justice
- Business
- The military
- Fairness
- The legal process
- Building friendship
- Hate
- The leaders
- Parents bringing up their children, etc.

Further, the notion of the complex negotiation processes needs to be explored from the standpoint of systems-thinking, Strategic Assumptions Surfacing and Testing (SAST), and Critical Systems Heuristics (CSH).

An exploration of the systems-thinking approach is essential to the non-linearity of thought, channels of communication and relationship of various actors found in complex negotiation processes and conflicts. Reductionist thinking portrays individuals' tendency to perceive reality of the situation in a fragmented manner, disregarding the connectedness of the fragmented elements in one way or another [15]. As a result, conflict resolution strategies often produce only short-term solutions, by addressing the symptomatic problems alone, while the underlying causes are being overlooked and neglected [15]. Systems thinking discovers the importance of 'metanoia', which is a fundamental shift of an individual's mind toward inner awakening and innovative thinking through the collective learning undertakings [15]. A radical shift of a mind could be a lengthy process, requiring a strong commitment from the modern society to develop a robust duty of care for each other, recognizing diversity as well as promoting tolerance and understanding.

Further, Strategic Assumptions Surfacing and Testing (SAST) addresses the complex problems characterized by lack of clarity and purpose, interconnectedness, uncertainty in conflicts and societal constraints. SAST, however, is often focused instead on the matters of problem-structuring, rather than the resolution itself [16]. The methodology focuses on four main stages:

Group formation - where various stakeholders are formed into groups for the purpose of exploring each other's perspectives. A key feature is to divulge their assumptions on certain issues of concern and at the same time take into account divergency of views as an advantage rather than a disadvantage;

Assumpton surfacing - refers to uncovering and analysis of key assumptions concluded by groups in the previous stage;

Dialectical debate -brings the debate forward by choosing the best possible strategy of a particular group presented by a spokesperson;

Synthesis - summarizes the debate by finding compromise of presented strategies and finding a possible solution to the problem/issue [16].

To be specific, SAST suggests a social system design where the views and contributions of disadvantaged citizens, who normally are not included in discussion of events or the process of planning and decision making that impact on their lives, are integrated [16].

An intergration occurs in the form of 'polemical employment boundary judgement' and hence any perceived limitations with respect to disadvantaged citizens' ability to contribute in an objective way is tackled by adopting boundary judgement strategies. As a consequence, it eliminates unfairness with respect to those who are disadvantaged, and thus provides them with an equal opportunity to contribute and challenge existing norms of systems' design. Moreover, CSH gradually transcends into 'Critical Systems Thinking for Citizens', aiming to fully emancipate marginalized citizens' views and provides a framework which will enable them to participate in decisional processes [16].

Furthermore, the conceptualization of boundary judgement lies in the notion of the ability and agility of various stakeholders, at a decision level, to frequently change and modify boundaries to allow continued accommodation of viewpoints of those who are external to the process of decision making [16]. In the context of social decision-making, as a rule there are four basic modes such as force, hierarchy (judication), coalition and negotiation. If looking at negotiations, one can affirm adversaritities of two parties exist at bilateral level and complex affairs at multilateral negotiation levels.

Senge states that leaders often tend to focus exclusively on their own views and visions of the future, and that these seldom or never manage to transcend into shared vision [15]. With that in mind, CSH methodology empasizes the fact that by choosing a solid partner to future dialogues, one should also include ordinary citizens into the process of planning and decision-making, understand and respect their sense of purpose, since they are the ones potentially to be affected by the decisions made.

6 Complex systems approach to leadership

A key question to be addressed with regard to complex systems is the role of top-down leadership. Addressing international disputes has mainly been conducted by representatives of the United Nations (UN), GATT, WTO, or the leaders of countries, if one considers hostile disputes between countries [22, 23].

So how do we reconcile the fact that complex systems are self-organizing, which is essentially a bottom-up activity. Mitleton-Kelly [7] points out that self-organization, emergence and the creation of new order are three of the key characteristics of complex systems. Kaufman [18] points out that complex systems exhibit order spontaneously. Westley et al. [17] argues that:

'*Bottom-up behaviour seems illogical to Western minds, as we have a hierarchical bias against self organisation, which is displayed in our common understanding of how human change happens, especially in organisations. Our popular management magazines are filled with stories of the omniscient CEO or leader who can see the opportunities or threats in the environment and leads the people into the light. However, self-organisation is critical to achieving change.*'

The art of leadership in complex environment relies on the leaders' ability to provide favourable conditions for elements within the system to organically interact and form novel ways of reality. Leaders play a role of self-organizing facilitators, who freedom the potential and hidden abilities of people within the system. Leadership in complex systems is more concerned with the nature of a human being [19], and should embrace features of learning, creativity, agility and adaptability [20].

After introducing the concepts of behavioural and complex systems' approach to leadership style, the question arises as to how influence hierarchical approach and implement a different framework of leadership style to the real world of negotiations. Despite an outline as to the need of a different leadership approach to existing, control inherent and traditional leadership approaches. The system still, in its bureaucratic ways, resists the dynamics of self-emergent leadership paradigm.

With that in mind, the role of training and personal development would be of a high importance and fundamental in this instance. One of the core ingredients of the process would be to encourage ordinary citizens to look at the same reality in a novel way. Further, it would be also vital for leaders to include ordinary citizens in creation of vision and co-jointly make that vision a reality [19].

7 Conclusions

The following has been found:

The literature review clearly suggests that there is an urgent need to understanding the management of negotiation in all its complexity. Therefore, the advancement of complexity methodologies would represent a beneficial approach to negotiating parties, from the standpoint of:

- Recognition by key practitioners of the complexities, interdependencies and interconnectivities surrounding negotiation processes as opposed to traditional negotiation modes;
- Recognition by the leaders of the differences between the reductionist thinking and systems thinking, as well as the soft systems methodologies, such as SAST and CSH need to be tested;
- Addressing the importance of the leaders' personal characteristics, such as their emotional intelligence, agility and adaptability to various environments serving as the key tool for determining the positive dynamics and ultimate attainment of successful peace settlements.

Complexity science emphasizes non-hierarchical, but self-organizational paradigm. A novel leadership framework is fundamentally based on adaptive capacities of leaders of being able to self-emerge, create and learn to achieve change [20]. Based on the findings, the critical point of difference is that the top-down leadership style, which is widely recognized in Western societies, may not effectively operational in complex negotiation processes.

References

[1] V. A. Brown, and J.A. Harris, and J. Y. Russell, Tackling Wicked Problems Through the Transdisciplinary Imagination, Earthscan p. 62, 2010.
[2] S.S Batie, and D.B. Schweikhardt, Societal Concerns as Wicked Problems: The Case of Trade Liberalisation, OECD 2010.
[3] L. Crump, and W.Zartman, Multilateral Negotiation and the Management of Complexity, Kluwer Law International, Netherlands, pp.1-4, 2003.
[4] L. Crump, Multiparty negotiation: what is it?, *ADR Bulletin*, Vol 8, No.7, pp.2-4, Jan.2006.
[5] M. Jackson, Systems Thinking - Creative Holism for Managers, John Wiley and Sons, Chichester, UK, pp.137-138, 2003.
[6] R.O. Mason & I.I. Mitroff, Challenging Strategic Planning Assumptions, Chichester, UK, John Wiley & Sons, 1981.
[7] E. Mitleton-Kelly, Complex Systems and Evolutionary Perspectives on Organisations, The Application of Complexity Theory to Organizations, Elsevier, pp.3-20, 2003.
[8] J.Tallberg, Formal Leadership in Multilateral Negotiations: A Rational Institutionalist Theory, *The Hague Journal of Diplomacy*, Martinus Nijhoff Publishers, pp.121-123, 2006.
[9] W. Zartman, Negotiation as a Choice of Partners, The Processes of International Negotiation Program, *Network Newsletter* 33, pp.13-16, 2009.
[10] Helmsman, A Comparison of Project Complexity between Defence and other Sectors, The Helmsman Institute, Sydney, Australia, 2010.
[11] D. Hendrick, Complexity Theory and Conflict Transformation: An Exploration of Potential Implications, University of Bradford, Department of Peace Studies, p. 21, 2009.
[12] B. Ramalingam, H. Jones, T. Reba, & J. Young, Exploring the science of complexity ideas and implications for development and humanitarian efforts, London: ODI, Working Paper 285, pp.44-52, 2008.
[13] P.M. Senge, The Fifth Discipline, Australia, Random House, 1992.
[14] Y. Bar-Yam, The Characteristics and Emergent Behaviors of System of Systems, NECSI, 2004.
[15] P.M. Senge, The Fifth Discipline, The Art & Practice of the Learning Organization, Doubleday New York, pp.11-46, 2004.
[16] M.C. Jackson, Systems Thinking: Creative Holism for Managers, John Wiley & Sons, Ltd, pp.137-229 2003.
[17] F. Westley, and B. Zimmerman, and M.Q. Patton, Getting to Maybe: How the World is Changed, Toronto: Random House, 2006.
[18] S.A. Kauffman, At Home in the Universe: The Search for Laws of Self-Organization and Complexity, New York, Oxford University Press, 1995.
[19] A. Keene, Complexity Theory: the changing role of leadership, Industrial and Commercial Training, Emerald, Vol.32 Iss: 1, pp.17-18, 2000.
[20] M. Uhl-Bien, and R. Marion, and B. McKelvey, Complexity Leadership Theory : Shifting Leadership

from the industrial age to the knowledge era, The Leadership Quarterly, Elsevier, 2008.
[21] B. Williams, Soft Systems Methodology, The Kellog Foundation, 2005.
[22] E.R. Robles, Political & Quasi-Adjudicative Dispute Settlement Models in European Union Free Trade Agreements, Is the Quasi-Adjudicative model a trend or is it just another model?, Economic Research and Statistics Division, World Trade Organization, p.3, 2006.
[23] Handbook on the Peaceful Settlement of Disputes between States, Office of Legal Affairs Codification Division, United Nations Publication, New York, 1992.

'Kony2012' Movement Through a System of Systems Engineering Lens

Ryley Smithson
Masters of Applied Project Management
ECIC, University of Adelaide
Adelaide, SA, Australia
ryley.smithson@student.adelaide.edu.au

Abstract - *The 'Kony2012' movement set the record for the fastest growing viral video in history. In 30 minutes it turned many thousands of people around the world, with little to no prior knowledge of the case, into political activists. The mission of the system was to create autonomous and geographically diverse groups that would perform various tasks in order to place pressure on decision makers to take action against Joseph Kony, wanted for war crimes by the International Criminal Court (ICC). Despite some initial success the system was fatally flawed and the emergent response of individuals 'against' the movement lead to self-destruction. Lessons that can be learn from the 'Kony2012' movement include; system transparency, garnering mass public response, and influencing the online community.*

1 Introduction

'Kony 2012' was a movement created by Invisible Children and released on various Internet video and social networking sites [20]. The film promoted the 'Stop Kony' campaign, highlighting the war crimes of Joseph Kony. The video has received a combined viewing of over 100 million and was the fastest growing viral video in history, toping 100 million view in 6 days [38]. The intent of the video was to create exposure - a traditional linier project with a clear cause and effect. However, once the video reached viral status it was met with emergent responses. This paper examines the 'Kony2012' movement though a systems of systems lens by firstly providing an overview of the movement and a sequence of events surrounding the case. Secondly, the movement is dissected using various systems of systems models. Thirdly, lessons that can be learned from the movement will be outlined.

2 'Kony2012' Overview

2.1 High level Diagram

The high level diagram (Figure 1.) shows the major elements and communication involved in the 'Kony2012' movement.

Invisible Children (IC) – Communicated with individuals for and individuals against using social media (Twitter, Facebook, etc) and shared videos (YouTube, Vimeo, etcetera).

Individual For – The primary sharers of the Kony movement influenced their friends to either partake in the movement. They also, at the direction of IC, contacted various celebrities encouraging them to share the message

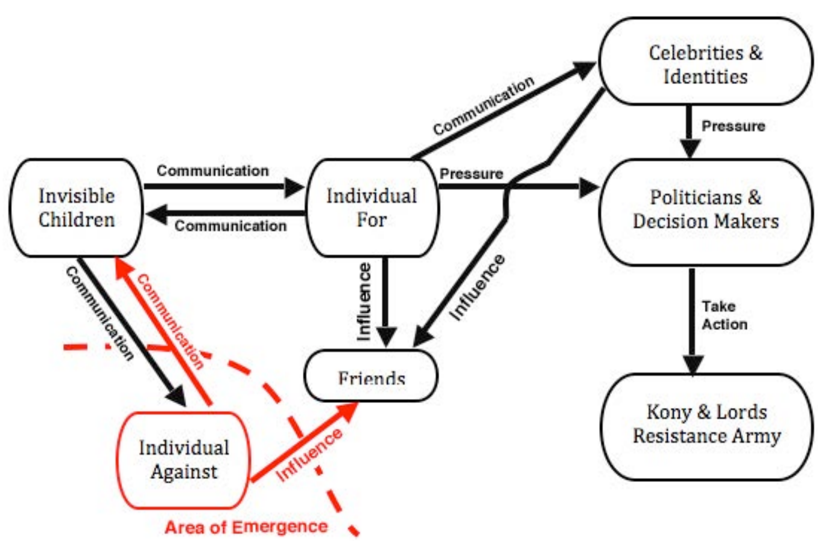


Figure 1. A high level diagram of the 'Kony2012' movement.

and pressured politicians and decision makers.

Individuals Against – Reacted to the message from Invisible Children and influenced their friends to do the same. This was the primary area of indeterminable emergence.

Friends – Came into contact with the movement via individuals for or against. They reacted in three ways; becoming part of the individual against, becoming part of the individual for or reacting in neither way and remaining within the friend section.

Celebrities and Identities – As part of the video, IC encouraged viewers to contact key celebrities and identities. They in turn influenced their large group of friends and followers, and put pressure on politicians and decision makers.

Politicians and Decision Makers – This group contains anyone with the power to take action against Kony and the Lords Resistance Army (LRA).

Kony and the Lords Resistance Army (LRA) – On the far end of the diagram is the group that IC is trying to reach.

2.2 Mission of the system

The high level diagram (Figure 1) gives a clear image of the mission of the system and the steps put in place by IC in order to reach their goal. The video and mission were posted online and shared with friends, the friends reacted and a certain amount re-shared. People contacted celebrities and influential individuals, further increasing the swell of people and putting pressure in decision makers to take action.

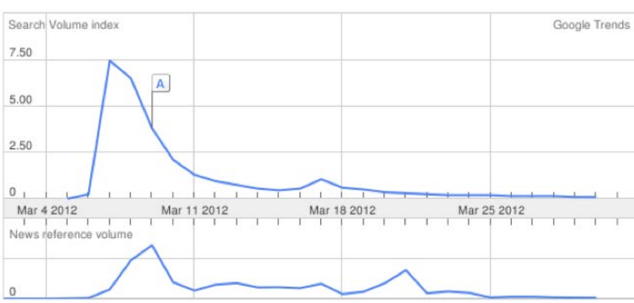


Figure 2. Google trend results for the search term 'Kony' graphed against the relative searches of the term over time [18].

2.3 Sequence of Events

The following is a recount of key events in the movement's history.

The Kony2012 YouTube video was uploaded on March 5th 2012. Following the uploading of the video, there was an immediate reaction with Google results for 'Kony' increasing by 7.5 times in the days following (Figure 2).

Three days after the first video was uploaded, March 7th, the term 'Kony' reached its peak searches on google.com. March 7th was also the date that fast moving bloggers began to collate and share credible sources highlighting the campaigns misgivings - information about the misrepresentation of information in the video and the charities credibility [13, 25, 28, 17]. By the 9th the criticism had begun to increase dramatically [9, 29, 37].

Six days after its release, and faster than any other viral campaign in history, the video, its responses and duplicates reached 100 million views. Of the top ten fastest campaigns to reach 100 million views, the Kony video has the longest runtime, 29 minutes, and it is the only video within the human rights field [4, 15, 38].

The 'Kony2012' movement was designed to culminate with a mass graffiti spree on the 20th of April 2012, aptly titled 'Cover the Night'. Before the date there were questions as to the commitment of the thousands of people that had join the social groups pledging their support to the campaign [8, 10]. On the morning of the 21st the result was clear, the 'Cover the night' events drew only a small amount of supporters [6, 30, 32].

2.4 Was the System Successful?

How you gauge the success of the Kony2012 campaign depends on the mission. If the mission was simply to 'make Kony famous,' than the campaign was a success prior to the 'Cover the Night' through the attention the video created. If the mission was to build support in order to put pressure on decision makers to launch military action, President Obama had already deployed US forces to the area in October 2011 [12, 23, 33]. Furthermore, it is clear that the vision to disarm and disband the LRA from power has been building for quite some time, thus it is impossible to state what impact the campaign has had on policy makers.

The most recent development has been the capturing of a senior LRA commander [2, 3, 22]. Despite the capture taking place months after the Kony2012 campaign started it is unlikely to be a direct result of the campaign. Long term, we many never know the effect that the campaign has had due to it's social complexity.

3 'Kony2012' through a SoSE lens

3.1 The movement through a Gorod, Sauser and Boardman model lens [1]

3.1.1 Autonomy

The video created localised autonomous, self-governing and self-organising groups driven to carry out the spreading of the message. The movement called for people to purchase 'action kits' containing poster, bracelets and other paraphernalia to 'make Kony famous' [20]. It is not the first time that this has occurred; there are organisations such as Critical Mass [27] and Flashmob [36] that also pertain to this franchised activism.

The result pulls towards the independence side; people acted independently to pertain to either side of the movement. The result is not fully to the independence side as the people on the 'for' side largely conformed to the instructions given to them by the Invisible children agency.

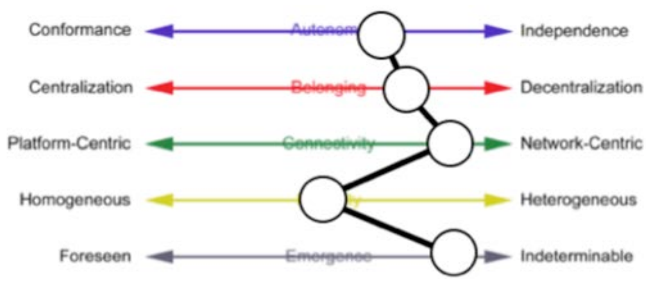


Figure 3. SoSE attributes of the 'Kony2012' movement [1].

3.1.2 Belonging

This is the key to the movement; the video showed individuals that if they joined the movement they would feel a sense of belonging and purpose. Similarly, as the crowd 'for' the campaign banded together to spread the word, the crowd 'against' the campaign also gathered. Motivated in a similar way, but for a different cause, people on the opposing side banded together along shared belief lines, sharing and swapping information for [5, 26]. Overall the people involved were a decentralized group, they made the choice to belong to either side of the movement partial to their belief. This places the result on the decentralized side. As the for side was told to organize themselves along geographical lines, it would be necessary for the result to reflect this by moving slightly towards the centralized side.

3.1.3 Connectivity

The high level of connectivity provided by social networks is the sole reason that the video was able to achieve viral status so fast. The rate of speed of which this sharing is occurring is increasing as technology and connectivity increase [38]. Without the network the sharing couldn't have occurred, placing the result firmly on the network-centric end.

3.1.4 Diversity

The group that reacted to the video's content were demographically and geologically diverse [4, 5, 14, 15, 18]. The group that was in support of the video was homogeneous; they accepted the content and their role in the system. The group that stood apposed to the video was heterogeneous, creating sub-groups structured around beliefs [24]. From one viewpoint the system was homogeneous, from the opposite it was heterogeneous, this places the result in the centre of the scale.

3.1.5 Emergence

Invisible Children were extremely unprepared for the emergent response that followed the viral status of their video. Within hours of the video being posted, the crowd had collected and shared information about the charities financial history, ranking, transparency, etcetera [11]. The emergent reaction status and emergent response of the crowd lead to one of the members of the charity being hospitalised after having a public break down [31]. To combat the negative response to the video, Invisible Children set up a Twitter feedback loop using the tag '#askICanything' [34]. The emergent properties of the against side were completely unpredictable and given the number of people involved it would be impossible to predict their behaviour beyond saying they will make decisions based on the information provided to them. The above reasons place the result firmly on the indeterminable side of the scale.

3.2 The movement through a Shred and Mostashari model lens[35]

3.2.1 Structural Complexity; Size

The project contained many millions of nodes; examples include the directors, people within the video and anyone who watched the video. As people who watched the video shared their opinions and found common ground with other people, they created more dense nodes. People who accepted the content of the video organized themselves into geographical groups to take action - sceptics of the video quickly banded together in a different way, forming groups based on common beliefs of the video.

3.2.2 Structural Complexity; Connectivity

Because of the high level of connectivity of social networks, people were able to quickly share information and facts about the campaign. The majority of swam occurred within a two-day period (Figure 2).

3.2.3 Structural Complexity; Architecture

The layout of the system is quite complex, involving many individual users and screens. Common elements in the system were video websites like Vimeo.com and YouTube.com, and social networks such as Twitter and Facebook.

3.2.4 Dynamic Complexity; Short-term

The short-term popularity of the video exposed many problems, such as various issues with the financing and ethical structure of the Invisible Children agency.

3.2.5 Dynamic Complexity; Long-term

The long-term implications of the 'Kony2012' campaign are not yet known.

3.2.6 Socio-Political Complexity

There is an immense socio-political complexity to the 'Kony2012' movement. The campaign also deals with several governments and various sensitive political issues [16].

4 Lessons Learned

4.1 What went wrong

Despite the initial views and shares of the video, the video failed to transfer into long term support. It was clear that the emergent response against the 'Kony2012' campaign was unexpected, and caught Invisible Children unprepared. Given the number of views on their other YouTube videos, highlighted above, it is understandable.

In hindsight, there were several things that IC could have done to improve the performance of the campaign; a slow and gradual uptake on the video would have created a longer lasting result and having the 'cover the night' event closer to the date of release of the video would have ensured that the feelings and motivation were fresher in peoples minds. However, both of these do not address the source of the negative emergent reaction.

The overriding factor that would have improved the reaction to the 'Kony2012' campaign was if the charity was more transparent in their processes, accounting, filming method and information.

4.2 Transperency

The transparency of IC and the 'Kony2012' movement was the major downfall of the campaign. Transparency is thought of as how open and accountable a charity, government, media source, etc. is. Transparency is important as it reduces corruption and adds to an overall feeling of quality [19].

The lack of transparency in the campaign created footholds for bloggers and critics to attack the charity soon after release. The level of transparency also affected the support that the campaign received from the media. Positive media reaction would have validated the ideas and lead to greater support from the general public.

Areas where the campaign lacked transparency can be found by dissecting the blog posts of critics that were posted within the first few days of the videos release; Oyston [17], Baker [25], Wilkerson [28]. These articles are not particularly scholarly - but unlike journalists, who took several days to gather information and judge the newsworthiness of story, they were written when the video was just building momentum, as society reacted. These blogs were the beginning of the emergent response that negatively affected the campaign.

Common themes among all the blog posts are: misinformation contained within the video, film techniques, Invisible Children's unclear finances and accountability. Misinformation in the video relates to several incorrect or misrepresented facts – one such is the location of Kony and the LRA; the video paints them as being in Uganda, however they are thought to be currently residing somewhere in Sudan. Secondly, the video used film techniques that were deemed to be to 'slick'[15] and corporate; they were designed to be emotive and felt more like a movie than a documentary. The final misinformation relates to the finances of Invisible Children. There were many questions raised about the funding and what benefits people were receiving from the money raised by IC, some of the blogs highlight IC's low accountability and transparency rating [11].

4.3 Altruistic Impulse

The most interesting characteristic of the system is how people feel belonging and the source of validation of this belonging. Bernoff and Li [7] call this the altruistic impulse, in the same way that people give blood because they think they should, users shared information because they felt they should. The video empowered them, either in a positive or negative way, to share information with their friends. The video not only created a way for people to belong but it also set up a framework for people to connect and feel validated by performing a physical action.

4.4 Online community as components in a collaborative System of Systems

In a limited amount of time and using social media as the tool, the online populous can be turned into *productive* components in a collaborative System of Systems.

Prior to the Kony movement it would have been inconceivable to motivate a large percentage of the online population towards a common cause. Large scale crowd support can now be built into a complex project. This is particularly important for social ventures.

5 References

[1] A. Gorod, B. Sauser and J. Boardman. "System-of-Systems Engineering Management: A Review of Modern History and a Path Forward." *IEEE Systems Journal*, vol 2, no. 4, pp. 484-499, Dec 2008.

[2] AFP. "*Uganda captures senior colleague of Lord's Resistance Army warlord Joseph Kony.*" Internet: http://www.theaustralian.com.au/news/world/uganda-captures-senior-colleague-of-lords-resistance-army-warlord-joseph-kony/story-e6frg6so-1226354465507, May 14, 2012 [May 30, 2012].

[3] BBC News Africa. "Uganda captures LRA senior commander." Internet: http://www.bbc.co.uk/news/world-africa-18051522, May 13, 2012 [May 30, 2012].

[4] BBC News. "Kony 2012 Movement." Internet: http://www.bbc.co.uk/news/technology-17599401, Apr 3, 2012 [April 22, 2012].

[5] C. Conti. "*KONY 2012- ADELAIDE*." Internet: http://www.facebook.com/events/248280081928359/, Mar 19, 2012 [Apr 1, 2012].

[6] C. Hingston and A. Lowe. "Kony event invisible under cover of night." Internet: http://www.theage.com.au/victoria/kony-event-invisible-under-cover-of-night-20120421-1xd60.html, Apr 21, 2012 [May 31, 2012].

[7] C. Li and J. Bernoff. "*Groundswell.*" Bostin, Mass: Harvard Business Press, c2008, pp. 61.

[8] C. Paine. "It's KONY night! Wait, Kony. Who's that again?" Internet: http://www.news.com.au/national/its-kony-night-wait-kony-whos-that-again/story-e6frfkvr-1226334418948, Apr. 20, 2012 [May 31, 2012].

[9] C. Rawlinson and K. O'Toole. "Kony social media campaign sparks scepticism." Internet: http://www.abc.net.au/local/stories/2012/03/08/3448740.htm, Mar. 8, 2012 [Apr. 20, 2012].

[10] C. Ray. "What Do Young People Think of Kony 2012 Now?" Internet: http://ideas.time.com/2012/04/20/what-do-young-people-think-of-kony-2012-now/, Apr. 20, 2012 [May 31, 2012].

[11] Charity Navigator. "Charity Rating;Invisible Children." Internet: http://www.charitynavigator.org/index.cfm?bay=search.summary&orgid=12429, Apr 2012 [April 1, 2012].

[12] CNN Wire Staff. "Obama orders U.S. troops to help chase down African 'army' leader." Internet: http://articles.cnn.com/2011-10-14/africa/world_africa_africa-obama-troops_1_obama-orders-south-sudan-central-african-republic, Oct. 14, 2011 [May 30, 2012].

[13] D. Goodman & J. Kron. "Online, a Distant Conflict Soars to Topic No. 1." Internet: http://www.nytimes.com/2012/03/09/world/africa/online-joseph-kony-and-a-ugandan-conflict-soar-to-topic-no-1.html, Mar 8, 2012 [April, 20 2012].

[14] D. Ratnam. "Young Indians gun for an African." Internet: http://articles.timesofindia.indiatimes.com/2012-03-25/people/31236430_1_invisible-children-facebook-user-posters, Mar. 25, 2012 [April, 20 2012].

[15] E. Orden and E. Steel. "How 'Kony' Clip Caught Fire Online." Internet: online.wsj.com/article/SB10001424052970204781804577271692294533870.html, Mar. 9, 2012 [Jun. 8, 2012]

[16] F. Opolot. "Response to International Discourse of LRA Activity." Internet: http://www.mediacentre.go.ug/details.php?item=1602&catId=3, Mar. 9, 2012 [Jun. 8 2012].

[17] G. Oyston. "We Got Trouble." Internet: http://visiblechildren.tumblr.com/post/18890947431/we-got-trouble, Mar. 7, 2012 [Jun. 8, 2012].

[18] Google, Inc. "Trends; Kony." Internet: http://www.google.com.au/trends/?q=kony&ctab=0&geo=all&date=mtd&sort=0, [Apr 1, 2012].

[19] International NGO Charter of Accountability Ltd. 1]"INGO Accountability Charter." Internet: http://www.ingoaccountabilitycharter.org/wpcms/wp-content/uploads/INGO-Accountability-Charter_logo1.pdf, Dec. 20, 2005[May 20, 2012].

[20] Invisible Children Inc.. "Kony 2012." Internet: http://www.youtube.com/watch?v=Y4MnpzG5Sqc Mar. 5, 2012 [Mar. 31, 2012].

[21] Invisible Children Inc.. "Invisible Children." Internet: http://www.invisiblechildren.com, [Apr. 1, 2012].

[22] Invisible Children Inc.. "Top LRA commander captured alive." Internet: http://blog.invisiblechildren.com/2012/05/14/top-lra-commander-captured-alive, May 14, 2012 [May 30, 2012].

[23] J. Ryan. "Troops making progress in hunt for Kony." Internet: http://www.usatoday.com/news/world/story/2012-05-29/joseph-kony-hunt/55260364/1?csp=34news, May 29, 2012 [May 30, 2012].

[24] J. Travis. "Kony 2012: Why It Worked." Internet: http://www.huffingtonpost.com/john-travis/kony-2012-why-it-worked-_b_1407099.html, Jun. 6, 2012 [Apr, 8 2012].

[25] K. J. M. Baker. "Think Twice Before Donating to Kony 2012, the Charitable Meme du Jour." http://jezebel.com/5891269/think-twice-before-donating-to-kony-2012-the-meme-du-jour, Mar. 7, 2012 [Jun. 8, 2012].

[26] L. Rozen. "Kony 2012: Invisible Children's viral video sparks criticism that others say is unfounded." Internet: http://news.yahoo.com/blogs/envoy/kony2012-invisible-children-viral-video-uganda-conflict-sparks-183106657.html, Mar. 8, 2012 [Mar. 31, 2012].

[27] M. Bluejay. "Critical Mass." http://critical-mass.info, 2009 [April 1, 2012].

[28] M. Wilkerson. "Guest post: Joseph Kony is not in Uganda (and other complicated things)." Intenet: http://blog.foreignpolicy.com/posts/2012/03/07/guest_post_joseph_kony_is_not_in_uganda_and_other_complicated_things, Mar. 7, 2012 [May 30, 2012].

[29] N. Mawby. "Social media sensation KONY 2012 cops criticism for 'cashing in'." Internet: http://www.heraldsun.com.au/news/more-news/its-all-over-the-webs-but-what-is-a-kony/story-fn7x8me2-1226293063302, Mar. 9, 2012 [Apr. 20 2012].

[30] N. Miller. "Catch Kony campaign loses couch potatoes." Internet: http://www.smh.com.au/technology/technology-news/catch-kony-campaign-loses-couch-potatoes-20120421-1xdqc.html, Apr. 22, 2012 [May 30, 2012].

[31] P. Harris. "Kony 2012 campaigner Jason Russell detained for public rampage." Internet: http://www.guardian.co.uk/world/2012/mar/16/kony-2012-campaigner-detained, Mar. 17, 2012 [Apr. 1, 2012].

[32] R. Carroll. "Kony 2012 Cover the Night fails to move from the internet to the streets." Internet: http://www.guardian.co.uk/world/2012/apr/21/kony-2012-campaign-uganda-warlord, Apr. 21, 2012 [May 31, 2012].

[33] S. Collinson. "Obama sends troops to help fight Africa rebels." Internet: http://www.google.com/hostednews/afp/article/ALeqM5g_k7HxM6K32WKaWD6GNF1QQf9NNw, Oct 14, 2011 [May 30, 2012].

[34] S. Gardiner. "Be less invisible: hundreds take to Twitter demanding anti-Kony group reveal its agenda." Internet: http://www.smh.com.au/technology/technology-news/be-less-invisible-hundreds-take-to-twitter-demanding-antikony-group-reveal-its-agenda-20120313-1uxai.html, Mar. 13 2012 [Jun. 8, 2012].

[35] S. Sheard and A. Mostashari. "Complexity Considerations for Systems Engineering." *CSER Conference; Stevens Institute of Technology*, 17-19 March, pp. 170-179.

[36] SYDMOB. Internet: http://www.sydmob.com, Nov. 10, 2006 [Jun. 8, 2012].

[37] T. McCarthy and P. Curtus. "Kony 2012: what's the real story?" Internet: http://www.guardian.co.uk/politics/reality-check-with-polly-curtis/2012/mar/08/kony-2012-what-s-the-story, Mar. 8, 2012 [May 30, 2012].

[38] Visible Measures. "Kony Social Video Campaign Tops 100 Million Views." Internet: http://corp.visiblemeasures.com/news-and-events/blog/bid/79626/Update-Kony-Social-Video-Campaign-Tops-100-Million-Views, Mar. 12 2012 [Apr. 1, 2012].

Dynamic Modularity: A Distributed decision mechanism in System of Systems[1]

Babak Heydari
Assistant Professor
School of Systems and Enterprises,
Stevens Institute of Technology,
Hoboken, NJ, 07093
babak.heydari@stevens.edu

Kia Dalili
Postdoctoral Associate
School of Systems and Enterprises,
Stevens Institute of Technology,
Hoboken, NJ, 07093
kia.dalili@stevens.edu

Abstract- *Effective use of modularity in distributed systems is a key to accommodate complexity arising from having multiple stakeholder requirements and increased environmental uncertainties and facilitates efficient dynamic resource allocation of the system. In this paper, a formal framework that links various classes of systems modularity to distributed decision complexities is presented by formulating modularity as an emergent phenomenon rising from environmental heterogeneity. The work also demonstrates behavioral and mathematical formulation of the problem as well as agent-based simulation of sample networked systems to verify the results.[1]*

Introduction

Modularity is a common feature of many complex networked systems. Study of modularity is crucial from the point of understanding the evolving nature of networked systems, community detection in networks and systems complexity measures. Understanding the driving forces of modularity in complex systems also helps us in design and governance of networked systems.

We view modularity as an emergent property of evolving complex networks. Profit maximizing nodes playing a network formation game in a heterogeneous environment organize themselves in a manner to balance the benefits of a more complex network against the costs of such complexity. In this setup modularity emerges as mechanism for complexity management. An engineered system on the other hand deviates from the most efficient design by practical constraints and by decisions made during the design process.

The goal of this paper is to study the behavior of independent agents in network formation games, and show that in modularity emerges naturally as a consequence of limited resource processing capacity of nodes within a heterogeneous environment.

[1] *This work was supported by DARPA contract NNA11AB35C as a part of System F6 Program.*

A basic model of cooperative games was developed in (Jackson and Wolinsky, 1996) assuming a fixed cost and benefits that scaled through network distance. (Johnson and Gilles, 2000) generalized the model by introducing a cost topology. We will generalize the set up further to allow for both cost and benefit to depend on the environmental heterogeneity, and for incorporating limited link formation capacity of nodes.

Finally to measure the degree of modularity in a network, we compare its clustering coefficient (Watts and Strogatz, 1998) against the expected clustering coefficient of a random network with the same link density. We will follow the notation and language of complex networks from (Jackson, 2008).

Let us summarize the content of this paper. In section two we will discuss the theory of emergent modularity (TEM). We will discuss the basics rules of interactions amongst agents and describe how modularity emerges as a mechanism to control the linkage costs when agents have a limited capacity but require access to resources in a heterogeneous environment.

In section three we provide the details and the results of our agent based simulations and show that modularity does emerge as anticipated in section two.

Theory of Emergent Modularity (TEM)

Emergence of Modularity- Evolving real networks are rarely formed in isolation, via nodes that have no connection to their environment. In practice, nodes of a network experience resource exchange with the environment. The decision of what nodes are considered as environment can vary depending on the context, the desired level of abstraction in studying a particular networked system and the choice of system boundary. The key point is that the state vector of systems' nodes are subject to change throughout the network formation cycles while environmental nodes are assumed exogenous to the model. The influence of the environment on the link formation-elimination is indirect and through changes it causes in the internal state vectors of the nodes.

Furthermore, each node is considered to maximize its pay-off based on certain notion of cost and benefit, developed in this work. In the presence of heterogeneous environment, networks are faced with a fundamental dilemma regarding the aggregate heterogeneity of their connections. On one hand, maximizing the diversity of connections- direct or indirect- is desirable because it ensures an access to a larger pool of possible responses to (expected) changes of environment. On the other hand each node, when considered to have finite resource processing capacity (RPC), can only handle certain level of heterogeneity in its direct connections. The reason is that each link imposes a transaction-cost on the related nodes that is a function of expected heterogeneity of them in relation to each other. The effect of the environment further amplifies this dilemma. This is because more heterogeneous

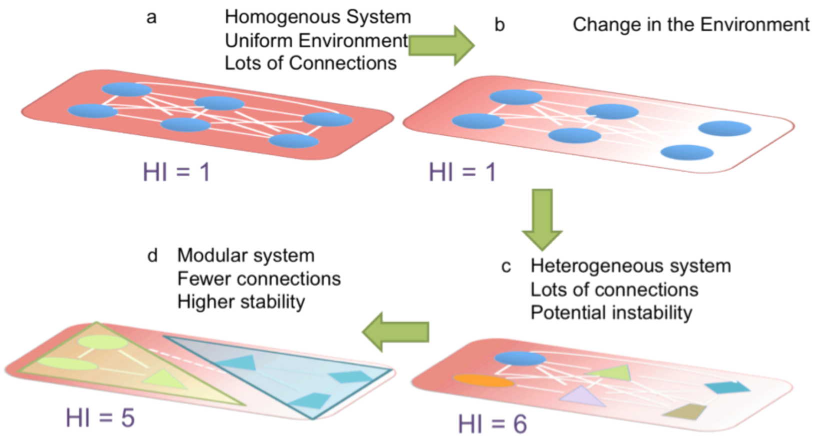


Fig. 1. Effect of environmental heterogeneity on network modularity

environments increase the expected benefit for nodes from a given diversity in their connections. Consequently, nodes benefits in this model need to be normalized to the expected environment heterogeneity that nodes are experiencing.

At the same time, this change in the environment also increases the transaction-cost of links, on average, as a result of widening the overall diversity of nodes state vectors. In this situation and intuitively speaking, modular communities emerge when each node maximizes the indirect connection diversity while keeping the total transaction- costs within its resource processing capacity. This is achieved by acquiring indirect benefits through direct connection to nodes with higher RPC who can manage to have high level of direct connection heterogeneity.

As a special case of this model, networks in which nodes have unlimited resource processing capacity will show little modularity regardless of the heterogeneity in the environment nodes. Similarly, even in the presence of limited resource processing capacity, network modularity will still be insignificant if the "transaction-cost" among nodes are kept small. The transaction-cost of a link depends on the state difference of the two related nodes and is modeled using Shannon conditional entropy measures. For other cases, the network becomes modular and its level of modularity is a function of the distribution of resource processing capacity of nodes as well as expected heterogeneity in the environment.

Influencing Factors and Their Interactions- The internal state of a node, affects the cost of link formation. It is more costly for an agent to connect to a node that is far apart compared to connecting to a node with similar characteristics. E.g. consider the communication network, it is more costly to create a direct connection to a far away node than to a node in ones vicinity. As another example consider a social network; it is easier to connect to a person with similar interests and beliefs than to one with a radically different belief system and little shared interest.

Similarly, the internal state of a node influences the benefits of link formation. Connecting to node with different internal state gives an agent a greater opportunity for resource exchange through the link. For instance in the communication network example creating a direct link to a far away node grants fast access to a new geographic location, and in a social network befriending a person with different beliefs and interests provides a new set of information to an agent, while befriending ones identical clone provides no new information or opportunity.

Two factors, the environment, and the neighbors of a node, in turn influence its internal state via resource exchange. In the first approximation we restrict our attention to the environments effect on the internal state. Therefore once the environment is set and the internal state of a node is determined it remains constant unless there is a change in the environment.

The link formation capacity for each node is a characteristic of the node. A node cannot maintain connections with total cost more than its capacity. Furthermore as the total cost of a nodes connections approach its capacity it becomes unstable and its utility rapidly deteriorates.

Lastly we incorporate two additional observations into our model. First a network should have some measure of memory; it should not reorganize instantaneously, rather it should evolve over time to a better configuration if one can be located via agent actions. Second, a random element is present in the evolution of the network. Nodes do not have the resources to examine all possible connections. We incorporate these two restrictions by randomly choosing candidate pairs for link formation (one at a time) and only allowing a link to be established if end points can accommodate the cost of link formation and benefit from it while severing fewer than one of their old connections.

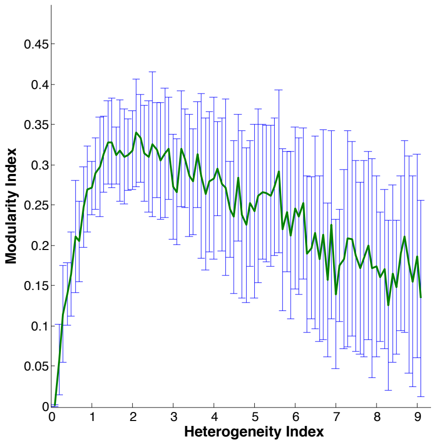
Lastly a word on the utility function utilized in our model. The guiding principles come from the following observations: Firstly costs and benefits in network formation are rarely of the same type, e.g. an agent spends time to receive information, an agent uses available bandwidth to establish a communication link with other agents and benefits from the resources available to the new neighbor. Secondly link formation costs affect agents differently depending on how much processing capacity they have left. Even when a link has a fixed formation cost, it affects the utility of an agent close to her processing capacity limit much more severely than that of an agent with a large amount of processing capacity to spare.

3 Agent-Based Simulation Results

To test our theory we set up and agent based simulation environment using Object Oriented MATLAB. As a first steps node properties are populated randomly, this includes internal states, and maximum resource processing capacity of each node. The distribution of the internal states at this stage reflects the level of environmental heterogeneity being tested.

At each time step afterwards, nodes attempt to improve their pay off by forming a new link, removing a connection, or creating a new connection while severing an old one. We sample the network after either it reaches a stable state, or after a suitable amount of time has passed and measure various network properties. Notably we measure the networks modularity index defined as the difference between the clustering coefficient of the network and the expected clustering coefficient of an Erdős-Rényi random network with the same number of links.

In the first experiment we vary the heterogeneity index of the environment while keeping all other

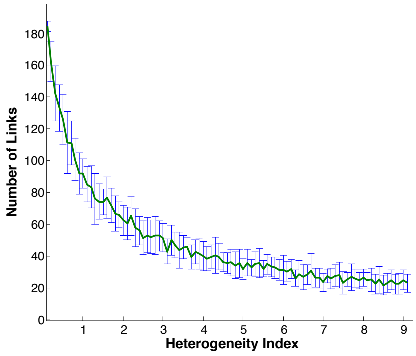


parameters fixed. The next figure show the resulting change in the modularity index of the network graphed against the heterogeneity index of the environment.

As predicted the modularity index rises rapidly with heterogeneity of the environment, but as the environmental

Fig. 3. Environmental heterogeneity results in network modularity

heterogeneity surpasses the node capacity, the network deteriorates and modularity index declines again. Another expected result is the rapid decline of the number of links as the heterogeneity index increase. This is shown in figure 3.



For the second experiment we keep the environmental heterogeneity fixed while varying the resource processing capacity of the nodes. The results are shown in figure 4 and follow closely

Fig. 2. A heterogeneous environment hampers the links ability to form links rapidly

our models prediction, i.e. when nodes have negligible resource processing capacity the network is fragmented and no structure is visible. As the resource processing capacity for nodes increases links start appearing and as the network exits the fragmented state a modular structure emerges.

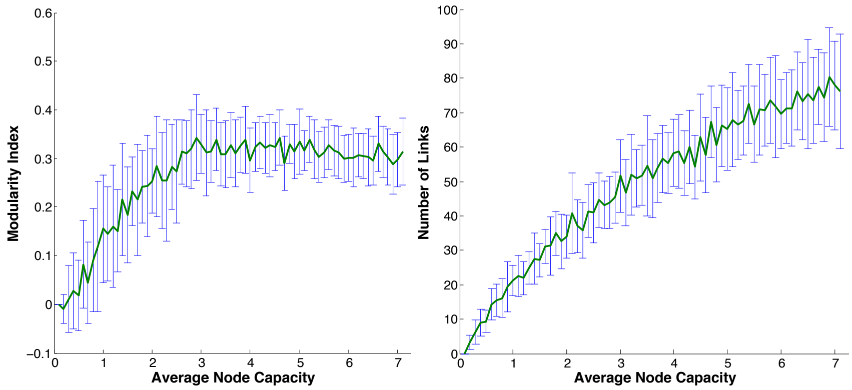


Fig. 4. As the average node capacity increases, the nodes connect to more and more neighbors and the number of links in the network increases. Similarly the network goes from a sparse network with little structure to one with a modular structure.

Bibliography

Jackson, M. O. (2005), A survey of models of network formation: Stability and efficiency, in: Demange, G., Wooders, M. (Eds.), "Group formation in economics: networks, clubs and coalitions." Cambridge University Press, Cambridge, UK, 11–57

Jackson, M. O. (2008), Social and Economic Networks. Princeton NJ: Princeton University Press.

Jackson, M. O. and Wolinsky, A. (1996), A Strategic Model of Social and Economic Networks, Journal of Economic Theory, 71, No. 1, 44—74.

Johnson, C. and Gilles, R.P. (2000), Spatial Social Networks, Review of Economic Design 5, 273–300.

Watts, D.J. and Strogatz, S.H. (1998), Collective dynamics of small-world networks. Nature 393, 440–442.

Watts, A. (2001), A Dynamic Model of Network Formation, Games and Economic Behavior 34, 331–341.

ARCNET: A System-of-Systems Architecture Resource-Based Collaborative Network Evaluation Tool

Jean Charles Domerçant, Ph.D.
Aerospace Systems Design Laboratory
Georgia Institute of Technology
Atlanta, GA, U.S.A.
jean.domercant@asdl.gatech.edu

Dimitri N. Mavris, Ph.D.
Aerospace Systems Design Laboratory
Georgia Institute of Technology
Atlanta, GA, U.S.A.
dimitri.mavris@ae.gatech.edu

Abstract – *Rapid advances in information technology and the increased emphasis on joint operations results in highly networked, interoperable Systems-of-Systems (SoS). The units comprising the SoS must act in collaboration with each other to establish and maintain information superiority and battlespace dominance. The problem that arises, however, is that there are few Modeling & Simulation tools that exist for quantifying the benefits of collaboration during a military mission. To address this need, this paper details the application of an Architecture Resource-Based Collaborative Network Evaluation Tool (ARCNET). ARCNET provides an estimation of the impact on mission effectiveness resulting from changes in collaboration between military units. These changes may be due to differences in interoperability, resource exchanges, and force structure. Ultimately, this will lead to more robust decision making during military SoS architecture selection.*

Keywords: System-of-Systems, complexity, architecture, interoperability, collaboration.

1 Introduction

Rapid advances in information technology and the increased emphasis on joint operations have led to the rise of Network Centric Warfare (NCW) as a dominant military doctrine [3, 4]. NCW relies upon highly networked, interoperable systems or Systems-of-Systems (SoS) acting in collaboration with each other to establish and maintain information superiority. This information superiority is intended to provide increased shared awareness and synchronization of forces to realize force multiplication effects and enhanced battlespace dominance.

Though the benefits that can be achieved through collaboration are desirable, there also negative consequences. For instance, individual systems/platforms may encounter information overload under certain conditions. Also, there are costs incurred when maintaining increasingly complex network infrastructures that enable greater levels of collaboration [11]. Thus, a balance must be achieved.

The problem that arises, however, is that there are few methods and Modeling & Simulation (M&S) tools that exist for quantifying the potential benefits of collaboration during a military mission. The need for such tools is especially critical during the conceptual design phase. This phase of design is where many of the SoS architecture decisions are made that have the furthest-reaching impact on SoS cost, schedule, and performance [2, 9].

To address this need, an Architecture Resource-Based Collaborative Network Evaluation Tool (ARCNET) is developed, which is also a revision to that presented in Reference [6]. ARCNET utilizes a method proposed by Perry for measuring the benefits of increased collaboration between military systems as they exchange resources such as data or information [11]. Perry's method relies upon Information Theory, with information entropy concepts developed by Shannon to assess the amount of knowledge available to a group of collaborating systems [11, 13]. Information entropy (also referred to as *Shannon entropy*) states the amount of information in the occurrence of an event is inversely proportional to the likelihood the event will occur [6].

This paper provides an overview of the adaptation of Perry's method to military SoS architectures in particular, where a SoS is defined as a "set or arrangement of systems that result when independent and useful systems are integrated into a larger system that delivers unique capabilities" [1]. Specifying an architecture is important because an architecture includes not only systems and their functions, but data flow and communications protocols, key SoS functions, as well as end-to-end functionality [5].

2 Background

Collaboration can be defined as "a process in which individuals work together to achieve a common goal. Shared information is an essential ingredient to ensure effective collaboration" [11, 12]. In general, more effective collaboration and information sharing has a direct impact on the speed and quality of decision making that occurs [6]. Perry notes that "Traditional measures of effectiveness (MoEs) usually ignore the effects of information and decision making on combat outcomes" [11]. However, to fully and accurately assess the effectiveness of a network-centric SoS architecture means that the M&S environment must be able to capture any potential differences in effectiveness that arise due to collaboration.

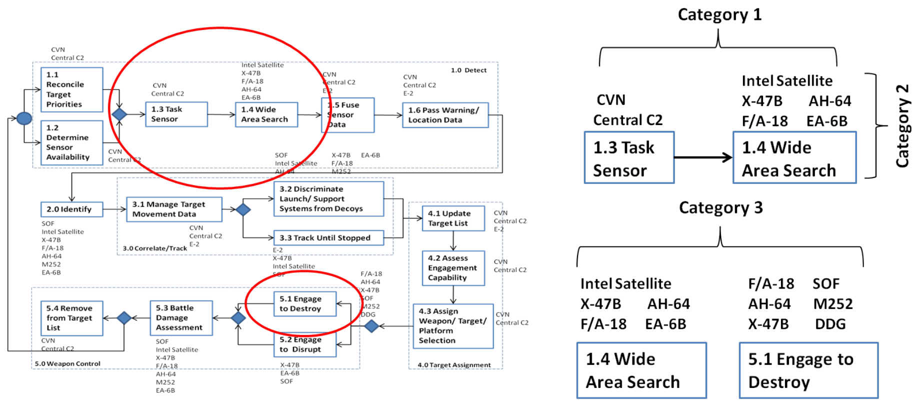


Figure 1: Example SoS Collaboration Categories.

A military mission or operation is often defined in terms of a task sequence like the one depicted on the left-hand side of Figure 1. This information helps describe the architecture in terms of the functionality that must be achieved in order for the mission to be executed. Using this task sequence and a mapping of different systems within the SoS that are responsible for performing each task, different patterns of collaboration may arise. In general, there are three distinct types of collaboration that can take place [6]:

- Category 1 — Required Task-to-Task Collaboration: Resource exchanges occur between systems performing direct sequential tasking.
- Category 2 — Shared Task Collaboration: Resource exchanges occur between systems with shared tasking.
- Category 3 — Non-specific Collaboration: Resource exchanges occur between systems performing indirectly linked tasking.

In a platform-centric SoS architecture collaborations are often limited to Category 1. Traditionally, understanding and modeling this type of collaboration has been the main focus of military M&S efforts. When Category 1 collaborations are inadequate, SoS functionality is either severely degraded or may not exist at all. Once the necessary resources have been exchanged between two systems performing different tasks, an independent assessment of system performance for a particular task can be conducted.

In a network-centric SoS architecture, Category 2 and Category 3 collaborations may be present in addition to Category 1 collaborations. Category 2 collaborations allow units to share task-related information and work together rather than independently. Finally, Category 3 collaborations typically involve the widespread sharing of data & information so that a Common Operational Picture (COP) may emerge.

Depending upon the tasks involved and the specific operational scenario, the impact of Category 3 collaborations may or may not be directly evident. Adequately assessing the impact of Category 3 collaborations on mission performance requires a fuller understanding of modeling situational awareness for interacting agents. This is an area of ongoing research by the authors. With this in mind, the focus will be to quantify the impact of Category 2 collaborations on attaining high levels of mission effectiveness.

Task 1.4 from Figure 1 will be used as an example to illustrate the usefulness of this approach. Task 1.4 is to conduct a wide area search, and may be performed by a combination of different military platforms working either independently or in concert. These include a military intelligence satellite and various manned and unmanned aerial platforms such as an X-47B Unmanned Combat Aerial Vehicle (UCAV), F/A-18 & EA-6B military jets, and an AH-64 attack helicopter. For brevity, the analysis will focus on utilizing F/A-18's and X-47B's.

3 Methodology

The first step in Perry's methodology is to determine the primary sources of uncertainty that affect mission requirements and MoEs, then to quantify this mission uncertainty using a probability distribution, $f(x)$. Once $f(x)$ is assumed, information entropy provides a measure of the average amount of information in $f(x)$. Next, a normalized knowledge function is created by mapping entropy onto a [0,1] knowledge scale. Once the amount of knowledge that can be gained from independent operations is estimated, the next step is to model the impact of collaboration on overall knowledge. A total system collaboration factor is then calculated. This is used to quantify the expected increase in knowledge due to collaboration, while also factoring in negative effects such as those due to increased network complexity [6, 11].

It is assumed that collaboration between different search platforms will result in greater probabilities of detection (P_d). The primary source of uncertainty that exists when conducting the wide area search is the number and location of enemy targets located within the search area. Therefore, to determine $f(x)$, the search area is divided into individual sensor grids. To simplify the analysis, the individual sensor grids are chosen to be small enough that multiple enemy units will not be co-located within a single search grid. Consequentially, locating an enemy unit in one search grid is independent of finding an enemy unit in another grid. This allows modeling the search as a Poisson process, and an exponential distribution is chosen for $f(x)$. Equations (1 & 2) show the probability density function and expected value, respectively.

$$f(x) = \lambda e^{-\lambda x} \tag{1}$$

$$E(X) = \frac{1}{\lambda} \tag{2}$$

$E(X)$ can be interpreted as the average number of sensor grids that must be searched between detections. As the number of enemy units increases, then E(X) should decrease and vice versa. It is important to note that the exponential distribution also possesses the *memoryless property*, meaning that if the random variable X measures the time until a certain event occurs and the event has not occurred by time x_o, the *additional* waiting time for the event to occur beyond x_o has the same exponential distribution as X. If any of the original assumptions are changed, the memoryless property may not hold, and a more suitable distribution should be chosen [7].

Next, an independent sensor coverage parameter, λ_i is defined for each search platform, taking into account the following factors:

- Search Area (A)
- Sensor Sweep Width (s)
- Platform Search Speed (v)
- Effective Search Time (T)

Equation (3) combines these factors into a parameter λ_i for each aircraft. The effective sensor coverage parameter is simply the summation of the independent sensor coverage parameters for each asset taking part in the search.

$$\lambda_i = 1 - e^{-(svT/A)} \tag{3}$$

Once $f(x)$ has been defined, the information entropy of $f(x)$ can be calculated using Equation 4:

$$H(x) = -\int_{\infty}^{\infty} ln\left[f(x)\right] f(x)dx = ln\left(\frac{e}{\lambda}\right) \tag{4}$$

This allows the mapping of entropy onto a [0,1] knowledge scale by selecting an upper bound on the entropy. The knowledge associated with the exponential distribution is defined in Equations (5) - (7):

$$K(\lambda) = \begin{cases} 0 & \text{if } \lambda < \lambda_{min} \\ ln\left(\lambda/\lambda_{min}\right) & \text{if } \lambda_{min} \leq \lambda < e\lambda_{min} \\ 1 & \text{if } \lambda \geq e\lambda_{min} \end{cases} \tag{5}$$

Where:

$$e \times \lambda_{min} = 1 \tag{6}$$

And:

$$\lambda_{min} = \frac{1}{e} = 0.368 \tag{7}$$

A plot of the normalized knowledge function is presented as Figure 2. This reflects the relative amount of information that can be gained from independent search operations. A saturation point is reached when the combined sensor coverage is greater than one, as the presence of additional search aircraft yields no net improvement in P_d.

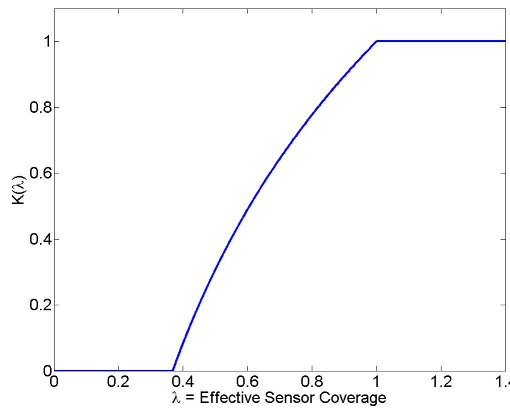


Figure 2: Normalized Knowledge Function.

The next step is to model the impact of collaboration. Perry's method assumes a statistical reliability model. The collaboration between a pair of systems takes the general form of Equation (8), where $r(s)$ is called the failure rate function and is dependent on the nature of the collaboration.

$$c_{ij}(t) = 1 - e^{-\int_0^t r(s)} ds \tag{8}$$

When there is no time to collaborate, *i.e.*, $t = 0$, then $c_{ij}(t) = 0$. Additionally, the time at which successful collaboration occurs between two systems depends on the form of the failure rate function for that collaboration, shown in Equation (9).

$$c_{ij}(t) = 1 - e^{-\theta t} \text{ for t} \geq 0 \tag{9}$$

A constant is selected so that earlier successful collaboration can be modeled by simply increasing the constant value. For ARCNET, this constant is chosen using interoperability as a guideline. Interoperability is formally defined as "The ability to operate in synergy in the execution of assigned tasks" [8]. Moreover, for the direct and satisfactory exchange of information or services between communications-electronics equipment, "the degree of interoperability should be defined when referring to specific cases" [8].

The degree to which two military systems interoperate to exchange information and services is often expressed as an interoperability level, or IOL. Various scales for defining IOLs exist. Ref. [10] provides an IOL scale for an X-47B. Also, a generalized IOL scale is presented in Ref. [6]. Typically, the minimum value on an IOL scale corresponds to little or no resource exchange. As additional resources are exchanged to support the mission on hand, IOL increases in value.

While IOL provides an indication of the *amount* of resources exchanged, the *satisfactory* exchange of resources may best be represented in terms of reliability. Thus, a reliability probability can be used, so that θ takes on values between zero and one. The exact value of θ depends on how often a failure occurs while attempting to satisfactorily exchange resources between two systems. More research is needed in this area to define a formal methodology for determining interoperability in terms of reliability probabilities, taking into account the amount of variability that exists under changing operational conditions.

Since Equation (9) is a cumulative probability, the probability density function can be calculated:

$$f_{ij}(t) = \theta e^{-\theta t} \tag{10}$$

The probability density function is an exponential distribution with $1/\theta$ being the mean time for systems i and j to collaborate with a variance of $(1/\theta)^2$. The entropy calculation for the exponential distribution with parameter θ is:

$$H(t) = -\int_{t=0}^{\infty} ln\left[\theta e^{-\theta t}\right]\theta e^{-\theta t} dt = ln\left(\frac{e}{\theta}\right) \tag{11}$$

The collaboration entropy function can now be used to develop a measure of knowledge by assessing the "certainty" in the density function. An approximate upper bound is assigned to H(t), the equivalent to assigning a maximum expected time to complete a collaboration. Letting $(1/\theta)_{max} = \theta_{min}$ represent the maximum expected time, then a measure of certainty or knowledge can be written as:

$$K(t) = ln\left(\frac{e}{\theta_{min}}\right) - ln\left(\frac{e}{\theta}\right) = ln\left(\frac{\theta}{\theta_{min}}\right) \tag{12}$$

Perry notes that $K(t)$ is a dimensionless quantity and therefore can be used directly to influence combat MoEs. He also states that it is desirable to normalize $K(t)$. This can be accomplished by noting that when $\theta = \theta_{min}$, $K(t) = ln(1) = 0$ and when $\theta/\theta_{min} = e$, $K(t) = ln(e) = 1$. This suggests the following definition for the knowledge gained from the collaboration between systems i and j:

$$K_{ij}(t) = \begin{cases} 0 & \text{if } \theta < \theta_{min} \\ ln\left(\theta/\theta_{min}\right) & \text{if } \theta_{min} \le \theta < e\theta_{min} \\ 1 & \text{if } \theta \ge e\theta_{min} \end{cases} \tag{13}$$

Where the following values are chosen for this example:

$$\theta_{max} = e \times \theta_{min} = 0.95 \tag{14}$$

And:

$$\theta_{min} = \frac{\theta_{max}}{e} = \frac{0.95}{e} = 0.35 \tag{15}$$

In Equation (13), for small values of θ, the mean and variance are large, thus implying great uncertainty and therefore little knowledge. For large values of θ, the opposite is true and therefore considerable knowledge is gained. In this way, $K_{ij}(t)$ models the positive effects of having more time, on average, to reliably collaborate. The resulting collaboration curve that can be plotted using Equations (13) - (15) is similar in shape to that of Figure 2, though the input to the collaboration curve is now θ rather than the effective sensor coverage.

After this, a total system collaboration factor is determined that accounts for all pairs of collaborating systems. Perry uses an inverse reliability model for sequential overall system collaboration. This results in the following:

$$K_M(t) = 1 - \prod_{[i,j]} K_{ij}(t) \tag{16}$$

Once $K_M(t)$ is calculated, the effects of collaboration are represented using the following linear model:

$$K_C(\lambda) = K_M(t)\left[1 - K(\lambda)\right] + K(\lambda) \tag{17}$$

Equation (16) assumes the that the collaboration effect from each collaborating pair is equal in value. Also, As IOL increases, so does the value of θ. Larger values of θ equate to larger $K_{ij}(t)$ between systems. For a given number of systems, using Equation (16) would result in a smaller value of $K_M(t)$ and thus a smaller value of $K_C(\lambda)$. Thus, ARC-NET utilizes a modified method of combining the effects of localized collaboration instead. The modified method begins with defining an $n \times n$ collaboration matrix, CK. Each entry of CK, or ck_{ij} is equal to the $K_{ij}(t)$ between the pairs of systems that collaborate to detect, identify, and track enemy units. Equation (18) provides an example of three search platforms in collaboration.

$$CK = \begin{bmatrix} 0 & .9 & .7 \\ .9 & 0 & .8 \\ .7 & .8 & 0 \end{bmatrix} \tag{18}$$

Next, the maximum absolute eigenvalue of CK ($\lambda_{max}^{(CK)}$) and the number of systems (n) is used to determine $K_M(t)$:

$$0 \le K_M(t) = \frac{\lambda_{max}^{(CK)}}{(n-1)} \le 1 \tag{19}$$

Since the CK matrix consists of actual systems and not *system types*, the diagonals of the CK matrix are zero (collaboration is defined as occurring with another external system). Normalization is therefore performed by dividing by $(n-1)$ nodes. Using the CK matrix in Equation (18), $K_M(t) = 0.8$. $K_M(t)$ will vary with the number of search units, their patterns of collaboration, and how reliably they interoperate.

The last step is to include the negative effects of collaboration in the model. Perry attempts to capture this as the complexity that results as the total number of connections between systems increase. However, Perry acknowledges that "complexity alone, as defined by the number of connections in a network, is clearly not enough to assess the effectiveness of network-centric operations" [11]. Reference [6] details a method for capturing the complexity of military SoS architectures, from both a functional standpoint and in terms of the complexity that arises from sharing and processing resources. Therefore, instead of simply using the number of connections, ARCNET uses the techniques described by Reference [6] instead.

More specifically, since IOL gives an estimate of the amount of resources that are exchanged, ARCNET utilizes this information to define a resource processing matrix, RP. RP is a symmetric $n \times n$ matrix that defines the relative amount of information or resources shared between each pair of collaborating systems. In this example, IOL is used. If available, other metrics such as bytes of data may be used instead. Force structure must be taken into account as well, so if there are multiple systems of a given type, this is reflected in the RP matrix. The total number of systems and logical interfaces help directly define the complexity of the resource sharing network. Equation (20) provides an example RP matrix for three systems with varying IOLs.

$$RP = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 2 \\ 2 & 2 & 0 \end{bmatrix} \tag{20}$$

Reference [6] also details how the specific nature of each resource that is being exchanged across a network influences complexity. Thus, the collection of resources can result in a very simple or a highly complex *resource state space*. For the wide area search example, this could mean that the information shared between platforms ranges from simple coordinates to annotated, overlaid imagery mapping transmitted across multiple time scales and with widely varying security classification. During this particular analysis, the resource state space will be considered homogeneous and not a contributing factor to network complexity.

Similar to what was done for the collaboration matrix CK, a normalized network complexity factor C_n can be calculated in the following manner:

$$C_n = \frac{(1/IOL_{max}) \times \lambda_{max}^{(RP)}}{n} \times \left(1 - \frac{1}{\sqrt{n}}\right) \tag{21}$$

$$0 \leq C_n \leq 1 \text{ for } n \geq 1 \tag{22}$$

By taking the normalized eigenvalue of the RP matrix, a weighted connection density is obtained. This weighted connection density is then corrected to avoid penalizing architectures with small force structures the same as architectures with large ones. Without the correction factor, a network consisting of two systems with maximum interoperability would have the same normalized network complexity factor as a network consisting of 100 nodes all operating at maximum IOLs. For the example RP matrix, $C_n = 0.095$, assuming that $IOL_{max} = 5$. In comparison, $C_n = 0.8910$ for a $n = 100$ network with the maximum amount of connections and all systems collaborating with an IOL of five. The final knowledge equation, taking into account complexity effects is then:

$$K_{CC}(\lambda) = [1 - C_n] \times \{K_M(t)[1 - K(\lambda)] + K(\lambda)\} \tag{23}$$

Equation (23) is now used to determine the total knowledge obtained by the systems collaborating to perform the wide area search. $K_{CC}(\lambda)$ is then directly equated with the P_d of enemy units within the search area. This aids the overall M&S effort in determining SoS capability.

4 Results & Analysis

With the method formally defined, an analysis using ARCNET is conducted. The collaboration between one F/A-18 and varying numbers of unmanned X-47B's is explored. The F/A-18 is given a $\lambda_i = 0.3$, while each X-47B is assigned a $\lambda_i = 0.2$. Intermediate calculations are given in Table 1.

Table 1: ARCNET Calculations

Num. X-47B's	Eff. Sensor Coverage (λ)	$K(\lambda)$	C_n IOL = 1 (All)	C_n IOL = 2 (All)
1	.50	.31	.03	.06
2	.70	.64	.06	.11
3	.90	.89	.08	.15
4	1.1	1.0	.09	.18

P_d is determined for different combinations of force structure, θ, and IOL. The same value of θ is applied for all search platforms, thus $K_M(t) = \theta$. Figure 3 shows the results for two collaboration alternatives as well as the case where each platform searches independently.

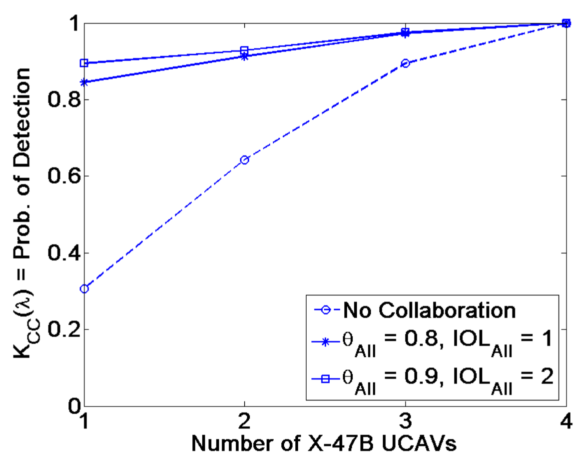


Figure 3: F/A-18 and X-47B Collaboration Results

The example problem depicted here is a relatively simple one, but ARCNET allows for the analysis of larger and more diverse architectures as well. This analysis includes varying force structure and interoperability as resources are exchanged, which are common trades for military architectures. ARCNET also allows insight into the effects of different network structure/collaboration patterns between different platforms. For example, during independent operations, when only one F/A-18 and one X-47B are used, the P_d is extremely low. However, collaboration yields a substantial benefit, and captures the impact of massing effects rather than force. It is not until three X-47B's along with an F/A-18 are used independently that they surpass the P_d of one F/A-18 collaborating with just one X-47B.

From Figure 3, the impact of collaboration is readily visible, but there are diminishing returns even for this simple task. It remains to be seen, however, if additional benefits may still exist for Category 3 collaborations when this point is reached. Collaboration also provides additional value when one considers potentially losing platforms due to enemy action or unavailability. During independent operations, the loss of each X-47B causes a severe drop in detection capability. In contrast, greater capability can be retained in the face of casualties during network-centric operations. An area of further research is to confirm the potentially large gains in capability that can be achieved through collaboration, possibly using agent-based modeling. Also, other methods of combining the factors comprising Equation (23) should be investigated. Finally, a formal examination of different task types and developing proper uncertainty distributions associated with each may result in a fully automated process.

5 Conclusions

Overall, the incorporation of Perry's method into the M&S environment provides the first steps in adequately assessing the impact of collaboration on combat outcomes. ARCNET allows system architects to examine tradeoffs in interoperability, resource exchange, and force structure for SoS architectures. Using an example problem consisting of a fighter aircraft in collaboration with varying numbers of unmanned aerial combat vehicles, ARCNET proves useful in assessing the benefits and limitations of collaboration. ARCNET not only aids M&S efforts by providing an estimate of the increase in performance due to collaboration, but also illustrates the point at which the gains achieved from collaboration diminish as both force structure and network complexity increase.

The principal advantage of ARCNET is that it is intended for use by system architects during the early phases of the conceptual design effort, and therefore does not rely upon highly detailed network models. These models tend to focus on the physical/technical implementation details of the resource exchange network. This information may not be present during conceptual design, though the framework should prove flexible in accommodating this information if available. ARCNET is more useful in determining the required patterns of collaboration and resource exchanges that fundamentally influence the basic network structure. Ultimately, this will lead to more timely, robust decision making during military SoS architecture selection.

References

[1] "Defense acquisition guidebook," Internet Publication, last Accessed October 20, 2010. [Online]. Available: https://dag.dau.mil/Pages/Default.aspx.

[2] *Pre-Milestone A and Early-Phase Systems Engineering: A Retrospective Review and Benefits for Future Air Force Acquisition*, Air Force Studies Board, Division on Engineering and Physical Sciences, Committe on Pre-Milestone A Systems Engineering, 2008.

[3] D. S. Alberts, J. J. Garstka, and F. P. Stein, *Network Centric Warfare: Developing and Leveraging Information Superiority*, 2nd ed., ser. CCRP Publication Series. DoD C4ISR Cooperative Research Program, Aug. 1999.

[4] A. K. Cebrowski, "Network centric warfare: Its origin and future," in *Proceedings of the Naval Institute*, vol. 124, Number 1, January 1998, pp. 28–35.

[5] J. Dahmann, G. Rebovich, R. Lowry, J. Lane, and K. Baldwin, "An implementers' view of systems engineering for systems of systems," in *Systems Conference (SysCon), 2011 IEEE International*, april 2011, pp. 212 –217.

[6] J. C. Domercant, "ARC-VM: An architecture real options complexity-based valuation methodology for military systems-of-systems acquisitions," Ph.D. dissertation, Georgia Institute of Technology, December 2011.

[7] A. J. Hayter, *Probability and Statistics for Engineers and Scientists*, 3rd ed. Duxbury (Thomson, Brooks/Cole), 2007.

[8] *Joint Publication 1-02: Department of Defense Dictionary of Military and Associated Terms*, Joint Chiefs of Staff, April 12 2001, (As Amended Through 31 July 2010).

[9] D. N. Mavris and O. Bandte, "A probabilistic approach to multivariate constrained robust design simulation," *SAE Technical Paper*, no. 975508, 1997.

[10] *STANAG 4586*, 1st ed., NATO, April 2004.

[11] W. Perry, R. W. Button, J. Bracken, T. Sullivan, and J. Mitchell, *Measures of Effectiveness for the Information-Age Navy: The Effects of Network-Centric Operations on Combat Outcomes*. RAND, 2002.

[12] W. Perry, D. Signori, and J. Boon, "Exploring information superiority," RAND, Tech. Rep., 2004.

[13] C. E. Shannon and W. Weaver, *The mathematical theory of communication*. University of Illinois Press, Urbana, 1949.

A Mathematical Model for Formulating Interdependence of Autonomy and Belonging in Systems of Systems

Darabi H. R. Author
School of Systems and Enterprises
Stevens Institute of Technology
Hoboken, NJ, USA
hdarabi@stevens.edu

Mansouri M. Co-author
School of Systems and Enterprises
Stevens Institute of Technology
Hoboken, NJ, USA
mo.mansouri@stevens.edu

Gorod A. Co-author
The University of Adelaide
Adelaide, SA, Australia
alex.gorod@adelaide.edu.au

Abstract - *Among different proposed characteristics of a system of systems, autonomy and belonging are widely accepted as definitive. Since these two characteristics were perceived as independent from one another, their underlying relationship has not been explicitly explored by scholars.*

This paper proposes that autonomy and belonging are interdependent characteristics according to the purpose of a system of systems and its constituent elements. The coherence between the purposes of the constituent elements and the system of systems defines the level of belonging, and the independence of these purposes is conceptualized as autonomy. The mathematical formulation of this interdependence is presented to analyze the correlation between these two characteristics.

Keywords: Autonomy, Belonging, SoS Characteristics, SoS Modeling, System of Systems Simulation.

1 Introduction

A system of systems (SoS) engineering is a modern engineering concept, which was developed to embody the complexity of emerging interconnected myriad of systems. Although the SoS perspective provided a new platform for understanding complex systems dynamics, a comprehensive set of definitions, methods and tools is still being continually developed. An important part of this process is to clarify the definition and characteristics of a SoS.

Autonomy and belonging are established as two definitive characteristics of a SoS [1, 2]. However, the SoS literature reveals that no significant attention has been paid to examining the interdependence between autonomy and belonging because these two characteristics are considered as independent [3].

In this paper, as part of an ongoing research, we aim to present a mathematical model to formulate the interdependence between autonomy and belonging. To achieve this objective, the definitions of autonomy and belonging in the literature are attuned to be suited for a mathematical formulation. This formulation provides a foundation to explore the interdependence between autonomy and belonging. This interdependence is based on the purpose of the constituent elements and the purpose of the SoS. Therefore, according to the definitions, autonomy is the ability of constituent elements to follow their own purpose, while belonging is their ability to adapt to the purpose of the SoS to which they belong [1].

The subsequent section of this paper reviews the current definitions of a SoS, the role of definitive characteristics in defining a SoS, and the meaning of autonomy and belonging in the literature. The developed mathematical formulation of the definitions is presented in Section 3. A case-study of swarm robots is used to show the application of the concepts and their relevance to real world problems. This case is presented in Section 4. The fifth and final part of the paper includes conclusion and suggestions for further studies.

2 Literature Review

2.1 A System of Systems (SoS)

In order to address the dynamic nature of modern, complex and interdependent systems, the term SoS has become widely used. For example, in explaining the usage of the term in the military domain, Mnthrope Jr. describes a SoS as the interactions and synergism of command, control, computers, communications, and information (C4I) systems with intelligence, surveillance, and reconnaissance (ISR) systems [4].

Other definitions for a broader purpose of a SoS have been proposed by several scholars. Shenhar, a pioneer in systems engineering, has defined a SoS as "an array [which] is a large collection of systems functioning together to achieve a common purpose" [5]. According to Shenhar, a conjunction or conglomeration of systems construct a SoS [5].

Kotov considered a SoS as a communication structure. As he indicates, "by systems of systems (SoS) we mean large-scale concurrent and distributed systems, the components of which are complex systems themselves. Communicating structures are hierarchical structures that represent SoS in a uniform, systematic way as composition of a small number of basic system objects" [6].

In developing an enterprise-wide system of systems engineering (SoSE) framework, Carlock and Fenton described a SoS as a large-scale, distributed, and concurrent system, which is comprised of complex elements [7]. They further attempted to provide a three layered framework for a SoS design and architecture.

Federation of systems is the counterpart term for a SoS, which is used by Sage and Cuppan. They define these federations of systems as "systems that are themselves comprised of other component systems, and where each of the component systems serves organizational and human purposes" [8].

Although the aforementioned definitions have provided a body of knowledge to understand a SoS, some scholars sought to identify the distinguishing characteristics of a SoS. For example, in Maier's view, the SoS characteristics are: (1) operational independence of the elements, (2) managerial independence of the elements, (3) following an evolutionary development path, (4) exhibiting emergent behavior, and (5) being geographically distributed [9].

Bar-Yam analyzed different disciplines such as biology, military studies, and sociology to integrate the characteristics of a SoS in these fields. He subsequently found that common characteristics between the different definitions include: (1) evolutionary development, (2) emergent behavior, (3) self-organization, (4) adaptation, (5) complex systems, (5) individual specialization, and (6) synergy [10].

Boardman and Sauser proposed five distinguishing characteristics to separate a system from a SoS. These characteristics are: (1) autonomy, (2) belonging, (3) connectivity, (4) diversity, and (5) emergence [1]. Gorod et al. mentioned that different combinations of these characteristics create different types of SoSs[2].

A study by Bjelkemyr et al. states that a SoS shows the same characteristics of complex adaptive systems. These characteristics are (1) evolutionary behavior, (2) self-organization, (3) heterogeneity, (4) emergent behavior, (5) small-word and scale-free network attributes [11].

The two definitive characteristics of Boradman and Sauser are explored in this paper. A comprehensive literature review about the history of a SoS is thoroughly presented in [3].

2.2 Autonomy & Belonging

The managerial and operational independence of a SoS's constituents, proposed by several scholars, illustrates their autonomy [9, 10]

In examining a SoS in the risk management context, Sage emphasizes the inherent tension in a SoS. This tension arises because of simultaneous independence and interdependence of the constituent elements in a SoS [12].

Boardman and Sauser's definition of autonomy states that "autonomy is the ability to make independent choices; the right to pursue reasons for being and fulfilling purposes through behaviors" [13]. As they put it, constituent elements in a SoS exercise autonomy to "fulfill the purpose of system of systems" [14]. They define belonging as "happiness found in a secure relationship" [13]. In their view, the reason for belonging in a SoS is based on the cost/benefit analysis of constituent systems [14].

2.3 Decision-Making in a SoS

Multi-criteria decision analysis (MCDA) provides a comprehensive framework for analyzing decisions. This framework is used in this paper to formulate the interdependence of autonomy and belonging in a SoS. It enables the decision makers to create a structure of the problem and to model their own values and judgments. It further helps them to synthesize the course of action [15].

In the framework, in any single decision, the objective of the decision-maker is to select between a set of different available options, or a range of options. The decision-maker is interested to maximize, or minimize, a set of different criteria or values, which is usually represented by a vector $v = \{v_1, v_2, v_3, \dots, v_n\}$. Therefore, the problem can be formulated as maximizing or minimizing with a weighted summation of the values being subject to the restrictions of the problem. The decision is the choice between the available options based on the values and formulation [15].

There are different methods to solve multi-criteria decision-making problems. The most well-known methods are multi-attribute utility theory (MAUT), analytical hierarchy process (AHP) and outranking [16]. Since the MCDA framework is only used to formulate the interdependence between autonomy and belonging, the description of these methods is omitted.

3 Interdependence of Autonomy and Belonging

The premise of this paper is that in a SoS, the autonomy of the constituent elements and their belonging to the SoS are two interdependent characteristics. To facilitate the discussion, we propose working definitions of autonomy and belonging, which are adjusted for later mathematical formulation:

Autonomy *is the ability and desire of a constituent element of a SoS to pursue its own purpose.*

In other words, the more the constituent elements in a SoS are autonomous, the more they are able to go for their own good. A less autonomous system means that there are more restrictions to the behavior of constituent elements to follow their individual will. Likewise,

Belonging *is the ability and desire of a constituent element to pursue the purpose of the SoS.*

Putting it differently, belonging of a constituent element to the SoS is the level that the purpose of that constituent element is in alignment with the purpose of its SoS.

Following these definitions, the constituent elements' *autonomy from* and *belonging to* the SoS, is determined by the purpose of the constituent element and the SoS. The more the purposes of the constituent elements are convergent to the purpose of the SoS, the more belonging they have and vice versa.

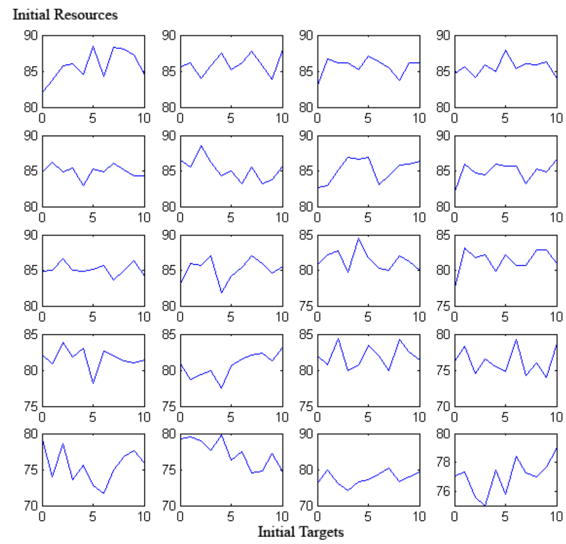


Figure. 2. Targets found in different environments

Similar to the model we have presented in our earlier work [17], each constituent element in the SoS is pursuing a set of purposes [18]. Although this purpose creates value for the stakeholders of that constituent element, it is not necessarily in alignment with the purpose of the SoS as the SoS stakeholders are different. Subsequently, the belonging of constituent elements in a SoS is the correlation between their purpose and the SoS's purpose. On the other hand, the autonomy is the independent part of their own individual purpose from the purpose of the whole.

To demonstrate the mathematical formulation, we assume that each constituent element in a SoS is facing a multi-criteria problem in each decision it needs to make. The decision variables are a set of *leverages* for that constituent element that are shown by $\{v_1, v_2, v_3, \dots, v_n\}$, and based on the utility that each choice creates. This decision-making problem is formulated as:

$$U_t^a : R^n \to R \qquad U_t^a = f(v_1, v_2, v_3, \dots, v_n) \tag{1}$$

Since the purpose can change through time, the indices *t* is used to represent the possible changes.

The SoS is faced with a similar multi-criteria decision making problem, which can be shown as:

$$U_t^{SoS} : R^n \to R \qquad U_t^{SoS} = g(v_1, v_2, v_3, \dots, v_n) \tag{2}$$

Using this formulation, we are able to define belonging as a parallel component of a constituent element's purpose to the purpose of the SoS:

$$Belonging\ X = \overrightarrow{U_t^{x\parallel}} = (\overrightarrow{U_t^x}.\overrightarrow{U_t^{SoS}}) \times \frac{\overrightarrow{U_t^{SoS}}}{\left|\overrightarrow{U_t^{SoS}}\right|} \quad (3)$$

As a result, the overall level of belonging to the SoS is the addition of the belonging of constituent elements, as described in the following formula:

$$\sum_x \overrightarrow{U_t^{x\parallel}} = \sum_x (\overrightarrow{U_t^x}.\overrightarrow{U_t^{SoS}}) \times \frac{\overrightarrow{U_t^{SoS}}}{\left|\overrightarrow{U_t^{SoS}}\right|} \quad (4)$$

According to the definitions, autonomy will be the independent component of the constituent element utility function:

$$Autonomy = \overrightarrow{U_t^{x\perp}} = \overrightarrow{U_t^x} - \left(\overrightarrow{U_t^x}.\overrightarrow{U_t^{SoS}}\right) \times \frac{\overrightarrow{U_t^{SoS}}}{\left|\overrightarrow{U_t^{SoS}}\right|} \quad (5)$$

The autonomy of two constituent elements can be in two independent directions. Therefore, unlike the belonging, the mathematical vectors of the autonomy of different constituent elements are not necessarily in the same direction. Consequently, the overall level of autonomy in the SoS is the summation over the absolute magnitude of the autonomy of the constituent elements:

$$\sum_x \left|\overrightarrow{U_t^{x\perp}}\right| = \sum_x \left|\overrightarrow{U_t^x} - \left(\overrightarrow{U_t^x}.\overrightarrow{U_t^{SoS}}\right) \times \frac{\overrightarrow{U_t^{SoS}}}{\left|\overrightarrow{U_t^{SoS}}\right|}\right| \quad (6)$$

These two relationships are illustrated in Figure 1. The interactions between the constituent elements and the SoS is simulated and presented in a swarm robot case study.

4 Case study: SWARM Robots

To exhibit the applicability of the proposed definitions and the interdependence between autonomy and belonging, a case study of swarm robots is presented. A set of interconnected robots is modeled based on a swarm intelligence of insects in the nature,. According to Hinchey et al. “Swarms consist of many simple entities that have local interactions, including interacting with the environment. The emergence of complex, or macroscopic, behaviors and the ability to achieve significant results as a team result from combining simple, or microscopic, behaviors” [19].

As mentioned in [20, 21] a swarm of robots is an instance of a SoS. Hosking and Sahin provided a set of discrete-event simulation models with XML-based message exchange protocol to simulate swarm robots [21] and a SoS in general [22]. In this research, agent-based simulation is used instead of discrete-event simulation to analyze the dynamic of interactions of robots in a swarm.

For illustrative purposes, the concept of the model is simplified. A set of swarm robots searches for a number of targets, which are dispersed in the environment. To find the targets, they require energy to move in the environment and to survive. In the initialization of the model, each robot owns a number of resources. During the model execution, each robot should search for the resources, which provide the energy for survival and movement. These resources are also dispersed in the environment.

We assume that the robots in this model are adaptive systems and the SoS consists of these multiple integrated systems, or robots, working jointly towards a common goal. As in any complex adaptive system (CAS), the outcome of the model is dependent on the relationships and the interactions of the constituent elements [23]. Different learning strategies are modeled to analyze the impact of adaptive behavior on the autonomy and belonging in the SoS. Two worthwhile analyses can be conducted on a macro and/or micro-level(s). First, the overall autonomy and belonging in the SoS could be measured based on the behavior of agents. It is also possible to analyze each constituent element level of autonomy and behavior based on its adaptation strategy.

This model provides a very useful conceptualization of autonomy and belonging in a SoS. In this example, the belonging of the robots to the SoS is determined by their ability to find the targets. Maximizing the number of the targets found in the environment is the ultimate purpose of the SoS. As a result, serving this purpose determines the level of belonging of one constituent element to the SoS as a whole.

Likewise, autonomy of the robots is their ability to serve their own requirements and purpose. In this model, the requirement of the robots is to survive. Therefore, the ability of the robots to survive is an indicator of the autonomy in the SoS. The discretion is required to realize that the survival of a larger number of robots does not necessarily mean finding more targets in the environment.

The impact of different initial conditions should be measured to ensure that the model does not replicate an imposed behavior. The imposed behavior is a result of defining simple principles for adaptive agents. As a result of imposed rules, the outcome of the whole SoS is an aggregate behavior of its constituents. In this model, because the robots have the ability to learn and to adjust their behavior based on their cognition of environmental conditions, the model does not reflect an imposed behavior.

The preliminary output of the model is presented in Figure 2. The number of the targets found in each round of simulation is depicted in different conditions and settings. This output can show the impact of environmental conditions on the autonomy and belonging in the SoS.

5 Conclusion

The objective of this paper is to illustrate the interdependence between autonomy and belonging in a SoS. The proposition of this paper is that this interdependence is related to the purpose of constituent elements and the purpose of the SoS. The more in alignment these two purpose are, the higher the level of belonging is and vice versa.

To provide a mathematical representation of this interdependence, we used a multi-criteria decision-making framework to formulate the purpose of a SoS and the purpose of its constituent elements. Having this formulation, belonging is the parallel component of the purpose of the constituent element to the purpose of the SoS. In a similar analogy, autonomy is the interdependent component of the constituent element utility function to the SoS's utility function.

The relationship between autonomy and belonging is illustrated using a model of swarm robots as a case study. Through modeling, we will estimate the autonomy and belonging of a SoS as well as its constituent elements. The research will shed some light on understanding internal dynamics of a SoS, providing a better method for measuring the SoS's autonomy and belonging.

The current paper lays the foundation for future research on measuring the level of autonomy and belonging in a SoS and designing engineering mechanisms to achieve a desirable level. It is also an important future research to provide mathematical definitions of other characteristics of SoSs and explore the possibilities of their interdependencies too..

References

[1] J. Boardman and B. Sauser, "System of Systems - the meaning of "of"," presented at the IEEE Conference on System of Systems Engineering, Los Angeles, CA, 2006.

[2] A. Gorod, et al., "System of Systems management: a network management approach," presented at the IEEE International Conference on System of Systems engineering, San Antonio, TX, 2007.

[3] A. Gorod, et al., "Sytem of Systems engineering management: A review of modern study and a path forward," IEEE systems journal, vol. 2, pp. 484-499, 2008.

[4] W. H. J. Manthorpe Jr., "The emerging joint system of systems: A systems engineering challenge and opportunity for APL," John Hopkins APL Technical Digest, vol. 17, pp. 305-313, 1996.

[5] A. J. Shenhar and Z. Bonen, "The new taxonomy of systems: toward an adaptive systems engineering framework," in IEEE Transactions on systems, man, and cybernetics - part A: systems and humans, 1997, pp. 137 - 145.

[6] V. Kotov, "Systems of Systems as Communicating Structures," Hewlett Packard1997.

[7] P. G. Carlock and R. E. Fenton, "System of Systems (SoS) enterprise systems engineering for information-intensive organizations," Systems Engineering, vol. 4, pp. 242–261, 2001.

[8] A. P. Sage and C. D. Cuppan, "On the systems engineering and management of system of systems and federations of systems," Information Knowledge Systems Management, vol. 2, pp. 325–345, 2001.

[9] M. W. Maier, "Architecting principles for systems-of-systems," Systems Engineering, vol. 1, pp. 267–284, 1998.

[10] Y. Bar-Yam, et al., "The characteristics and emerging behaviors system of systems," NECSI: Complex Physical, Biological and Social Systems Project, 2004.

[11] M. Bjelkemyr, et al., "Definition, classification, and methodological issues of system of systems," in System of system engineering: principles and applications, M. Jamshidi, ed Boca Raton: CRC Press, 2009, pp. 191-206.

[12] A. P. Sage, "Conflict and of risk managment in complex system of systems issues," in IEEE International Conference on Systems, Man and Cybernetics, Hague, Netherlands, 2003, pp. 3296 - 3301.

[13] B. Sauser and J. Boardman, "Taking hold of system of systems management," Engineering Management Journal, vol. 20, pp. 3-8, 2008.

[14] B. Sauser, et al., "Systomics: toward a biology of system of systems," IEEE Transactions on systems, man, and cybernetics—part A: systems and humans, vol. 40, pp. 803-814, 2010.

[15] J. Figueira, et al., Multiple Criteria Decision Analysis: State of the Art Surveys vol. 78. Boston: Springer, 2005.

[16] I. Linkov, et al., "Multi-Criteria Decision Analysis: A Framework for Structuring Remedial Decisions at Contaminated Sites Comparative Risk Assessment and Environmental Decision Making." vol. 38, I. Linkov and A. Ramadan, Eds., ed: Springer Netherlands, 2005, pp. 15-54.

[17] H. R. Darabi, et al., "A framework for decision making in extended enterprises: the FAA NextGen case," in International Congress on Ultra Modern Telecommuncation and Control Systems (ICUMT), Moscow, Russia, 2010.

[18] A. Gorod, et al., "Paradox: Holarchical view of system of systems engineering management," in International Conference on System of Systems Engineering, 2008 (IEEE SoSE '08), Singapore, 2008, pp. 1-6.

[19] M. G. Hinchey, et al., "Swarms and Swarm Intelligence," Computer, vol. 40, pp. 111-113, 2007.

[20] F. Sahin, "Robotic swarms as system of systems," in System of Systems Engineering: Innovations for the 21st century, M. Jamshidi, Ed., ed Hoboken: John Wiley & Sons, 2009.

[21] M. Hosking and F. Sahin, "Testability of a Swarm Robot using a System of Systems Approach and Discrete Event Simulation," in 5th International Conference on System of Systems Engineering, United Kingdom, 2010.

[22] F. Sahin, et al., "A system-of-systems simulation framework and its applications," in System of Systems Engineering: principles and applications, M. Jamshidi, Ed., ed Boca Raton: CRC Press, 2009, pp. 107-132.

[23] J. H. Holland, "Complex Adaptive Systems," Daedalus, vol. 121, pp. 17-30, 1992.

Integrated Approach and Decision Making Algorithms for Complex Systems Effectiveness Evaluation

Simeone M. Solazzi, Francesco Ciambra, Michele Sinisi
SELEX Sistemi Integrati S.p.A. – A FINMECCANICA Company
ssolazzi@selex-si.com, fciambra@selex-si.com, msinisi@selex-si.com

Abstract - *The modern operational concept of naval integrated Combat System (C/S) focuses on a key performance indicator that is the System Effectiveness as an integrated vision of the capability of the System to perform its mission and the probability for the System to be available to perform its mission. An innovative integrated approach based on Model Based Systems Engineering (MBSE) and Integrated Logistic Support (ILS) Engineering has been defined for the System Effectiveness evaluation. The C/S Effectiveness can be evaluated through algorithms based on a combination of the Integrated Functional Breakdown (IFB) and an FMECA approach at C/S level.*

Keywords: Systems Effectiveness, Model Based Systems Engineering, Decision Support Systems.

1 Integrated approach for complex systems effectiveness evaluation

A naval integrated Combat System (C/S) is a complex system composed of sensors, navigation system, weapon systems, telecommunications equipment and services and of the Combat Management System (CMS), connected each other through an internal networking system. The modern operational concept of naval integrated C/S focuses on a key performance indicator that is the *System Effectiveness* as an integrated vision of the capability of the System to perform its mission (that is in the Systems Engineering domain) and the probability for the System to be available to perform its mission (that is in the Logistic Engineering and ILS domain) [1]. This definition of the key performance indicator of a C/S means that there is a relationship between two events: 1) the Operational C/S (consisting on the functional integration of the Main System and the Logistic Support) is full performance according to its functional requirements at the beginning of a mission and 2) the C/S is able to guarantee that full performance for the whole mission time. A specialization for the general formulation is reached when the System Effectiveness is evaluated depending on the C/S mission profile and Logistic Support scenarios. This new vision of the main key performance indicator for an Operational System has been explained and refined by the new NIILS Directive (*Normativa Interforze per il Supporto Logistico Integrato*) of the Italian MoD for the harmonization of the acquisition process [2], by means of the following main features:

- To adopt a Systems Engineering approach to manage the life cycle of the Operational System.
- To increase the effectiveness of the ILS Process through the use of appropriate tools to manage and integrate System design data and ILS Process data recorded in a database called the Product Common Source Data Base (PCSDB).
- To define relationship between System Views and ILS Views, throughout the life cycle of the Operational System.

To answer to the best practices identified by this performance-oriented directive, an improvement on industrial processes is needed and this is what can be obtained with an integrated approach based on Model Based Systems Engineering (MBSE) principles, methodologies and data and including also ILS Engineering data, to define and to capture relationship between the System Views (e.g. Architectural View, Functional View, etc.) and the ILS Views (i.e. Logistic View, Maintenance Tasks View, Manuals View, etc.), and considering in particular the Logistic View and logistic support requirements as input in the MBSE process. As shown in Figure 1, we consider the core of the MBSE process is the Integrated Systems Functional Model (ISFM) [3] [4], a model based system design method, suitable for the complete representation of the architecture, state transitions and dynamic behaviour of a complex systems. The ISFM plays a fundamental role throughout all the life-cycle of the System being the repository where to collect and to manage information and data needed for the specification, design, integration, verification and validation of the System. The ISFM is compliant with the guide lines described in the NIILS directive for the realization of the Product Common Source Data Base (PCSDB). The PCSDB is an information tool (a database in a formal language that is the Unified Modelling Language [5]) through which the Program Management Office (PMO) of the MoD and its Industrial Partner can share information in a standard modelling notation about the Operational System throughout its life-cycle.

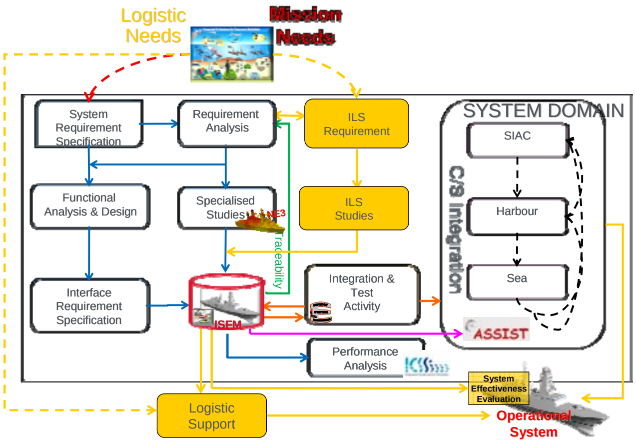


Figure 1: Model Based System Engineering e ILS Engineering Integrated Approach

The ISFM captures the whole system behaviour in a single object-based database, coded in the SysML language [6], integrating the functional model of each Sub/System, including the complete model of Combat Management System (CMS) and the Human Factors design items.

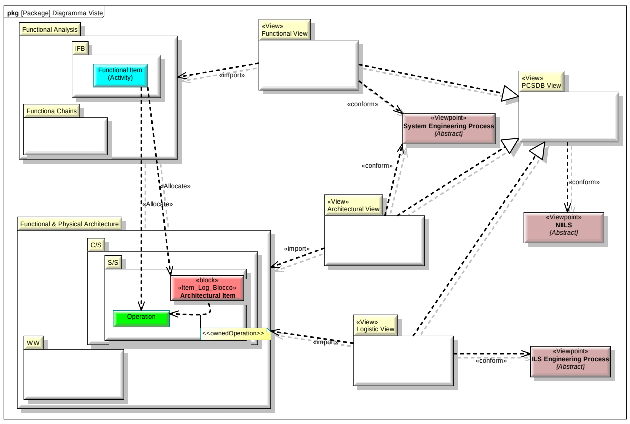


Figure 2: Views diagram in the ISFM

The ISFM can characterize, capture and manage the complexity of the Operational System connecting the Functional Views to the Architectural Views and to the additional Logistic View afferent to the ILS Engineering (Figure 2). According to this approach, we will explain in this paper how the System Effectiveness can be evaluated using a combination of the Integrated Functional Breakdown (IFB) part of the Functional View of the System in the ISFM and output of the Functional Analysis discipline [7], and an FMECA-like approach at C/S level (i.e. for the whole complex system): lost and/or degradation on the performance of Sub/Systems (S/S) due to critical failures are analyzed with respect to their consequences on S/S functions and then they are propagated with a bottom-up Bayesian approach through the IFB, allowing both the evaluation of the System Effectiveness (the root of the functional tree), and also the forecast of its evolution.

2 Complex systems basics diagnostics

Equipment and subsystems of a complex system should be able to perform automatic diagnostics (Built-In-Test, BIT) using internal embedded resources. They make use of standard sensors and devices (Built-In Test Equipment, BITE) to provide a comprehensive source of data to carry out an accurate diagnostics at the unit level. External resources, both hardware and software, connected on-demand to the Sub/System, e.g. maintenance PC with embedded maintenance applications and utilities, can be considered as BITE. Automatic diagnostics can be defined in the following way referring to the S/S operational state during its execution:

- On-Line BIT performs diagnostic checks not interfering with system operations. On-line BIT is performed when the system is in the operational state to find out failures in real-time. Failures detected on the S/S through the On-Line BIT can be distributed at C/S level by updating Health Status and Full Status messages accordingly.

- Off-Line BIT performs diagnostic checks not well compatible with the operational state of the system (i.e. they are usually performed when the system is in the Maintenance or Off state).

The Failure Mode and Effects Analysis (FMEA) is a methodology within the ILS Engineering to find all failure modes within equipment or sub/systems and their effects on the equipment functionalities. This analysis enables to identify and isolate which can degrade the mission or the safety of the equipment. This kind of analysis can be executed following either a physical/hardware approach or a functional approach. The physical approach can be used when the hardware components of the equipment can be univocally identified and the FMEA analysis follows a bottom-up approach with a hierarchy structure from the lower elements and defining for these parts all the possible failure modes and analyzing the effects of each one on the components and on the equipment hierarchy. The functional approach can be used when the elementary parts of the systems can not be univocally or easily identified and in this case the FMEA analysis follows a top down approach considering the functionalities of the equipment and of all of its parts. The criticality analysis (CA) is a technique to extend the qualitative output of the FMEA analysis with quantitative results. All the failure modes identified by the FMEA are classified considering the severity, the effects, the probability to occur, etc. The FMECA, a technique combining FMEA and CA, makes possible to establish the effects of failures on the equipment and to extract the criticalities of failure modes.

Typically, diagnostic information and data of an equipment or Sub/System are usable through a local console. When the equipment or Sub/System is integrated within a complex system the diagnostic information should be also distributed at system level, i.e. through Health Status and/or Full Status messages, sent periodically or on event through the internal network. These diagnostic data allow to monitor and to get statistical information about the operational availability of the whole complex systems and of its Sub/Systems and equipment.

3 Integrated Functional Breakdown

The Functional Analysis is the methodology of the System Engineering [7] used to explain how a complex system works. The basic idea for the Functional Analysis is that the system is viewed as computing a complex function (or, more generally, as solving a complex information processing problem) [3], [4]. Functional Analysis assumes that such processing can be explained by decomposing this complex function into a set of simpler functions that are computed by an organized system of sub-processors. When this type of decomposition is performed, the sub-functions that are defined will be simpler than the original parent function, and as a result will be easier to explain. The Functional Analysis produces the Integrated Functional Breakdown (IFB) of the complex system. The IFB is structured by levels. Passing through one level to the following, C/S functions are more detailed until the reaching of the S/S functions or operations at the lowest level. In our model of IFB we can identify some levels of functional decomposition:

Level 0 (C/S Class of Functions): is the classification of all C/S functions according to the functional requirements of the C/S.

Level 1 (C/S Capability): is the complete list of C/S capabilities for each functional class defined at Level 0.

Level 2 (C/S Sub-Capability): if necessary, this level of the IFB can be used to represent the decomposition of the related C/S capability in more than one sub-capability.

Level 3 (C/S Function): this is the first level at which it is possible to build and show relationships between functions.

Level 4 (C/S Sub-Function): if necessary, it represents the decomposition of the related C/S function in more than one sub-function.

Level 5 (S/S Function): the lowest level of the IFB represents the atomic functions of the C/S that are allocable to an equipment or S/S of the C/S (Figure 3).

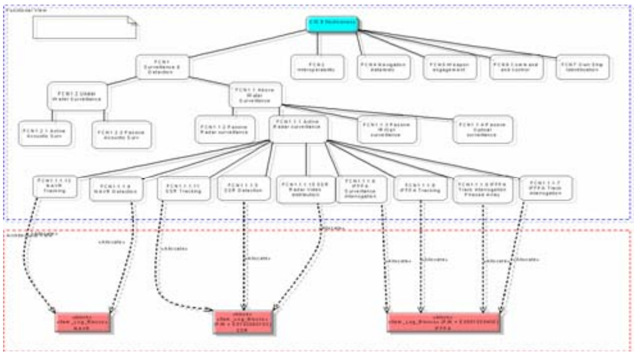


Figure 3: Allocation of C/S functions to equipment and S/S

If necessary the IFB can contain more intermediate levels in order to have the breakdown of the C/S functions at the right degree of detail that is useful in C/S definition and design. Anyway at the lowest level of the IFB the functions of the C/S correspond to the capabilities of the equipment or S/S to perform its mission with respect to the integrated C/S context. Note that such capabilities are obviously affected by the operative status of the equipment or S/S: if the equipment or S/S is affected by critical failures so that it is no more able to perform its mission as required within the integrated C/S, then it is the C/S itself to be degraded for its mission. In this way considering the IFB of a C/S along with the C/S diagnostic, with the on-line or off-line analysis and monitoring of the operative status of S/S or equipment, we get the vision of the C/S as an Operative System.

4 FMECA at C/S Level

As seen before in this paper, FMEA and FMECA are methodologies (owned by the ILS Engineering) designed to identify potential failure modes for equipment or S/S, to assess the risk associated with those failure modes, to rank the issues in terms of importance and to identify and carry out corrective actions to address the most serious concerns. These methodologies, when implemented through the use of appropriate sensors or devices, provide basic but accurate diagnostic data for a S/S or equipment. The major limitations of these diagnostic models is that either they are based exclusively on failure space or, also when a Functional Diagnostic Model is used, at least they should be brought back into failure mode before they can be used to implement on-line diagnostics. It lacks the ability to correlate the failure space with the functional space in which the capabilities of the equipment or S/S are placed. The additional step we must introduce to realize a modern concept for a C/S is to evaluate how equipment or S/S capabilities are affected by the occurrence of critical failures and how the effect of a critical failure can propagate through the IFB to affect the performance of the whole C/S. We define this approach *FMECA at C/S Level* (i.e. *Failure Effects on Functional Analysis*). Note that if the IFB is an output of the System Engineering methodologies, in the same way the FMECA at C/S Level uses methodologies and data of the ILS Engineering. So an Integrated Project Team between these two departments of the Company is appropriate for a joint analysis of functional performances and operational availability of the C/S.

5 "Expert judgment" decision making algorithms

The Functional FMECA at C/S Level is based on the elements of the complex system which have some system capabilities allocated to their operations according to the Architectural View and to the Functional View. The failure effects on the components of the system are analyzed respect to their effects on the system functions (i.e. functional-node) at the lower level of the IFB. This kind of analysis is propagated through the IFB to the functional-node at the higher level and to the root of the IFB, to evaluate the system effectiveness. The FMECA analysis at system level has been modelled is SysML using parametric diagrams [7] (Figure 4): this is another aspect of the integration of the Logistic View in the ISFM, in compliance with NIILS Directive.

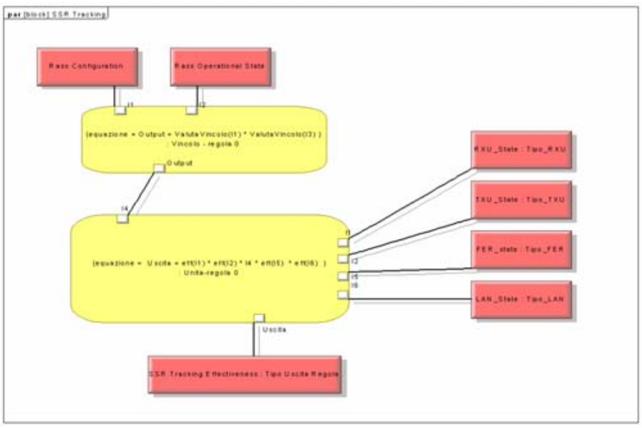


Figure 4: SysML parametric diagram for effectiveness evaluation of a functional-node of the IFB

The algorithm has been modeled in an expert judgment processor that works as a support decision making algorithms. The main principles are the following:

- Within an integrated C/S the diagnostic information of S/S and equipment are distributed through the internal networking system by means of Health Status and/or Full Status messages, which are sent periodically or on event. These diagnostic data can be received by a dedicated processing unit and then collected in a centralized repository to monitor and to get statistical information about the operational availability of S/S, equipment and of the whole integrated C/S.

- The core of the expert judgment algorithm is to evaluate how equipment or S/S capabilities are affected by the occurrence of critical failures and how the effect of a critical failure can propagate through the IFB of the integrated C/S to affect its performance respect to the mission of the C/S.

- With the collection of the diagnostic data of S/S and equipment of the integrated system stored in a database we can predict the evolution of System Effectiveness as the risk that some degradation can occur respect to the current value.

The IFB released as the output of the Functional Analysis of the C/S design shall be converted as a look-up table to be implemented in the repository of the expert judgment processor. Lost of performance on C/S equipment and S/S due to failures or degraded operation conditions, can be analyzed respect to their consequences on sub-system capabilities and then propagated through the IFB allowing the evaluation of C/S Effectiveness, as the root of the IFB, using a bottom-up aggregation process. Additional conditions to take into account in the C/S Evaluation algorithm are the asset and the configuration of the S/S in the C/S: the IFB requires that the S/S shall work with the mastership of the Combat Management System to be controlled and directed to perform the C/S mission. C/S functions or S/S capabilities at the lower level of the IFB are mapped in a numeric indicator in the range [0, 1] according to the output of the FMECA process at C/S Level previously described. The Effectiveness evaluation process for each function at any other level of the IFB is a succession of consecutive weighted sums which allow to determine parent-function (at the i-th level of the IFB) effectiveness from children-functions effectiveness (at the (i-1)-th level of the IFB). If a functional-node at the i-th level of the IFB is parent of N children functional-nodes then we can define the branch weights $\alpha 1, \alpha 2, \dots, \alpha N$ with the constraint $\alpha 1 + \alpha 2 + \dots + \alpha N = 1$ (constraint C1). Weights are used to express the C/S Mission dependency for the C/S Effectiveness evaluation: the branch weights $\alpha 1, \alpha 2, \dots, \alpha N$ can be defined differently according to the real effort of each functional-node of the IFB for the C/S Mission. If the Effectiveness evaluation of children-functional-nodes results in Ec1, Ec2, … , EcN with Eci in [0, 1], then the aggregation rule can be written in the following way (Eq. 1):

$$E_P = \sum_{i=1}^{n} \alpha_i E_{ci} \tag{1}$$

Weights αi are user defined with a Bayesian approach according to the mission of the C/S allowing to obtain different values for System Effectiveness evaluation from the same data collected in the system repository. Some algorithms have been implemented to help the user to assign weights to each branch of the IFB. With the manual quantitative or qualitative method the user can directly assign weights paying attention to satisfy the constraint C1. On the other hand the *Pair Comparison Analysis* is a semi-automatic method to support user in the weights decision process providing a pair comparison matrix useful to evaluate the importance of each function-node compared to other function-nodes at the same level of the IFB and referring to the same parent-function; starting from the user opinions the system automatically computes the weights vector satisfying the constraint C1. The weights vector v resulting from the pair comparison analysis is the dominant eigenvector of the pair comparison matrix (Eq. 2):

$$A\mathbf{v} = \lambda^* \mathbf{v} \tag{2}$$

where A is the pair comparison matrix and λ* is the dominant eigenvalue, that is the element with the highest module in the eigenvalues vector λ = (λ1, λ2, … , λN). The pair comparison matrix A is consistent if for every element a of the matrix the condition aik = aij * ajk (constraint C2) is satisfied. Since users are usually to define inconsistent pair comparisons matrix (that aik ≠ aij * ajk) the eigenvector will be the solution that minimizes the inconsistency error. A measure of inconsistency introduced by users in its weights decision process is an output of the algorithm with the indicator I (Eq. 3):

$$I = \sum_{j=1}^{n} \sum_{i=1}^{n} \sqrt{\left(a_{ij} - \frac{v_i}{v_j} \right)^2} \tag{3}$$

At every run of the Pair Comparison Analysis the algorithm points out computing the indicator I the evaluation or the evaluations by which users has introduced the main inconsistency introduced by the user in the pair comparison matrix: so it suggests to users which decision is recommended to be correct.

With a collection of the Health Status of the equipment of the C/S stored in a database we can predict the evolution of C/S Effectiveness as the risk that some degradation can occur on C/S Effectiveness respect to the current value. We call this prediction algorithm "*Effectiveness at Risk*" (EaR) in analogy with financial mathematics and risks management concept for risks measure. This probabilistic evaluation is based on the forecast of the System Effectiveness distribution generated by a Monte Carlo process on a future time interval [t1, t2] (t1 = forecast start time, t2 forecast stop time) and a confidence level (these two parameters are user-defined). The algorithm can be described as follows:

1. To compute the probability of the status transition (e.g. from Normal to Fault) and the mean time of that transition, for all equipment of the integrated C/S and for all of their possible status transitions.

2. To generate N possible future scenarios into the time interval [t1, t2] defined by user, according to the transition probability and the mean time of that transitions computed at the previous step 1.

3. To compute a mathematical distribution of the Effectiveness for each functional-node of the IFB.

4. To compute the Effectiveness value for each functional-node of the IFB which is the percentile of the mathematical distribution of the Effectiveness associated to the confidence level defined by the user (Figure 5).

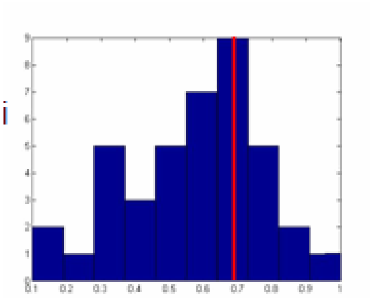


Figure 5: System Effectiveness distribution (based on real data) and confidence level for Effectiveness at Risk

EaR algorithm can compute the probability of the status transition and the mean time of that transition from historical data collected on a database respectively by a frequency analysis and by a distribution analysis under the assumption of normal distribution for the status transitions. The same parameters can be computed from a-priori data e.g. the Mean Time Between Failure (MTBF) and the Minimum Time To Repair (MTTR) of each equipment. Note that these parameters are parts of the Logistic View of the System. With a Monte Carlo process it is possible to generate many different classes for acceptable future scenarios for C/S Effectiveness starting from data computed at step 1. The scope of the Monte Carlo runs is to generate a number of scenarios and to analyse statistical properties of the generated classes. The Monte Carlo process is applied for all functional-node of the IFB, the same as step 3 and step 4 computations.

6 Conclusions and follow-on

The modern operational concept of naval integrated Combat System (C/S) focuses on a key performance indicator that is the System Effectiveness as an integrated vision of the capability of the System to perform its mission and the probability for the System to be available to perform its mission. This concept is highlighted in the NIILS Directive of the Italian MoD to manage efficiently the logistic acquisition within the overall defence system acquisition process. Consequently, new methods and practices are required to Defence Industry to integrate functional and performance requirements (in the System Engineering domain) with supportability requirements (in the ILS Engineering domain) at C/S level and at CSE level.

The algorithms for System Effectiveness evaluation (and Effectiveness at Risk algorithm for prognostics) have been first simulated in a *Expert System Model* using Matlab® as the modelling tool and then implemented in a diagnostic processing unit: this prototype will be integrated in the near future in the C/S of a Naval Unit of the Italian Navy for a testing mission. The new approach described at the beginning of this paper and for System Effectiveness evaluation are the foundation for a modern state-of-the-art functional and operational decision support system whose application can be easily extended from a naval C/S to all distributed complex systems.

Other applications of System Effectiveness evaluation algorithms can be the use in an expert decision support systems to suggest corrective and preventive maintenance actions when the System Effectiveness is less than 1. Algorithms can run in different ways starting from the Functional View of the ISFM, to identify the branch (or the branches) of the IFB where the biggest lost of effectiveness has happened or starting from the Architectural View, to give a maintenance actions schedule on units, equipment or sub-systems.

7 References

[1] Solazzi, S., Ciambra, F., Sinisi, M. "*Integrated Diagnostic Processor for Naval Combat System Effectiveness Evaluation*", Paper presented at the Maritime Systems and Technology (MAST) Conference, Rome (Italy), 2010.

[2] ———. NIILS – Normativa Interforze per il Supporto Logistico Integrato (ILS – Integrated Logistic Support), SGD – G – 018.

[3] Ciambra, F and Nardini, M. "*Naval Combat System Design: System Engineering approach and complexity management*", Paper presented at the Fourteenth Annual International Symposium of INCOSE, 2004.

[4] Ciambra, F. Nardini, M. Protano, A. Tocci, A. "*Integrated Systems Functional Modelling in Naval Combat Systems Design, Integration & Validation*", Paper presented at the Maritime Systems and Technology (MAST) Conference, Rome (Italy), 2010.

[5] OMG Unified Modeling Language (OMG UML™) Version 2.0.

[6] OMG Systems Modeling Language (OMG SysML™) Version 1.1.

[7] ———. 2010. *Systems Engineering Handbook: A Guide for System Life Cycle Processes and Activities*. Version 3.2. Revised by M. Krueger, D. Walden, and R. D. Hamelin. San Diego, CA (US): INCOSE.

Self-aware architecture to support partial control of emergent behavior

Leo Motus
Research Lab for Proactive Technologies
Tallinn University of Technology,
Estonian Academy of Sciences
Tallinn, Estonia
leo.motus@akadeemia.ee

Merik Meriste
Research Lab for Proactive Technologies
Tallinn University of Technology
Tartu, Estonia
merik.meriste@ut.ee

Jürgo-Sören Preden
Research Lab for Proactive Technologies
Tallinn University of Technology
Tallinn, Estonia
jurgo@preden.ee

Raido Pahtma
Research Lab for Proactive Technologies
Tallinn University of Technology
Tallinn, Estonia
raido.pahtma@dcc.ttu.ee

Abstract – *System of systems comprises interacting, heterogeneous, autonomous components with incomplete information about their inner states, and about the surrounding environment. Many interactions are often not rigorously defined, and change dynamically. System of systems usually exhibits emergent behavior that cannot be predicted by analyzing static properties of the components, and is not always permissible. This paper suggests that the designer can improve system's behavior by substituting (part of) regular interactions with smart mediated interactions that bolster up shared situation awareness of the system's components and thus strengthens system's capability to monitor and partially control its emergent behavior. This paper discusses smart mediated interactions that focus on awareness of temporal features and on estimates of spatial location of the components. Interactions are assembled into proactive middleware that forms a backbone of system of systems.*

Keywords: System of systems, mediated interactions, self-aware architecture, partial control of emergent behavior

1 Introduction

System of systems usually results from networking of stand-alone systems in order to coordinate their behavior, and to benefit from concerted operation of stand-alone systems. Formally, system of systems can be described as a heterogeneous network of collaborating agents that strive to improve their joint performance by harmonizing their individual behaviors. It is characteristic to the system of systems that part of the autonomy of collaborating agents is maintained and agents have been left with some freedom in choosing their interaction partners. It is hoped that such loose architecture will facilitate synergy that further increases the efficiency of networked stand-alone systems.

Software intensive artifacts, biological creatures and their communities, pervasive computing systems, and wireless sensor networks illustrate the variety of potential agents and multi-agent systems. Interactions between the agents, and between agents and parts of environment, are not definitively fixed during the design stage – the interaction pattern can be dynamically changed by interacting partners and by the system manager during system operation.

Ideally, the strength of a system of systems stems from the fact that each component receives persistent feedback information from the other components and from the environment. They strive to exchange this information in order to jointly reason about the ways to improve the system and influence its environment to achieve the overall goals of the system. More advanced systems exhibit cognitive and proactive behavior, although they cannot fully control nor avoid emergent behavior.

Proactive behavior means anticipatory, usually self-initiated behavior striving to achieve (pre-set or dynamically selected) goals of the system. Cognitive and proactive behavior cannot always be controlled by applying conventional methods of computation, in many cases one needs to rely on non-classical paradigms of computation to analyze and verify the system of systems behavior [1, 2, 3].

The instant of appearance of emergent behavior and its impact on system's overall performance cannot be predicted. In the cases when emergent behavior can be detected in due time, one might be able to partially control it, or mitigate its impact. The closest analog to emergent behavior in conventional artificial systems (i.e. no autonomy of components, strictly limited self-organization, completely controllable), is exception handling subsystem.

This paper suggests extending the role of system's architecture by enabling its active assistance in monitoring and conducting the system's evolution (e.g. changes in interaction pattern between the components, modifying the composition of components and/or their functionality). This extended role provides system components with the situation awareness and with the ability to share situation awareness across the system. The shared situation awareness serves as the informational basis for deciding how to reorganize system's architecture and/or functionality of its components.

The architecture that actively participates in evolution and adaptation of the system is called self-aware architecture. Self-aware architecture supports detection of emergent behavior and estimating its impact on the required system's behavior. It also enables on-line behavior verification and reasoning about the effects of dynamic architectural modifications on the overall performance of the system.

2 Interactions

Historically, conventional computer science has considered direct, ordered interactions – e.g. to describe superposition of algorithms, and composition of software modules. A new and decisive role of interactions in describing computing systems was pointed out by R. Milner in [4], and elaborated in [5] and [2]. Indirect interactions, although used widely for describing operation of natural systems (especially those related to modeling swarm intelligence), were introduced to computer science much later as a necessary extension to describe distributed concurrent computations [6, 7]. Indirect interactions have obtained more attention in [8, 9, 10].

Disordered and dynamic interactions [9] have been introduced quite recently to capture specific communication problems in mobile ad hoc networks. Those interactions are ideologically close to mediated interactions and are essential for on-line engineering of complex systems that comprise autonomous components with evolving interaction pattern, exhibit emergent behavior, and possess self-organizing properties.

2.1 Mediated interactions

A vast majority of research into mediated interaction is related to studying human-machine, or human-machine-human communication. In this paper we focus on artificial systems, and primarily on machine-machine interactions – whereas machine is the initiator of interactions and mediator's role is also carried out by a machine (or respective software). Mediation here denotes, in the most cases, possibility to subscribe to messages on certain topic with well-defined (e.g. temporal) restrictions that are modifiable by the subscriber during system's operation and will automatically be used for building filters on the interaction.

When designing artificial systems the designer can pragmatically substitute some of (direct or indirect) interactions, operating in the original system, with suitable mediated interactions. Those mediated interactions do not modify substantially the original interaction, but additionally provide on-line access to authorized (in-system, environmental, or human operator) agents who decide when to impose new rules of engagement, or modify the existing rules that guide the interactions, or filter the contents of transmitted messages. The inserted mediated interactions enable simulation and foster capability for on-line engineering of system of systems – e.g. they enable to observe, analyze, and verify system's behavior, and partially to control system's emergent behavior. The system can be engineered on-line either by human operators, or preferably by the system itself – partly because the required response time to changing conditions is often too short for human reaction, and partly because of the human's physical inability to digest large amounts of rapidly changing situational information during long period.

Mediated interaction is a smart, potentially proactive one-to one interaction whose functionality and operation is dependent on situational information and on goals of the system. The mediated interaction enables dynamic filtering of transmitted messages, or modifying the mapping carried out by interaction. The mediation can be triggered by one of the interacting partners (usually by the consumer of messages), or by an authorized agent from the environment of interacting agents. An indirect mediated interaction (i.e. influencing via environment) has turned out to be inevitable to get a realistic simulation of systems with swarm intelligence [7,10].

Please note that mediated interactions as used and studied in human-machine-human are passive (they convey the message and do not modify its substance actively)). In the case of technological applications mediated interactions could play multiple roles – they convey the message, filter out excessive data, validate the contents, and attempt to detect emergent behavior. As a by-product, proactive mediated interactions help to reduce the required in-system communication bandwidth.

2.2 Modeling of mediated interactions

Mediated interactions between agents have been earlier modeled by specific coordination artifacts (e.g. blackboards) [11] that explicitly assign operating instructions to interacting partners, but are as capable as agents. In this paper mediated interactions are considered as regular agents – they possess active properties (e.g. autonomy, proactiveness, rationality). The model adopted here for describing system of systems comprises two types of equally important components – actor agents providing (local) functionality, and link agents executing and validating cooperation of actor agents.

The behavior of system of systems, and its model cannot be adequately analyzed relying on theory and tools stemming from Turing machine paradigm. In this paper the description and analysis is based on a prototype of a multi-stream interaction-centered computation model [1, 12].

A system of systems is described as a pair (P, Σ), where $P = \{p_1, p_2, \ldots, p_i, \ldots, p_n\}$ is a set of actor agents, and $\Sigma \subseteq PxP$ is a set of link agents, $\Sigma = \{\sigma_{ij}\}$, where i and j = 1,2, …, n. In this paper we discuss the case of time-aware systems, and time- and location-aware systems. The list of explicitly considered situation variables can be extended, if necessary, e.g. by considering stress level of human operator, or health condition of the system, by suitably elaborating the set of metadata -- T(p), and L(p,t) – that will be used to tag variable values in val p.

An actor agent in a time-aware system is a mapping p

$$\mathbf{p: T(p) \times dom\ p \rightarrow val\ p,} \qquad (1)$$

where T(p) is a well-ordered set of time instants when the mapping execution starts. An actor agent in time- and location-aware systems is a mapping p

$$\mathbf{p: L(p, t) \times dom\ p \rightarrow val\ p,} \qquad (2)$$

where L(p, t) is a set of pairs (location coordinates, time instant), defining the position of agent at the instant when execution of the mapping starts

$$\mathbf{L(p, t) = \{(l(p, t), t);\ t \in T(p)\},} \qquad (3)$$

where l(p, t) computes location of agents at instant t.

A link agent in time-aware system is a mapping $\boldsymbol{\sigma_{ij}}$ that conveys a specified (by the mediated interaction) part of the producer agent $\mathbf{p_i}$ value range to the consumer agent's $\mathbf{p_j}$ domain of definition

$$\sigma_{ij} : \mathbf{T(p_i) \times T(p_j) \times val\ p_i \rightarrow proj_{val\ pi}\ dom\ p_j,} \quad (4)$$

with channel function filtering out the data that is not required by the consumer agent

$$\mathbf{K(\sigma_{ij}, t) \subset T(p_i), and\ t \in T(p_j).} \qquad (5)$$

Please note that each interacting agent can have its own time counting system (for metric time) and can also use (independently of the others) time concepts that fit its purposes the best, e.g. fully reversible time, strictly increasing time, relative time with moving origin, etc.

A link agent in time- and location-aware system is a mapping **σij**

$$\mathbf{\sigma_{ij} :\ L(p_i,t_i) \times\ L(p_j,t_j) \times val\ p_i \rightarrow proj_{val\ pi}\ dom\ p_j,} \qquad (6)$$

with channel function

$$\mathbf{K(\sigma_{ij}, s(t_j)) \subset L(p_i,t_i), s(t_j) \in L(p_j,t_j),}\ \text{and}$$

$$\mathbf{L(p_i,t_i) = \{(l(p_i,t_i),t_i)\ ;\ t_i \in T(p_i)\}.}$$

Please note the possibility that each actor agent can work in its own time-counting system loosening thus clock synchronization task in large scale systems. The mediated interactions as described in (1) – (6) enable to achieve intrinsic situation-awareness of the system and to reason on-line about many aspects of the system's behavior and its architecture without external advice (e.g. from human operators), see some pilot examples from [1,13].

The innovative formalism for time- and location-aware actor agents (formulae (1) and (2)) and link agents (formulae (4) and (6)) enables to build theories for formal detection of inconsistencies and/or contradictions related to values of temporal and spatial constraints, and to expected operation of link agents. A theory, focused on time-aware systems has been discussed in [12]. As a side effect, one needs a novel understanding of the nature and role of time in modeling and analyzing software intensive system of systems – some of the details can be found in [14].

2.3 Proactive middleware

A middleware that supports creation of a team situation-awareness in a multi-agent system has been built based on the concept of mediated interaction, [13]. Two aspects make information exchange between agents in system of systems cumbersome – there is excessive amount of situational information available, and system's configuration is changing unexpectedly.

The proactive middleware concept addresses both of those aspects by encouraging local processing of situational information (as much as feasible), and by tagging messages with time and location tags that facilitates immediate on-line verification of the message contents before sending the message to consumer agent (time and location selective communication), and enables to avoid using handshake based message exchange between the producer and consumer. The impact of configuration change is relieved by the fact that consumer subscribes to particular variable values, not to messages from a particular message producing agent.

System of systems is often composed of already existing stand-alone systems, it is reasonable to assume that distributed stand-alone systems have their own middleware that are to be integrated into the larger system. In the following is discussed a middleware prototype implementation for ad hoc sensor network.

Every data provider and data consumer is viewed as an agent. Every agent is associated to middleware interface; both the agent and the interface to middleware are running (usually) on the same computing platform (see Figure 1). In the prototype the platform is an 8 bit embedded computer (a smart dust mote). Every smart dust mote is equipped with a wireless communication capability that enables agent-to-agent communication via (mobile) *ad hoc* network.

Data communication follows a producer-consumer model, where the data consumers request data from the data producers, the requests being in the form of data subscriptions sent from the data requester to the data provider. The subscription is created by the mediator component on the data consumer and processed by the mediator component on the data producer. Each subscription contains the data type requested and also the constraints on the required validity of data items requested by the consumer. The constraints can be in various dimensions, at the moment temporal and spatial constraints are used. This means that the data requester can specify the temporal and spatial validity constraints on the data, i.e. where and when the data required by the consumer must be valid.

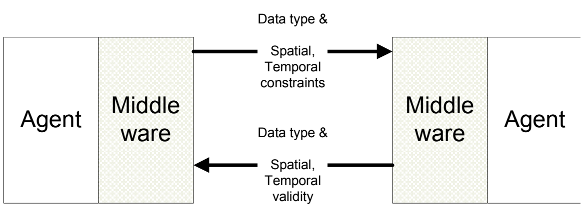


Figure 1 Data exchange between consumer and producer

The middleware interface associated with the agent is responsible for handling all the communication to and from the agent. This communication has two facets: data requested by the agent and data provided by the agent. In the incoming data flow (data requested by the agent) the mediator handles requests for data coming from the agent associated with the mediator, propagating those requests to potential data providers, preliminary validation of the incoming data and delivery of data to the agent. In the other direction (data provided by the agent) the data produced by the agent is made available to the other agents via the mediators: incoming data requests from the other agents are processed by the local mediator. If the request can be satisfied by the current agent, the request is accepted and data is delivered to the requester, after having checked that the requirements set forth by the data requester are satisfied.

In order to exchange the required data between agents a new data encoding format, suitable for sensor networks, has been developed. The objective of development of new data encoding format was to become able to seamless transition of data from the agents with low processing power to computationally more powerful agents without a need for additional data interpretation in the intermediate steps. In the case of computationally powerful agents (e.g. PCs) the natural communication protocol and encoding would be XML. Departing from this observation an encoding that enables direct transition of data from the agents of low processing power to XML has been developed that eliminates the need to converting the data itself. Part of a sample subscription to data in XML format is presented in Figure 2.

Due to hardware and resource constraints majority of motes in *ad hoc* sensor networks are able to support only a limited and fixed set of XML tags. Hence, the XML messages need to follow a strict structure where each tag can only have one parameter value. The value of the parameter must be a 32 bit integer with sign. Larger values can be transferred by appending additional tags and appropriately structuring the messages. Content outside the value parameter is not supported. The number of tag-value pairs in a message is formally limited to 255, but practically WSN node radio messages will fit much less. Each XML tag needs to have a unique integer identifier associated with it in the converter database.

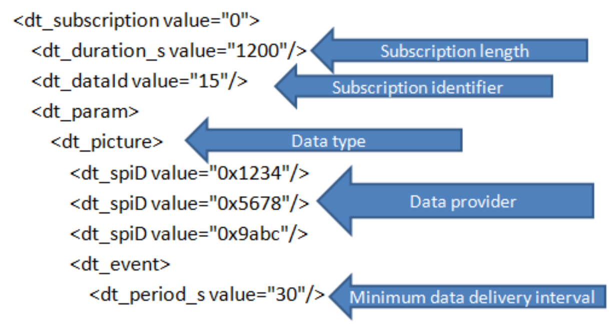


Figure 2 Sample subscription in XML format

The smart dust motes on which the proactive middleware prototype is implemented are Defendec's Dnode motes based on Atmel ATMega128RFA1 with 128 kBytes of ROM and 16 kBytes of RAM. The middleware implementation on embedded devices is quite resource demanding -- an implementation that is able to support 5 incoming subscriptions and 3 outgoing subscriptions requires approximately 26 kBytes of ROM and 3.2 kBytes of RAM. For comparison, the entire software image of an operational device running the TinyOS operating system and communicating via the radio requires 9 kBytes of ROM and less than 500 bytes of RAM.

3 Work-in-progress

Situation-awareness, as a first step in building self-X systems has recently become a buzzword. However, many research efforts focus on rather simple stand-alone systems or components of larger systems. This could be partly explained by the fact that for those applications the approximate solutions provided by tools and methods based on conventional Turing machine paradigm provide sufficiently good results.

Pervasive computing systems tend to be dependent on (at least) time and location constraints and persistent feedback from the environment is therefore essential for those systems. Hence they interact immediately with physical processes in the natural and/or man-made environment. The networked pervasive computing systems are even more sophisticated and demanding, and should be able to handle autonomous components, complex interactions, sensitivity to temporal and spatial constraints, indefinitely on-going operation, practical impossibility of clock synchronization, truly (a.k.a. forced) parallel processing of interacting data streams, required self-X properties of systems, etc. Majority of those new characteristics cannot be properly handled, within the framework of conventional Turing machine based theories, with sufficient accuracy. At the same time non-classical models of computation are still developing and there is no commercially available and widely accepted set of methods and tools.

Our lab proceeds in studying theoretical issues stemming from elaboration of multi-stream interaction machine paradigm [1], and simultaneously strives to acquire feedback from reality – by building, and analyzing the behavior of networked pervasive computing systems, and methods for advancing shared situation-awareness as the basis for self-organizing systems. Our experimental work is illustrated by the following project titles:

- Self-organizing Intelligent Middleware Platform for manufacturing and Logistics Enterprises (ARTEMIS JU project)
- Asymmetric Threat Environment Analysis (EDA project)
- Information Interoperability and Intelligence Interoperability by Statistics, Agents, Reasoning, and Semantics (EDA project)
- Self-organizing systems with on-line monitoring and diagnostics (IMECC project)

4 Conclusions

This paper has made an explicit attempt to apply non-classical computing paradigm for upgrading the situation awareness in a system of systems, and build up a self-aware architecture that can actively support on-line evolution of system of systems. The suggested approach caters for temporal and spatial constraints based features. Further extension of constraints is possible.

The approach has been practically tested with temporal constraints in the industrial environment and is being tested with temporal and spatial constraints. The latter experiments are performed in industrial environment and in a mixed human-machine system of systems – e.g. crisis mitigation and management, and security related applications.

All the listed applications are essentially complex systems that generate emergent behavior. Prediction of the emergent behavior is not theoretically possible. Hence one can only partially control the impact of emergent behavior on the system (in order to keep a system within a security envelope). The sooner emergent behavior is detected the better it can be responded.

The concept of mediated interactions is a promising tool for monitoring the system's status, detecting emergent behavior, partially controlling the detected emergent behavior, and supporting the control of system's evolution. Intrinsic complexity of such systems can be better handled if one relies on theories and tools based on non-classical model of computation.

References

[1] L. Motus, M. Meriste, W. Dosch, "Time-awareness and Proactivity in Models of Interactive Computation", *Electronic Notes in Theoretical Computer Science*, vol. **141**, 69-95, 2005, www.elsevier.com/locate/entcs

[2] S. Stepney S., S. Braunstein, J.A. Clark, A. Tyrrell, A. Adamatzky, R.E. Smith, T. Addis, C. Johnson, J. Timmis, P. Welch, R. Milner, D. Partridge, "Journeys in Non-Classical Computation. A Grand Challenge for Computing Research", *Int. Journal of Parallel, Emergent and Distributed Systems*, vol.20, no.1, 5-19, 2005

[3] A. Sloman, "The irrelevance of Turing Machines to AI", in M. Scheutz (Ed.) *Computationalism: New Directions,* MIT Press, 87-127, 2002

[4] R.A. Milner, "A Calculus of Communicating Systems", *LNCS*, vol. **92**, Springer, 171 pp., 1980

[5] P. Wegner, "Interactive foundations of computing", *Theoretical Computer Science*, vol. **192**, 315-351, 1998

[6] R.A. Milner, "Communicating and Mobile Systems: the π-calculus", Cambridge University Press, 161 pp., 1999

[7] T. Hoverd and S. Stepney, "Environment orientation: an architecture for simulating complex systems", CoSMoS workshop, 67-82, Luniver Press, York, U.K., 2009

[8] Jiming Liu, Xiaolong Jin, Kwok Ching Tsui, "Autonomy oriented computing: from problem solving to complex systems modeling", Springer (former Kluwer Academic Publishers), 216 pp., 2005

[9] Pena J., R. Levy, R. Corchuelo, "Towards clarifying the importance of interactions in Agent-Oriented Software Engineering", Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial, no.25, 19-28, 2005

[10] S. Stepney, "Embodiment", in Darren Flower, Jonathan Timmis, (Eds.), *In Silico Immunology*, chapter 12, Springer, pp. 265-288, 2007

[11] M. Viroli, A. Ricci, "Instruction-based Semantics of Agent Mediated Interaction", Proc. Conf. on Autonomous Agents and Multi Agent Systems, vol.1, doi10.1109/AAMAS.2004.148, 102-109 pp., 2004

[12] L. Motus and M.G. Rodd, "Timing analysis of real-time software", Elsevier/Pergamon, 212pp., 1994

[13] J. Preden, L. Motus, M. Meriste, A. Riid, "Situation awareness for networked systems", IEEE Intern. Multi-disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support, CogSIMA 2011.

[14] L. Motus, „*Modeling metric time"*. In B. Selic, L. Lavagno, G. Martin (Eds), *UML for Real: Design of Embedded Real-time Systems,* Kluwer Academic Publ., Norwell (2003), 205–220.

Application of Component Engineering to the Design of Holistic Spell Checking Algorithm

Leena J. Alhussaini
The University of Edinburgh
EH8 9AB Edinburgh, United Kingdom
s0460867@sms.ed.ac.uk

Abstract – *In the field of human languages, a correct spelling of words is pivot in communication activity. Spell checking of many words at once is the invention of this work. It is a requirement in activities like: book spell checking before publication, plagiarism detection in a thesis against its references. We design a holistic spell checking algorithm system using Koala component model. This system contains three basic sub-systems: input system which is a decision making system to manage user words, and correctly spelled words as rule-out data for user data; behavior system which is a dynamic system of application of Markov model on trees; and output system which is a dynamic system to manage output of ranked suggestion list for user misspelled data. The design model presents an explicit architecture meeting Koala component model requirements. The holistic spell checking system presented a novel problem with a novel application design.*

Keywords: spell checking, holistic, software design, component engineering, algorithm.

1 Introduction

To study human language [1], there are three main parts to consider. One is language form. Another is language meaning. A third one is language in context. Considering language form, we study the shape of a word in terms of its building blocks of: affixes, suffixes, roots, etc. A sentence in a language is a composition of words according to certain rules used to communicate meanings. The correction of a word spelling is pivot to deliver the accurate purpose of communication. In this work, we shed light on the pivot of any language, and it is correct spelling of words.

Current spell checker algorithms consider single word to spell check at a time. This work invention addresses a novel problem with a novel solution. It is the problem of many words spell checking at once (holistic spell checking). We give the name ZIPPER to the project as it simulates the act of the zippers clip: one tick, two ticks, or as many as there are. ZIPPER can spell check one word, two words, or as many as there are words at once. The design of ZIPPER software uses component-based software engineering. ZIPPER algorithm has three basic models: input model, which is decision tree model; behavior model, which is Markov model on trees; and output model, which is clique tree model. The limitation of this work resides in handling spelling errors in first character. We will first illustrate the project design, and then discuss the component engineering of the algorithm.

2 Application

Spell checking of many words at once is the invention of this work. It is a requirement in activities like: book spell checking before publication, and plagiarism detection in a thesis against its references. Also, in other cases where many words are managed at once in organizations like: Information Center of a country containing the nations full names, and roads names. Furthermore, accumulating data from different nodes in a grid network before spell checking them at once.

3 Related Work

As the problem of this work is unapproached, there is not a specific spell checker that works holistically. Most spell checkers [2] spell check one word at a time. Thinking of parallelizing unit spell checker may be time inefficient.

4 Holistic Spell Checking Algorithm System Engineering

We concentrate on the design issue of the holistic spell checking system. We specifically address the requirements and design phases of software development life-cycle. The software architectural style is Object-Oriented. We use Unified Process [3], as our process model, as it is designated to component-based software design. Within the Unified Process model, there are four phases: (1) Inception, (2) Elaboration, (3) Construction, and (4) Transition. Each phase is accomplished by a number of component-based life-cycle steps. The component-based software life-cycle contains eight steps: (1) Requirements Analysis and Definition, (2) Component Selection and Evaluation, (3) System Design, (4) System Implementation, (5) System Integration, (6) Validation and Verification, (7) System Operation Support and Maintenance; and (8) System Development Process. Even though, each phase of the process model iterates over the eight steps, each concentrates on specific steps. For the Inception phase: (1) Requirements Analysis and Definition: we define system boundaries, system architecture, and component definition; (2) Component Selection and Evaluation: we specify the internal and external components after evaluating them for their applicability. For the Elaboration phase: (3) System Design: we use Koala component model [4] to design the system as we plan to embed the software in a consumer electronic product. For the Construction phase: (4) System Implementation: we use JavaBeans [8] as the implementing language as a bean means component; (5) System Integration: we use Enterprise JavaBeans (EJB) [9] to integrate system components; (6) Validation and Verification: with validation we try to meet component specification and with verification we try to meet customer expectations. For the Transition phase: (7) System Operation Support and Maintenance: we maintain and support component and system as two different entities; and (8) System Development Process: considers the development of the overall system including all components.

In the rest of the paper, we elaborate on the specification of the system design, and its corresponding components engineering.

4.1 Background: The Koala Component Model

Koala [4] [5] [7] is an architectural definition language (ADL) used to detail the design of the algorithm. This model consists of three major requirements. The first requirement ensures that the design of the components eases their bindings. The second requirement ensures the design is targeted to a resource-constrained environment, like electronic devices. The third requirement enforces the clarity and explicit of an architecture design to handle future complexity at ease. From these requirements, we get the notion of binding, interfaces, and glue code. According to Figure 1, we can see three components: C1, C2, and C3.

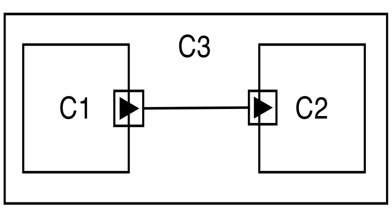


Figure 1: Koala Components

The small squares with triangle are known as interfaces. The interface at C2 is known as *provides interface* as it accepts invocation of methods from outside the component. The interface at C1 is known as *requires interface* as it binds with provides interfaces to pass data and invoke methods. The link between component C1 and component C2 is known as *binding*. Component C3 is known as *compound* (*composite*) component for its containment of other sub-components. The description of components is influenced by C language syntax. For example, as in Figure 2, to describe component C2, we write:

```
interface I {
Object  i (Object x); }

Component C2 {
provides I i;
contains module C2_impl present;
connects i = C2_impl; }
```

Figure 2: Koala ADL

The description of component C2 illustrates the definition of interface I, which contains one method. The type of interface is *provides*. Component C2 self-contains implementation known as *module C2_impl present*. The module implementation connects with interface *i*.

4.2 Zipper System Design

The algorithm consists of the three basic models: input, behavior, and output; see Table 1.

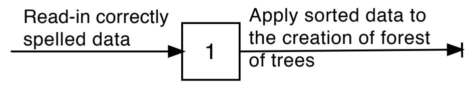
For the input model, we use a decision making tree system. Inputs for this model are user data and correclty spelled data. Output from this model is a forest of rooted trees that contains user misspelled data and correctly spelled data. Correclty spelled data can be a dictionary or any authorized list of data that are used to rule-out user data. At step 1, we read-in correctly spelled data, then sort them and create a forest of rooted trees. At step 2, we read-in user data. At step 3, we sort them and apply them on the forest of rooted trees with the correctly spelled data. At step 4, we decide on misspelled data to remain in the forest and exclude user correct data.

For the behavior model, we read-in the forest of rooted trees with the inclusion of user misspelled words. The ouput of the behavior system is a populated forest of rooted trees after application and processing of Markov model [10]. In this model, we use Markov model applied on trees. Markov model is a probabilistic model. It exploits the dependencies amongst nodes in a graph to compute marginal probability. It supports tree representation. The system at the behavior model is cooperative and competitive distributed, complex, and dynamic system. It is cooperative and competitive distributed system as each rooted tree is processed by a single thread. Threads of each single rooted tree are cooperative and competitve to find processing results. It is complex in the way a single tree is structured. The width and height of a tree grow with no limit to space. Each tree has the following properties: (a) unbalanced: words grow the way they grow; (b) rooted: there are a number of trees forming a forest, where each root is a single letter in the alphabet; (c) ordered: roots of the trees in the forest are ordered using the 26 English alphabetical letters for ease of processing. It is dynamic by Markov model property on trees which states that every link between a pair of characters is independent;

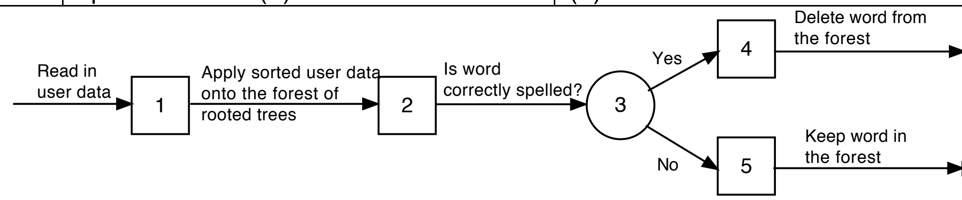
Table 1: ZIPPER System Design

INPUT MODEL: DECISION TREE

Step 1: Read-in correctly spelled list of data

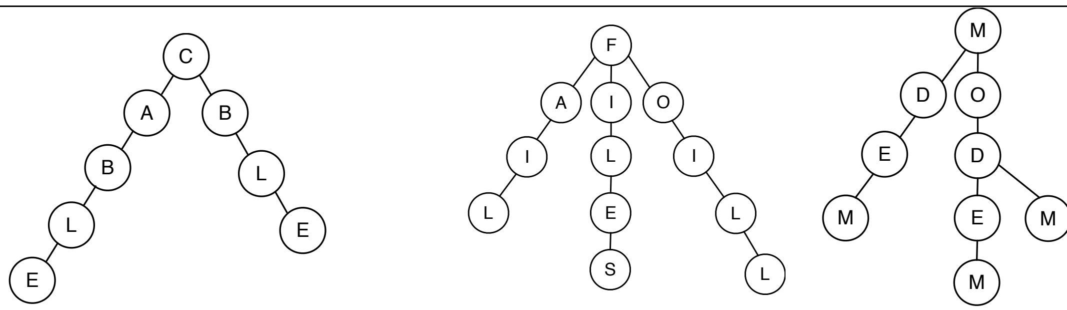


Step 2: Read-in User Data at (1)	**Step 3:** Sort user data and apply it onto the forest with the correctly spelled data at (2)	**Step 4:** Delete correctly spelled data from forest and keep misspelled data at (3), (4), and (5)



BEHAVIOR MODEL: MARKOV MODEL ON TREES

Step 5: In parallel over forest of rooted trees: (a) apply Markov model, (b) quantify links, (c) quantify nodes



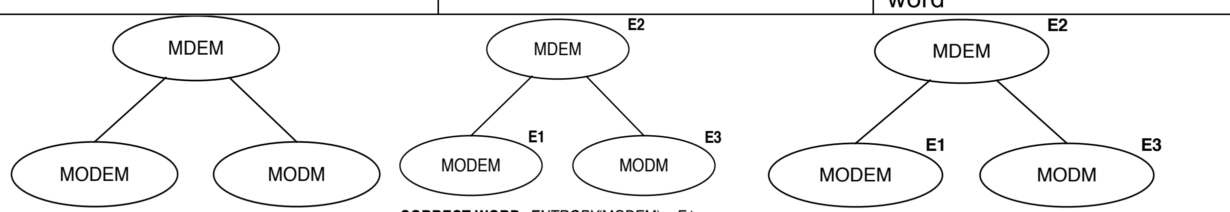
CORRECT WORD: CABLE
USER MISSPELLED WORDS: CABL, CBLE

CORRECT WORD: FILES
USER MISSPELLED WORDS: FAIL, FOILL

CORRECT WORD: MODEM
USER MISSPELLED WORDS: MODM, MDEM

OUTPUT MODEL: CLIQUE TREE

Step 6: Construct clique tree	**Step 7:** Compute entropy per word	**Step 8:** Compute mutual information value between correctly spelled word and user word

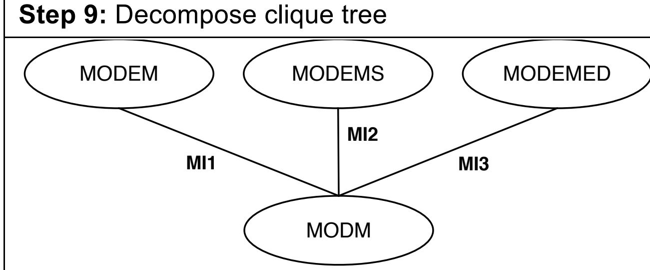


CORRECT WORD: MODEM
USER MISSPELLED WORD: MDEM, MODM

CORRECT WORD: *ENTROPY*(MODEM) = E1
USER MISSPELLED WORD: *ENTROPY*(MDEM) = E2, *ENTROPY*(MODM) = E3

MUTUAL INFORMATION(E1(MODEM), E2(MDEM)) = MI1
MUTUAL INFORMATION(E1(MODEM), E3(MODM)) = MI2

Step 9: Decompose clique tree	**Step 10:** Print ranked suggestion list



RANKED SUGGESTION LIST FOR (<u>MODM</u>):
MODEM (MI1)
MODEMS (MI2)
MODEMED (MI3)

thats a future state is only dependent on current state. At step 5, we parallelize the processing of rooted trees in the forest using Java threads. At step 5(a), we apply a propability distribution over the trees along with Markov property. After that, at step 5(b), we quantify the links in the trees using an information theory metric which is point-wise mutual information (PMI) [1] as given in equation (1).

$$PMI = log\frac{P(x,y)}{P(X).P(Y)} \quad (1)$$

The PMI is a measure of uncertainty of points in a distribution. It computes the amount of information of event y given event x. At step 5(c), we quantify the nodes using Belief Propagation using Message Passing Paradigm (BP) [10]. This paradigm is a tool accompanied with Markov Model to compute nodes marginal probability (a.k.a. belief) in a graphical representation using messages; (see equations 2 and 3).

$$b_i(x_i) = k\Phi_i(x_i)\Pi_{j\in N(i)}m_{ji}(x_i) \quad (2)$$

$$m_{ji}(x_i) = \sum_{x_j}\Phi_j(x_j)\Psi_{ji}(x_j,x_i)\Pi_{k\in N(i)\backslash j}m_{ki}(x_i) \quad (3)$$

For the output model, we read-in populated trees after application of Markov model and processing of edges and nodes. We output a ranked suggestion list for the user to correct the misspelled words. For the processing of the output model, we use clique trees [10] which decompose the words previously entered into the tree of the behavior system into their original words for evaluation. Each node in the clique tree represents the original word firstly entered in the forest of rooted trees in the input system. A node is also a set of collective characters in the behavior system that represents a single word. Correctly spelled words are distinguished from user words. At step 6, the clique tree is constructed. The tree contains three nodes: Mdem, Modem, and Modm. Modem is a correctly spelled word. Mdem and Modm are user misspelled words. At step 7, we compute entropy [1] for each node as in equation (4).

$$H(X) = -\sum_{x\in X} f(x)logf(x) \quad (4)$$

Entropy is a measure of uncertainty firstly used in signal processing to quantify the loss of the amount of information to the other site. For example, typing the characters of a word in a dictionary also has an amount of loss of certainty either due to typing errors, or due to lack of knowledge of correct spelling. Node *modem* is given entropy E1, node *mdem* is given entropy E2, and node *modm* is given entropy E3. At step 8, we compute mutual information [1], as in equation 5, in a cartesian-product matrix as in Table 2.

$$I(X;Y) = \sum_{x\in X}\sum_{y\in Y} f(x,y)log\frac{f(x,y)}{f(x).f(y)} \quad (5)$$

Mutual information (MI) is the reduction in uncertainty of variable X due to the knowledge of variable Y. Each user

Table 2: Cartesian-product Matrix

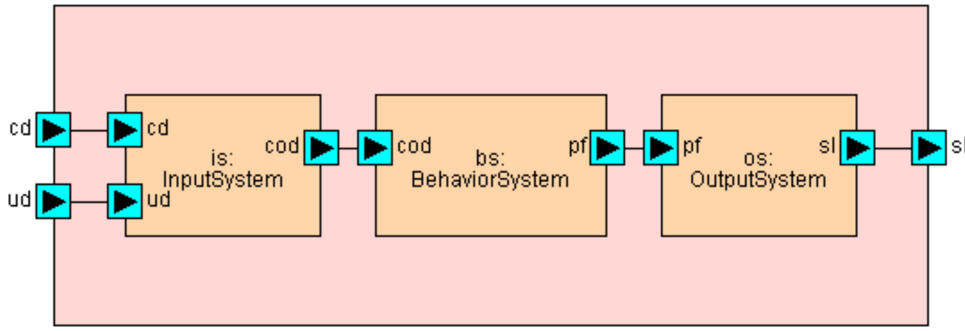
Correct Word / Misspelled Word	(1) Modem	(2) Modems	(3) Modemed
(a) Mdem	MI (a1)	MI (a2)	MI (a3)
(b) Modm	MI (b1)	MI (b2)	MI (b3)

misspelled word is joint with the set of correctly spelled words for computation of mutual information value to see later which correctly spelled word is likely to be the replacement of the user misspelled word. As in Table 2, user misspelled word (a) *mdem* is joint with correctly spelled words : *modem*, *modems*, and *modemed*. Mutual information value is computed for each joint. As for user word *mdem*, we have *MI(a1)* with *modem*, *MI(a2)* with *modems*, and *MI(a3)* with *modemed*. At step 9, we decompose the clique tree. Each decomposed tree contains a user misspelled word with a number of correctly spelled words where each relationship between a correct and misspelled word is given a value for ranking the suggestion list. Finally, at step 10, for each user misspelled word we rank the suggestion list of words that likely to replace the misspelled word and present them to user for choice of correct word. For example, for user misspelled word modm, we have : *modem* with rank *MI1*, *modems* with rank *MI2*, and *modemed* with rank *MI3*.

4.3 Zipper Component-based System Engineering

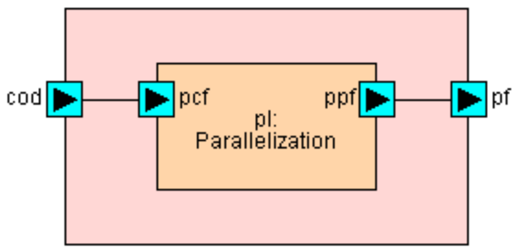
We present a case study of the application of component-based system engineering to the design of holistic spell checking algorithm. With reference to Figure 3, the graphs in the figure are generated using Koala Viewer [6] and the definitions are hand-written as devised by Koala Component Definition Language (CDL).

The overall system, in Figure 3a and definitions below it, has three major components. It also has two input interfaces and one output interface. The input interfaces are *CorrectData cd* and *UserData ud*. The ouput interface is *SuggestionList sl* for misspelled user data. The definitions of the overall system state two provides interfaces (*provides CorrectData cd*; *UserData ud*) and one requires interface (*requires SuggestionList sl*). It also defines three basic components as : *contains component InputSystem is*, *component BehaviorSystem bs*, and *component OutputSystem os*. Besides component definitions, we define the connections amongest components. As the main system inputs get connected to the InputSystem which, after processing input data, transfer control to the Behavior System: *cd = is.cd*; *ud = is.ud*; *is.cod = bs.cod*, where cod stands for condition of data. The behavior system then populates the forest by applying



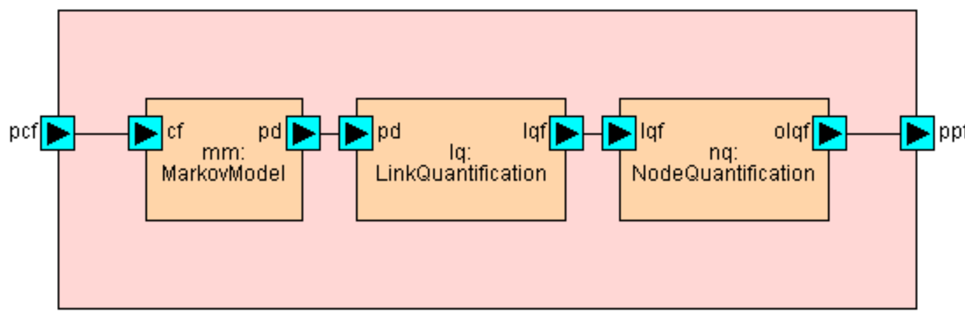
(a) ZIPPER overall system

```
component ZIPPER {
provides  CorrectData cd; UserData ud;
requires SuggestionList sl;
contains component InputSystem is;
         component BehaviorSystem bs;
         component OutputSystem os;
connects cd = is.cd; ud = is.ud;
         is.cod = bs.cod;
         bs.pf = os.pf;
         os.sl = sl; }
```



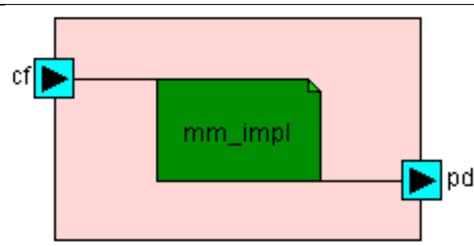
(b) Within Behavior System

```
component BehaviorSystem{
provides ForestInput cod;
requires PopulatedForest pf;
contains
   component Parallelization pl;
connects
   cod = pl.pcf;
   pl.ppf = pf;
 }
```



(c) Parallelization Compound Component

```
component Parallelization{
provides ForestInput pcf;
requires PopulatedForest ppf;
contains
   component MarkovModel mm;
   component LinkQuantification lq;
   component NodeQuantification nq;
connects
   pcf = mm.cf;
   mm.pd = lq.pd;
   lq.lpf = nq.lqf;
   nq.olqf = ppf;
   }
```



(d) MarkovModel Component with Module

```
interface ProbabilityDistributedForest
{ void cf (Object data[][]);
  Object[][] pd (void);}

component MarkovModel{
provides ProbabilityDistributedForest cf;
requires  ProbabilityDistributedForest pd;
contains module mm_impl present;
connects  cf = mm_impl;
          mm_impl = pd;}
```

Figure 3: Some Components with their Components and Interfaces Definitions

and further processing Markov model. It then outputs data through interface *PopulatedForest pf* and connection *bs.pf = os.pf*. It executes each tree in the forest in parallel with other trees and applies three functions to each tree. Firstly, it applies a probabilistic distribution using Markov model with Markov property using *Markov model* component. It does so through *mm : MarkovModel* component and connections *pl.pcf = mm.cf* and *mm.pd = lq.pd*. Secondly, it quantifies links in each tree using *link quantification* component. It does so through *lq : LinkQuantification* component and connections *lq.pd* and *lq.lqf*. Thirdly, it quantifies nodes using *node quantification* component. It does so through *nq : NodeQuantification* component and connections *nq.lqf* and *nq.olqf*. The output system processes the populated forest and produces the suggestion list *sl* which is connected to the overall system output interface *sl* as *os.sl = sl*.

As the major system components are compound components, well illustrate how it looks like by taking the behavior compound component as an example. Examining Figure 3b and the definitions on the right-side of the figure, we can see that it contains another component and it is the parallelization component. This component has one provides and one requires interface and connections between the behavior system and the parallelization component.

The parallelization component parallelizes three actions on the forest of data. Thus, it is also a compound component. Once clicked, in the Koala viewer, youll get figure 3c. The definitions of the figure is shown below the figure. Each component within the parallelization component contains a module of the actual implementation. For example, component *MarkovModel mm*, once clicked in the koala viewer gives us the figure in 3d. The definitions of MarkovModel component is given on the right-side of the figure.

Describing Markov model component (see Figure 3d), we have one interface with two functions in this component. The interface is *ProbabilityDistributedForest* with functions *cf* that reads-in the forest as a matrix, and function *pd* that outputs the forest after application of a probability distribution. Using these functions, we get in the description of MarkovModel component: *provides ProbabilityDistributedForest cf*, *requires ProbabilityDistributedForest pd*. The component, itself, contains its implementation (*contains module mm_impl present*) which is connected to its interfaces, as: *cf = mm_impl*, and *mm_impl = pd*.

5 Further Work

ZIPPER further work is to rebuild the system components into services that function in a distributed complex system like Grid network [11]. We plan to integrate large sets of data over the Grid network for spell checking holistically.

6 Conclusion

We presented in this work a novel problem with a novel solution. It is the problem of many words spell checking at once; holistic spell checking. We also illustrated the design of the algorithm of spell checking and the software engineering using component-based system. We used Koala as our component-based system model. The future of the system is to operate on Grid network as services.

7 Acknowledgements

To the knowing of knowledges.

References

[1] H. Scutze and C. D. Manning, *Foundations of Statistical Natural Language Processing*, MIT Press, 1999.

[2] [K. Kukich. *Techniques for automatically correcting words in text*, ACM Computing Surveys, Vol. 24, Issue 4, Dec 1992.

[3] I. Crnkovic and M. Larsson, *Building Reliable Component-based Software Systems*, Artech House, 2002.

[4] R. van Ommering and F. van der Linden, *Koala, a Component Model for Consumer Electronics Product Software*, Springer Berlin, Vol. 1429, 1998.

[5] R. van Ommering and F. van der Linden, *The Koala Component Model for Consumer Electronics Software*, IEEE Computer, March 2000.

[6] Koala Softwares. http://www.hitech-projects.com/SAE/koala.

[7] K.-K. Lau, Z. Wang, *Software Component Models*, IEEE Transactions on Software Engineering, Vol. 33, No. 10, 2007.

[8] C. Horstmann, *Big Java*, John Wiley & Sons Inc, 2002.

[9] R. Monson-Haefel and B. Burke. *Enterprise JavaBeans 3.0*, Fifth edition, OReilly Media, May 2006.

[10] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, first edition, 1988.

[11] I. Foster and C. Kesselman, *The Grid 2: Blueprint for a New Computing Infrastructure*, Morgan Kaufmann, second edition, 2003.

Proc. of the 2012 7th International Conference on System of Systems Engineering, Genoa, Italy - 16-19 July 2012

The Method of Analyzing Mapping between Capability and Performance Index Based on DSM/DMM Models

Ke-wei Yang
Department of Management
National University of Defense Technology
Changsha, Hunan Province, China
kayyang27@nudt.edu.cn

Yan-jing Lu
Department of Management
National University of Defense Technology
Changsha, Hunan Province, China
yanjinglu_nudt@yahoo.com

Jie Mao
Department of Management
National University of Defense Technology
Changsha, Hunan Province, China
maojie8818@hotmail.com

Zhi-wei Yang
Department of Management
National University of Defense Technology
Changsha, Hunan Province, China
zhwyang88@hotmail.com

Long Li
Department of Management
National University of Defense Technology
Changsha, Hunan Province, China
ericlonglee@gmail.com

Qing-song Zhao
Department of Management
National University of Defense Technology
Changsha, Hunan Province, China
zqszqr@163.com

Abstract - *Performance Index is the foundational and essential indices of measurement of SoS(System of Systems) capability. We propose ten sorts of mapping relationships description between kernel elements of SoS architecture based on the DSM/DMM model. The formalization of model transformation, combination and transposing operators are described as the foundation of proposition of formulation to measure the dependability of capabilities and performances of SoS. The mapping N-N Matrix for capabilities and performances provides a necessary foundation to evaluate capability of certain SoS.*

Keywords: SoS Capability; Performance Index, DSM/DMM; Mapping

1 Introduction

Capability of System-of-Systems (SoS) is the behavior shown by SoS in macro level, they measure the ability of component systems collaborating and interoperating in a dynamic environment to accomplish the certain tasks. Performance indices are measurements of system's characteristics or functions. Figure1 describes the basic structure of performance indices of SoS capability. In Figure.1, performance indices of the equipments measure the physical characteristics of SoS capability. In a single system, the relationship between performance indices and equipments systems is N-1, namely, an equipment system contains many performance indices and a performance index only belongs to an equipment system. However, in a SoS, the relationship between performance indices of SoS capability and equipment systems is N-N. That is to say, an equipment system is correlative with performance indices of SoS capability, while a performance index of SoS capability may be correlative with many equipment systems. Therefore, the key factor to build a framework for assessing SoS capability is to construct the mapping relationship between SoS capability and performance indices.

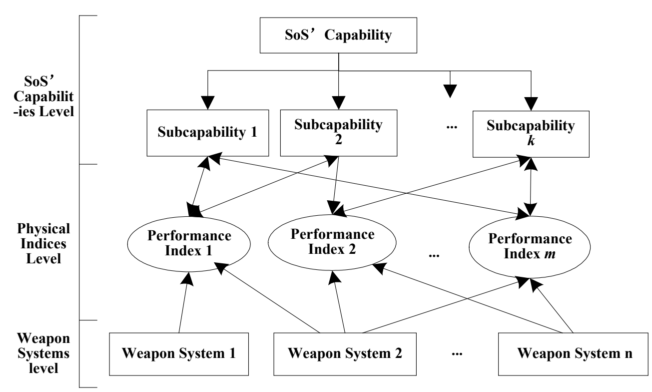


Figure. 1 Evaluating SoS Capability accumulating Multiple level elements

There are four vital factors in the multi-view models of SoS architecture: activity, capability, system and technology [1]. The stakeholders could describe their requirements which focus on certain factors from different perspectives. Currently, several leading methodologies, including DoDAF, provide the mechanism to model relations between the four factors. But these relationship data are separated in different view models, which can't be gained from view models directly, neither can be easily analyzed. In order to analyze the mapping between SoS capability and performance indices, performance indices (PI) are considered as one of the core elements of SoS architecture. Then, based on analyzing and visually expressing of mapping between these five core elements, by giving definitions to transpose, transfer and combination operations of DMM mapping models, we propose a model to compute the mapping between SoS capability and performance indices. And this computing model provides a foundation to evaluate and analyze the capability of certain SoS.

2 Diagram of Relationship between Modelling Kernel Elements in SoS Architecture

The Architecture Framework (AF), which includes elements and interactive relationships, provide a formalized modeling method and standard framework for SoS application development. And the AF-described models are blueprint of describing the basic framework [7]. Based on the method of multi-view in AF modeling, capability, system, activity and technology are the core elements in relationship analysis considered for SoS model. Capability describes the stakeholder's understanding for the expected state of SoS developed under AF, which is the general requirements for SoS model to be able to complete a series of specific mission and represents the demand of the development and construction of AF; Activity represents the external environment of SoS running and concludes the description of operational demands and the relationship in operational information exchange which are required for completing operational tasks; System is an important component of SoS, which represents the final physical state of AF modeling; Technology describes the technology needs in realizing the AF and reflects the challenges faced in current and future expected period in the process of SoS development.

The capability demands which come from the operational analysis provide constraints for the architecture design and optimization. The completion of capability calls for the completion of a corresponding System. The realization of operational mission is supported by the corresponding capability. System serves activity and the completion of activity needs to be supported by the corresponding system. Technology roots in the key technologies in the research of system. And the forecast level of technical standards is constrained by the development itself which need also satisfy the demands for future combat capability.

Measure of the Performance (MoP) is a quantitative description of certain system, focusing on single or multiple properties specification. MoP is the foundation and precondition of evaluating certain system operational effectiveness indicators. But MoP does not consider the influence of environment where certain system runs. The MoP of a single System reflects the capability for System to complete the desired task in accordance with its internal intrinsic properties. U.S. Department of Defense (DoD) sets up a series of common parameters for system general description, including quantity, quality, scope, timeliness and availability [1]. In order to analyze the mapping between system capacity and MoP, MoP was regarded directly related to System as one of the core elements for SoS architecture modeling.

The mapping relationship and operation flow of five core elements which consist of Capability, System, Activity, Technology and MoP are shown in Figure 2.

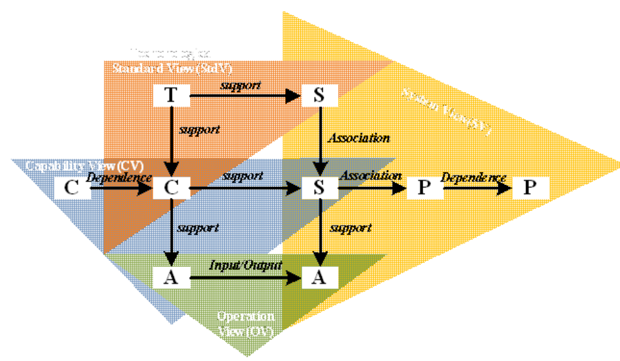


Figure. 2 the Development of the Relationship between Modelling Kernel Elements

3 Transformation form AF Viewpoint Product to Mapping Relationship

3.1 Design Structure Matrix and Domain Mapping Matrix

Design Structure Matrix (DSM)[2] and Domain Mapping Matrix (DMM) [3][4] are system analysis tools based on matrix which are usually used to describe dependence between elements in complex systems analysis. DSM method was first proposed by Steward[8], which describes the dependence between elements of the same domain. So DSM model can be represented by a N×N matrix. Columns and rows represent the same kinds of elements, then the crosses represent their communication or dependence relationship. Danilovic and Browning investigated the dependence between activities from the development of cross domain products, Domain Mapping Matrix (DMM) [3][4] model was proposed based on DSM model. DMM considers dependence between elements from multi-domains, which differs from DSM method that considers dependence from only one domain. Therefore, DSM model is a N×N matrix and DMM model is a N×P matrix (N equals to P or not). Actually, DSM model and DMM model complements each other, the former concentrates on relationships between elements from single domain, while the latter concentrates on relationship between elements form multi-domain. Charles Dickerson[5] applied DSM method to SoS requirements management and proposed a new SoS requirement traceability method based on DSM.

3.2 Reflection of View Product to Mapping Relationship

The mapping relationships between performance indices and four modeling essentials (capability, equipment, activity and technology) are implicit in different DoDAF viewpoint models, which can't be described totally in a single model. DSM/DMM model provided a formalized model to describe mapping relationships between different core essentials. In fact, as shown in figure 2, there are ten classes of mapping relationships among the five core essentials existing in viewpoint models. The ten classes of mapping relationships with DSM/DMM models and their original models are given in Table 1.

Table 1 Description Models of Ten Mapping Relationships and Original Viewpoint Product Models

	Code	Entity of Mapping	Relationship	Model of Description	DoDAF ViewPoint
1	C-C	From Capability to Capability	Dependence	DSMCC	CV-2 CV-4
2	C-S	From Capability to Weapon System	Support	DMMSC	CV-7
3	C-A	From Capability to Operation	Support	DMMAC	CV-6
4	S-S	From System to System	Interface	DMMSS	SV-3 SV-6
5	S-A	From System to Operation	Support	DMMAS	SV-5b
6	A-A	From Operation to Operation	Input/Output	DMMAA	OV-5b OV-6b OV-6c
7	T-C	From Technology to Capability	Support	DMMCT	TV-6[6]
8	T-S	From Technology to System	Support	DMMST	TV-1
9	S-P	From System to Performance Index	Association	DMMPS	SV-7a
10	P-P	From Performance Index to Performance Index	Dependence	DSMPP	SV-7

Notice: viewpoint models marked with * are from document [6], others from literature [1].

Models of the ten classes of mapping relationships can be extended to produce a hypercube mapping model representing multiple-level mapping matrix, which is shown in figure. 3.

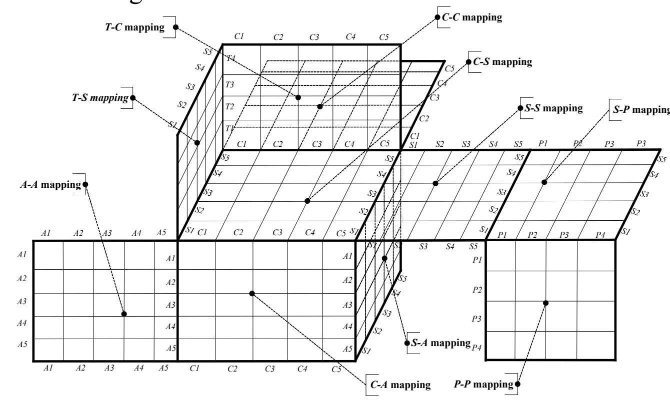


Figure. 3 Hypercube model of multiple-level elements relationships

4 Three Formal Operators for DMM Model

4.1 Formalization of DMM

In order to formalize the analysis of mapping between capability and performance indices, we define DMM model as:

$$DMM = <<H_{class}, H_{entities}>, <V_{class}, V_{entities}>, [B]> \quad (1)$$

H_{class} is the type of the row element. $H_{entities} = \{h_1,h_2,...,h_n\}$ represents a list of the row elements. V_{class} is taken as the type of column element. $V_{entities}=\{v_1,v_2,...,v_m\}$ is a list of column element. $[B]=\{b_{ij}\}_{n\times m}$ represents the relationship of the model which is a $n\times m$ matrix. The values of b_{ij} is defined as:

$$b_{ij} = \begin{cases} 1 & \text{if there exists the reflection from } v_j \text{ to } h_i \\ 0 & \text{else} \end{cases} \quad (2)$$

DSM is a special case of model shown above:

$$DSM = <E_{Class}, E, [A]> = <<E_{Class}, E>, <E_{Class}, E>, [A]> \quad (3)$$

The details of definition to three basic operators for DMM models are shown as follows, it is Transfer Operator Communication Operator and Transpose Operator.

4.2 Transfer Operator

The transfer operation operator is based on the transfer of mapping relationship. Define the transfer of mapping relationship as: if A maps to B and B maps to C, then A maps to C. Assume DMM model $DMM_{AB}=<<A_{Class}, A = \{a_1,a_2,...,a_n\}>, <B_{Class}, B=\{b_1,b_2,...,b_k\} >, \{r_{ij}\}_{n\times k}>$, and DMM_{BC} =<<B_{Class}, B ={ $b_1, b_2, \ldots, b_k$ }>, < C_{Class} , C={$c_1, c_2, \ldots, c_m$} >, $\{r'_{ij}\}_{k\times m}$ >, The transfer operator $\otimes$ is defined as:

$$DMM_{AC} = DMM_{AB} \otimes DMM_{BC} = << A_{Class}, \{a_1, a_2, \ldots, a_n\}>, <C_{Class}, \{c_1, c_2, \ldots, c_m\}>, \{\tilde{r}\}_{k\times m}> \quad (4)$$

$\tilde{r} = \bigvee_{s=1}^{k}(r_{is} \wedge r_{sj}) = \max_{s}\{\min\{r_{is}, r_{sj}\}\}$ DMM_{AB} is the original mapping model, DMM_{BC} is the middle mapping model. DMM_{AC} is the goal mapping model. Element B is the transfer carrier.

The model transfer operator must obey the follow conditions:

(1) The transfer carrier B of original mapping model and middle mapping model must be the objects of the same type.

(2) The numbers of transfer carrier B must be even. If not, define B of DMM_{AB} as B_1 and B of DMM_{BC} as B_2. According to the following rule, we can get DMM_{AC}:

a) $\tilde{B} = B_1 \cap B_2 = \{\tilde{b}_1, \tilde{b}_2, \ldots, \tilde{b}_{k'}\}$

b) Translate DMM_{AB} to $DMM_{A\tilde{B}}$, $DMM_{A\tilde{B}}$=<<A_{Class}, $A = \{a_1, a_2, \ldots, a_n\}$>, <$B_{Class}$, $\tilde{B} = \{\tilde{b}_1, \tilde{b}_2, \ldots, \tilde{b}_{k'}\}$>, $\{r_{ij}\}_{n\times k}$>, in which

$$b_{ij} = \begin{cases} 1 & \text{if there exists the relection from } b_j \text{ to } a_i \\ 0 & \text{else} \end{cases} \quad (5)$$

c) According to the rule b), translate DMM_{BC} to $DMM_{\tilde{B}C}$

d) $DMM_{AC} = DMM_{A\tilde{B}} \otimes DMM_{\tilde{B}C}$

4.3 Combination Operator

The combination operator deals with two object models whose type of row and column elements are the same. If two DMM Models are DMM_{AB} = <<A_{Class}, $A = \{a_1, a_2, \ldots, a_n\}$>, <$B_{Class}$, $B=\{b_1, b_2, \ldots, b_m\}$>, $\{r_{ij}\}_{n\times m}$ >, $DMM_{A'B'}$= << A_{Class}, $A' = \{a'_1, a'_2, \ldots, a'_s\}$>, < B_{Class}, $B' = \{b'_1, b'_2, \ldots, b'_t\}$>, $\{r'_{ij}\}_{s\times t}$>, then define the combination operator $\oplus$ of the two models as the following:

$$DMM_{AC} = DMM_{AB} \oplus DMM_{A'B'} = < A_{Class}, \tilde{A} = \{\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_{\tilde{n}}\} >, < B_{Class}, \tilde{B} = \{\tilde{b}_1, \tilde{b}_2, \ldots, \tilde{b}_{\tilde{m}}\} >, \{\tilde{r}_{ij}\}_{\tilde{n}\times\tilde{m}} > \quad (7)$$

There into, $\tilde{A} = A \cup A'$, $\tilde{B} = B \cup B'$, $\tilde{n} = length(\tilde{A})$, $\tilde{m} = length(\tilde{B})$,

$$b_{ij} = \begin{cases} 1 & \text{if there exists the relation from } \tilde{b}_j \text{ to } \tilde{a}_i \\ 0 & \text{else} \end{cases} \quad (8)$$

4.4 Transpose Operator

The transpose operator is an operation of exchanging a row of the model to its column. Assumed the model is DMM_{AB} = <<A_{Class}, $A = \{a_1, a_2, \ldots, a_n\}$>, <$B_{Class}$, $B = \{b_1, b_2, \ldots, b_m\}$>, $\{r_{ij}\}_{n\times m}$>,the transpose model is DMM_{BA}, the transpose operator ® is defined as the following:

$$DMM_{BA} = ®DMM_{AB}® = <<B_{Class}, B=\{b_1, b_2, \ldots, b_m\}>, <A_{Class}, A = \{a_1, a_2, \ldots, a_n\}>, \{r_{ji}\}_{m\times n}> \quad (9)$$

The priority of three formal operations above is: $® \succ \otimes \succ \oplus$.

5 AnalyzingMapping between Capabilities and Performance Indices

The performance index is correlative directly with weapon systems, and this relationship can be attained only by mapping matrix for capability and weapon systems. Actually, mapping between capability and performance index is not only concerned with weapon systems, but also have indirect relationship with operation activities and technology, which respectively represent the response to SoS capability for the design of weapon systems, the requirement of operation activities and the dependent of technology development. According to the mapping between core elements described in Figure.2, there are three calculation procedures from SoS capability to performance indices.

1) C->S->P: SoS capability reflects design demand of equipment system

2) C->A->S->P: SoS capability reflects requirement of operation activities

3) C->T->S->P: SoS capability reflects requirement of technology development.

The calculation process from capability to performance indices are shown as Fig. 4.

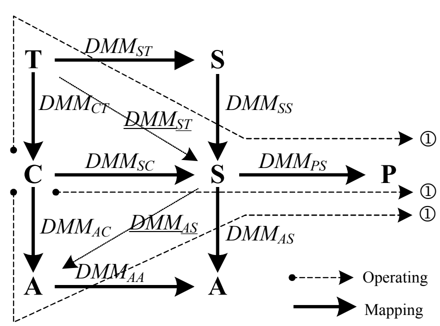


Figure4 Calculation Process from Capacity to Performance Indices

The communication relationship between each weapon systems throw each interface is described by mapping model DMM_{SS}. While calculating the mapping between capability and performance indices, DMM_{SS} becomes the basis of mapping analysis between elements such as systems and capability, systems and operation activities, systems and technology, which are embodied in DMM_{SC}, DMM_{ST} and DMM_{AS} models.

In summarize, mapping model DMM_{PC} that describe one capability (C) has direct connection to which performance indices can be gained as follows.

$$DMM_{PC} = DMM_{PS} \otimes [DMM_{SC} \oplus (DMM_{ST} \otimes DMM_{CT}{}^{\Phi}) \oplus (DMM_{AS}{}^{\Phi} \otimes DMM_{AC})] \quad (10)$$

6 Conclusions

Mapping relationship between SoS capability and performance indices is connected in SoS architecture product modes based on multi-views mode. It is codetermined by performance indices associated with items such as weapon systems devoting to accomplish capability, operation activities supported by capability and key technology necessary to capability development. By using DSM/DMM models, mapping between elements like capabilities, operation activities, weapon systems, technology and performance indices contained in view manufactures can be visually described. Based on this foundation, mapping between SoS capability and performance index can be achieved by transform, combination and transpose operation of DMM models, then provide a necessary foundation to evaluate the capability of certain SoS.

Acknowledgements

This research was funded by the National Science Foundation of China (Grant No. 70901074, 71001104, 71031007).

References

[1] DoD Architecture Working Group. *DoD Architecture Framework Version2.0 Volume 2: Architectural Data and Models*, U.S. CIO US DoD Chief Information Officer, Washington, 2009.

[2] Steward D. "The Design Structure System: a Method for Managing the Design of Complex Systems," *IEEE Transaction Engineering Management*, Vol 28, No. 3, pp. 71-74, Mar. 1981.

[3] Danilovic M, Sandkull B. "The Use of Dependence Structure matrix and Domain Mapping Matrix in Managing Uncertainty in Multiple Project Situations," *International Journal of Project Management*, Vol 23, No. 3, pp. 193-203, 2005.

[4] Danilovic M , Browning T R. "Managing Complex Product Development Projects with Design Structure Matrices and Domain Mapping Matrices," *International Journal of Project Management,* Vol 25, No.3, pp. 300-314, 2007.

[5] Charles Dickerson , Ricardo Valerdi. Using Relational Model Transformations to Reduce Complexity in SoS Requirements Traceability:Preliminary Investigation, 2010 5th International Conference on System of Systems Engineering, IEEE, pp. 1-6, June. 2010.

[6] Ren Changsheng, Ge Bingfeng, Chen Yingwu. "Technical Architecture Description Method of Weapon System-of-Systems", *Ordnace Industry Automation*, Vol 29, No. 10, pp. 42-46, Oct. 2010 [in chinese]

[7] DoD Architecture Working Group. DoD Architecture Framework Version2.0 Volume 1: Introduction, Overview, and Concepts ,Washington:U.S. CIO US DoD Chief Information Officer, 2009.

[8] Steward D. "The design structure systems", No.67APE6 San Jose:1967.

Embedded Concurrent Computing Architecture using FPGA

Muataz H. Salih
School of Computer & Communication Eng.
Universiti Malaysia Perlis
Perlis, Malaysia
moutazsaleh@yahoo.com

R. Badlishah Ahmad
School of Computer & Communication Eng.
Universiti Malaysia Perlis
Perlis, Malaysia
badli@unimap.edu.my

Abid Yahya
School of Computer & Communication Eng.
Universiti Malaysia Perlis
Perlis, Malaysia
abidyahya@unimap.edu.my

Mohd. Rizal Arshad
School of Electrical & Electronic Eng.
Universiti Sains Malaysia
Penang, Malaysia
rizal@eng.usm.my

Abstract - *Simultaneous multithreading by use of embedded parallel systolic filters is a novel technological approach to achieve multiprocessing. It is important for the designers to ensure that FPGA chips that are fully operational. There is great emphasis on the design area, performance, challenges and opportunities posed by multi-tasking as a result of the huge number of inputs and outputs required by the design. The Embedded Concurrent Computing Architecture proposed is implemented on a FPGA chip. There are expected speedups in the implementation based on the results shown in this proposal. Synthesis has been used in gathering of the results with implementation being achieved by use of low complexities in the FPGA usage and frequency. The efficiency of the new model is over 75% with the performance of the design is secured for a tolerance of 2 m for 25 m range. The Particle filter tolerance is less than 1m with an operating frequency of 212 MHz or thereabouts.*

Keywords: Embedded system design, parallel processing, simultaneous multithreading, underwater detection filtering.

1 Introduction

Embedded concurrent computing is a technology that uses several techniques in order to design a multiprocessor. The multiprocessor is able to conduct several tasks simultaneously thereby enhancing its speed and efficiency. The efficiency of multiprocessors is determined by the speed compared to the resources used in designing and implementing the multiprocessor [1]. This paper is a design and implementation proposal of an embedded concurrent processor that utilizes embedded parallel systolic filters in its architecture in simultaneous multithreading.

The design identifies the major materials and the design to be adopted. There are several multithreading techniques but the one discussed herein is based on simultaneous multithreading [2]. The embedded architecture is designed such that to ensure that the memory management unit is able to multitask. This is achieved through parallel computing of multiple processors and distributed computing networks of processors.

The design implementation is done on FPGA chip that makes use of VHDL language where pixel and parameter calculation is possible through use of frame and parameter buffers. The proposed design uses concurrent architecture platform with the chip design selected being Altera® Cyclone II FPGA that has several logical resources and memory balance capabilities. An AUV sensor with thirty two sound sensors in all directions is chosen for the design whose implementation for underwater applications. The paper also addresses the design area, performance and validation.

2 Multiprocessor

A processor system that uses two or more processors that are interconnected is a multiprocessor [3]. A multiprocessor system has multiple processors that work to perform a many tasks at once [4, 5]. This is possible through three major ways: the first is the use of a single chip that has processors connected together through Network on Chips (NoC) or bus, the second is the use of multiple chips connected by bus with each chip being a multiprocessor chip, and finally having a multiple processor working with more than one computer [6].

A parallel system which utilizes more than one threads allow the processor to multitask with minimal idle time differences. This requires the designing of a processor that has special-purpose IP cores [7]. This means that for an equivalent number of processors in the system, there is an equivalent threads thereby ensuring that each processor is functional at all times. This requires sharing of memories by the processors and ensuring that none of the processor access same value at once in ensuring thread safety. This is possible through the use of two major architectural designs, which are Chip Multi-Processors (CMPs) and Simultaneous Multi-Threading (SMT) [8].

CMPs’ processor cores use single threads in moderating parallelism amounts in a single thread in order to execute multiple threads [2]. A single program is used to execute multiple instructions in a single cycle. Exploiting the Thread-Level Parallelism (TLP) in multiprocessors is possible through parallel execution of different threads in different processors [7].

Simultaneous multithreading of processor consumes thread-level and instructions-level parallelism. This is achieved through parallel programs multithreading or independent or individual multithreading programs [9, 10]. Each single thread or program provides instruction-level parallelism. This makes SMT processors to be successful and efficient for they consume resources efficiently and have high throughput and speedups instructions.

3 Proposed Design

The proposed design utilizes Embedded Parallel Systolic Filters (EPSF) in the architecture of an Embedded Concurrent Computing (ECC). The advent of control units and microcontroller leading to widespread components are possible with systematic development. However, these components are limited in terms of the tasks they accomplish and inherent rigidity in interfacing. This has put constraints in applications in other scenarios. The introduction of modern programmable circuits including FPGA and CPUs has changed the situation gradually [11, 12].

In the proposed embedded processor, the core contains "a dual-issue, superscalar, pipelined processing unit" together with other elements whose function help implement solutions for an embedded SoC. Memory management and timers are other functions that the component is achieved. The processor offers separate tasks and data units. The local organization of the embedded processor is illustrated in the Figure 1.

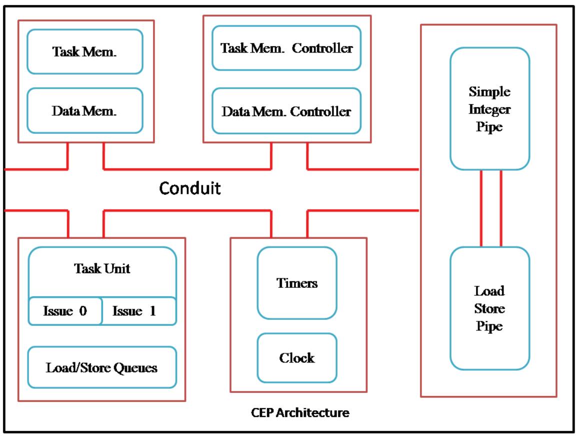


Figure 1. Block diagram of embedded processor

The embedded processor has abilities to separate tasks and control data thereby allowing concurrrent access and minimiztion of pipeline stalls. The proposed design is more focused on advancing multiprocessor architecture in terms of the embedded multiprocessor that has features such as core processor with two concurrent embedded processor, dual port RAM, four parallel systolic filters that are embedded in the design, internal input/output buffers, signal emulator, and a view and control sub-system. The design is fabricated on a single FPGA chip as illustrated in Figure 2.

Computing and processing in the new architecture is through two strategies namely:

1. Parallel computing-multiple processors are used simultaneously to solve a task and are all on a single architecture.

2. Distributed computing-network processors with capabilities of being viewed as computers in their own right are used in solving problems with each processor being assigned a specific task.

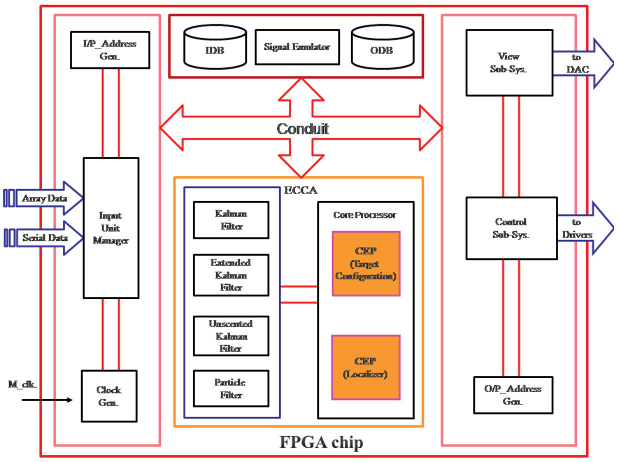


Figure 2. The proposed architecture

The design platform utilizes concurrent architecture since it is a realistic hardware environment. The chip selection and the methodology are aimed at synthesizing the concurrent system. They also are running the platform on a design custom board as illustrated in Figure 3. The chip for the design platform from Altera® Cyclone II FPGA.

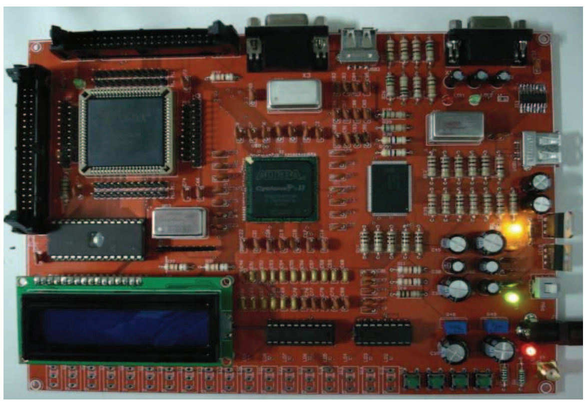


Figure 3. Designed custom board

4 Experimental setup and case study

Figure 4 shows the components for the experimental platform and their relationship. There is the use of three sets of parameters in the full description of concurrent experimental designs. The embedded processor is defined in the architectural parameters represented in the first set.

The parallel unit is defined by the architectural parameters represented in the second set. The third set has system-wide parameters. The SoC design tool is fed with the parameters to generate the RTL code of the entire system. FPGA programming file is produced by RTL code using the synthesis tool. The platform's software consists of software compilation tool and embedded machine generator that generates codes.

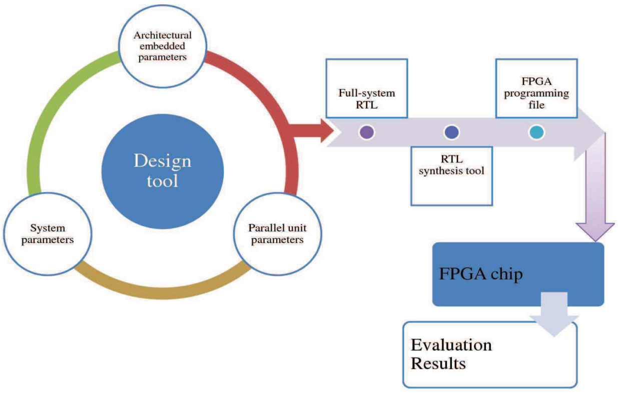


Figure 4. Experimental platform

In exploring the concurrent architecture design for the navigation and tracking in AUV, the application involves building a system capable of taking N inputs and outputs that process in real-time. This design poses challenges of its own despite being parallel. It is difficult to perform reliable localization and navigation in highly unstructured underwater environment. In order to ensure that repeatable measurements are taken correctly, it is important to know the position and distance moved by AUV [13, 14]. There are several techniques used and proposed in estimating vehicle motion that can be categorized as either acoustic or vision-based [15].

The project utilizes thirty-two sound sensors in array to the AUV covering all directions. The design system is represented by the receiver signals. The signals enter EPSF module directly as shown in Figure 2 before going to the core processor module after filtering takes place in order to get specific processing. The distributions of AUV's hydrophone sensors are:

- Six on the left side
- Six on the right side
- Six at the front
- Six at the back
- Four on top
- Four at the bottom

The arrangement of these sensors is two-dimensional matrix [4 x 8] in the design. The location of each sensor is represented by 8-bit from a fast ADC chip. Specification of the location of each array sensor is further done by use of 6-bit array of sensors using a two dimensional matrix [4 x 8]. This has 3 bit for direction and 3-bit for the position of the sensor coming from a voltage sensor in an identification circuit. A (010100) for instance represents a backward side and fifth sensor. These matrices represent location information for underwater vehicle. Grounding the truth in visual odometry system has been difficult in previous works due to close proximity required in sea floor with GPS operating on the surface. The array sensors' signals enter the 8-bit ADC in order to be converted from analogue to digital before entering the embedded processors to estimate distances, speed of objects and direction in underwater environment. The speed of sound in the environment, the flexibility of the architecture proposed and the response is real time due to the high speed of FPGA chip. The architecture is applicable to other media services with modifications in the FPGA design.

Underwater navigation and tracking has several challenges. It is critical to obtain simultaneous range of measurements in objects. This demands a large AUV motion between the ranges of measurements. There have been the designing of three independent embedded processors in the recent past. Each of the three has four input data channels that emanate from the EPSF module and there is manual selection of the filtering type in the design. Applying a technique that gives automatic selection of filtering autonomy in underwater environment is the next step in the design.

There is need to generate different pattern of input signals in order to verify the design. This requires generation of sound sensors in the custom board and oversees the behavior design. The pattern achieves a major goal in that the global positions are identified in order to estimate the position and orientation of the AUV through solving projective geometry. This arrangement provides a clear view for the AUV in each direction thereby simplifying localization, mapping and tracking as shown in Figure 5. Finding neighborhood in terms of detected signals is the next phase in localization of systems. Compensation of distortion is the first step that helps counter the affected position of the detected signal in sensor planes. Ideal projective geometry makes the new status of sensor's signals. The embedded processor builds lines in real world so that the lines appear as images. This property and relative distances and angles make viewing of landmarks in water possible. The AUV must have at least two sensors on each side in order to detect the landmarks; however this reduces accuracy and lowers the precision of a target position. Finding neighborhood of each sensor is the next step.

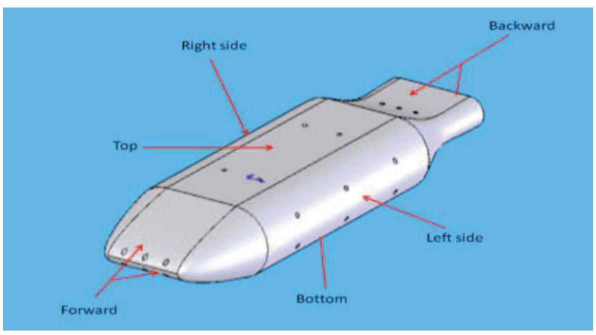


Figure 5. Sensor arrangements on the AUV

Each position estimation of AUV requires a set of sequential tasks. Detecting the signal of the sensors is the first phase. Acquired signals are digitalized, but calibration of voltage must be done prior to this. This is because there is non-uniformity in the sensitivity of the sensors in the field of view. Sound echo is used for calibration since it has robust digitalization. The algorithm originates from the center before proceeding to others. In order to assign all sensor neighbors, a recursive algorithm is adopted. This uses distances and angles in between the sensors. After all the sensors have been assigned, an image joining the neighboring sensors is drawn. This technique gives the system the chance required in processing enough data resources for efficient decision-making.

There is accurate estimation of the AUV positioning and orientation in the sensor system, sensor-based localization presented in this paper. Imperfections in the signal pattern generated are the main sources of error that are likely to affect the system. Based on studies on sources of errors, there was an assumption that localization system behaves like a real process whereby the estimates coincide with the robot's real position [16]. The estimation is carried out through global positioning of the marks identified by the sensors. In clean and fresh conditions, tracking of targets and detection hardly fails, thereby showing the absence of drift in estimates [17]. Unscented Kalman Filter (UKF) sends processed sensor signal illustrated in Figure 6. Also, Figure 7 represents the same signal from an Particle Filter (PF).

Particle filter has better and accurate filtered signals compared to the unscented Kalman filter as shown by both Figures 6 and 7. The system also gets an opportunity to compare its processed data differently. This yields landmarks in localization thereby indicating different landmarks and targets that shows that nonlinearities in the system are a possible cause of errors. Even if any of the filters fails, the system continues to operate and detect due to the presence of too large nonlinearities due to linearization errors. The system is also capable of dealing with measurements that correspond with different points of reference.

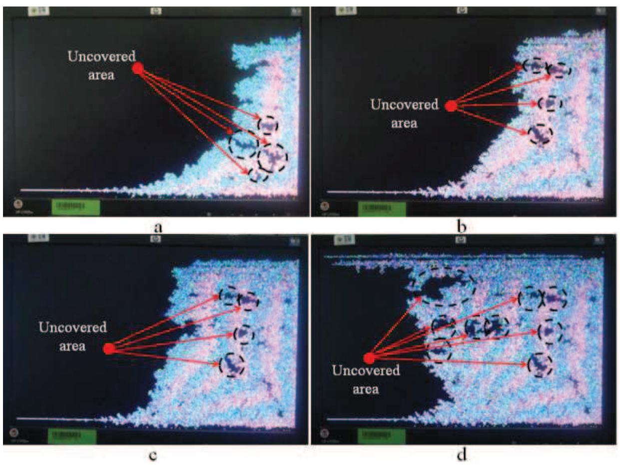


Figure 6. Processed sensor signal coming from the UKF

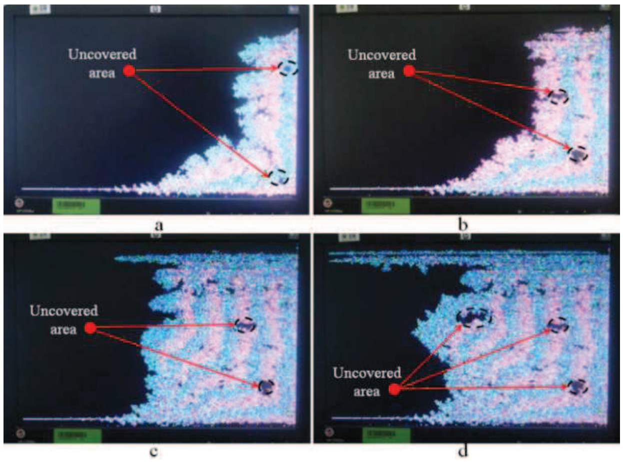


Figure 7. Processed sensor signal coming from the PF

5 Design area and performance

Synthesis process is used in gathering a lot of data in the system such as maximum frequency and size. Optimization with different goal and effort is possible by use of synthesis in order to get uniform results that are comparable. Speed is used as a goal and a normal in optimization while synthesizing since they give the best results with speed and area being factors in the tree structures. the stated maximum frequencies coincide with Altera® Cyclone II FPGA the synthesizing tool model number EP2C35F672C6. The main used models and their sizes in elements are illustrated in Table 1.

The design's sizes and maximum frequency is shown in Table 2. Data gathered is based on systems positioned at a top level design since it was the only way to have data collected from tools. Therefore, the data gathered do not fulfill 1022 elements in the system neither does I/P manager fill 214 elements used in the system. The size of the module is described only when used with another.

Table 1 Size of modules

Module	No. of elements
Kalman Filter	1262
Extended Kalman Filter	1738
Unscented Kalman Filter	1529
Particle Filter	1599
Core processor	2274
Input manager	214
Buses	1022
Signal emulator	101
View sub-system	171
Control sub-system	109

Table 2 Size and system speed

Area	Max frequency
12,257	212MHz

One of the core objectives of this paper is to determine the effectiveness of the novel embedded processor by evaluating its performance. Because of the device target that has been used in this research project from the family of Altera®, a comparison takes place Nios® II soft processor and the designed embedded processor. The Nios® II soft processor [18] was originally designed for implementation in Altera® FPGAs and is configurable. The performance parameters for both the processors are illustrated in Table 3.

Table 3 Performance comparison between designed embedded processor and NIOS II

	Designed Processor	Nios II/f	Nios II/s	Nios II/e
f_{MAX} (MHz)	212	140	110	195
LE Usage	1137	1,600	1,030	540

6 Design validation

Range information is the main important condition that needs to be tested in the design. In order to obtain range, onboard synchronized clocks and ping are utilized at specified intervals. Ping's echoes from different directions are listened by the sensors before being computed in the range by use if round trip times. Ping signal frequency represents important parameters usable for range finding. There is a reverse relation between signal penetration and frequency because of characteristics of water.

There is high absorption in high frequency leading to low penetration compared to infra frequency that has high penetration [17]. 1 Hz is chosen for design validation in order to get range finding. This range enables self-localization of the system in the AUV. There is good performance in the system based on the results thereby showing tolerance is less than 2m for 25m range. The minimum tolerance obtained of 1m or less from PF module is logical since it is a precise error minimizing as illustrated in Figure 8. Additional experiments are performed to determine the position of the moving targets as shown in Figure 9; the data is collected during 20 minute run of the proposed design.

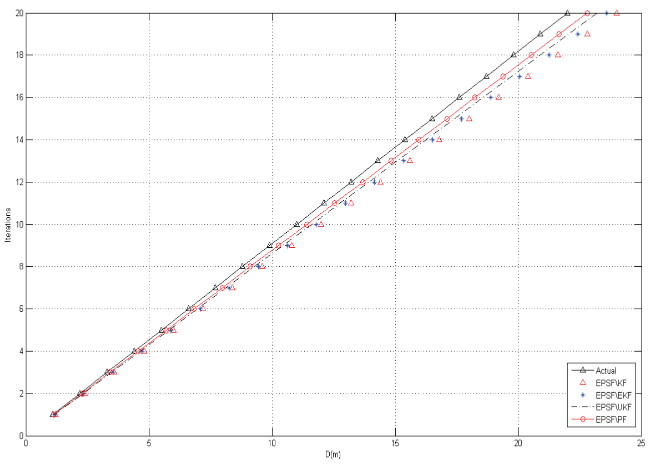


Figure 8. Movement of oncoming target to AUV for 20 iterations

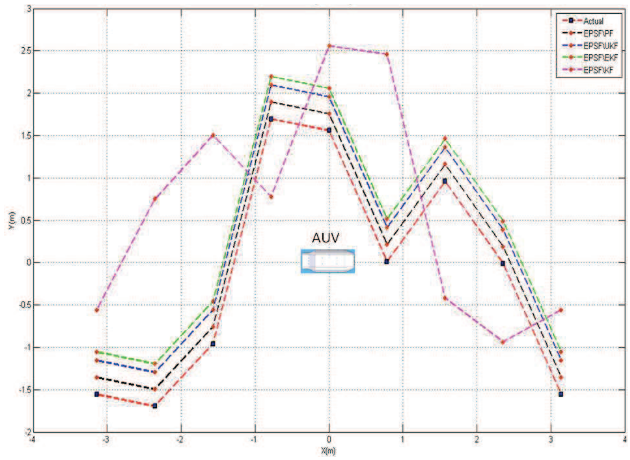


Figure 9. Tracking results of moving target

7 Conclusion

This paper focused on the designing and implementation of concurrent architecture by use of the new ECC architecture. There is an emphasized on the embedded core processor with a goal for implementation being twofold. The first goal was the design of a full architecture exploration while achieving high performance in a complexity-efficient manner. This required consumption of minimal chip resources to achieve high operating frequency.

There are better performances in PF as presented in this paper covering a sensing area of over 95% compared to UKF that has less than 75% cover . This strategy is an opportunity to apply reconfigurable technology applicable in underwater sensor imaging application. There are investigations of improving processing algorithms that make them more robust and are able to process data without the need of human input. The proposed system, also, has good performance based on calculated range and tolerance that is less than 2 m for a range of 25m. Particle Filter module has a minimum tolerance of less than 1m which is logical based on its precise error minimization capabilities. There are acceptable size logical elements in the implementation 12,257, lower complexity with more optimization in terms of FPGA resource usage and operating frequency which is 212 MHz.

References

[1] Nurmi, J. 2007. Processor Design: System-On-Chip Computing for ASICs and FPGAs. New York: Springer p.300-548.

[2] Beveridge, J and Wiener, R. 1996. Multithreading Applications in WIN32: The Complete Guide to Threads. San Francisco: Addison-Wesley Professional.

[3] Hubner, M and Becker, J. 2010. Multiprocessor System-on-Chip: Hardware Design and Tool Integration. New York: Springer p. 201-245.

[4] Geer, D. 2005. "Chips Makers Turn to Multicore Processor," Computer, 38(5): 10-12.

[5] Flynn, M. and Luk, W. 2011. Computer System Design: System-on-Chip. San Francisco: Wiley p. 123-213.

[6] Anderson, D. and Shanley, T. 1995. Pentium Processor System Architecture. New Jersey: Addison-Wesley Professional.

[7] Chen, J., and Hong, S. 2004.Real-Time and Embedded Computing Systems and Applications: 9th International Conference, RTCSA 2003, Tainan, Taiwan, February 18-20, 2003. Revised Papers. New York: Springer. p. 546-601.

[8] Kumar, R., Tullsen, D., Jouppi, N. and Ranganathan, P. 2005. "Heterogeneous Chip Multiprocessors." Computer, 38(11): 30-36.

[9] Cathey, C., Bakos, J. and Buell, D. 2006. "A Reconfigurable Distributed Computing Fabric Exploiting Multilevel Parallelism" In Proceedings of the 14th Annual IEEE Symposium on Field-Programmable Custom Computing Machines (FCCM' 06). IEEE Computer Society Press.

[10] Chang, C., Wawrzynek, J and Brodersen, R. 2005. "A High-End Reconfigurable Computing System" IEEE Design and Test of Computers, 22(2): 114-125.

[11] Casalino, G., Giorgi, F. Turetta, A. and Caffaz, A. 2003. Embedded FPGA-based control of a multifingered robotic hand, IEEE Int. Conf. on Robotics & Auto. M. p. 2787-2789.

[12] Hamblen, J., Hall, T. and Furman, M. 2007. Rapid Prototyping of Digital Systems: SOPC. New York: Springer p.327-313.

[13] Dalgleish, F., Tetlow, S. and Allwood, R. 2004. "Vision-based navigation of unmanned underwater vehicles: a survey part 1: Vision-based cable-, pipeline-, and fish tracking," Journal of Marine Design and Operations, no. B7, p. 51–56.

[14] Graham, R. Woodall, T. and Squyres, J. 2005." Open MPI: A Flexible High Performance MPI." In Proceedings, 6th Annual International Conference on Parallel Processing and Applied Mathematics, Poland: Poznan.

[15] Amidi, O. 1996. "An autonomous vision-guided helicopter," Ph.D. dissertation, Dept of Electrical and Computer Engineering. Pittsburgh, PA: Carnegie Mellon University, 15213.

[16] Cortadella, J., Yakovlev, A., and Rozenberg, G. 2003. Concurrency and Hardware Design: Advances in Petri Nets. New York: Springer.

[17] Dunbabin, M. Usher, K. and Corke, P. 2006. "Visual motion estimation for an autonomous underwater reef monitoring robot," in Field and Service Robotics: Results of the 5th International Conference, P. Corke and S. Sukkariah, Eds. Springer Verlag, 2006, vol. 25, p. 31–42.

[18] AlteraTechnicalSupport (July 2011). Nios II Performance Benchmarks: Altera Corporation.

Proc. of the 2012 7th International Conference on System of Systems Engineering, Genoa, Italy - 16-19 July 2012

Features of CML: a formal modelling language for Systems of Systems

J. Woodcock*, A. Cavalcanti*, J. Fitzgerald†, P. Larsen‡, A. Miyazawa*, and S. Perry§

*University of York, United Kingdom, {jim.woodcock, Ana.Cavalcanti, alvaro.miyazawa}@york.ac.uk
†Newcastle University, United Kingdom, john.fitzgerald@newcastle.ac.uk
‡Aarhus University, Denmark, pgl@iha.dk
§Atego, United Kingdom, simon.perry@atego.com

Abstract — We discuss the initial design for CML, the first formal language specifically designed for modelling and analysing Systems of Systems (SoSs). It is presented through the use of an example: an SoS of independent telephone exchanges. Its overall behaviour is first specified as a communicating process: a centralised telephone exchange. This description is then refined into a network of telephone exchanges, each handling a partition of the set of subscribers (telephone users). The refinement is motivated by a non-functional requirement to minimise the cabling required to connect geographically distributed subscribers, who are clustered. The exchanges remain as independent systems with respect to their local subscribers, whose service is unaffected by the loss of remote exchanges.

Keywords: architecture, SysML, modelling, specification, refinement, evolution, formal analysis, VDM, *Circus*, CML, semantics, UTP, enslavement pattern.

1. Introduction

The design of products and services that exploit Systems-of-Systems (SoS) technology is in its infancy. It is hampered by the complexity caused by the heterogeneity and independence of SoS constituent systems. State-of-the-art SoS engineering lacks models and tools to help developers make trade-off decisions during design and evolution, and to assist in working out and recording precise contracts between constituents and the global SoS. This leads to sub-optimal design and expensive rework during integration and in service.

COMPASS (Comprehensive Modelling for Advanced Systems of Systems)[1] is augmenting existing industry tools and practice with a modelling language in which SoS architectures and contracts can be expressed. A formal semantic foundation—the first to be developed specifically for SoS engineering—enables analysis of global properties. The language and methods will be supported by an open tools platform [1] with prototype plug-ins for model construction, dynamic analysis by simulation and test automation, static analysis by model checking and proof, and links to an established architectural modelling language, SysML [2]. These strengthened foundations and tools will support enhanced methods that help users embed this new technology in industrial practice.

Our approach is based on CML (COMPASS Modelling Language) with formal semantics, and methods and tools that take advantage of this. It uses an integration of the Systems Modelling Language, SysML, and CML, where developers can start if they wish from a graphical architectural view that is readily communicated to stakeholders.

CML is founded on the well-established *Circus* [3] (which in turn is based on Z [4] and CSP [5]) and VDM [6] formalisms, and includes SoS-specific aspects, such as contracts for constituent systems. A contract is a specification that describes the assumptions and guarantees of a CML model, and whose compliance is checked using refinement. A CML model is a collection of process definitions; each process encapsulates a state and operations written in VDM and interacts with the environment via synchronous communications, like in *Circus*. Using CML, many different kinds of analyses can be conducted, and some will be presentable at the SysML level. The semantics of CML is currently being developed using UTP [7].

In Section 2, we describe the approach to modelling in CML; in Section 3, we describe the

[1]COMPASS is a EU FP7 project. The COMPASS website is www.compass-research.eu.

CML model of our telephone exchange SoS; and in Section 4, we draw some conclusions from the work.

2. Modelling in CML

A central notion is that of conformance between models: one model C is correct with respect to another model A, if every behaviour of C is also a behaviour of A. When conformance holds, then C must inevitably pass every test carried out with respect to A. This notion of behavioural refinement offers a way of defining contracts: we would like A to be a simple presentation of all acceptable behaviours (the contract); C can then be more ingenious about how to implement the contract. A formal notion of refinement allows us to check the correctness of C against the contract defined by A.

CML can be used to model existing systems, compose them into an SoS, define suitable contracts for this composition, and check that the contracts are fulfilled. In this model-based development, an SoS specification can be traceably decomposed into an architecture and a collection of requirements on constituent systems expressed as contracts. The constituent systems can then be further decomposed or implemented by procurement of specific systems. Other requirements can be shown to emerge as a consequence of executing constituent systems.

The overall contract must describe the behavioural properties of the SoS, as well as specific policies and constraints for coordinating constituent systems and their workflow. It will provide global invariants that may be inexpressible at a lower level, and so can be used to constrain emergent behaviour. As an example, consider an SoS that manages the clearing system for a group of banks. A global invariant would reconcile the amount of money coming into the clearing system, the money moving between banks, and the money leaving the clearing system. This is a clear and intuitive invariant of the global view of the system, but it is obviously not an invariant of any individual bank.

The architectural description needs to represent the topology of the SoS and the interactions between the different constituent systems. These channels may require additional properties of bandwidth, delays, and potential faults, as well as the more obvious behavioural protocols for correct operation. Some SoSs will be dynamically evolving, and CML will provide mechanisms to describe mobile channels and mobile processes with inspiration from [8].

3. Example: Telephony System

We describe a small but realistic example using CML. We restrict ourselves to a high-level informal overview of the models involved, omitting formal details and the use of SysML to document the architecture at key phases. Although the development described results in a homogeneous SoS, the pattern is equally applicable to heterogeneous constituents.

A. Abstract Model

We start with a model of the functionality of an automated telephone exchange for connecting subscribers' calls. We assume that a caller identifies the recipient when initiating a call in a single, atomic action. This simplifies our presentation; there is no conceptual difficulty with separating this into seizing a line and supplying the recipient's identity, and even decomposing this into dialling a number digit by digit. Only the caller can clear a telephone call: if a recipient tries to clear a call, then it becomes merely suspended; if the recipient lifts the receiver again, then the call is re-established.

The exchange is specified as a single CML process, encapsulating a state described in VDM, recording the status of every subscriber and every call currently in progress. The operations on this state are made reactive with CML's CSP notation, linking events to the effect on the state. The model gives us a contract for the service provided by the exchange, and we can augment this model with guarantees about quality of service, such as the time to connect a call to the engaged or ringing tone.

B. System of Systems

Subscribers are geographically distributed, but clustered. The initial model requires extensive cabling, and a non-functional requirement is to reduce cabling costs. One way to do this is to install a separate exchange at each cluster, and provide trunk cabling between these exchanges. A subscriber then makes a call to the nearest exchange; the call is either serviced locally or routed to a remote exchange, as appropriate. This suggests an SoS architecture embedding instances of the simple exchange in a suitable topology, with the addition of trunk signalling between exchanges. This SoS can then be shown to refine the original simple exchange. This demonstrates that the required service is being delivered correctly, in spite of a more elaborate implementation. Crucially, the subscriber need not be aware of the way that the service is implemented, either as a single exchange, or as a collection of exchanges with trunk signalling between them.

An emergent property of the new architecture is its fault tolerance. It may be that a single fault in an exchange could cause the exchange's entire service to be lost; but in the new architecture, the only loss would be the telephone calls to and from subscribers local to that exchange. This capability also supports evolution of the system in a way that the single, centralised model does not. An individual exchange can be taken down and replaced while most subscribers can continue calling each other.

C. Architectural Considerations

The exchange is implemented by decomposing it into a number of similar exchanges, each handling a partition of the set of subscribers. This requires a system architecture, which is a set of structures for reasoning about a system: it does not simply explain how to wire up the individual constituent systems, but rather it explains the consequences of doing so. We want to re-use each exchange without modification, and deduce that each exchange continues to provide its existing service without interfering with additional behaviour added to extend the collective capabilities. To do this, we use a particular architectural pattern called enslavement.

To explain this pattern, consider two systems P and Q, where every external event of P is shared with Q, so that P is entirely controlled by Q, but that Q may do other things. P is then Q's slave: P can communicate with no one other than its master Q. Suppose further that P uses two channels in and out. In our example, Q has many similar slaves—the telephone exchanges—so Q communicates with P using channels labelled by some name, say s: the channels would then be s.in and s.out. So there's an asymmetry between the two of them: Q has to be aware of P's identity (the use of the label s), but P does not need to know who Q is.

We use this architectural pattern in our SoS, where we have a collection of identical copies of the single exchange. They differ only in the set of subscribers they are connected to: they serve different customers, and they do not communicate with each other. They form an unconnected group of systems whose behaviours are merely interleaved. We then put a transport layer above this group. Here, all subscribers talk to a single process, which then relays signals from subscribers to the right exchange slave. Similarly, when the slave responds, its signals are relayed back to the subscriber.

In the first model, all subscriber and call information is centralised, but this new architecture distributes this to local exchanges, and correctness can be proved as a refinement. This is necessary, but does not achieve a geographical separation, since the transport layer is a single service. So we decompose the transport layer into a number of nodes, one per exchange. A subscriber now connects to its nearest local node, which either relays messages to the local exchange or sends something to another node to cause it to interact with its local exchange. This is similar to trunk signalling in telephony, and must provide the same service as the centralised description. A suitable architecture for this transport system is to arrange the nodes in a ring, or to take advantage of geographical considerations to arrange them hierarchically, and we adopt the former.

In the next two sections, we use CML to demonstrate two example scenarios: the first shows the exchange of messages involved in a telephone call using a single exchange. The second shows the start of a call involving two separate exchanges.

D. Single Exchange

Suppose that `Jim` wants to phone `Ana` and that both of these subscribers are linked to a single telephone exchange. We assume in this example that `Jim` is not engaged in any other telephone call. He starts the call by sending a message to the exchange: `call(Jim,Ana)`. This and the subsequent sequence of messages are displayed in Figure 1, which contains a SysML sequence digram for the entire call. Once the exchange has received this call, it acknowledges it by replying with a `callok` message. Internally, the exchange now creates a record to keep track of the call, which is in the `connecting` state, and the identity of the subscriber to whom `Jim` is connecting, `Ana`. The exchange also knows `Ana`'s state; we assume for this example that `Ana` is not busy. The exchange sends the message `startringing` to `Ana`; if `Ana` had been busy, then the exchange would have instructed `Jim`'s equipment to start the engaged tone. After a while, `Ana` answers the call by sending an `answer` message to the exchange, which then acknowledges this and instructs `Jim`'s equipment to start receiving speech packets (which are not described in this model, as they are not part of the telephone signalling protocol). A little while later, `Ana` hangs up. As she is the recipient of the call, she does not own it, and so hanging up suspends the call rather than clearing it, and so `Ana` sends a `suspend` message to the exchange, which acknowledges it. A little while later, Ana picks up the phone again and re-establishes the call by sending an unsuspend message to the exchange, which again acknowledges it. Finally, Jim hangs up, clearing the call, since he owns it as the call initiator. This clear message causes the exchange to delete the call record and start a new one for Ana to register the fact that she is still off-hook. When she eventually hangs up, then that final fragment of the call can be deleted.

This informal description of the telephone call, and the accompanying SysML sequence diagram, are based on a mathematical model in CML. As well as explaining what is going on, the model can also be used to predict events and situations: 1) Is it ever possible for the telephone exchange to deadlock? 2) Given the real-time properties of the various components, what are the response and connection times? 3) Are all connected telephone calls one-to-one (no sharing)? 4) What happens when Jim calls himself? 5) What happens when Ana calls him during one of his self-obsessed calls?

We give just a flavour of the CML specification of the exchange. It has three instance variables:

```
public status: map SUBS to STATUS;
public number: map SUBS to SUBS;
public subs: set of SUBS;
```

The status and number variables record information for each call in progress; subs records the set of subscribers linked to this exchange; status is a mapping from subscribers to the status of their calls, such as connecting, speech, or ringing, while number is a mapping from subscribers to subscribers. These two mappings record all necessary information, so an invariant requires that they record the same set of subscribers. In our example, number maps Jim to Ana from the point at which Jim initiates the call until he clears it down. The call passes through the sequence: connecting, ringing, speech, suspended, and speech (again), before being deleted.

Each message received by the exchange triggers a state operation. For example, a new call is handled in the following way, specified in VDM:

```
Call (s,t: SUBS)
frame wr status, number
      rd subs
pre
  s in set subs and
  t in set subs and
  s in set free(status,number,subs)
post
  status = status~ ++ {s |-> <connecting>} and
  number = number~ ++ {s |-> t};
```

The Call operation takes two subscribers s and t as parameters, and has write access to the status and number instance variables and read access to the subs variable. The precondition requires that s and t belong to the set subs, which means that they are subscribers linked to this exchange. Additionally, the precondition requires that s is also free. The function free returns the set of subscribers who are neither initiators nor recipients of calls in progress. The postcondition describes the effect on the instance variables, which are updated to record the new call, which is between s and t and is in the connecting state.

The Call operation describes the changes in the internal state of the exchange that take place following the reception of a call message. The reactive aspect of the operation is specified in CSP. Here, we see the first part of the behaviour in the definition of the Exch process:

```
Exch =
  call?s:(s in set subs)?t:(t in set subs) ->
    ( if s in set free(status,number,subs)
      then callok -> Call(s,t)
      else callerror -> SKIP ); Exch
...
```

A request can be received, via the call channel, to establish a connection between two subscribers, s and t, both of which are linked to this exchange. Following the reception of this message, the exchange tests to see if s, the initiator, is actually free. If it is, then a positive acknowledgement is sent (callok), and the state operation Call(s,t) (above) is invoked. Otherwise a negative acknowledgement is sent (callerror). After this, the process repeats its main loop.

E. Multiple exchanges

Figure 2 describes the interactions between participants in an SoS of just two telephone exchanges, using a SysML sequence chart. We assume in this very simple instantiation there are only two clusters of subscribers, one in York and one in Aarhus. A subscriber in York, Jim, wants to put a call through to a subscriber in Aarhus, Peter. Jim initiates the call with the message call(Jim,Peter), which the SoS receives at the York node. York detects that this call is not to another subscriber in York, but to someone in Aarhus. So it makes a virtual call to Aarhus and sends a vcall message along the mid channel to the York link. This is the process responsible for all messages going out on the ring from York; it can buffer messages up to some limit. The link then forwards this message to Aarhus, the next node in the ring. Aarhus detects that this message is indeed destined for Aarhus, and so takes it off the ring. Aarhus now completes the virtual call by first adding Jim as an Aarhus subscriber and then making a local call from Jim to Peter. The Aarhus exchange responds with an acknowledgement that the call is in progress. It will not have been connected yet, since the exchange needs to find out if Peter is free or busy, but that will follow. In the meantime, the Aarhus node needs to relay this acknowledgement back to Jim in York, so it sends the acknowledgement along its mid channel to the Aarhus link. The link then forwards this along the ring, with the York node recognising and intercepting the message, which can finally be delivered to Jim.

Again, this model can be analysed to prove various properties. Unlike the centralised single exchange, the SoS is a distributed system and the problems of deadlock and livelock are more acute. For example, the links must have a bounded amount of buffering, so what happens when buffers fill up: does the system deadlock? Messages pass around the ring: is it possible for a message to be perpetually travelling? Or we could ask if the SoS of exchanges provides the same functionality as a single exchange would. That is, is the SoS a refinement of the single exchange? If it is, then it has all the properties that the single exchange has: functionality, liveness, real-time performance, and so on. This can be checked using a refinement model checker. This kind of tool would explore all the behaviours of the SoS and make sure that they were all behaviours of the single exchange. For this to make sense, CML provides a way of hiding certain channel communications from view. For example, the communications between constituent nodes are not part of any behaviour in the single exchange.

A model checker has two very strong advantages: it is a push-button, fully automatic tool; and when the model checking fails (as is most often the case in debugging a specification), it produces an explicit counterexample that may be useful is working out why the refinement does not hold. But there is a drawback: there is a limit to the number of behaviours that can be checked in this way, and so the size of the CML models becomes an issue. This can be overcome by using a symbolic model checker that replaces specific values by symbolic names and proceeds to check whole classes of behaviours at once. Such tools can even check infinite behaviours. An alternative to model checking is theorem proving, which can in principle prove properties regardless of the number of behaviours.

4. Conclusions

Our main contribution in this work is in the design of a language specifically for modelling and reasoning about SoSs. It is suitable for describing both constituent systems and the architecture needed for composing them into an SoS. The formal basis of this language is suitable for building powerful analysis tools for verification and validation.

The complete development of CML—its syntax, semantics, refinement technique, connection to SysML, and tools platform—is the subject of ongoing work in COMPASS. At the moment, given the wealth of experience available in the use of VDM, CSP, Circus, and their semantics in the UTP, we are confident of the soundness of our approach. In particular, we can rely on the mature tools of VDM and CSP for restricted reasoning while the CML tools are under development [1].

References

[1] J. W. Coleman, A. K. Malmos, P. G. Larsen, J. Peleska, R. Hains, Z. Andrews, R. Payne, S. Foster, A. Miyazawa, C. Bertolinik, and A. Didier, "COMPASS Tool Vision for a System of Systems Collaborative Development Environment," in The 7th International Conference on System of System Engineering, IEEE SoSE 2012, July 2012.

[2] J. Holt and S. Perry, SysML for Systems Engineering. IET, 2008.

[3] J. Woodcock and A. Cavalcanti, "The semantics of Circus," in Proceedings of the 2nd International Conference of B and Z Users on Formal Specification and Development in Z and B, ZB '02, (London, UK, UK), pp. 184–203, Springer-Verlag, 2002.

[4] M. Spivey, The Z Notation – A Reference Manual (Second Edition). Prentice-Hall International, 1992.

[5] T. Hoare, Communication Sequential Processes. Englewood Cliffs, New Jersey 07632: Prentice-Hall International, 1985.

[6] J. S. Fitzgerald, P. G. Larsen, and M. Verhoef, "Vienna Development Method," Wiley Encyclopedia of Computer Science and Engineering, 2008. edited by Benjamin Wah, John Wiley & Sons, Inc.

[7] T. Hoare and H. Jifeng, Unifying Theories of Programming. Prentice Hall, April 1998.

[8] C. B. Nielsen and P. G. Larsen, "Extending VDM-RT to Enable the Formal Modelling of System of Systems," in The 7th International Conference on System of System Engineering, IEEE SoSE 2012, July 2012.

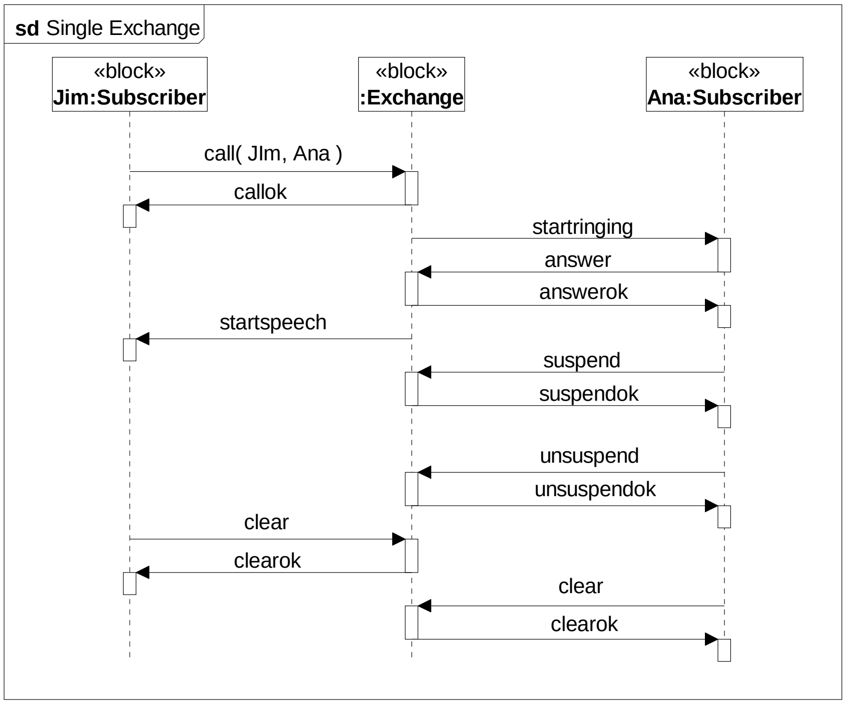


Figure 1. Telephone call in a single exchange.

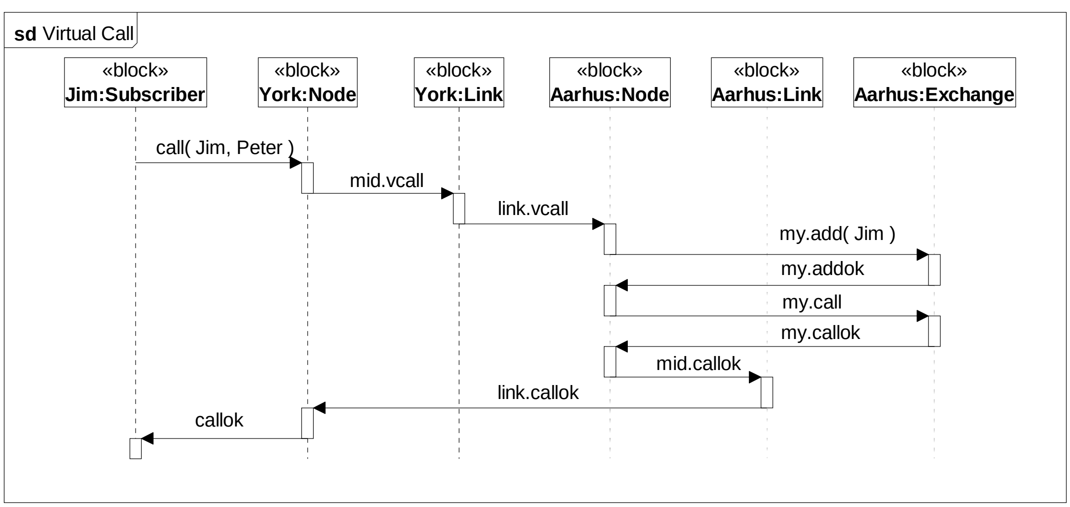


Figure 2. Virtual call.

COMPASS Tool Vision for a System of Systems Collaborative Development Environment

Joey W. Coleman*, Anders Kaels Malmos*, Peter Gorm Larsen*, Jan Peleska†, Ralph Hains‡, Zoe Andrews§, Richard Payne§, Simon Foster¶, Alvaro Miyazawa¶, Cristiano Bertolini‖, and André Didier‖

*Aarhus University School of Engineering, Denmark, {jwc,akm,pgl}@iha.dk
†Bremen University, Germany, jp@informatik.uni-bremen.de
‡Atego, United Kingdom, ralph.hains@atego.com
§Newcastle University, United Kingdom, {zoe.andrews,richard.payne}@newcastle.ac.uk
¶University of York, United Kingdom, {simon.foster,alvaro.miyazawa}@york.ac.uk
‖Universidade Federal de Pernambuco, Brasil, {cbertolini,alrd}@cin.ufpe.br

Abstract — *It would be useful to have a tool platform that supports systematic engineering of Systems of Systems, especially focused on the case where multiple parties are collaborating on the development. We attempt to provide a vision for a tool that allows collaboration on models developed jointly and that can be systematically analysed. The focus is on the challenges that make this kind of tool –a Collaborative Development Environment– different from a traditional Integrated Development Environment. Finally the paper describes the plans of the COMPASS project to address these challenges.*

Keywords: Systems of Systems, Collaborative Development Environments, COMPASS project, tool support

1. Introduction

Present in all fields of engineering is the need for modelling technology that can assist engineers with the exploration and evaluation of different design trade-offs early in the process. Without such technology the design decisions are made based on experience and intuition and this does not give a rigorous method when developing something new and complex. The field of System of Systems (SoS) engineering is in its infancy and is therefore lacking such technology. The goal of this paper is to present a vision of how such technology for SoS engineering could function.

The challenges in SoS engineering supersede those of systems engineering due to inherent characteristics that define an SoS. In [1] five characteristics are identified which distinguish a SoS from a mere system:

- operational independence,
- managerial independence,
- evolutionary development,
- emergent behaviour, and
- geographic distribution.

Each of these characteristics presents a set of challenges that must be addressed by a tool supporting collaborate development for SoSs.

Operational and managerial independence presents a great challenge for the tools to be developed, since there is no guarantee that a single organisation has full control of the development process. Further exacerbating this, the individual organisations developing the constituent systems are often sited at different locations and do not necessarily have any direct contact with each other.

This imposes the requirement that the tool be a Collaborative Development Environment (CDE) that supports models that are jointly produced and can still be systematically analysed even though a development team might not want (or be able) to disclose every part of their model [2]. The tools therefore need to support private and public parts of the model and a mechanism for extracting the public parts of the model which still should be usable for other organisations.

The paper will provide a vision of how this can be addressed, as this is a goal of the COMPASS project [3][1]. Here the intent is to produce a new modelling language called the COMPASS Modelling Language (CML) [4]. This language is intended as the first generic modelling language specially targeting the SoS context.

[1]COMPASS is an acronym for "Comprehensive Modelling for Advanced Systems of Systems" and more information about the project can be found at http://www.compass-research.eu/.

2. Modelling systems of systems

At an abstract level, an SoS model is just a collection of constituent systems and a description of the specified/desired SoS behaviour. Note that the behavioural specification is very likely to be *incomplete* — it seems doubtful that it is possible, in general, to give complete behavioural specifications for all cases.

Each constituent system must contain a specification of its relevant behaviour (relevant to the SoS), and enough state information to allow the constituent system to be verified, simulated, and so on. This internal state need not be exactly the same as the actual system specification, but it must represent a consistent abstraction of the specification.

Each constituent system must also, in principle, have a specification that is completely independent of the overall SoS. If it is not possible to create such a specification, then it may not be a system in its own right.

An SoS model should also be able to include infrastructure, modelled as constituent systems. These parts of the overall specification are used to either bind the component systems into the SoS, or to provide a bit of extra functionality beyond what is provided by any of the constituent systems. Full specifications of these constituent systems must be provided in detail to all members of the SoS.

3. Tool support

We propose three main areas for tool support of collaborative development of SoSs. At the high-level, a tool that supports an architectural notation like SysML is necessary to describe the abstract structure of the overall SoS and to clarify the desires, requirements, and capabilities of all of the constituent systems. At the detailed level, a tool that supports rigorous formal modelling of constituent systems provides a way of formally verifying the behaviour of constituent systems, and provides a concrete basis for the generation and verification of implemented systems. After the details are modelled, a tool to exercise the SoS model and build confidence and insight into its behaviour, doing so by providing comprehensive test automation for the constituent systems and their interaction in the overall SoS.

Note that this three-tiered setup is generally not present in IDEs, but becomes necessary for CDEs especially in situations where the development environment is a federation of standalone tools.

Aside from these three areas for tool support we also propose to enable the use of tool plugins at all tiers, though we will focus on the development of plugins at the detailed level.

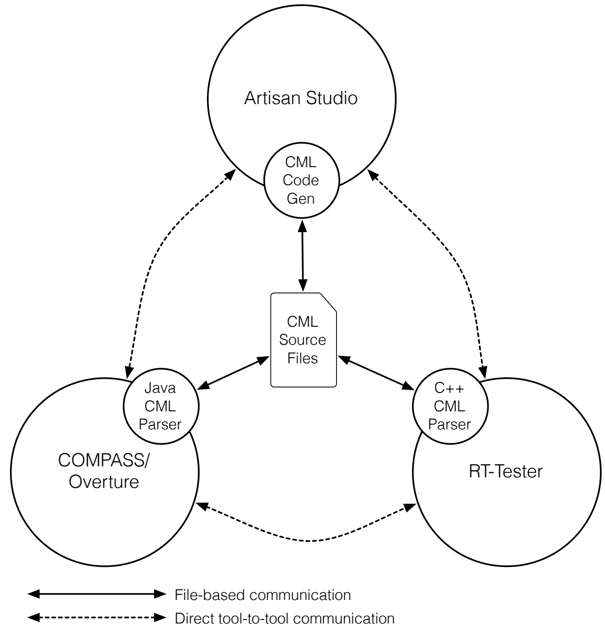


Fig. 1. Overview of the COMPASS Toolset

We also note that our proposed tool is, itself, an SoS in miniature. The three specific software packages detailed in the sections below are each an independent system in their own right.

The high-level overview of the proposed tool architecture is given in Figure 1. The figure shows the main three software packages that form the basis of the tool support proposed for SoS development, with the arrow indicating modes of communication between them. Solid arrows indicate communication paths between the packages by the use of flat files; all three packages will be able to read and write CML source files. There is also the potential for direct communication between the tools to allow for synchronisation of their user interfaces.

A. Architectural modelling in SysML with Artisan Studio

To meet the high level modelling need the COMPASS project will use Artisan Studio [5]. Studio already supports system and requirements modelling using SysML as well as software modelling using UML and code generation. COMPASS will use Studio's well established extension mechanisms to extend traditional Systems modelling as needed to model SoSs, and support the options to add any detail that may be needed to unambiguously generate a CML model. Studio's

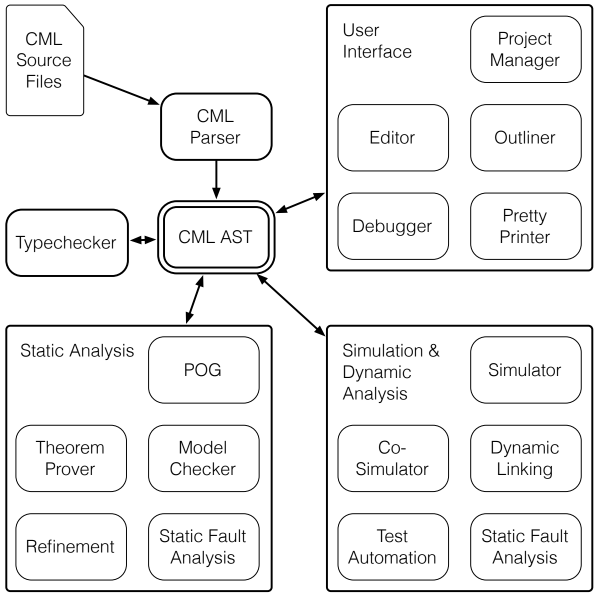


Fig. 2. Overview of the COMPASS/Overture components

TDK (Transformation Development Kit) will be used to develop a new code generator that will transform such an augmented SysML model into the textual format of a CML model, and also to roundtrip changes between the SysML and CML models.

B. Detailed SoS modelling in CML

To provide tooling for the detailed level, the COMPASS project is developing a language called CML, using the deep experience of its members in the VDM state-based formal method and the CSP/Circus process-based formal methods. CML will combine the best elements of these formal methods and be implemented on the Overture Tool platform [6] (itself based on the Eclipse tool platform).

We give an indication of the eventual set of components that will be a part of the COMPASS tool built on the Overture platform in Figure 2. Starting with CML source files, they are read by the parser generating an Abstract Syntax Tree (AST) of the detailed model. The AST is typechecked, ensuring a basic level of validity of the model. At this point the AST is available for use by the various plugins (described below): the usual user interface elements; various static analysis plugins; and various simulation and dynamic analysis plugins.

C. Test automation with RT-Tester

For test automation tasks the RT-Tester tool is used as a starting point; it will be enhanced with respect to SoS-specific test and simulation techniques during the COMPASS project. Using formal concurrent real-time models of expected System Under Test behaviour as input, RT-Tester identifies test cases, traces them to requirements and generates concrete test data by means of an integrated constraint solver [7]. As a crucial capability for SoS testing, the current RT-Tester version already supports model-in-the-loop tests (test data is exercised against a simulation of the model), software tests and hardware-in-the-loop tests, the latter being important for checking the appropriateness of the HW/SW integration.

D. Extending tool support with plugins

The software packages above will provide the core functionality required to support SoS development. However, large parts of the overall goals for tool support will be provided by plugins for the software packages. This allows us to implement the necessary functions in a modular manner and, where appropriate, make the functionality available for (re)use in other contexts as well.

1) Proof Obligation Generator: Proof obligations are properties which, when proven, demonstrate the consistency of a model. For a CML model to have a valid meaning, all such obligations must be proven. Proof obligations arise in several places in CML models, such as function/mapping domain checking and implicit operation satisfiability. A proof obligation generator plugin shall be developed for the COMPASS tool which analyses the CML model, producing a set of proof obligations to be verified. The proof obligation generator plugin of the Overture tool will be a basis for the development of the COMPASS plugin. The proof obligations generated by the plugin act as inputs to the theorem prover plugin — consisting of a set of hypotheses and a goal to be proven. The result from the theorem prover should be returned to the plugin which will inform a user of the correctness of the obligations.

2) Theorem Proving: Theorem proving support will be provided by connecting an existing theorem prover system to Overture and facilitating automated proofs about CML. Our tool of choice for proving theorems about an SoS is the interactive theorem prover *Isabelle/HOL* [8]. Isabelle is a well-established theorem prover with comprehensive support for describing and reasoning about mathematical models and as such has been used extensively in program verification. Isabelle provides powerful facilities for semi-automated reasoning based on higher-order unification. Furthermore, Isabelle has recently been extended with support for external automated first-order theorem provers, including Microsoft's powerful SMT solver *Z3* [9], through the

Sledgehammer [10] tool.

Our aim is to mechanise proof about CML by way of a denotational semantics based on Hoare's *Unifying Theories of Programming* (UTP). UTP provides a unified semantic basis for reasoning about a wide variety of formal models, including alphabetised relations, process calculus and object-oriented systems. We are therefore extending existing work on mechanising the UTP [11] with the aim of providing a comprehensive automated reasoning packages for CML. Our hope is that by making use of the various automated reasoning facilities which Isabelle provides we can verify the correctness of specifications with minimal effort from the user.

3) Model Checking: As CML is based on Circus and CSP, the COMPASS model checker will be based on refinement checking theory [12] of CSP and the Circus model checker [13]. It will be indeed a refinement checker. That is, instead of using some kind of temporal logic, two CML models will be compared according to the refinement theory developed for CML. In this way, we will be able to check SoS system designs evolution.

4) Model Refinement: Initially, tool support for model refinement will be provided in the form of a refinement editor for CML similar to that available for *Circus* [14]. The approach supported by the tool will be based on an algebraic refinement calculus for CML in the style of that in [15], which will be induced and justified by the UTP semantics of CML. With that result, a refinement editor can provide guidance for the compositional and stepwise refinement of CML models. The refinement editor will support isolation and identification of components, selection of applicable laws, transformation by law application, and generation of proof obligations to be discharged using the CML theorem prover. Soundness of the development will rely on the soundness of the refinement laws. Whilst the approach in [16] and [14] leaves this as open question, using the approach in [17], we can have assurance of soundness of laws built in on top of the mechanised semantics of CML.

Another interesting challenge will be the description of refinement laws in terms of SysML notation, for usability, allowing formal development at the graphical level and hiding the underlying CML models. This can be pursued once we have a clearer picture of the CML semantics of SysML, the opportunities to use CML inside SysML diagrams, and the connection between diagrams. Future work will also explore tactics of refinement as in [18].

5) Fault Modelling: To develop a dependable SoS it is important to understand how constituent systems fail, and the impact of such failures on the SoS as a whole. Numerous approaches exist for the analysis and development of dependable systems [19]. The primary aim of fault modelling in COMPASS is to support fault tree analysis (FTA) for SoS. FTA is an industry standard for deductive (top-down) fault analysis, which raises interesting challenges when applied to SoS. FTA aims to find the cause of undesirable events down to a very low level (e.g. transistor failures) where failure rates are known. However, in SoS the internals of constituent systems may be confidential, thus the occurrence of faults at the interface level must be determined (and if possible quantified) instead. The propagation (or tolerance) of errors between constituent systems must also be modelled to assess the impact on SoS level behaviour.

Due to the above challenges, we propose two levels of tool support for fault analysis. First an extension to Artisan Studio is proposed to model and analyse faults at the architectural (SoS) level. A SysML profile for fault modelling will be developed that guides the modelling process and enables a plug-in to Artisan Studio to generate fault trees. At the constituent system level, extensions to CML will be used to model the internal fault behaviour of constituent systems and determine the nature and frequency of faults at their interfaces. For this a plug-in to the COMPASS tool will be developed that generates fault trees from annotated CML models. The analysis of the fault trees will be performed by some external tool such as HiP-HOPS [20].

4. Challenges

Here we outline some specific challenges that we intend to address during the construction of our tool.

A. *Confidentiality during collaboration*

A challenge to the development of a CDE for SoS design involves modelling parts of constituent systems that are confidential. The confidentiality may arise for any number of reasons, including competitiveness (as some partners involved in an SoS design may also be market competitors), trade secrets, obligations to third parties, and so on. Because of this we may run into a situation where a partner in the SoS design has created a complete model of their constituent system, but is unable to share it with the other partners.

It is clear that, to test and verify an SoS design, we need to be able to assemble a model of the entire SoS and as such, proper testing becomes very difficult when missing a partner's model of their constituent system.

There are a few possible approaches to overcome this. First, it may be possible to relax the restrictions that prevent sharing; this requires non-technical intervention. Second, it may be possible to release an obfuscated version of the model. However, even were it to satisfy the initial confidentiality constraints, it depends on people being unwilling to expend the effort to reverse-engineer the model; assuming that it is impossible to reverse-engineer is not a safe assumption. Finally, it may be possible to create an alternate specification of the system that abstracts away from the confidential details and only reveals non-confidential behaviour.

We discard the first two approaches: the first is not of technical interest, and the second is not a general solution.

As a solution to this problem we advocate the replacement of concrete models by *contracts*, that is, abstracted specifications of the behaviour of constituent system that is visible to other constituent systems. The creation of such an abstract behavioural model of a constituent system for sharing among all SoS partners requires a few things. Both contracts and complete models must be formal: it must be possible to reason about the models in a rigorous way. The contract must be consistent with the complete model, and it must be provably so. This relies on the first requirement –that the models be formal– and means that any behaviour of concrete model that is visible to the other constituent systems must correspond to a behaviour exhibited by the contract.

An overview of this is given in Figure 3. The development of constituent systems relies upon the common published abstract models and both abstracts and publishes its own abstract model into the common repository. A given constituent system may depend on any of the published common models, and may use them to help verify its behaviour.

Technically, these requirements are feasible, and we plan to support them within the COMPASS tool. In the COMPASS methodology and tools we expect to build support for generation of CML models from SysML models, and then refinement of CML models. The largest questions involve the degree of automatic support for generation of abstract models, the degree of verification possible with abstract models, and how one would verify that confidential detail is not included in the abstract model.

B. Evolution of an SoS

Another challenge is how the tool supports change of the overall SoS structure that all the constituent systems rely on. This presents no challenges when developing a directed/acknowledged SoS. However, when developing a collaborative SoS there must be some

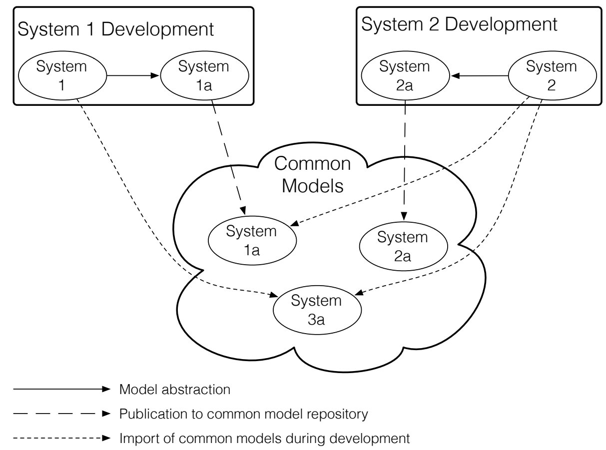


Fig. 3. COMPASS SoS development model

agreement between all the constituent system owners regarding proposed changes. Solutions to this issue can never be purely technical since they involve human interaction. The tool will therefore provide the necessary technical support for socio-technical processes that manage change to the SoS.

The support of changes in SoS requirements and architecture may even occur in a dynamic way during SoS operation. This may call for runtime testing of the novel SoS configuration, in parallel to its proper operation, and it will require dynamic adaptation of test goals: new requirements imply new test cases, while others may become invalid [21]. As a consequence we envisage the CDE to be an integral part of the SoS itself, supporting its life-long evolution.

C. SoS Test and Verification Strategies

The size of typical SoSs poses the general problem that verification and test methods suitable on component level will generally not scale up to SoS V&V. While, for example, it may be possible to cover all equivalence classes of an SoS component during HW/SW integration testing, it will certainly not even be possible to test all relevant combinations of such classes from each concurrent component during SoS system testing. This is the case due to the complexity of this task and it depends exponentially on the number of concurrent components involved. As a consequence, test methods requiring the generation of a global SoS model as the product of component models will generally fail when applied to SoS system testing: in [22] the authors state the hypothesis that even for functional system testing in the automotive domain, where the size of the model is far smaller than for SoSs in general, it will be infeasible to apply any test automation method relying on a global model representation as a product automaton or similar construct of component models. This problem is reflected in an analogous way in formal SoS verification by model checking: even with deeply abstracted component models, the combination of the latter in a global SoS model will be infeasible in most cases due to the size of the state space.

Our strategy to tackle this problem is to systematically incorporate knowledge about V&V results obtained for constituent systems and their components into SoS system-level verification and test campaigns. This strategy extends initial suggestions described in [23] and relies on stimulating pairs of SoS components during system-level tests having direct or indirect writer-reader relationships. In order to justify the strength of the strategy we currently investigate the conjecture that it will "converge" to a correctness proof when – hypothetically– tests are continuously refined by extending the combinations of sets of concurrent writers and associated readers involved in the tests.

5. Conclusions

We have described the goals we propose to address in the COMPASS project with respect to building methodology and tool support for designing and verifying SoSs. Relevant to this is its connection to CDEs, and how they must provide technical support for socio-technical processes, such as approval of changes that affect the entire SoS.

Acknowledgements

The work reported in this paper has been carried out in the COMPASS project which is funded by the European Commission 7th Framework Program.

References

[1] M. W. Maier, "Architecting Principles for Systems-of-Systems," *Systems Engineering*, vol. 1, no. 4, pp. 267–284, 1998.

[2] G. Booch and A. W. Brown, "Collaborative development environments," *Advances in Computers*, vol. 59, pp. 1–27, 2003.

[3] J. Fitzgerald, P. G. Larsen, and J. Woodcock, "Modelling and Analysis Technology for Systems of Systems Engineering: Research Challenges," in *INCOSE*, (Rome, Italy), July 2012.

[4] J. Woodcock, A. Cavalcanti, J. Fitzgerald, P. Larsen, A. Miyazawa, and S. Perry, "Features of CML: a Formal Modelling Language for Systems of Systems," in *The 7th International Conference on System of System Engineering*, July 2012.

[5] Atego, "Artisan Studio." http://www.atego.com/products/artisan-studio/, March 2012.

[6] P. G. Larsen, N. Battle, M. Ferreira, J. Fitzgerald, K. Lausdahl, and M. Verhoef, "The Overture Initiative – Integrating Tools for VDM," *ACM Software Engineering Notes*, vol. 35, January 2010.

[7] J. Peleska, E. Vorobev, and F. Lapschies, "Automated Test Case Generation with SMT-Solving and Abstract Interpretation," in *Nasa Formal Methods, Third International Symposium, NFM 2011* (M. Bobaru, K. Havelund, G. J. Holzmann, and R. Joshi, eds.), (Pasadena, CA, USA), pp. 298–312, NASA, Springer LNCS 6617, April 2011.

[8] L. C. Paulson, *Isabelle: A Generic Theorem Prover*. Springer-Verlag LNCS 828, 1994.

[9] L. De Moura and N. Bjørner, "Z3: an efficient SMT solver," in *Tools and algorithms for the construction and analysis of systems*, vol. 4963 of *LNCS*, pp. 337–340, Springer, 2008.

[10] S. Böhme and T. Nipkow, "Sledgehammer: Judgement day," in *Automated Reasoning* (J. Giesl and R. Hähnle, eds.), vol. 6173 of *LNCS*, pp. 107–121, Springer, 2010.

[11] M. Oliveira, A. Cavalcanti, and J. Woodcock, "Unifying theories in ProofPower-Z," in *Unifying Theories of Programming*, vol. 4010 of *LNCS*, pp. 123–140, Springer, 2007.

[12] A. W. Roscoe, C. A. R. Hoare, and R. Bird, *The Theory and Practice of Concurrency*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1997.

[13] J. C. P. Woodcock, A. L. C. Cavalcanti, and L. Freitas, "Operational Semantics for Model-Checking *Circus*," in *FM 2005: Formal Methods* (J. Fitzgerald, I. J. Hayes, and A. Tarlecki, eds.), vol. 3582 of *Lecture Notes in Computer Science*, pp. 237–252, Springer-Verlag, 2005.

[14] M. V. M. Oliveira, A. C. Gurgel, and C. G. de Castro, "CRefine: Support for the *Circus* Refinement Calculus," in *6th IEEE International Conferences on Software Engineering and Formal Methods* (A. Cerone and S. Gruner, eds.), pp. 281–290, IEEE Computer Society Press, 2008.

[15] A. L. C. Cavalcanti, A. C. A. Sampaio, and J. C. P. Woodcock, "A Refinement Strategy for *Circus*," *Formal Aspects of Computing*, vol. 15, no. 2–3, pp. 146–181, 2003.

[16] A. Freitas, C. Nascimento, and A. Cavalcanti, "A refinement tool for Z," in *Formal Methods and Software Engineering* (J. Dong and J. Woodcock, eds.), vol. 2885 of *Lecture Notes in Computer Science*, pp. 396–415, Springer Berlin/Heidelberg, 2003. 10.1007/978-3-540-39893-6_23.

[17] M. Oliveira, F. Zeyda, and A. Cavalcanti, "A tactic language for refinement of state-rich concurrent specifications," *Science of Computer Programming*, vol. 76, no. 9, pp. 792–833, 2011. Special issue based on the 2008 Refinement Workshop (Turku, 27 May 2008).

[18] M. V. M. Oliveira and A. L. C. Cavalcanti, "ArcAngelC: a refinement tactic language for *circus*," *Electronic Notes in Theoretical Computer Science*, vol. 214C, pp. 203–229, 2008.

[19] A. Avizienis, J.-C. Laprie, B. Randell, and C. Landwehr, "Basic Concepts and Taxonomy of Dependable and Secure Computing," *IEEE Transactions on Dependable and Secure Computing*, vol. 1, pp. 11–33, 2004.

[20] Y. Papadopoulos, M. Walker, D. Parker, E. Rüde, R. Hamann, A. Uhlig, U. Grätz, and R. Lien, "Engineering Failure Analysis and Design Optimisation with HiP-HOPS," *Engineering Failure Analysis*, vol. 18, pp. 590–608, 2011.

[21] A. Gonzalez, E. Piel, H.-G. Gross, and M. Glandrup, "Testing challenges of maritime safety and security systems-of-systems," in *Proceedings of the Testing: Academic & Industrial Conference — Practice and Research Techniques*, (Washington, DC, USA), pp. 35–39, IEEE Computer Society, 2008.

[22] J. Peleska, A. Honisch, F. Lapschies, H. Löding, H. Schmid, P. Smuda, E. Vorobev, and C. Zahlten, "A real-world benchmark model for testing concurrent real-time systems in the automotive domain," in *Testing Software and Systems. Proceedings of the 23rd IFIP WG 6.1 International Conference, ICTSS 2011* (B. Wolff and F. Zaidi, eds.), vol. 7019 of *LNCS*, (Heidelberg Dordrecht London New York), pp. 146–161, IFIP WG 6.1, Springer, November 2011.

[23] Q. Liang and S. H. Rubin, "Randomization for testing systems of systems," in *IEEE IRI 2009: Proceedings of the International Conference on Information Reuse & Integration*, pp. 110–114, IEEE, 2009.

Extending VDM-RT to Enable the Formal Modelling of System of Systems

Claus Ballegaard Nielsen* **and Peter Gorm Larsen***
*Aarhus University School of Engineering, Denmark, {clausbn,pgl}@iha.dk

Abstract — *When one wish to properly engineer dynamically evolving System of Systems (SoS) it is important to have tools that enable the different stakeholders to understand the consequences of diverse design decisions. Instead of merely using natural text and/or graphical diagrams we propose to make use of slight improvements to existing technology from the formal methods area to enable simulation of a model of the SoS. Here, we present how an extension of the tool support for the VDM Real Time (VDM-RT) technology enabling dynamic reconfiguration can form the basis for properly validating evolving SoS.*

Keywords: System of Systems, Formal Modelling, Dynamic Reconfiguration, Evolving Systems, VDM-RT

1. Introduction

The field of System of Systems Engineering (SoSE) is still in its early stages and it is apparent that the existing system engineering methods and tools have been challenged by some of the key SoS characteristics. Characteristics such as; independence, distribution, evolution and as an effect hereof emergent behaviour [10] is challenging to capture. Given that each independent constituent in the SoS is operational and managerial independent their behaviour is uncontrolled and challenging to predict during the design phase. Furthermore, as the SoS is always evolving the overall system architecture becomes difficult to anticipate. The lack of proper tool support induces uncertainty in the design and the development process, resulting in a decrease in the confidence that stakeholders normally receive from such tools. New methods and tools aimed at supporting the engineering of SoS are needed to explore and evaluate these characteristics during the early design phases.

One existing technique used for the engineering of complex systems is that of formal methods. Formal techniques have been used in the analysis and validation of software systems for over four decades [5] and it has been widely encourage and extensively researched in both academia as well as in industry [2], [19]. While formal methods have traditionally been used in connection with the design and development of complex critical systems their application in a SoS context is still in its infancy, although some initial advances have been performed [14], [1]. The foremost reason for this is possibly the youth of the entire SoSE field and accordingly the lack of tools to support the formal description and analysis of SoS. A great deal of the general research conducted within formal methods has revolved around verifying system behaviour with a focus on algorithms, state and data structures. However, being able to express algorithms and data structures is not sufficient for describing a SoS and consequently the existing formal modelling tools are inadequate for modelling SoS in their current state.

From a SoS point of view formal tools are needed that can express system architectures and at the same time incorporate SoS characteristics such as distribution and evolution. The tools that are particularly interesting in this context are those that support; (1) the object-modelling of distributed systems, (2) the interconnection between constituent systems and (3) the communication within the system

Some formal modelling techniques have been developed which allows for a stronger focus on the system architecture of distributed systems [6], [17]. They do however have limitations and syntactical expression that does not directly line up with SoS. In order to make the model more expressive and comprehensible the formal method must include notations and functionality that supports the language and technical fundamentals of the SoS engineering field.

An object-oriented formal technique such as VDM-RT (VDM Real-Time) gives the possibility of exploring and simulating the system design and behaviour of a distributed system by defining a system of CPUs and laying out the network topology connecting them [16]. Many types of distributed system architectures can naturally be described in VDM-RT by assuming a static set of hardware and software configurations. However, these static configuration sets become an issue when modelling systems with highly dynamic behaviour, such as found in SoS. Similarly the notations used in VDM-RT are focused on embedded systems and operate at the lower-level of CPUs and busses. In this paper we present an extension to VDM-RT which will enable the engineering of executable models of SoS expressed in a language that is closer to the SoS domain. The extension includes: (a) minor notational changes to the formal language in order to harmonize better with the SoSE field, and (b) it enables dynamic reconfiguration of the distributed system architecture meaning that network topology, number of channels and number of constituent systems can be adjusted during the run-time execution of a model. The intention is to enable a VDM-RT model to capture the fundamentals of dynamic reconfiguration in SoS and not to incorporated specific reconfiguration approaches. The goal is not to deliver completeness or proof of an entire system functionality or fulfilment of system requirements, but to aid system design decisions by creating an overview and identifying potential functionality and design flaws during the development phases.

The paper is organized as follows; Section 2 contains the extension of VDM-RT and introduces the notations and dynamic reconfiguration functionality enabling the modelling of SoS. A case study to which the extension is applied is described in Section 3 and the results of the extension are evaluated through a model of the case study in Section 4. Related work is addressed in Section 5 and finally future work and concluding remarks are presented in Section 6.

2. Modelling of Evolution in SoS

In this paper a SoS is considered as a collection of autonomous and independent constituent systems that themselves are distributed heterogeneous parts of an adaptive complex system, that achieve a higher functionality through the interaction between these constituents. While we acknowledge the managerial and socio-technical aspects of SoS these are not considered in this paper because of its technical focus.

System developers have been utilizing engineering techniques such as formal methods as part of their development process to build models that will enable the analysis and evaluation of overall system functionality. Using models of both existing systems and new systems enable system experimentation and verification without affecting real running systems. When creating models that can express SoS architecture and behaviour there are three characteristics which are the particular focus of this paper. First, the independence and autonomy of the constituents making it difficult to determine the joint behaviour of SoS. Secondly, the dynamicity which entail frequent changes in the system architecture with additions, exclusions and alternations of constituents systems and communication channels. Thirdly, the emergent properties of the SoS based on the synergistic collaboration between constituents. These are all characteristics that raise challenging demands for the models ability to express change and adaptability of the modelled system.

The Vienna Development Method (VDM) is one of the longest established model-oriented formal method techniques for the development of computer-based systems and software [3]. It has its foundation in the ISO standardized VDM Specification Language (VDM-SL). VDM-RT is a dialect of VDM-SL with a focus on modelling real-time distributed systems. A subset of VDM is executable, which allows models to run and simulate system behaviour. VDM is supported by an industry-strength tool set; VDMTools, owned and developed further by SCSK Systems [4] and an open source tool called Overture [8]. The VDM-RT extension being proposed in this paper has been implemented into a branch of the Overture tool Java based interpreter [9].

In current VDM-RT models, different software implementations encapsulated in application objects can be deployed and processed on specific simulated **CPU**s. Simulated **BUS** connections or networks can be defined between the different **CPU**s to enable communication between the distributed objects. When the model is executed the interpreter will do checks to ensure that all objects communicating are actually connected through a **BUS**, otherwise an error will be issued. Communication between the distributed software implementations is performed by remote operation invocations, meaning that data formatting and message transfer over the network is handled by the tool. A VDM-RT model has a global notion of time on which all processing depends, meaning that all systems and communication execute in parallel in relation to a global time. If application objects are deployed to different **CPU**s they will run concurrently, while if the application objects are deployed to the same **CPU** they will share processing time. Application objects can contain separate threads that will run on the constituent system that objects are deployed to.

The existing VDM-RT capability of describing static real-time systems must be extended to allow for the alternation of the overall system architecture during the run-time execution of a model. Enabling these dynamic changes allow models to express the dynamicity and evolution of SoS.

Three types of dynamic architectural changes are needed to enable the modelling of highly dynamic systems. The dynamic run-time changes include dynamic run-time: (1) addition and removal of new constituent systems, (2) addition and removal of communication channels, (3) changes of the network topology, meaning changing how constituent systems are connected to channels.

Besides the dynamic reconfiguration functionality, small notational adjustments are introduced into the VDM-RT language to bring the models closer to the SoS engineering field. The overall system architecture of the SoS is defined and encapsulated in a special **system** class. The **system** class can be given a specific name like any other class, but it is special as it is defined with the **system** keyword instead of **class**. The **system** class can contain instances of a **Constituent** type that represents constituent systems in the overall architecture and a **Channel** type that is used for establishing communication links between the constituent systems. **Constituent** and **Channels** are extensions of the **CPU** and **Bus** types described above. An instance of a **Constituent** is defined with a specific processing performance, in terms of executed instructions with relation to time, and multiple applications can be deployed and run in parallel on each **Constituent**. Instantiations of a **Channel** are defined with the specific channels transmission capacity, meaning the max. data amount that can be moved in a time unit. The constituent systems initially connected to a given channel can be defined by supplying a set of constituents during the creation of the channel. An example architecture using these components is illustrated in Figure 1.

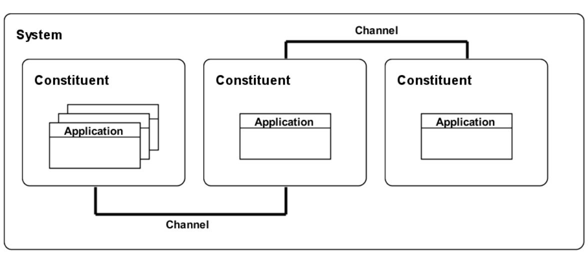


Figure 1: Example architecture using the introduced components

As a whole the overall structure of the SoS is defined in the **system** class that will both define the initial architecture of the system and also be used as a manager of dynamic changes. The dynamic changes to the system architecture will be performed through invocation of special reconfiguration operations defined in the **system** class.

The reconfiguration operations defined in the **system** class are listed in Table I.

Operation	**Description**
connect(Constituent, Channel)	Connects a Constituent with a channel.
disconnect(Constituent, Channel)	Disconnects a Constituent from channel.
addConstituent(Performance)	Adds a new Constituent with specific performance capabilities to the system.
removeConstituent(Constituent)	Removes an existing Constituent from the system.
addChannel(Capacity, Constituent)	Adds a new communication channel with a defined transmission capacity, and a set of Constituent's being connected to the channel.
removeChannel(Channel)	Removes an existing communication channel.

Table I: Dynamic reconfiguration operations

Besides operations for adding and removing **Constituent**s and **Channel**s during run-time, operations are defined for dynamically altering the network topology by connecting and disconnecting **Constituent**s to and from **Channel**s. The difference between disconnecting a **Constituent** from a **Channel** and removing a **Constituent** is that a disconnected **Constituent** still resides within the architecture and is still processing, but a **Constituent** that is removed is permanently deleted from the architecture and will cease to process. This allows the modelling of **Constituent**s that become temporarily isolated but still function locally. It should also be noted that a **Constituent** can be connected to multiple channels, so being disconnected from one channel does not necessarily mean that the system is isolated.

Changing the network topology or removing constituents and channels may result in disruptions of on-going communication, either by the connection being torn down or by a constituent leaving the network. In such a case all affected messages will be lost and thus introduces the risk of data loss. As VDM-RT has exception handling natively, exceptions will be used to report message loss. An exception is thrown by using the **exit** statement, which make the exception propagate upwards until an exception handler is encountered. An exception handler is defined by using the **trap** statement which wraps the statement in which the exception may occur. A run-time error will occur if an exception is not handled at all. When a constituent is removed all messages currently in transit will result in exceptions being raised at the senders of the messages. Any message sent from the removed system is lost without exceptions being raised in the removed system. When a channel is removed all messages in transit will be lost and exceptions will be raised at the senders of the messages. Constituents will not be notified of the channel being removed, unless they are communicating or attempt to communicate over the removed channel.

3. Case Study

To examine the effect of the extension a model of a case study system has been created. The case study revolves around a system named Vehicle Monitoring (VeMo) which is designed to improve road safety by increasing the traffic information available to motorists. The increased information flow is built on an intelligent traffic infrastructure along with open collaboration between motorists that join co-operative networks in which information can flow. Participants in the co-operative networks are called *VeMo entities*, which can be vehicles or traffic infrastructure.

Figure 2 depicts the basis of the system, which is a communication technique described as Inter-Vehicle Communication [15], in which a wireless network is used for communication as vehicles get physically close to each other or infrastructure such as intelligent traffic lights.

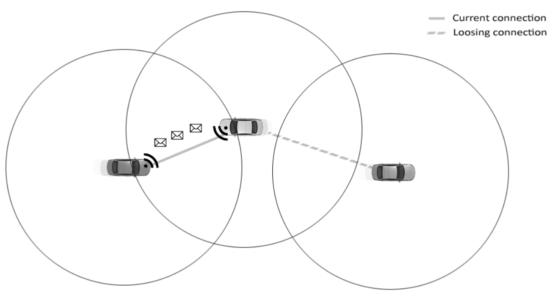


Figure 2: Communication and network between vehicles

The communication network is established rapidly and the information exchange occurs for as long as possible within the limited time frame where constituents are in range. The vehicles are a vital part of the system because it is their autonomous movements that make the system come to life; information will be spread out as vehicles pass other vehicles or infrastructure, which in turn will share the information with others.

The VeMo case study is a system in which the autonomous movement of vehicles results in high interoperability and changes in the relations between the constituents in the SoS. As the VeMo system does not control where the vehicles go, the information originating from a single geographical position can be spread out in an unpredictable fashion. This creates an unrestrained web of relations to a single event and as there is no central entity to control communication and information flow, each constituent in the system must be able to determine what information to share and what information to discard. The purpose of the case study is to examine how well the extension can be used to create a model of a SoS that is both ample and communicative and as a result can improve the reasoning and confidence in system behavior.

4. Modelling and Simulation

An executable VDM-RT model of the VeMo system which models the infrastructure, movement of vehicles, potential traffic hazards and the exchange of information has been created. The model simulates vehicles moving around in a grid-like system where they establish connections when they get in the vicinity of each other or infrastructure. Input events can be given to the model in order to simulate traffic hazards, which will make vehicles or infrastructure report traffic warnings.

While the constructed model itself has a considerable size, this section will only focus on the particular lines that relate to the proposed language extension (For the complete model please refer to [11]). The intent is to show how the extension can be used to model a SoS that evolve and change throughout its lifetime. The **system** class is named VeMo after the modelled system itself and thus it will be through this class that the dynamic reconfiguration operations are called. Listing 1 shows the relevant fractions of the VeMo class. The **system** keyword indicates that `VeMo` is our main system, and in its initial state this contains a single **Constituent** system. In this simple model the initial constituent is a traffic light for which the control software is define in the `TrafficLight` class. In the `VeMo` system constructor the traffic light software is deployed to the constituent system and a mapping is create between VeMo entities (`ve`) and constituents (`consti`) to keep track of which constituent system the software is deployed to.

```
system VeMo
instance variables
consti1 : Constituent := new Constituent(1E6);
tl1 : TrafficLight := new TrafficLight(999,
                        new Position(20, -70));
public ve2consti : inmap VeMoEntity to
                         Constituent := {|->};
operations
 public VeMo: () ==> VeMo
 VeMo() ==(
 consti1.deploy(tl1);
 ve2cons:= VeMo`ve2consti munion
 {tl1 |-> consti1} );
end VeMo
```

Listing 1: System class defining one constituent and the deployment of a traffic light application to it.

In this particular model an `Environment` class is used to handle changes in the system surroundings, such as adding or removing vehicles.

Listing 2 shows the segment that adds a new vehicle to the system.

```
class Environment
...
let constit= VeMo`addConstituent(1E6) in
let vehicle= new VehicleController(id, coords)
in (
 constit.deploy(vehicle);
 VeMo`ve2consti := ve2cons munion
         {vehicle |-> constit};
 start(vehicle));
...
```

Listing 2: Adding a new constituent and deploying the Vehicle VeMo Controller application to it.

In the `VeMo` **system** class the `addConstituent` operation is invoked which introduces a new constituent into the system, which is ready to operate immediately. The software used in the vehicle is instantiated and deployed to the constituent. Like in Listing 1 a mapping is create between the vehicle and the new constituent to keep track. Finally the vehicle software is started using the **start** keyword, making it start executing in the constituent system. Once a vehicle is added to the model it runs on its own until removed.

Vehicles are not connected to any communication channel when they are initially added to the model. To simulate the establishment and use of wireless communication between the vehicles and infrastructure their position on the grid is calculated and once in range the communication channel is set up, and reversely torn down once out of range. The fragment in Listing 3 shows how the proposed extension is used to establish a connection between two VeMo entities that have come in range of each other.

```
--once in range
let newChnl = VeMo`addChannel(1E6) in (
VeMo`connect(VeMo`ve2consti(entity),newChnl);
VeMo`ve2channel := ve2channel ++
                         {entity|-> newChnl}
VeMo`connect(VeMo`ve2consti(entity2),newChnl);
VeMo`ve2channel := ve2channel ++
                         {entity2|-> newChnl});
```

Listing 3: Creating new channel and using it to connect two VeMoUnits.

The `addChannel` operation is invoked on the `VeMo` **system** class with an argument indicating the capacity of the channel. Afterwards the `connect` operation is invoked for each of the constituents to connect them to the new channel. The `ve2consti` map in the listings above is used to find the constituent related to that entity, and likewise a mapping is created between the entity and the channel it is connected to.

To provide a better perception of the modelled system a graphical representation of the simulation has been built on top of the model using a technique that enables VDM to interact with Java [12]. A screenshoot of the illustration is shown in figure 3 where the connection and exchange of traffic information is illustrated.

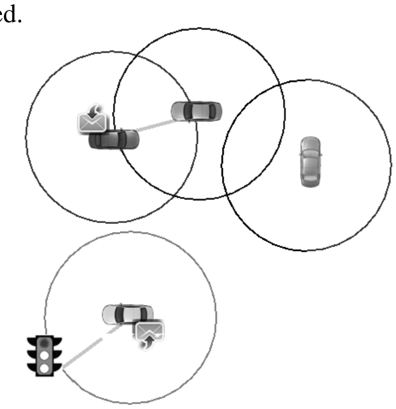


Figure 3: Simulation of the modelled SoS

Having the ability to simulate the model is a great advantage when investigating emergent behaviour as one can easier observe the results from the dynamic changes occurring in the evolving system.

5. Related Work

Kotov [7] presents the Communicating Structures Library (CSL) which is an object-oriented framework aimed at the modelling and analysis of SoS. In CSL system components are represented as *nodes*, which maintains a *memory* of *items*, the connection between nodes is represented by sets of *links*, denoted *nets*, and *items* acts as data traffic across the *nets*. Nodes can represent a range from microprocessors to computer clusters, and nets can represent anything from peer-to-peer network to busses. Procedures known as *Processes* are the drivers of the models and can send items between *Nodes*, while *transfer* functions determine how the transfer is performed. These building blocks enable models of SoS to be constructed and simulations of system behaviour, interaction, message flow and system scalability to be conducted. The elements presented in this paper have a close resemblance with the basic elements and approach presented by Kotov. The differences are in how the main processing and behaviour is enclosed in the systems themselves and there is a stronger focus on changing the topology dynamically than found in CSL. Oppositely, CSL provides interesting capabilities for compositional system structures through the use of subnodes, which cannot currently be performed in VDM-RT. Another difference is that CSL is based on C++ while our technique centres on a model-oriented formal method which entail different approaches in the development process.

Sahin et al. [13] presents a framework for the architectural representation and simulation of an SoS based on a combination between DEVS and XML. DEVS is a discrete-event formalism for building hierarchical systems by defining components and their interconnection, and simulating the interactions between them including the run-time addition/removal of couplings and components. DEVS represents the structure of a SoS, while an XML based data representation is used as a common language for wrapping the data communication between SoS constituents. DEVS and VDM are both capable of modelling discrete event systems but they use dissimilar approaches. DEVS has an origin in systems theory and has a strong focus on state transitions, where VDM comes from a formal methods origin with a Set-theoretic foundation.

6. Future Work and Concluding Remarks

This paper has proposed a language extension of VDM-RT which we believe enables the modelling of evolving SoS. Models have come closer to the SoS domain by making minor language adjustments and enabling dynamicity as constituents and channels can be added and removed during run-time. The presented work can be enhanced by including an automated approach for keeping track of the relations between deployed applications and the constituents, as well as the relations between constituents and channels. If this knowledge could be obtained via the language itself the task of manually creating mappings to keep track could be avoided. Additionally, it will be relevant to investigate how other types of SoS characteristics can be included into the language.

We believe that the experience reported here may be fed into the definition of the COMPASS Modelling Language (CML) [18]. From a tool perspective we believe that easy extensions enabling visualization of a SoS in operation may make it substantially easier to communicate the general design ideas behind the engineering of a SoS model between different stakeholders. We hope that this work can be a valuable start, enabling a generic solution for validating models of SoS and communication between different kinds of stakeholders such that a common understanding can be reached in an efficient manner.

References

[1] J.B. Caffall, D.S.; Michael. Formal methods in a system-of-systems development. In *Systems, Man and Cybernetics, 2005 IEEE International Conference on*, pages 1856–1863. IEEE, October 2005.

[2] Edmund M. Clarke and Jeannette M. Wing. Formal Methods: State of the Art and Future Directions. *ACM Computing Surveys*, 28(4):626–643, 1996.

[3] J. S. Fitzgerald, P. G. Larsen, and M. Verhoef. Vienna Development Method. *Wiley Encyclopedia of Computer Science and Engineering*, 2008. edited by Benjamin Wah, John Wiley & Sons, Inc.

[4] John Fitzgerald, Peter Gorm Larsen, and Shin Sahara. VDMTools: Advances in Support for Formal Modeling in VDM. *ACM Sigplan Notices*, 43(2):3–11, February 2008.

[5] R.W. Floyd. Assigning Meanings to Programs. In *the Symposium on Applied Mathematics*, volume 19, pages 19–32. American Mathematical Society, 1967.

[6] Einar Broch Johnsen, Reiner Hähnle, Jan Schäfer, Rudolf Schlatte, and Martin Steffen. ABS: A Core Language for Abstract Behavioral Specification. In *Proc. 9th Intl. Symp. on Formal Methods for Components and Objects (FMCO 2010)*, volume 6957 of *LNCS*, pages 142–164. Springer-Verlag, November 2011.

[7] Vadim. Kotov. Systems-of-systems as communicating structures. Technical Report HPL-97-124, Hewlett Packard Computer Systems Laboratory Paper, 1997.

[8] Peter Gorm Larsen, Nick Battle, Miguel Ferreira, John Fitzgerald, Kenneth Lausdahl, and Marcel Verhoef. The Overture Initiative – Integrating Tools for VDM. *ACM Software Engineering Notes*, 35(1), January 2010.

[9] Kenneth Lausdahl, Peter Gorm Larsen, and Nick Battle. A Deterministic Interpreter Simulating a Distributed Real Time System using VDM. In *ICFEM 2011*, October 2011.

[10] Mark W. Maier. Architecting Principles for Systems-of-Systems. *Systems Engineering*, 1(4):267–284, 1998.

[11] Claus Ballegaard Nielsen. Modelling Dynamic Topologies via Extensions of VDM-RT : A Case Study of an Evolving System. Technical Report ECE-TR-9, Aarhus University, School of Engineering, July 2012.

[12] Claus Ballegaard Nielsen, Kenneth Lausdahl, and Peter Gorm Larsen. Combining VDM with Executable Code. In *ABZ 2012, LNCS 7316*, pages 266–279, Heidelberg, 2012. Springer.

[13] Sahin, F. and Jamshidi, M. and Sridhar, P. A Discrete Event XML based Simulation Framework for System of Systems Architectures. In *System of Systems Engineering, 2007. SoSE '07. IEEE International Conference on*, April 2007.

[14] Scott A. Selberg and Mark A. Austin. Toward an Evolutionary System of Systems Architecture. Technical report, Institute for Systems Research., INCOSE, 2008.

[15] Pravin Varaiya. Smart Cars on Smart Roads. Problems of Control. *IEEE Transactions on Automatic Control*, 38(2):195–207, February 1993.

[16] Marcel Verhoef. *Modeling and Validating Distributed Embedded Real-Time Control Systems*. PhD thesis, Radboud University Nijmegen, 2009.

[17] Marcel Verhoef, Peter Gorm Larsen, and Jozef Hooman. Modeling and Validating Distributed Embedded Real-Time Systems with VDM++. In *FM 2006: Formal Methods*, Lecture Notes in Computer Science 4085, pages 147–162. Springer-Verlag, 2006.

[18] J. Woodcock, A. Cavalcanti, J. Fitzgerald, P. Larsen, A. Miyazawa, and S. Perry. Features of CML: a Formal Modelling Language for Systems of Systems. In *The 7th International Conference on System of System Engineering*, July 2012.

[19] Jim Woodcock, Peter Gorm Larsen, Juan Bicarregui, and John Fitzgerald. Formal Methods: Practice and Experience. *ACM Computing Surveys*, 41(4):1–36, October 2009.

Proc. of the 2012 7th International Conference on System of Systems Engineering, Genoa, Italy - 16-19 July 2012

System of Systems: "Defining the System of Interest"

Mr. Andrew Kinder
Lockheed Martin UK IS&S Ltd,
Chester House, Farnborough Aerospace Centre, GU14 6TD, UK.
Andrew.M.Kinder@lmco.com

Dr. Vishal Barot, Prof. Michael Henshaw, Dr. Carys Siemieniuch
School of Electronic, Electrical and Systems Engineering,
Engineering Systems of Systems (ESoS) Group
Loughborough University, LE11 3TU, Loughborough, UK.
V. Barot@lboro.ac.uk, M.J.d.Henshaw@lboro.ac.uk,
C.E. Siemieniuch@lboro.ac.uk

Abstract - *This paper proposes a multi-dimensional framework approach to defining the System of Interest (SOI) for System of Systems (SoS). A number of dimensions are described in detail, enabling development of a more comprehensive model that illustrates the relationships between these dimensions in SoS. The resultant framework model is evaluated using 2 practical case studies in the domains of Defence and ICT to indicate the applicability and suitability of the proposed approach. The potential benefits of the framework are described within the context of the case studies, in conjunction with the limitations of this initial framework. A brief description of intended further research is provided giving some insight into future development of the framework.*

Keywords: System of Systems, System of Interest, dimension, emergent behaviour, systems boundary, SoS lifecycle

1 Introduction

A distinguishing feature between Systems Engineering and System of Systems Engineering is the difficulty of defining the System of Interest (SOI) [1]. But failure to properly identify the SOI is one of the causes of unpredicted emergent behaviour which usually arises from the combined actions and interactions of the constituents of a SoS. A framework approach for comprehensive and sufficient definition of the systems boundary for SoS of all four types [2] is proposed in this paper. For a SoS the boundary is rarely static; this framework is intended to capture the dynamic nature of this by clearly defining the relationship between the SoS lifecycle and other dimensions. The definition of the SoS boundary will contribute to the ability to predict some aspects of emergent behaviour.

This article defines SOI through a number of dimensions; component systems (including specific and general system types), interactions (including types), lifecycle, variability, classification, functions, system owners and operations, concept of operations / use / employment, and nature of relationships. Using these dimensions a conceptual model has been constructed illustrating the relationships between them. The generic approach to SOI definition, and the associated model, enables a rich, dynamic definition of a SoS providing a common framework in which related SoS Engineering activities may be undertaken.

2 System of Interest (SOI)

The International Council on Systems Engineering (INCOSE) defines the SOI for a system [4] as; "*The system whose life cycle is under consideration*".

Wasson elaborates upon the definition [5]; *"The system consisting of a MISSION SYSTEM and its SUPPORT SYSTEM(s) assigned to perform a specific organizational mission and accomplish performance-based objective(s) within a specified time frame."*

ISO/IEC 15288 [7] takes a hierarchical view; *"the top system in the system structure is called a system-of-interest and consists of lower level systems. Except the lowest level is identified as being made up of system elements."*

Hitchens [8] provides an alternative definition; *"It [the SOI] contains intra-connected sub-systems, which are systems in their own right, existing within their own environment."*

Whilst these system level definitions can be interpreted with reference to the SoS SOI (particularly with reference to multiple component systems and system lifecycle), they do not fully address the relationship between emergent behaviour and the SOI. An extension to the INCOSE SOI definition is therefore proposed;

SoS SOI: *The system of systems whose life cycle is under consideration described by all dimensions that contribute to the resultant emergent behaviour.*

3 The SoS SOI Dimensions

The challenge to be addressed is to identify the relevant parameters through which a SoS is adequately defined. Whilst trivial to state, the complexity of many SoS make this challenge a remarkably difficult undertaking and it is not clear that there is a generalised approach that may be applied to all SoS. To attempt the derivation of a generalised approach, this research has simultaneously used a top down approach, in which the many definitions of SoS [3] have been analysed, and a bottom up approach based on analysis of specific SoS case studies. The proposed SoS SOI definition makes reference to dimensions; these are now described in more detail, allowing the initial high level statement to be developed into a more comprehensive model.

3.1 Component Systems (including specific and general system types)

By definition, specific component systems (also known as constituent systems) will be identified for the directed SoS ("*those in which the integrated system-of-systems is built and managed to fulfill specific purposes"*) [6] and to some extent the acknowledged SoS ("*the constituent systems retain their independent ownership, objectives, funding, and development and sustainment approaches")* [6] earlier in the SoS lifecycle than those classified as collaborative (*the component systems interact more or less voluntarily to fulfill agreed upon central purposes)* [6] and virtual (*lack a central management authority and a centrally agreed upon purpose for the system-of-systems*) [6], however, it is possible that generic system types may be identified for all classifications. Indeed, for a collection of systems collaborating on an unplanned basis, it may not be possible to identify all specific systems at any point in time but in order for the systems to interact there must be a certain commonality between them allowing an abstracted general system type to be inferred. The level of abstraction is dependent on the scope of the commonality and the system purpose. The "*Component Systems*" dimension captures this distinction by allowing the definition of both specific systems and generic system types.

The generic system types are dependent upon SoS under consideration; e.g. for an SoS concerned with the processing and management of information [9] the generic component system types may include;

- Management information system.
- Executive information system.
- Executive support system.
- Decision support system.
- Group decision support system.
- Electronic meeting system.
- Organizational decision support system.
- Expert system.
- Office information system.
- Intelligent organizational information system.

3.2 Classification

The SoS classifications [6]; *directed, acknowledged, collaborative and virtual*, are used to populate this particular dimension. However, a particular SoS may exhibit the behaviours of more than one classification depending on the make-up of component systems and the current point in the lifecycle. As shown in the NHS case study (section 5.2), the classification may evolve over time.

3.3 Interactions (including types)

A SoS exists only because of interactions between constituent systems. With no interaction the SoS merely becomes a set of independent systems exhibiting no overall emergent behaviour, so the interactions must provide some definition of the SoS SOI.

The interactions may be initially defined at a generic level and subsequently at a more specific level which will identify both the interaction medium and the nature of the interaction. This allows a typical top down approach to be taken, progressing from the abstract to the specific where interaction contents, protocol and media are identified. Considering a financial SoS, a requirement to exchange customer bank account details may exist between component systems, this provides the general interaction type, i.e. "Customer Account Information". As the SoS is developed design decisions are taken which elaborate upon this general type until a point is reached whereby the precise data contents and format, communication protocol and medium are identified.

The interaction type will therefore be influenced by the requirements derived from the SoS purpose. Considering another example; a SoS whose purpose is to perform a Search and Rescue mission will require a combination of surveillance and co-ordination information exchanged between component systems. In a military environment this may be accomplished through the utilisation of Tactical Data Links (TDL) and voice communications. The SOI description may then be further refined to identify specific TDLs (e.g. Link 11 [10] or Link16 [11]) and voice communication frequencies enabling interoperability between component systems. The ability of the SoS to fulfil its purpose is dependent upon the effectiveness and availability of these interactions. Using the search and rescue case it is clear that the performance will be compromised if communications are degraded because component systems are not interoperable. Simplistically, this may occur if correct voice frequencies are not used or platforms have incompatible TDL fits with no "gateway" to link them, compromising the exchange of situational awareness data.

The availability of interactions is important and related to performance and agility, i.e. a SoS may be formed or reconfigured relatively quickly if component systems share a common means to interact. For two component systems to interact there must be at least one homogeneous interaction medium, if there is none then an intermediate system must act as a "translator" or gateway. Interactions can be also constrained by geographical dispersion, therefore this should be considered.

3.4 Nature of Relationships

The Nature of Relationships dimension is distinct from interactions (section 3.3) which are more concerned with the inter-connections between systems. This dimension is used to define the category of relationship between component systems, such as 'peer-to-peer', hierarchical control or distributed control.

3.5 Lifecycle

This dimension considers the lifecycle of the SoS as a whole as well as the respective timelines of the component systems, providing an insight into dynamism within the SOI. For some SoS, it may not be possible to define a *cycle*, as such, but rather an evolution in which the SoS passes through identifiable phases.

ISO 15288 [7] considers the lifecycle of single systems, it does allude to the lifecycle of “enabling systems” (e.g. a concept system, development system, production system, utilization system, support system or retirement system [7]) but these are distinct from SoS component systems. It lists typical lifecycle stages applicable to a single system as;

- Concept stage - Development stage - Production stage - Utilization stage - Support stage - Retirement stage

A directed SoS which is “*built and managed*”[6], and therefore developed in a more structured manner, aligns more readily to these suggested stages with the traditional System Engineering lifecycle being applicable to each component system. As the SoS classification moves away from the more managed type and to the ad-hoc type these stages become far less applicable. Alternative stages may include;

- Planning - Convergence - Collaboration - Execution - Evolution - Dispersion

However, where the SoS has no agreed central purpose and behaviour emerges through a process of self-discovery among the component systems (a virtual SoS [6]) the lifecycle becomes more an observation of the evolution of the SoS rather than an imposed structure permitting a degree of management.

It is also established that there is a link between the point in the lifecycle and the ability to populate a dimension. Table 1 shows the population of the dimensions for two case studies. For example, until the execution phase of a SoS the actual component systems may not be known but the types should be. Consider a military SoS whose purpose is to recover a downed pilot; it will require a surveillance component system. There are several possible systems that could fulfil this role, depending on suitability and availability this will be instantiated with a specific type when the SoS is executing its task, a potential type being an E3-D Sentry aircraft. It is argued that the lifecycle is an overarching dimension, affecting all other dimensions. As the SoS passes through the lifecycle phases the dimensions evolve both in relation to each other and to external influences.

3.6 Variability

The frequency of change of a SoS is considered to be a critical dimension. This is linked to stability and also agility. Variability is related to time in the sense that it may reveal some sort of characteristic frequency representing change in the SoS. Stability is linked to the behaviours of the SOI such as “Stable over time”, “Unstable, rapidly grows” and “Boundary changes”. Factors such as an evolving purpose or a response to environmental factors will influence the variability.

3.7 Functions

A SoS performs a combination of functions in order to accomplish its purpose. Functions may be performed by individual component systems, by subsets of component systems or by the entire SoS. The Function dimension defines both specific functions and more generic function types. The aggregate effect of the functions gives the SoS emergent behaviour, it is this behaviour that should fulfil the SoS purpose. Analysis of these functions may enable unexpected behaviour to be anticipated.

3.8 Systems Owners and Operations

This dimension is closely related to organisations, management and enterprise (multi-organisational) relationships. It is dependent on the classification type, for example within a directed SoS the owners and operations will be well defined but for a more ad-hoc SoS this dimension will not be so well defined. However, the definition of System Owners and Operations includes generic types enabling the dimension to be defined to a certain extent for less rigidly defined systems.

3.9 Concept of Operation / Use / Employment

Given that the individual systems can perform operations independently the use/mission of the SoS is critical for defining the SOI for the SoS as opposed to the individual constituent systems. This dimension is potentially the starting point for populating the SOI dimensions.

4 Relationships between Dimensions

The SOI dimensions should not be considered in isolation, Figure 1 illustrates the relationships between the dimensions. It is the combination of the identification of the individual dimensions and the relationships that gives this perspective its inherent power. There are numerous, well documented, accounts of SoS failures[1], with a common theme of systems and/or dimensions being developed independently using traditional SE techniques. The SOI provides an holistic view of the SoS rather than a set of unconnected viewpoints.

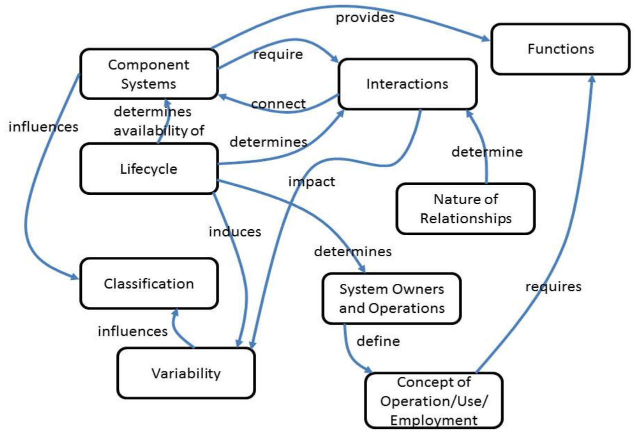


Figure 1: Dimension Relationships

5 Case Study Evaluation

Two case studies in the domains of defence and ICT have been initially selected to illustrate the applicability and significance of the proposed framework. These case studies are "*Counter Air Mission*" and "*National Programme for IT in the NHS (NPfIT)*". Each case study has been analyzed in terms of the SOI dimensions, Table 1 presents the results.

5.1 Counter Air Mission

This case study uses a typical example of a military Counter Air Mission [12], as shown in Figure 2, the purpose of which is to provide protection from airborne threats, either offensively or defensively. In this case the SoS will be required to perform offensive pre-emptive strikes against known hostile airborne assets as well reacting defensively to unpredictable airborne attack. The component systems exchange tactical information through a combination of voice and tactical data links (TDL), both Link 11 [10] and Link 16 [11], utilizing radio and satellite communications with gateways to extend range and provide interoperability.

[1] E.g. the National Programme for IT in the NHS which is examined as a case study in this paper.

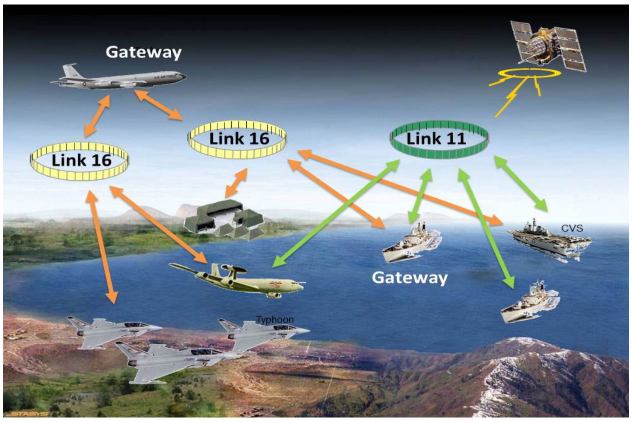


Figure 2 – Typical Counter Air Mission

5.2 National Programme for IT in the NHS (NPfIT)

The overall purpose of the NPfIT system [13], as shown in Figure 3, was to use Information Technology (IT) to help deliver better patient care by moving towards a single, centrally-mandated electronic care record for patients and to connect 30,000 GPs to 300 hospitals, providing secure and audited access to these records by authorized health professionals. The Core of the programme was going to be NHS Care Records Service (i.e. "the Spine"), which would have made relevant parts of a patient's clinical record available to whoever needs it to care for the patient. However, progress has been poor with completion date for some component systems now forecast to be 2014-15, at best. Other component systems are in use but only operating independently. Overall, this is an excellent example of SoS failure.

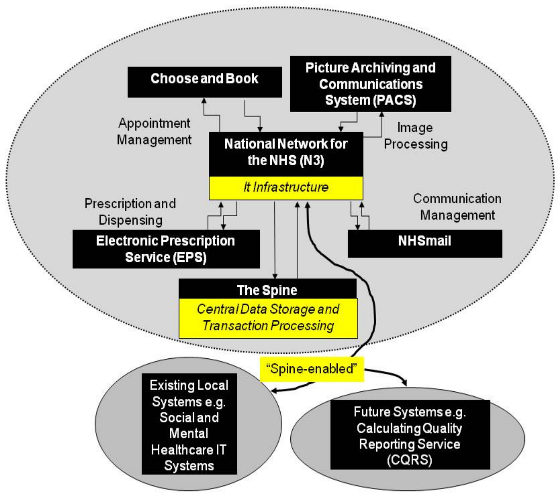


Figure 3 – National Programme for IT in the NHS

Dimensions	Case Studies	
	Counter Air Mission	**National Programme for IT in the NHS (NPfIT)**
Component Systems	Counter Air Aircraft (Typhoons) Command and Control Unit (E3-D, TACC) JRE Gateway (KC-135) TDL gateway (Type 45) Air Defence Artillery (Type 45) ISR (E3-D) CVS Aircraft Carrier Satellite (Skynet 5)	Choose and Book Electronic Prescription Service (EPS) National Network for the NHS (N3) NHS Mail Picture Archiving and Communication System (PACS) The Spine
Interactions	Digital Communications (Link 11/16, Satellite) Voice Communications Track Data Command and Control Imagery	N3 Network eMAIL Appointment Data Image Processing Patient records Prescriptions (electronic)
Lifecycle	Requirement, Planning, Assembly, Execution Dispersion	Concept, Design and Development , Operation, Reconfigure/Upgrade, Disposal
Variability	Assets may change during mission. The purpose may evolve as the mission progresses. This SoS may exhibit a high level of variability.	Intuitively one assumes variability is low because of the directed nature of this SoS. However, due to the poor definition of requirements component systems were frequently changed which resulted in a higher level of variability than would perhaps be expected.
Classification	Collaborative/Acknowledged. Directed elements, e.g. Link 11/16 networks	Originally Directed but later moved towards Acknowledged.
Functions	Surveillance, Reconnaissance, Targeting, Detect , Identify, Intercept, Destroy, Aircraft Control, Battle Management.	Choose and Book appointment, Change appointment, Produce prescription, Picture Archiving
Systems Owners and Operations	Nations (Governments), Services (Air force, Navy, Army), Overarching control, e.g. NATO.	Project Team, NHS Connecting for Health Local Authority / Strategic Health Authority Project Board, Board of Directors / PCT Executive Group
Concept of Operation	The SoS provides protection from air and missile threats.	Provide a single, centrally-mandated electronic care record for patients and to connect 30,000 GPs to 300 hospitals.
Nature of Relationships	Hierarchical (military C2)	Plural system of procurement. Client-Server.

Table 1: Mapping of Case Study Data to SOI Dimensions

6 Benefits of the Proposed SOI Framework

Having shown that the dimensional framework “fits” selected SoS case studies, we examine some of the benefits that application of this framework could bring in each case.

Regarding the first case study, a common problem that often besets military operations is a lack of interoperability between component systems. The proposed framework shows the relationship between component systems and interactions, emphasising the need for interoperability. Whilst this in itself can already be determined with current frameworks (e.g. MODAF (Ministry of Defence Architecture Framework)), the proposed framework allows further elaboration of each system, in accordance with the life cycle and the SoS purpose and shows the relationships to other aspects of the SoS.

The second case study regarding the National IT Programme for the NHS has not, as yet, entirely fulfilled its purpose for a number of reasons, some of which are;

a. Those closest to the healthcare delivery were not able to provide adequate input into the SoS requirements definition, resulting in specific local requirements not being met. The SoS SOI links Lifecycle, System Owners and Operations, Concept of Operation / Use / Employment and Functions.

This relationship shows that system owners indirectly influence the functions provided by the SoS. By definition all health trusts (System Owners) are part of the SOI and, therefore, in accordance with the proposed model, would be provided with input into the definition of Lifecycle, System Owners and Operations, Concept of Operation / Use / Employment.

b. Incompatibility between systems. For example, a Patient Administration System in a particular NHS trust was delayed when found to be incompatible with the Choose and Book System. The SoS SOI links Component Systems, Interactions and Lifecycle. The SOI allows component system types and interaction types to be defined early in the lifecycle and, if correctly modelled, should enable any incompatibilities to be detected and prevented. The relationship of Lifecycle in this case is intended to enable component systems at different phases of development to form part of the SOI.

c. A common issue was that the NPfIT system was implemented with insufficient flexibility in terms of range of locations from where data could be collected. As described in response to the issue ‘a’ above, the SOI should ensure that all System Owners have sufficient input to the system definition, allowing a greater range of locations to be specified. In addition, the interactions may identify a geographical element.

7 Limitations

The proposed framework has been exercised on a small number of case studies, two of which are given by example herein. Generalisability of the framework approach has clearly not been demonstrated in the work conducted so far. However, the purpose of this activity is to provide a practical and pragmatic approach for SoS Engineers that will enable such activities as modelling and simulation to be conducted with greater confidence. To this extent, the next stages of the research will be to formalise the approach and to exercise it on a wider set of case studies.

8 Conclusions

This paper shows the benefit of a multi-dimensional SoS SOI description for enabling the impact of inter-dependencies between the dimensions of a SoS to be considered through all phases of the SoS lifecycle. The use of factual case studies is used to assess the applicability of the dimensions beyond a theoretical environment into practice. Further research is planned to formalize the approach as a tool for SoS modelers as a means of enabling greater confidence in the application of modeling and simulation techniques for SoS.

9 Acknowledgements

The authors would like to acknowledge EU T-AREA-SoS FP7 support action (Grant agreement no: 287593), the Loughborough University Engineering SoS group and other collaborative experts for their kind support.

References

[1] Bjelkemyr, M., Semere, D.T., Lindberg, B. (2009) Definitions, classification, and methodological issues of systems of systems, ch. 7 in Systems of Systems Engineering Principles and Applications, Jamshidi, M.

[2] Dahmann, J., Baldwin, K.J. and Rebovich, G. (2009), Systems of Systems and Net-Centric Enterprise Systems, 7th ann. Conf. on Sys. Eng. Res., Loughborough.

[3] Jamshidi, M. (2005), Theme of the IEEE SMC 2005, Waikoloa Hawaii, UUSA, http://ieeesmc2005.unm.edu/

[4] INCOSE Systems Engineering Handbook v. 3.2.1, January 2011.

[5] System Analysis, Design, and Development Concepts, Principles, and Practices, Charles S. Wasson, John Wiley & Sons, Jan 2006.

[6] Dahmann, J., and K. Baldwin. 2008. "Understanding the Current State of US Defense Systems of Systems and the Implications for Systems Engineering." Paper presented at IEEE Systems Conference, 7-10 April, Montreal, Canada

[7] ISO/IEC 15288:2008 System life cycle processes.

[8] Basic Models for System Thinking, Hitchins, D, http://www.hitchins.net/SysMods.html.

[9] G Mentzas, “A Functional Taxonomy of Computer-based Information Systems”, *International Journal of Information Management* (1994), 14 (397-410).

[10] NATO Standardization Agreement TDL 5511 Ed 6.

[11] NATO Standardization Agreement TDL 5516 Ed 6.

[12] Counterair Operations, Air Force Doctrine Document 21.1, October 2008, http://www.dtic.mil/doctrine/jel/ service_pubs/ afdd2_1_1.pdf

[13] NHS The National Programme for IT in the NHS: an update on the delivery of detailed care records systems, Report by the Comptroller and Auditor General HC 888, Session 2010–2012, 18 may 2011

Proc. of the 2012 7th International Conference on System of Systems Engineering, Genoa, Italy - 16-19 July 2012

Contract-based Modelling and Analysis Technologies for Systems-of-Systems

Steve Riddle
School of Computing Science, Newcastle University
Newcastle upon Tyne, NE1 7RU, United Kingdom
steve.riddle@ncl.ac.uk

Abstract – *Despite the opportunities offered by Systems-of-Systems engineering, there is a lack of models and tools to support dependable decision-making during design and evolution. Such models should combine the rigour needed for reliable analysis with the abstractions needed to describe heterogeneous and independent constituent systems, and the contracts between them. Models should also provide a basis for communication between their diverse stakeholders. Suitable models and tools have the potential to enable the analysis of local and global properties by a range of techniques including simulation and formal verification. The papers in this invited session will provide participants with an understanding of the state of the art and the research directions in rigorous model-based approaches to SoS engineering, through the support of COMPASS, the European Commission's FP7 project on "Comprehensive Modelling and Analysis for Advanced Systems of Systems".*

Keywords: SoS, Models, Tools, Formal Languages, Requirements, Architecture, Contracts, Design, Analysis.

1 Introduction

This invited session at SOSE 2012 provides an opportunity for discussion of the potential, and challenges, of formal model-based approaches to the model-based engineering of Systems-of-Systems. SoS engineering suffers from a lack of suitable models and tools to support dependable decision-making during design and evolution. We see a need for models which combine the *rigour* needed for reliable analysis with the *abstractions* needed to describe heterogeneous and independent constituent systems, together with the contracts defined between their interfaces. Such models would provide a basis for communication between the diverse stakeholders, and have the potential to enable the analysis of local and global properties by a range of techniques including simulation and formal verification.

This session will provide participants with an understanding of the state of the art, and the research directions, in rigorous model-based approaches to SoS engineering. The topic is new and timely because advances in formal methods and tools are only now beginning to make it possible to contemplate trade-off analysis at the SoS level.

2 The COMPASS project

The session is supported by COMPASS[1], the European Commission's FP7 project on "Comprehensive Modelling and Analysis for Advanced Systems of Systems". COMPASS seeks to combine and extend existing modelling techniques for SoS, through a new purpose-built modelling formalism called CML (COMPASS Modelling Language). We will develop methods and tools that take advantage of the formal semantic foundations of CML, allowing developers to choose from different levels of description.

As a starting point, a graphical architectural view in SysML is easy for most stakeholders to understand. We will link SysML to CML, extending it with SoS-specific features to describe the assumptions and guarantees of constituent systems. This extended SysML will have semantics in pure CML, which can then be readily processed by static analysis tools including theorem provers and model checkers, allowing automated detection of inconsistencies, and of potential deviation from contract conformance. The CML representation can also form the basis of test generation and simulation in demanding operational scenarios.

Tool support will allow users to operate at the SysML level, or at the CML level, or both, with a tool set that can be extended with CML plugins for static fault analysis, model-checking, theorem-proving, test automation, and support for run-time contract checking. The open platform allows links to a range of architectural modelling tools to be developed.

3 Session papers

The session papers reflect the COMPASS project's agenda, covering: engineering of SoS requirements; architectural concerns such as interface contracts; the features of the formal language CML being designed for SoS modelling and analysis, and the vision for tool support for SoS development. A brief overview of the session papers is given below.

1. *Technical Challenges of SoS Requirements Engineering.* Stefan Hallerstede, Finn Overgaard

[1] www.compass-research.eu

Hansen, Jon Holt, Rasmus Lauritsen, Lasse Lorenzen, Jan Peleska.

Taking first the challenges of SoS requirements, this paper looks specifically at problems that appear in SoS engineering. One example problem is the need to deal with requirements validation and verification without compromising the continued evolution of the SoS, a challenge which requires strong support for requirements traceability.

2. *Model-based Requirements Engineering for System of Systems.* Jon Holt, Simon Perry, Mike Brownsword, Daniela Cancila, Stefan Hallerstede, Finn Overgaard.

   Model-Based Systems Engineering (MBSE) is a well understood technique at the system level, but there is a lack of application to the SoS level. This paper looks at an established approach to applying MBSE for requirements engineering, and assesses the suitability for its application at a SoS level.

3. *Interface Specification for System-of-System Architectures.* Richard Payne, Jeremy Bryans, John Fitzgerald, Steve Riddle.

   This paper examines the issues involved in establishing that a SoS architecture respects global SoS-level properties, through the use of interface contracts defined at the boundaries of constituent systems. Current architectural notations have only limited support for description of such interfaces. By combining a common standard architectural notation (SysML) with the formal specification language VDM, a rigorous approach to defining interface specifications can be developed. This paper identifies the research challenges involved, and illustrates the approach with a case study based on an emergency services SoS.

4. *Features of CML: a Formal Modelling Language for Systems of Systems.* Jim Woodcock, Ana Cavalcanti, John Fitzgerald, Peter Gorm Larsen, Alvaro Miyazawa, Simon Perry.

   To formally model SoS, the next step after investigating combinations of VDM and SysML is to develop a new purpose-built formalism, the COMPASS Modelling Language (CML). CML is founded on VDM and Circus, which together provide the necessary vocabulary to describe processes with state and operations that can interact with an environment via synchronous communications. This paper describes the CML language which will be integrated, in the COMPASS approach, with SysML so that developers can work with either formal text or with a graphical architectural view that is more readily communicated to stakeholders.

5. *Extending VDM-RT to Enable the Formal Modelling of System of Systems.* Claus Ballegård Nielsen, Peter Gorm Larsen.

   The final two papers consider tool support for the COMPASS approach. Tools must enable stakeholders to understand the consequences of diverse design decisions, and this paper describes the use of model simulation to support this understanding. An extension of the tool support for the VDM Real Time (VDM-RT) technology is described, with the aim of validating an evolving SoS.

6. *COMPASS Tool Vision for a System of Systems Collaborative Development Environment.* Joey W. Coleman, Anders Kaels Malmos, Peter Gorm Larsen, Jan Peleska, Ralph Hains, Zoe Andrews, Richard Payne, Simon Foster, Alvaro Miyazawa, Cristiano Bertolini, André Didier.

   The characteristics of SoS present unique challenges for a tool that is to support collaborative SoS development. With no guarantee that a single organisation has full control of the development process, and geographically dispersed organisations developing the constituent systems, a Collaborative Development Environment must be able to support joint production and analysis of models while not forcing teams to disclose every part of their model.

4 Conclusion and Future

This invited session provides an overview of the technological approach and early results in the COMPASS project. Further developments in languages and tools will be evaluated through a number of industrial case studies. In each study, a SoS development problem is addressed first using current best practice and then using SysML+CML. The main studies are an accident response SoS (for dynamic coordination of diverse healthcare services in an acute emergency) and an audio/video/home automation ecosystem (managing content and applications from diverse sources). A number of challenge problems are also being proposed by members of the COMPASS Interest Group (CIG), to stretch the formalism in a wider range of sectors.

5 Acknowledgements and partners

The COMPASS consortium is funded by the EU FP7 COMPASS project under grant agreement no.287829. Partners in the project are Newcastle University, Aarhus University, University of York, Bremen University, Universidade Federal de Pernambuco, Bang & Olufsen, Insiel and Atego.

Influence Strategies for Systems of Systems

Nirav B. Shah
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology
Cambridge, MA, USA
nbshah@mit.edu

Dr. Donna H. Rhodes
Engineering Systems Division
Massachusetts Institute of Technology
Cambridge, MA, USA
rhodes@mit.edu

Prof. Joseph M. Sussman
Department of Civil and Environmental Engineering
and Engineering Systems Division
Massachusetts Institute of Technology
Cambridge, MA, USA
sussman@mit.edu

Prof. Daniel E. Hastings
Department of Aeronautics and Astronautics
and Engineering Systems Division
Massachusetts Institute of Technology
Cambridge, MA, USA
hastings@mit.edu

Abstract - *Distributed decision making has been identified as a source of managerial complexity for the SoS engineer. A new framework, AIR (Anticipation-Influence-Reaction), is proposed to capture the feedback relationship between the decisions made by constituents and those made by the managers of the SoS. AIR is then used to develop a five-member set of basic influences that can bring about changes in constituent behavior thus modifying the SoS. These influences, the 5 I's, are Incentives, Information, Infrastructure, Integration, and Institutions. AIR and the influences are demonstrated through qualitative application to real-world SoS and quantitatively through a simulation of an inter-modal transport network. It is found that cooperation between competing constituents can be quite fragile and sensitive to the SoS context. Careful, dynamic planning of influence strategies is needed to maintain SoS behavior in the face of constituents who are driven by self-interest and a limited, local perspective of the SoS.*

Keywords: Systems of systems; Influence; Management; Complexity; freight; Simulation; Transportation; Intermodal; Game Theory; Distributed decision making; Anticipation; Reaction

1 Introduction

Since the mid 90's there has been a growing interest in how systems come together to form systems of systems (SoS). These coalitions of independently operated and managed systems can meet unforeseen needs in a timely and cost effective fashion. A challenge for systems engineers is to design, develop and manage constituent systems that are capable of successfully operating within a SoS. Traditional systems engineering theories and approaches do not fully address the technical and managerial challenges caused by this problem. This research focuses on developing better strategies for coping with the managerial complexity caused by the dynamic interactions between constituent systems within an SoS. By understanding these interactions, systems engineers and managers will be better able to develop engineering and management strategies to influence an SoS.

Current interest in these SoS can be traced back to the 1990's with the work of Maier [1, orig. published in 1996]. Maier defines two independence properties characteristic of SoS that have subsequently been used by many authors to define the class of systems termed SoS [2]. These properties, *operational independence* and *managerial independence* specify that both from a technical and a social perspective an SoS is composed of independent yet interacting entities. This formulation has been extended and refined, e.g., Boardman, et al., [3] define several dimensions upon which SoS can be differentiated from traditional systems. More recently, Karcanias, et al., [4] echoed Maier's claim stating:

> "The multi-agent dimension of SoS has characteristics such as:
> *Autonomy:* the agents are at least partially autonomous
> *Local Views:* no agent has a full global view of the system, or the system is too complex for an agent to make practical use of such knowledge
> *Decentralisation (sic.):* there is no designated single controlling agent, but decision and information gathering is distributed." [4]

This third characteristic, distribution of decision making, is a core challenge within SoS engineering. The design and management of SoS is a problem of coordinating the parallel development and operations of the SoS with its constituents. Such coordination can be externally imposed such as in an enterprise [5] or arise as consequence of interaction between the constituents [6].

SoSE is a two-sided problem. On the one hand, it is a technical problem of the determination of the appropriate interfaces [1] between constituent systems in order to accomplish SoS objectives. On the other hand, it is a social problem of convincing those who control the constituents to actually implement such interfaces [7]. Both challenges are recognized gaps in the theoretical SoS literature and each has been identified as a key components to SoS community's research agenda [8, 9].

A variety of frameworks have been proposed to describe the structure, operation and management of an SoS [10, 11, 12]. Of particular importance is that each constituent is trying to satisfy a locally specified value proposition, i.e., they are free to make decision that ensure their local needs are met. The extent to which these decision support a broader SoS agenda depends upon the alignment of these local needs with the SoS goals mediated by whatever influences that the SoS authority brings to bear upon the constituents.[1] As described by Bjelkemyr:

> "Each system within a SoS is a self-interested node in a network. These system nodes try to maximize their own utility under the influences of and in competition with the other nodes. The global SoS behavior thus emerges as a result of the actions at the lower levels of the SoS, down to the system element level." [12]

One can observe this challenge in real world SoSs. For example, peering disputes among the Internet service providers is an issue of choosing with which other systems one wishes to connect, i.e., with whom to collaborate. In October of 2005, Level 3 communications a Boston based tier 1 Internet service provider decided to terminate its peering agreement with Cogent communications, another tier 1 provider [13]. By refusing to peer with Cogent, Level 3 cut-off direct traffic flow between their respective networks. This forced routing via third-party network increasing congestion on those links. For some customers whose only connection was via Level 3, they were disconnected from those hosts whose only connection was via a Cogent network. The same was true in the other direction. After a few days, cooler heads prevailed and the peered connection was reestablished [14]. The underlying cause of the dispute was an imbalance in traffic flow between the two networks. Level 3 felt that Cogent was in violation of their contract when Cogent tried to make inroads into Level 3's market of selling access to Tier 2 providers. If a given Tier 2 provider, directly connected to Cogent instead of going through Level 3, this might create a traffic imbalance to Cogent's benefit.

The essential difference between the decision structure in traditional SE vs. SoSE is one of alignment. The SoS architect may need to influence the constituent decision makers to behave in a manner that is not necessarily locally optimal for them but does serve the interest of the SoS. This relationship between the SoS architect and the constituent decision makers is a principal-agent problem[2] [15]. In the SoS case, the principal is the central authority/SoS architect who wishes to effect some SoS behavior that they value via the actions of the agents, i.e., constituents. Given this framing, the central authority is referred to as an SoS principal. Note that constituents may be interacting with multiple such authorities at a given time (e.g. if they are participating in multiple SoS) and may also act as such an authority themselves with respect to other constituents such as in a collaborative SoS.

2 Anticipation–Influence–Reaction

The role of the SoS Principal (or Influencer) is one of coordinating constituent action to generate SoS behavior that the principal desires via influencing the constituents. This type of relationship is not new to the field of decision theory or organizational management. In logistics, for example, the problem is quite commonplace. Schneeweiss [16] extends the work done in logistics to more generic distributed decision making problems in organizations. The current work applies and extends his formulation to the SoS. While his focus was on the organizational relationships, the current research also includes connections between the systems they control. As a consequence, a broader selection of influence mechanisms are considered. While Schneeweiss, looks at direct incentives and information, the current research extends that to include technological and institutional mechanisms as well. Wernz [17] takes a similar approach in developing a theory of Multiscale Decision-Making.

The AIR Framework is shown Figure 1. The upper portion of the diagram represents the *social* interaction between the constituent decision makers and the SoS principal, while, the lower portion represents the *technical* interaction between the constituent systems in the SoS. Constituent action are changes made in the constituent systems by the constituent decision makers causing the SoS to change over time. The principal first observes current SoS behavior. This observation is used by the principal to capture the current state of the SoS and evaluate direct changes they could make to SoS entities under their direct control. Second, they anticipate constituent decision-making and interactions. The word 'anticipate' is used instead of 'observe' since, unlike system behavior, constituent decision-making process is not generally visible to the principal. As independent agents, constituents can make choices in private only revealing them through their actions. Therefore the principal must use their

[1] The situation is somewhat different in the case of directed SoS. The fact that a central authority has coercive influence on the constituents renders the problem of SoS and constituent alignment moot, however, it can also bring additional responsibility on the central authority to manage constituent needs.

[2] A classic example of this situation the the employer–labor relationship. In that case the employer, wishes to maximize the productive output of her firm. The output, however, is dependent upon the effort put forth by the employees. The employees wish to maximize their total wages while minimizing work hours. Each player, the employer and the employees, makes their own choices with regards to the variables they control.

best estimate of constituent actions and constituent objectives, i.e., the constituent decision problem in assessing influence strategies. Third, based upon observation of the systems and anticipation of constituent decision-making, influences are brought to bear upon the constituent with the aim of modifying their behavior.

Constituents respond to these influences in two ways. First, they take actions to modify the systems they control in response to the changes in their decision problem caused by the influences. If the influences were well-formed and the principal's understanding of the SoS and constituents accurate, then the effect of those changes in the systems will modify SoS behavior in an manner of value to the principal. The principal will observe the extent to which this has occurred (post-facto feedback). The constituents may also react directly to the influences, signaling their (dis-)satisfaction.

These three interactions, *anticipation, influence and reaction*, form the core social feedback mechanism between SoS principals and their constituents, and give the AIR framework its name.

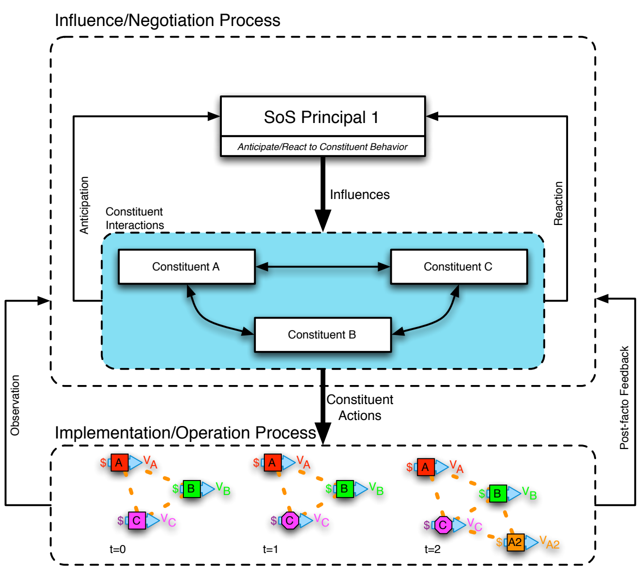


Figure 1. The AIR framework represented graphically

To illustrate this approach for a collaborative SoS, the AIR framework is applied to GEOSS, the Global Earth Observation System of Systems. GEOSS is an effort to combine and coordinate the collection, dissemination and exploitation of earth observation data[3]. A multinational effort, it is coordinated by the Group on Earth Observation (GEO), an intergovernmental organization with membership from 80+ countries. Each country contributes its own local data and expertise. As the assets that produce this data are all locally managed and operated, GEOSS is an SoS. It is also a federation of systems as defined in [18]. One area of focus for GEO has been the establishment of data sharing standards to allow re-use of data collected by various GEOSS constituent systems [19]. Khalsa [20] describes a pilot program by which GEO is establishing an information system (of systems) for data sharing. A key challenge in building this SoS has been been diversity of needs of the end-users combined with the distribution of decision making amongst globally (and, therefore, culturally) dispersed constituents [21]. GEO met this challenge by implementing a service-oriented architecture (SOA) for data sharing. The SOA allowed each constituent to chose which data they published and specified a common repository that served as a catalog for these data sources. The process by which this repository was established is a good example of the different pieces of the AIR framework in practice. The constituents are the data providers. These same providers formed a working group that serves as the SoS influencer or principal.

As described in Khalsa [20], the data interoperability pilot program proceeded in phases. In the first phase, recognizing, that in many cases, constituents were already exchanging data, an effort was made to document the *de facto* standards under which these exchanges took place. This corresponds to the arrow labeled 'Observation' on the left side of Figure 1. In the next phase, communities of potential users were formed to examine what new data exchange/normalization requirements needed to be developed to harmonize the de facto standards from the first phase. This is the anticipation step within which the influencer attempt to envisage how the constituents will respond to various influences and thereby find the influences that best induce the desired behavior of the SoS as a whole. In this case, the working group identified the new standards and protocols needed to enable the desired use-cases of the GEOSS members. The influence in this case is the offering of these new standards for adoption by the constituents. In the third phase, they will implement a demonstration version of the new data exchange service repository, thereby creating a opportunity for the users to try the new approach before broader deployment. This is a form of reaction, wherein private information to the constituents, i.e., the effect of the new standards upon them is revealed to the influencer through the constituents' participation in the demonstration.

3 Five Basic Influences (5 I's)

The SoS principal or influencer is trying to effect the choices being made by the constituents. Therefore, a natural starting point for developing strategies are the constituents' decision problems. Proposed below are five distinct ways that the influencer can exert influence upon the constituents' decision problems. Each influence mechanism impacts a different part of the constituent's decision problem. They are outlined in Figure 2. The constituent's utility function is u and x are their decision variables. Both are indexed by i as they are defined separately for each constituent. Estimates of the decision made by others that affect them are $\hat{x}$. Con-

[3] See http://www.earthobservations.org/geoss.shtml

straints are represented by g and h reflecting those that arise from social and technical concerns respectively. The basic influences are identified by an 'I' word and so are collectively known as the 5 I's. They represent a basis set of strategies and may be used in combination to create the desired effect.

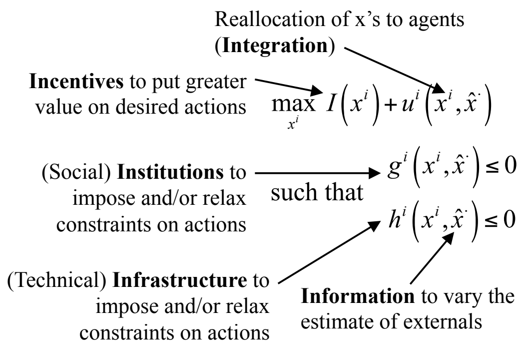


Figure 2. Influencing the constituent decision problem

Incentives reward or penalize constituents for particular behavior that they would not do otherwise. For example, one may contract a commercial communication network to carry SoS related traffic in addition to their normal operating load.

Information can be provided to constituents to modify decisions made under uncertainty. For example, interface specification documents provide potential SoS constituent advance notice of the necessary interfaces needed for future SoS participation.

Integration is the re-assignment of particular SoS components to different constituents. A common example would be combining two systems into one such as in a merger between firms.

Infrastructure refers to introducing new technology into the SoS that effects one or more constituents. An example would be a new high-speed data network to facilitate higher bandwidth inter-connection between constituents and thereby encourage the formation of new interfaces.

Finally, *institutions* refer to the rules and regulations that constituents follow. An example of this would be allowing collusion between ordinarily competing constituent to ensure more efficient allocation of shared resources.

These five influences, can be used to change both the actions of a particular constituent and the relationships between constituent. Examining the Internet peering dispute described earlier, Level 3 had several options available. Using the influence mechanism of incentives, Level 3 could have renegotiated peering agreement to Cogent require compensation from Cogent in the event that traffic is imbalanced. In terms of integration they could buy Cogent's business (or sell their business to Cogent). They could exchange additional information with Cogent, e.g., traffic data and projections to better allow planning by both parties and possibly avoid traffic imbalances. They could have modified their technical infrastructure to selectively reduce the quality of service for Cogent customers traversing the Level 3 network. This would encourage those customers to find alternate transport thereby reducing the traffic imbalance. Finally, they could use dispute resolution institutions provided for in a typical peering agreement. Each of these approaches has side-effects and one may leave Level 3 in a better position than another, however, they do demonstrate the variety of strategies available.

4 Case Study: Intermodal Transport

To demonstrate the AIR framework and 5 I's, they are applied to the problem of intermodal freight transport. Driven by both increasing demand and increased concern for externalities such as environmental damage and noise, one area of focus is making more efficient use of transportation networks. In terms of overland transport, there is much interest in greater use of intermodal routes [22]. Intermodal transport refers to transportation solutions that use two or more transport modes. While both passenger and freight intermodal over many mode combinations are active areas of study, the focus here is on freight transport using trucks and railroads. For inland transport in particular, research into rail-truck intermodal has revealed that combining rail with truck can lead to significant cost savings when compared to using trucks alone. This is a consequence of the greater efficiency of rail over long distances and carrying many carloads. In addition, using modern locomotive technology, rail can generate less pollution than trucks for the same move [23]. On both these accounts, increasing the use of rail via intermodal links to the trucking system appears to be a beneficial policy objective. The objective of the case study, therefore, is to apply the AIR (Anticipation-Influence-Reaction) framework to look at the influence strategies in an intermodal transport network. While the overall context being described in the case study is manufactured, the behaviors of the constituents are rooted in real-world examples.

4.1 Anticipation Phase

Recall that the anticipation phase consists of the SoS principal attempting to understand the behavior of the constituents (and by extension the SoS) so that he may look at potential influences. Many SoS are far too complicated to allow quantitative prediction of behavior. As such, building a predictive model is difficult.[4] Rather, the SoS principal should seek to understand the key behaviors of and interactions between the constituents and include those in a simplified model that can be used to better understand the dynamics that emerge when all the pieces interact. This type of 'behavioral' model is much easier to produce. Using concepts from the models proposed by Fernandez [24] and Gambardella [25], the following local (constituent-level) decision

[4] This can be due to complexity, scale of the SoS, limitations on available data to characterize past behavior and, conversely, inability to use past data when considering novel SoS forms.

makers (DM) are identified: (1) Shippers; (2) Trucking companies; (3) Railroads; (4) Terminal operators. SoS-level decision makers are (1) Coalitions of mode/terminal operators who offer intermodal service as a door-to-door offering as perceived by the shippers and (2) external influencers (aka SoS principal) who provide incentives. For this study, the SoS design problem is framed from the perspective of an external influencer who wishes to increase intermodal moves.

For this case, goods flow from two origin points to two destination points. Connecting these are a network of road and rail links with intermodal terminal between rail and road Figure 3. Traffic is simulated upon this network for a period of 15 years (60 quarters). While the underlying processes being represented are stochastic, under the assumptions used, the expected value and variance of the simulations variable can be computed analytically and so the model is deterministic.

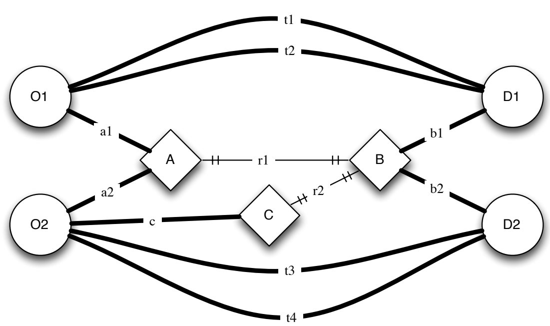


Figure 3. A simple intermodal network

The model represents the interaction between shippers who wish to use the transport network to manage the stock of a good at the destination point and carriers (railroad operators and truckers) who provide transport service between points on the network. The overall flow of the model is shown in Figure 4. For the shipper model, shippers are assumed to use

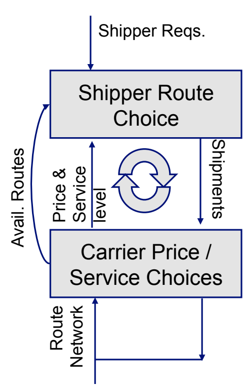


Figure 4. Overall transportation model flow

trigger-fixed quantity reorder inventory management strategy. Each quarter they attempt to find the transportation solution that minimizes their total logistics cost [26]. There are 50 shippers each of whom move 2000 TEU (twenty foot container equivalent units) per quarter. Rail pricing and operations (train frequency is the only operational variable considered) is modeled as profit maximization problem with exponential forecasting used to estimate future behavior of other actors (e.g. prices of competitiors). This is inspired by the pricing approach used by BNSF as described in [27]. Pricing and service frequency are re-evaluated each quarter. For truck carriers the Owner/Operator Independent Driver Association[5] cost model is used. Again, prices are set quarterly to maximize profits and exponential forecasting is used. Terminals are assumed to be fixed time delay transfers between transport modes incurring a fixed cost per container moved. In the baseline case, shippers form routes using a forwarding company to contract with the relevant carriers forming a complete chain from their desired origin to destination. The forwarding company as modeled is a simple monetary pass-through to the carriers.

Baseline results (Figure 5) show that after a brief period of variation, the market eventually settles to an almost even split between traffic using long haul truck routes (aquamarine) and traffic using intermodal routes (purple).

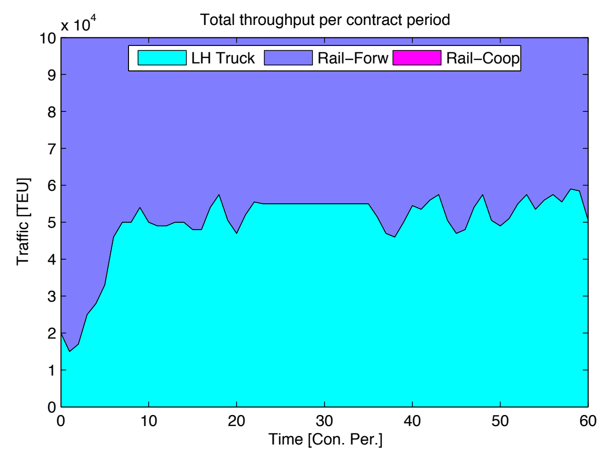


Figure 5. Baseline case: Traffic allocation by mode

4.2 Influence Phase

Two different influences are considered with respect to intermodal terminals. The first provided a subsidy to offset the additional cost borne by shippers when going through terminals, while the second posited an infrastructural improvement in terminal throughput [28]. Neither terminal improvement nor subsidy had a significant impact on shipper mode choice. The reason for this seems to be that, for the situation as modeled, terminal costs do not represent enough of a share of total logistics cost to cause a shift in shipper behavior, and, shippers compensated for higher (and more variable) transport delays by increasing inventory. As inventory costs are increased, these influences begin to have an effect,

[5] http://www.ooida.com/Education%26BusinessTools/Trucking_Tools/costpermile.shtml

however, quite large costs shifts (more than 40% of per TEU value per TEU held, on average, in inventory) are needed to see a change.

In the case of road travel, Janic [23] claims that 20% of the total cost can be attributed to externalities. Conversely, in the intermodal case, only 6% of the cost is the from externalities. Upon imposition of a tax to balance the difference in cost due to externalities, there is a shift of traffic from the uni-modal truck to the rail intermodal as expected. However, that shift is small resulting in only 60% of the traffic going on rail–better than the baseline, but not by much. Insufficient traffic is shifted to justify the railroad increasing capacity. The net effect was little shift in traffic in the long run and a large dead-weight loss to the shippers.

The final influence considered was an institutional change to allow the formation of cooperative routes through a negotiation between longhaul truckers and the railroad. This is also an example of integration as two entities that were separate are now acting together.Once formed, coalitions are kept in place for 4 contract periods (one year) and then re-negotiated (prices for these cooperative routes are re-evaluated every quarter to maximize the joint-profit earned by the coalition). Coalition formation is modeled using the Nash bargaining solution with service offering of non-involved parties forecasted as above. The results are shown in Figure 6. The influence was turned after 20 quarters had elapsed. There was an immediate and stark shift in traffic away from the truck (aquamarine in Figure 6) to rail intermodal – both from coalitions (magenta) and via forwarding companies (purple). Though coalitions come and go as the other parties adjust prices to make them no-longer advantageous, for almost 30 quarters, truck traffic is kept to under 20% of the total flow.

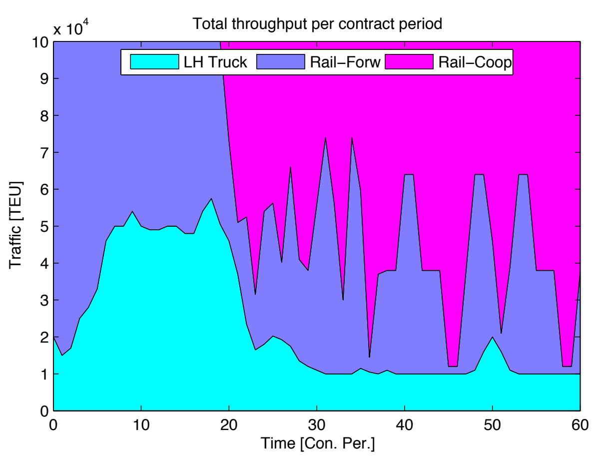


Figure 6. Cooperative routes allowed case: Traffic allocation

4.3 Reaction Phase

To look at reaction requires examining how each of constituent parties are effected by the influence mechanisms. Both the tax and cooperative routes resulted in shifting traffic from road to rail. However, the tax resulted in a significant cost increase (11%) that was passed on to the shippers. Either directly to the external authority (i.e. the government) or through the carriers, they are likely to protest such price increases. The cooperative route on the other hand resulted in a net cost decrease to the shippers of 10%. Looking just at this it would seem that the cooperative solution is the best. However, the truckers suffer in that situation. Moving so much traffic to rail greatly reduces their profits by almost 2/3. They would likely object to such a move. Thus, in implementation, the SoS principal would likely have to balance the protests from the truckers while not placing to large a burden on shippers. This is representative of the types of trade-offs that can occur when influencing SoS with many interacting constituents.

4.4 Observations, Limitations and Extensions

From a transportation perspective, the case study results could lead one to hypothesize that external market interventions such as taxes and subsidies can be less effective that mechanisms that exploit self-interest such as allowing cooperation. The strong effect of cooperation is consistent with empirical studies of intermodal transport networks [29]. Of course such mechanisms may not always be available, but when extant, they should be carefully considered. As formulated, however, the case study model is quite simplified and so its results should be taken as behavioral and not predictive beyond showing potential trends. It can be extended to include a, larger, more realistic route network and more varied shipper and carrier populations such as in [30] allowing better characterization of the effectiveness of the proposed strategies. Costs of implementing the influences was not considered as evaluating the cost of a social change in way that is comparable to a technological change is difficult. Nash bargaining is only one approach to look at cooperation between constituents. Other game-theoretic approaches such as those developed in [31] could be particularly useful as they can be applied to modeling the participation/cooperation decision in other SoS.

5 Limitations and Future Work

The preceding sections provide a glimpse into how the AIR framework and the 5 I's could aide an SoS principal. There are several significant limitations and opportunities for extension. With respect to the AIR framework, one must keep in mind that AIR, on its own, is not sufficient for managing an SoS. It is best used in the context of broader framework such as those cited earlier. AIR only helps formulate strategies for changing constituent behavior. It does not aide in determining what the desired constituent behavior should be. That is the design problem of SoS and progress towards it has been made in [32]. Simulation and modeling of SoS is required for AIR. Progress has been made there by [33, 34, 35]. As developed thus far, AIR assumes a fixed constituent set.

Changing this would require modeling a super-set of potential constituents and their respective life-cycles. In addition, scaling the agent-based modeling approach demonstrated in the case-study to very large numbers of constituents can be challenging. For such large numbers, constituents may need to be represented as member of a class whose behavior is characterized statistically instead of considering individuals. Systems dynamics can be helpful in such a situation as was shown in [36] where multiple satellite operators were aggregated. Determining the costs associated with influence mechanism, especially those that are social in nature is quite challenging. Research from political and organizational science should be used in assessing such costs and managing tradeoffs between constituents that arise during the reaction phase. Finally, influence strategies were discussed in isolation and were implemented as such in the case study. In reality, they will likely need to be used in combination to achieve the desired effect. How to form such combined strategies is an area for future research.

6 Impact on SE Practice

The AIR framework and 5 I's can have significant impact on systems engineering practice. They provide a simple, consistent representation of the key decision-making roles that controls an SoS. At the highest level, these are those of the constituent and the influencer. While the notion of constituent is not new, the notion of an 'influencer' is novel. SoS engineers often find themselves in this influencing role that can only indirectly effect the systems within the SoS. Traditional systems engineering is predicated on the ability of the highest level stakeholder to proscribe requirements which determine decision making at the lower levels. Such an approach would not work in SoS when there was a conflict between the needs of the system of systems engineer and that of the constituents. Rather strategies that account for the local needs of the constituents are required. The 5 I's are a first steps towards developing such strategies.

As is demonstrated in the case study, counter-intuitive results can occur when attempting to intervene in systems of such significant decision-making complexity. Therefore modeling such as the agent-based approach used in the case is crucial to gaining a sufficient understanding of the dynamics of the SoS before intervening in the real world. Examples of this are replete in case studies of real SoS. When trying to modernize document production in the DoD, the need of for common standards was identified [6]. In implementing these standards, however, problems arose given the diverse areas in which the standards needed to be applied. Furthermore, making such changes without disturbing on-going operations was quite challenging. Even though the end-state was much better than the status quo, there was a need to ensure local buy-in to make the transitions happen. AIR and the 5 I's can help the systems engineer think through such issues systematically before making changes in already operating systems.

7 Summary

A key challenge in the management of SoS arise from the operational and managerial independence of the constituent systems. They are free to make decisions based upon local concerns. As these concerns may not align with SoS needs there is no guarantee that constituent decisions will benefit the SoS. To help mitigate this, SoS principals must use influence upon the constituents to make those actions that support the SoS the preferred actions of the constituents, i.e., to ensure incentive compatibility [37]. The AIR framework and 5 I's can help with this task. By anticipating how constituents will behave, understanding how influences change their behavior and ensuring that mechanisms exist for constituents to react, SoS principals can gain a deep understanding of both the social and technical dynamics within their SoS along with the levers available to them to change it. The 5 I's, incentives, information, infrastructure, integration and institutions provide a basis set of influences from which many aspects of constituent decision making can be changed.

References

[1] M. W. Maier. "Architecting principles for systems-of-systems." Systems Engineering, vol. 1, pp. 267–284, 1999.

[2] C. Keating, R. Rogers, R. Unal, D. Dryer, A. Sousa-Poza, R. Safford, W. Peterson and G. Rabadi. "System of systems engineering." Engineering Management Journal, vol. 15(3), pp. 36–45, September 2003.

[3] J. Boardman and B. Sauser. "System of systems – the meaning of of." In "IEEE/SMC International Conference on System of Systems Engineering," pp. 118–123, 2006.

[4] N. Karcanias and A. G. Hessami. "System of systems and emergence part 1: Principles and framework." In "Fourth International Conference on Emerging Trends in Engineering & Technology," IEEE Comp. Soc., 2011.

[5] J. Morganwalp and A. P. Sage. "A system of systems focused enterprise architecture framework and an associated architecture development process." Information Knowledge Systems Management, vol. 3, 2003.

[6] A. J. Krygiel. Behind the Wizards's Curtain: An Integration Environment for a Systems of Systems. CCRP, July 1999.

[7] J. S. Dahmann and K. J. Baldwin. "Understanding the current state of us defense systems of systems and the implications for systems engineering." In "IEEE Systems Conference," IEEE, 2008.

[8] R. Valerdi, E. Axelband, T. Baehren, B. Boehm, D. Dorenbos, S. Jackson, A. Madni, G. Nadler, P. Robitaille and S. Settles. "A research agenda for systems of systems architecting." International Journal of System of Systems Engineering, vol. 1(1/2), pp. 171–188, 2008.

[9] DoD. Systems Engineering Guide for System of Systems. Office of the Under Secretary of Defense for Acquisition, Technology and Logistics (OUSD AT&L), v1.0 edn., August 2008.

[10] D. A. DeLaurentis. “Understanding transportation as system-of-systems design problem.” In “43rd AIAA Aerospace Sciences Meeting and Exhibit,” AIAA, 2005.

[11] A. P. Sage and S. M. Biemer. “Processes for system family architecting, design, and integration.” IEEE Systems Journal, vol. 1(1), September 2007.

[12] M. Bjelkemyr, D. Semere and B. Lindberg. “An engineering systems perspective on system of systems methodology.” In “1st Annual IEEE Systems Conference,” pp. 1–7, 2007.

[13] Team Register. “Level 3 depeers cogent.” The Register, October 2005.

[14] S. Cowley. “Level 3, cogent resolve peering dispute, renew deal.” COMPUTERWORLD, 2005.

[15] K. Binmore. “Taking charge.” In “Playing for Real,” Oxford University Press, chap. 20, 2007.

[16] C. Schneeweiss. “Distributed decision making–a unified approach.” European Journal of Operational Research, vol. 150(2), pp. 237–252, 2003.

[17] C. Wernz and A. Deshmukh. “Multiscale decision-making: Bridging organizational scales in systems with distributed decision-makers.” European Journal of Operational Research, vol. 202(3), pp. 828 – 840, 2010.

[18] A. P. Sage and C. D. Cuppan. “On the systems engineering and management of systems of systems and federations of systems.” Information Knowledge Systems Management, vol. 2, pp. 325–345, 2001.

[19] M. Rao, S. Ramakrishnan and C. Dagli. “Modeling net-centric system of systems using the systems modeling language.” In “CSER 2006,” CSER, 2006.

[20] S. Khalsa, S. Nativi and G. Geller. “The GEOSS interoperability process pilot project (ip3).” Geoscience and Remote Sensing, IEEE Transactions on, vol. 47(1), pp. 80 –91, jan. 2009.

[21] Eliot and Christian. “Planning for the global earth observation system of systems (GEOSS).” Space Policy, vol. 21(2), pp. 105 – 109, 2005.

[22] Transportation Research Board. Policy Options For Intermodal Freight Transportation. Transportation Research Board, 1998.

[23] M. Janic. “Modelling the full costs of an intermodal and road freight transport network.” Transportation Research Part D, 2007.

[24] J. Fernández L., J. de Cea Ch. and A. Soto O. “A multi-modal supply-demand equilibrium model for predicting intercity freight flows.” Transportation Research Part B: Methodological, vol. 37(7), pp. 615 – 640, 2003.

[25] L. M. Gambardella, A. E. Rizzoli and P. Funk. “Agent-based planning and simulation of combined rail/road transport.” SIMULATION, vol. 78(5), pp. 293–303, 2002.

[26] O. K. Kwon. Managing Heterogeneous Traffic On Rail Freight Networks Incorporating The Logistics Needs Of Market Segments. Ph.D. thesis, MIT Civil and Environmental Engineering, 1994.

[27] M. F. Gorman. “Intermodal pricing model creates a network pricing perspective at bnsf.” INTERFACES, vol. 31(4), pp. 37–49, July-August 2001.

[28] R. Stahlbock and S. Voss. “Operations research at container terminals: a literature update.” OR Spectrum, vol. 30, pp. 1–52, 2008.

[29] M. R. Van Der Horst and P. W. De Langen. “Coordination in hinterland transport chains: A major challenge for the seaport community.” Maritime Econ Logistics, vol. 10(1-2), pp. 108–129, 2008.

[30] J. Flodén. Modelling Intermodal Freight Transport. Ph.D. thesis, Göteborg University, 2007.

[31] K. Daniel Cooksey and D. Mavris. “Game theory as a means of modeling system of systems viability and feasibility.” In “Aerospace Conference, 2011 IEEE,” IEEE, pp. 1–11, 2011.

[32] W. A. Crossley and M. M. Nusawardhana. “Variable resource allocation using multidisciplinary optimization: Initial investigations for system of systems.” In “10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference,” AIAA/ISSMO, 2004.

[33] E. Sloane, T. Way, V. Gehlot, A. Levitin and R. Beck. “SoSE modeling and simulation approaches to evaluate security and performance limitations of a next generation national healthcare information network (nhin-2).” In “SoSE ’07,” pp. 1–6, 2007.

[34] P. T. Biltgen and D. N. Mavris. “Capability-based quantitative technology evaluation for systems-of-systems.” In “SoSE ’07,” pp. 1–6, 2007.

[35] C. Dagli and N. Kilicay. “Understanding behavior of system of systems through computational intelligence techniques.” In “1st Annual IEEE Systems Conference,” pp. 1–7, 2007.

[36] N. B. Shah, M. Richards, D. Broniatowski, J. Laracy, P. Springmann and D. Hastings. “System of systems architecture: The case of space situational awareness.” In “AIAA Space 2007,” AIAA, 2007.

[37] D. E. M. Sappington. “Incentives in principal-agent relationships.” The Journal of Economic Perspectives, vol. 5(2), pp. 45–66, 1991.

MODELLING SYSTEMS-OF-SYSTEMS: ISSUES AND POSSIBLE SOLUTIONS

Dr Craig Wrigley, Lockheed Martin UK IS&S Ltd, Chester House, Farnborough Aerospace Centre, GU14 6TD

I. ABSTRACT

This paper discusses a problem with current Systems modelling techniques when applied to Systems-of-Systems (SoS) – namely that in the SoS case, the composition and boundary of the SoS is often only defined transiently and that a System may also form part of several SoSs simultaneously. Therefore there is no longer a simple mapping from an operational view of requirements to the system of systems which enact the required capabilities. This calls into question whether current modelling methods and capability-based procurement paradigms are sufficiently dynamic and flexible to cover future SoS architectures. The paper discusses the attributes required of a modelling approach which attempts to address this problem.

II. AIM

The aim of this paper is to discuss the extent to which existing diagrammatic languages allow us to reason about System-of-Systems problems and to propose the basis for a more comprehensive approach. While there is a large body of analytical techniques which can be used to analyse *systems*, the System-of-Systems problem has proved more intractable, and as a result, the ability to analyse and specify requirements for systems which are to form part of a System-of-Systems is incomplete. As Dahmann and Baldwin [1] note "The literature says little about implications of the trend toward SoS for the engineering of systems. If the trend towards SoS continues, most new systems can expect to be constituents of one or more SoS throughout their service lives. There has been very little attention to how systems engineers should address the engineering of new systems so they are able to support current and future SoS."

III. DISCUSSION – CURRENT APPROACH

The current approach to capability –based system analysis is well documented (see for example, [2]). To paraphrase, a capability requirement is identified in terms of a desirable effect. The process is then to design and model possible system solutions, normally in the context of defined scenarios in order to refine the system-level requirements and then to apportion requirements between subsystems. This in turn motivates the system and subsystem interface requirements.

However, this approach frequently does not adequately reflect the agile and *ad hoc* nature of many current Systems of Systems, in which systems are employed based on their flexibility, availability and familiarity, possibly in combinations which were not originally considered. It is important for reasons of agility and cost-effectiveness to be able to improvise and adapt what is already available to achieve the desired effects, without necessarily waiting for the planned system configurations to be available. The current approach often limits system design decisions to only accommodate the most immediate project requirements, or those within a single context, without sufficient consideration of the need for flexibility and operation in a range of future contexts.

Future SoS are more likely to be *assembled* than *designed,* and only then on an as-needed basis. This means that we can no design a system element highly adapted to a specific purpose, but we need to take account of existing capabilities and their combinations, and to design system elements which are more adaptable and composable so that they can be reused in multiple contexts.

IV. IDENTIFICATION AND CLASSIFICATION OF SYSTEMS OF SYSTEMS

A. *SYSTEM OF SYSTEMS CHARACTERISTICS*

Maier [3] proposes that the following are characteristic of a System of Systems:

- Operational Independence of Elements
- Managerial Independence of Elements
- Evolutionary Development
- Emergent Behaviour
- Geographical Distribution of Elements

For the purposes of this paper we propose to focus on the type of SoS which results from assembling existing (ie already designed) systems to meet a particular capability requirement. We will call this class of SoS an "Assembled SoS" or *Assembly*. This is very similar to the notion of a "Play" in [4]. The following discussion will largely be directed at this type of SoS, because this type is likely to be the most prevalent in reality and presents some real difficulties for current analysis techniques. In the following we will focus on some derived attributes of the *Assembly* which most directly affect the design process – Temporality, Heterogeneity and Complexity.

B. TEMPORALITY

As a consequence of the first of Maier's three attributes we suggest that Temporality, ie boundedness in time is also an important consideration, following from the operational and managerial independence of the system elements, which means that they may not be permanently assigned to a single SoS, and the expectation the the system will evolve and change form over time. At any point in time, the component elements are defined by the signalling or information (or even material) flows which are currently active.

In the particular case of the *Assembly*, one of the main issues is that of system boundary or system composition. When the task or mission is over, the systems in the *Assembly* are typically reallocated to perform the next mission. In addition a system within the SoS may be participating in multiple missions concurrently, and so be part of several SoSs.

By contrast, the subsystems within a standard system (ie a System of SubSystems) are normally specified (and optimised) to support the specific range of tasks which that system is procured to achieve. They cannot generally be used outside the context of the system (unless specifically designed as network services). In this case the system boundary is clearly fixed and does not intersect with other system boundaries.

Therefore the modelling approach must allow for time-dependent mutable and intersecting SoS boundaries in addition to the traditional permanently fixed boundaries.

C. HETEROGENEITY

In the introduction we noted the problem resulting from the diversity of systems available to perform a particular military mission. Even when the operational nodes are strictly identified, they can be instantiated by a wide range of actual systems in multiple combinations. Therefore the model needs to allow for the generalisation of possible system configurations.

D. COMPLEXITY

SoSs are inherently more complex than standard systems, because of the extended connectivity possible between already complex system elements, each of which may support auxiliary goals in addition to the SoS. Karcanis [3] suggests that this is likely to result in emergent behaviours. It certainly requires sophisticated modelling techniques.

V. IMPLICATIONS FOR DESIGN PROCESS

The purpose of the classical design activity is to develop a system (often a bespoke system) which meets a clearly defined set of requirements, generally derived from a limited set of scenarios and use cases.

However, as already stated, this does not readily extend to the *Assembly* which is typically assembled from ready-designed systems. If the system elements are highly adapted to a particular scenario and context the extent to which they can be reused will be correspondingly reduced (the "adaptable" vs "adapted" situation in evolutionary biology).

The construction of the individual system elements takes place on a longer timescale than the specific operations which employ them so we need a means of designing system elements to be adaptable to a range of future scenarios, not just designed to meet the immediate problem. In this sense many conventional systems are "over-designed. A system which achieves a sub-optimal performance in a number of different contexts may be more cost-effective than one which is optimised for performance in a single context.

The proposal is that we no longer design monolithic systems of subsystems designed to achieve a limited range of objectives but design *component* systems of SoS (called system elements in the following) which are adaptable and can meet a wide range of future scenarios in multiple contexts. In the following we will discuss how this affects current modelling techniques and how the techniques may have to be adapted to meet this new challenge.

VI. PROPOSED FORMAL APPROACH

A. PARADIGM

The proposed new paradigm recognises that the design activity now has two phases. First there is the design of the individual system elements which will form the set of system elements from which we can select systems to perform a particular mission – the system "catalogue". This takes place well in advance of any operation and it may take months or years to develop the individual system elements.

The next phase occurs very near to the date of proposed operations and is conventionally known as planning, ie the assembly of those system elements required to meet a mission into an *Assembly*. Let us call these phases "System Design" and "SoS Assembly"

B. SYSTEM DESIGN

We have already suggested that the current approach to system design may become too specific too rapidly resulting in over-designed systems which are inflexible in application.

This suggests we should step back further from any specific scenario(s) and start by considering the most general types of behaviour to initiate the classification process, which can then be refined as required until we have enough resolution to define required system element behaviours and interfaces.

This is where a balance between the requirements of reusability and the needs of a specific project will have to be carefully balanced. SoS aspirations mean that system elements may have to be designed within a more general context and therefore have features which go beyond immediate needs, and the globally most cost effective solution will generally not be the same as the locally cost effective solution.

In network-centric and effects-based analysis a commonly used paradigm is that of "sensor-decider-effector". This immediately suggests a starting point for the behavioural hierarchy, ie Sensor, Decider and Effector classes (see Figure 1)

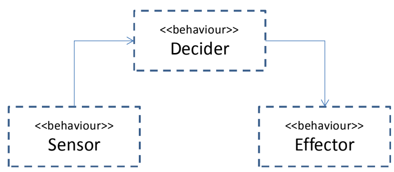


Figure 1 – Starting point for the Behavioural Hierarchy

System elements may belong to multiple classes. This is not necessary a problem as long as the behaviours are still distinct and accessible from outside the system. For example, the F-35 Joint Strike Fighter can act as both a sensor and an effecter. This is acceptable in a SoS context as long as both behaviours can be exposed to other systems. This could be represented by different, linked instances of this system element.

Further subclasses can then be defined, for example the Decider class may contain multiple levels of hierarchy. The Sensors class may be split into different types of sensor. Figure 2 shows an example decomposition of the behavioural classes, using UML as a convenient class notation.

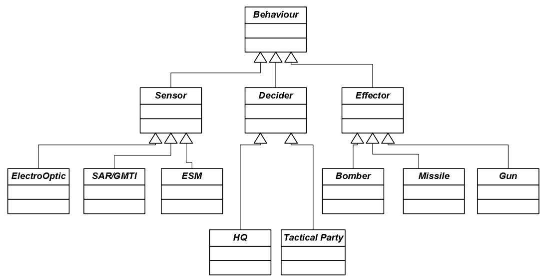


Figure 2 – Example of further refinement of Behaviour Hierarchy

Following a progressive refinement process we can classify system elements by required (or actual) system element behaviour within this hierarchy. Furthermore we can develop interface requirements based on these behaviours. This will be discussed in a future paper.

C. SYSTEM ASSEMBLY

In the system/ subsystem case it is clear that the subsystems fall within the boundary of the system. However, in the SoS case, different assemblies of systems are possible in order to achieve the same effect in different contexts. In this case the systems which form part of any *SoS* configuration may change and evolve with time. In particular, the capability of a *SoS* will generally be provided by a set of distributed systems which are linked only for the time needed to support the capability. Thus we will view SoS boundaries as temporary constructs unless they apply to a specific planned system.

Once a set of of system elements has been developed it needs to be configured to suit a particular mission. This requires a near-term planning activity. The design activity now requires the planner to break down the required effect into "sub-behaviours"and allocate Systems to fulfil each of these sub-behaviours. Since this is a SoS we do not necessarily expect a perfect match between the requirement and the solution, but a "satisficing".

Since we have already classified the system elements according to behaviour, the SoS can then be assembled from a knowledge of the component behaviours and a plan of how to combine the behaviours to achieve the objective.

Lewis [5] has suggested a mechanism for this derived from Service-Oriented Architecture considerations based on the development of a system capability catalogue.

VII. IMPLICATIONS FOR DIAGRAMMATIC TECHNIQUES

We do not intend to provide an exhaustive description of conventional system modelling techniques in this paper, but would refer the reader to [2], for example, which provides an excellent reference. We take conventional techniques to include MODAF, DODAF, and UML.

The main characteristics of these diagrammatic techniques are that:

- They follow a fixed "grammar" – often expressed as a metamodel
- They comprise multiple views dependent on context and depth of analysis
- They generally assume a successive decomposition of the requirement
- They often map system elements directly to operational "nodes"

An example of a suggested approach to modelling SoS using MODAF and BPEL is given in [6].

Many existing architectural frameworks are intended to support SoS modelling. However, the underlying concepts are often still based on single system thinking. In particular, the nodes on OV-x diagrams in MODAF represent capability node requirements within a given (often single) operational context. The capability requirements are typically used to directly imply requirements on constituent systems. However, in the SoS case, we have to consider the collection of systems supporting a capability may only be aggregated temporarily for that purpose and may only partly satisfy the capability while at the same time supporting other capability requirements. Therefore there is no unique system solution to the capability requirement. This becomes quite difficult to represent diagrammatically. As we have discussed, a change in thinking and representation is required, in which the capability is generalised to a behaviour rather than a more narrow set of requirements based on a specific capability need.

The OV concept from MODAF can still be reused, but the nodes more generalised, delocalised and decoupled from a particular scenario. To indicate the difference from the MODAF node concept we used a new <<behaviour>> stereotype in Figure 1 and also emphasis the generalisation and delocalisation by using a dotted line boundary.

VIII. POSSIBLE DEVELOPMENTS OF MODELLING TECHNIQUES

In order to do this we need a library of models of the existing set of system elements and proposed system elements and the ability to construct and simulate a range of Assemblies. The following sections discuss how some of the key modelling issues can be addressed

D. Temporality

A key concept for the dynamic behaviour of the *SoS* is *SoS*-assembly. This is the act of identifying the systems which are to form part of an *SoS* designed to support a specific mission. This is an act performed by a higher level (command) *SoS*, based on the commander's model of the situation and course of actions required to achieve the desired effects. In effect it draws the *SoS* boundary by "marking" those systems involved in the *SoS*. Typically this process provides a (temporary) context to identify the individual systems in involved (eg a mission callsign). For example, we can indicate that system S belongs to Assembly C2 for a bounded time period [t0,t1]:

$$S \in C2|_{t=t0}^{t1}$$

The systems belonging to an Assembly can be classified according to the behaviour(s) they satisfy. We propose the notation in Figure 3 to indicate the behaviour which system from the collection is satisfying in a particular Assembly (for a particular time period).

It is suggested that at this level the operational concept is trialled against a wide range of operational scenarios and contexts to provide a sufficiently general basis for a reusable capability[7]

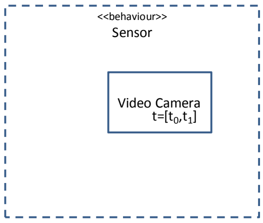


Figure 3 – Proposed notation to show temporality

E. Heterogeneity

It is tempting to define a model with multiple different functional system elements. However, it then becomes necessary to define a wide range of specialised items such as fusion systems and weapon allocators. Keus has developed a NetForce model along these lines [8]. However, for present purposes we require a more general and extensible model which can accommodate future net-centric SoS but also legacy systems. This means that we would prefer a model with a very simple uniform template which can be combined and modified to form complex system structures. This is discussed in the following section.

F. Complexity/ Emergence

A promising approach to handling the complexity of modelling a range of system elements, based on a simple template, is the use of a "cellular" model drawn from biological systems (as suggested in [9]). The basic units of the system (cells) are simple structures with simple behaviours, but when combined into more and more complex structures they can produce arbitrarily complex behaviours. This type of representation has also been used in a human-system context by Miller [9] and Heylighen [10] and is related to the Cybernetic approach to systems modelling.

In this particular context, the use of Agent-Based Modelling seems appropriate [4], combined with the concept of a simple formally defined unit which can be built into combinations of high complexity. We propose the use of an Agent with the properties described in UML notation as in Figure 4.

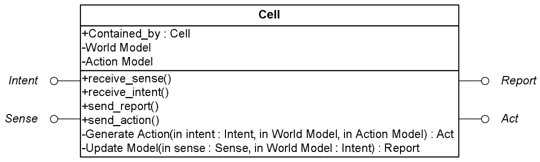


Figure 4 – Proposed UML representation of Agent

The operation of the Agent represented above is formalised as follows: the Agent holds a "World Model" which can be at different levels of abstraction. For example it may be a map with positions of important features marked on it. The stream of input from the Sense interface continuously updates the World Model (eg adds, deletes, moves features). The Agent is given an "Intent" through the Intent interface which may be a command such as "go to point x". The Agent then examines its Action Model for commands which are appropriate to meeting this goal (eg "turn steering wheel clockwise"). It then outputs this to the "Act" interface. The Report interface allows it to publish a report on its current World Model and any Actions it has performed.

This simple agent model can be used to represent a wide range of system elements and human actor roles (in a simplified way). As shown in the figure, the Agent may be part of a larger more complex Agent which can generate complex and emergent behaviours.

IX. CONCLUSION

We have made an assessment of how far existing modelling frameworks such as MODAF can assist in modelling SoS. We found that there are some specific features of SoS which are not well captured by the existing frameworks and which may cause designs within the frameworks to result in over-engineered and inflexible solutions. We have suggested some changes to the paradigm used to design and model the type of *Assembly* described in the paper and some generalisations to diagrammatic techniques which we hope will result in the ability to improve system design so that systems can more readily be assembled into SoS with consequent benefits for cost-effectivess and flexibility. We have proposed the form of a suitable Agent Based Model to support the modelling of the resulting *Assembly*.

References

[1] J Dahmann, K Baldwin, "Implications of Systems of Systems on System Design and Engineering", Proceedings of the 6th International Conference on System of Systems Engineering, Albuquerque, New Mexico, June 2011

[2] C Dickerson, D.N Mavis, "Architecture and Principles of systems engineering", CRC Press, 2010

[3] Maier, M.W., "*Architecting Principles for System of Systems,*" Systems Engineering, Vol. 1, No. 4, 1998, pp.267-284

[4] N. Karcanis, A.G. Hessami, "*System of Systems and Emergence*", presented at IEEE Fourth International Conference on Emerging Trends in Engineering and Technology, 2011

[5] Grace A Lewis, "*Service Orientation and System of Systems*", SEI Carnegie Mellon Presentation, Jul 2011

[6] X. Pan, B. Yin, J. Hu. "*Modeling and Simulation for SoS Based on the DODAF Framework*", presented at IEEE Fourth International Conference on Emerging Trends in Engineering and Technology, 2011

[7] E.N. Urwin et al.,"*Scenario-based Design and Evaluation for Capability*", presented at 5th IEEE International Conference on System of Systems Engineering, 2010.

[8] Hans E.Keus, "Netforce Principles – An Elementary Foundation of NEC and NCO", presented at 10th CCRT Symposium, 2005

[9] J G Miller, "*Living Systems*", McGraw Hill, 1978

[10] F.P Heylighen, "*The Global Superorganism: an evolutionary-cybernetic model of the emerging network society*", Social Evolution & History. Vol 6 No. 1,p. 58-119, 2007

A Semantic Mediation Framework for Architecting Federated Ubiquitous Systems

Georgios Moschoglou, Timothy Eveleigh, Thomas Holzer, and Shahryar Sarkani
Department of Engineering Management and Systems Engineering
The George Washington University
Washington, DC, U.S.A.
{geomos, eveleigh, holzert, g1usxs}@gwu.edu

Abstract - *Despite 20 years of research, ubiquitous systems have yet to become truly ubiquitous. A key challenge is the design for volatility and evolution experienced when those systems are deployed in more than one environment as well as for a substantial time period. The work presented here describes a proposed federated System of Systems (SoS) engineering approach for creating ubiquitous systems based on service-oriented principles. Service orientation is becoming more common for SoS implementation as it supports operational independence, managerial independence, and geographic distribution of constituent systems. However, in a virtual SoS, there is no central management authority and centrally agreed purpose, making interface standardization and integration of capabilities a difficult task. In this paper, we approach this problem by proposing a conceptual ontology-based semantic mediation framework to orchestrate the system engineering activities related to publishing constituent system capabilities during the design stage of the lifecycle, and enable automating capability discovery, selection, and composition at runtime.*

Keywords: Federated systems engineering, semantic interoperability, semantic web services, ubiquitous computing.

1 Introduction

The compelling vision of ubiquitous computing (UbiComp), first articulated by Mark Weiser twenty years ago, describes a world where interaction occurs with machines made to fit the human environment, instead of forcing humans to comply with the machine environment [1]. In this environment, intelligent software, processors and sensors, embedded in everyday objects, sense the behavior and needs of its inhabitants and assist them in day-to-day or exceptional activities. Many technological advances in communications, computation, storage, and mobile and sensor devices are gradually turning this vision into reality. However, despite these technological advances and the rich, multifaceted research work done in this area, UbiComp systems are still facing many challenges in order for them to become truly ubiquitous. A key challenge is creating ubiquitous systems which are not self-contained and bound to their building blocks and environment, but can be deployed in more than one environment and for substantial time periods by adapting their behaviors to the volatility and evolution experienced in those environments [2]. UbiComp systems must deliver functionality in our everyday world, in varying physical circumstances, that cause components routinely to make and break associations with peers of a new degree of functional heterogeneity [3]. We address this challenge by adopting a federated systems engineering approach to create ubiquitous systems. A "Federation of Systems" (FoS) can be described as a coalition of partners with decentralized power and authority and potentially differing perspectives on situations [4].

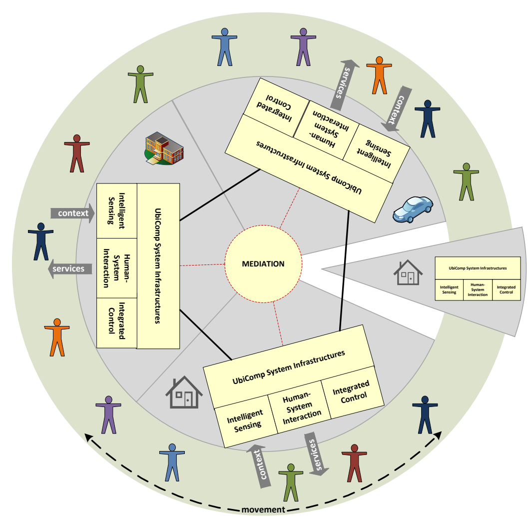


Figure 1. Federation of UbiComp systems.

A federation of UbiComp systems can be realized as a collection of smart enclaves [5] representing the federation entities (see Figure 1). Each enclave usually encompasses a heterogeneous set of intelligent objects and devices, a set of information and service behaviors, and some infrastructures upon which the objects and devices rely in order to offer meaningful services to the enclave's users. Each enclave is self-organizing and can function autonomously to satisfy its own mission objectives and requirements. Systems, subsystems, and components from each enclave can dynamically join or leave the federation without interrupting its operation. An intelligent mediation entity enables system components to seamlessly and spontaneously interoperate in order to combine partial information or service behaviors into comprehensive services that can meet user objectives and beliefs, desires and intentions (BDI). As people move from one enclave to the other, personalized services are composed and offered dynamically based on the user's needs and the current situation they experience during their interaction with the enclave's system. Information flows to and from any device in the user's vicinity, in the most appropriate fashion. A user can leave one enclave, while having pending services, and resume them in a different enclave. For this complex federated systems scenario, there is a need for systems engineering processes suitable to handle a "plug and play" approach. Service orientation is becoming more common for Systems of Systems (SoS) implementation because it supports operational independence, managerial independence, and geographic distribution of constituent systems [6]. However, in a "virtual" SoS [7], such as the conceptual Ubicomp systems federation described above, there is no central management authority and centrally agreed purpose, therefore, systems that have never come together will do so at runtime. Interface standardization and integration of capabilities then becomes a difficult task requiring tedious manual work, significant time and human involvement in capability discovery. In this paper, we address this problem by exploring the use of semantic web tools and techniques to orchestrate the system engineering activities related to: (1) rigorously publishing constituent system capabilities during the design stage of the "federated" system engineering lifecycle, and (2) automating capability discovery, selection, and composition at runtime. The primary research question is, "Does the use of semantic web principles and technologies enable partial or full automation of system capabilities discovery, selection, and composition between heterogeneous, autonomous and geographically dispersed UbiComp systems?" The rest of this paper is organized as follows: Section 2 describes the Semantic Web as it relates to systems interoperability. Section 3 proposes a conceptual semantic mediation framework for integrating autonomous and heterogeneous ubiquitous systems, based on semantic SOA approaches and reference implementations. And, finally, Section 4 concludes this work and discusses our progress and future directions.

2 Leveraging Semantic Interoperability

The semantic web is envisioned as "a Web of actionable information – information derived from data through a semantic theory for interpreting the symbols. The semantic theory provides an account of meaning in which the logical connection of terms establishes interoperability between systems" [8]. Semantic interoperability, though frequently referenced, is seldom explicitly defined in literature. For the purpose of this paper and over-arching research, we adopt the definition denoted in [9] and we add the service exchange component in addition to data exchange: *Semantic Interoperability is achieved when data and service exchange partners have a common understanding of the meaning of the shared data and services, data and service exchanges adhere to the shared understanding, and data and services are exchanged without misinterpretation.* This definition implies that semantic interoperability capabilities include three core functions: semantic description, semantic mediation, and semantic discovery [10].

2.1 Semantic Description

This capability describes information assets based on some form of conceptualization. Knowledge is organized by considering each "entity" a resource and finding a simple way to represent and share the resource [11]. UbiComp Systems can be broken down into resources representing low-level entities such as persons, sensors, mobile devices, and objects, as well as higher level conceptual entities, such as context and activities. Each resource has a unique name across the entire Internet, called a Uniform Resource Identifier (URI). URIs fit well in our federated approach because they have global scope. A resource associated with a URI means that anyone can link to it, refer to it, or retrieve a representation of it [8]. The Resource Description Framework (RDF) is a knowledge-representation language that provides the foundation for semantic interoperability by adding meaning to the resources identified by URIs in a machine-readable format. It can be conceptualized as the data model for resources and relations between them [11]. Integration of diverse and heterogeneous data and services is being achieved in large part through the adoption of common conceptualizations referred to as ontologies. An ontology, which provides greater expressivity than RDF, is a set of statements (triples) that define concepts, relationships and constraints. It is a formal explicit specification of a shared conceptualization in that it supports the capture and specification of domain knowledge with its intrinsic semantics through consensual terminology and formal axioms. This consensual terminology makes it possible to share and reuse data and knowledge across different components within a system as well as across different systems [12]. For example, [13] and [14] present shared ontologies that facilitate the capture of domain knowledge in ubiquitous and pervasive environments. The Web Ontology Language (OWL) extends the RDF vocabulary with additional resources which can be used to build more expressive ontologies for a particular domain. Ontologies can become distributed, as OWL allows ontologies to refer to terms in other ontologies [8].

2.2 Semantic Discovery

This capability locates information assets based on semantic descriptions. Knowledge stored in RDF statements can be discovered in three ways [15]: *Navigation* - Tools, such as semantic web and RDF browsers, are used to retrieve and visualize RDF data by dereferencing URIs to locate additional triples. Since the semantic web is focused on machine-based accessibility, system designers can use those tools for debugging, data verification or, free discovery. *Searching* – Builds on and goes beyond navigation by looking for answers (goal-oriented) in data that have been preprocessed by a system. Preprocessing includes discovering, sorting, filtering, indexing, and storing. *Querying* – Is based on formal syntax and semantics and allows for complex, explicit, and structured questions to be posed. Languages that can recognize and query RDF include SPARQL (a W3C Recommendation), RDQL, and SeRQL. Repositories, called triple stores, have been developed to store RDF statements.

2.3 Semantic Mediation

This capability resolves differences in conceptualization when searching for information assets. Its objective is to resolve semantic differences that may arise in the context of information exchange between participants in an interaction. Semantic mediation involves providing translation services to interacting system entities from different semantic domains [10]. It is the process of modeling the mediation between heterogeneous systems in terms of inter-ontology mappings. Ontology mapping describes the relationship between two ontologies. Ontology mapping between synonymous terms from different systems can indicate that the terms refer to the same entity type. Since ontologies represent entities and their relationships in a system, the interoperability of those systems, as they join the federation depend on ontology mapping. Often, ontology mapping follows an intermediate transformation activity that converts data generated from back-end databases to RDF. For example, GRDDL (Gleaning Resource Descriptions from Dialects of Languages) is a technique for obtaining RDF data from XML and XHTML documents. The semantic mediation process also involves underlying semantic engines, referred to as reasoners, for inference computations that support inferred mappings. The inferred mappings provide implicit relationships among terms in multiple systems. An inferred mapping is systems-generated based on mappings explicitly created by the system providers, and it does not physically exist within the provided mapping ontologies [16]. Rules-based inference engines support inference beyond what can be deduced from description logic in RDF statements. Rule-based formalisms, such as horn-clause logics and higher-order logics, have been extended and modified by Artificial Intelligence researchers to capture causal, temporal, and probabilistic knowledge [8]. Causal logic is intended to capture an important aspect of common sense understanding of mechanisms and physical systems. Temporal logic formalizes the rules for reasoning with propositions indexed to particular times. Probabilistic logics are calculi that manipulate conjunctions of probabilities of individual events or states. The most well-known of these logics are Bayesian, which are used to derive probabilities for events according to prior theories about how probabilities are distributed [8].

Based on the aforementioned semantic interoperability capabilities, the objective of our on-going research is to design a framework to orchestrate the system engineering activities related to architecting federations of ubiquitous systems. UbiComp systems will join the federation through the proposed semantic mediation framework that defines the rules for publishing constituent system structure and components (including physical assets) and discovering their capabilities.

3 Proposed Semantic Mediation Framework

In a UbiComp-based environment services are performed with multiple configurations which dynamically adapt to context changes in that environment and its inhabitants. In a smart home, for example, services can be appropriately configured in space, hardware, software and quality, so that a smart home can fulfill the comfort, convenience, and security requirements of its inhabitants [17]. Our approach under investigation leverages service-oriented architecture (SOA) principles to achieve service integration and interoperability in multiple heterogeneous UbiComp-based environments. However, existing SOA standards are not capable of dealing with web services at a semantic level of expressivity for properly representing and discovering service capabilities [18]. The lack of machine-understandable semantics limits the possibilities of runtime automation in our federated approach since services must be located and bound to service requesters at design-time. Extending SOA with semantics allows the definition of semantically rich and formal service models where semantics can be used to describe both services offered and capabilities required by potential consumers of those services [19]. The emerging concept of Semantic Web Services (SWS) extends Web service descriptions with rich semantic annotations using ontologies instead of XML and provides inference-based techniques for automating the discovery, selection, and composition of services. There are a number of SWS frameworks that define comprehensive specifications for creating suitable semantic descriptions for services. The main frameworks that most of the research is based upon include the Web Service Modeling Ontology (WSMO) [20], OWL-S [21], WSDL-S [22] and the Semantic Annotations for WSDL and XML Schema (SAWSDL) [23]. A comparison of the SWS frameworks reveals that WSMO is a more exhaustive framework that propagates a goal-driven approach along with integrated mediation facilities [19]. WSMO provides a rich conceptual model and a formal language (WSML) to semantically describe all relevant aspects of Web services. Its ultimate goal is to enable the automation of tasks (e.g. discovery, selection, composition, mediation, execution, and monitoring) involved in both intra- and inter-enterprise integration of web services [20]. WSMO identifies four top-level elements as the core concepts: ontologies that define formalized domain knowledge, goals that describe objectives clients want to achieve by using web services, semantically described web services, and mediators for handling potentially occurring heterogeneities [24]. Our proposed framework is based on the WSMO approach mainly because of its goal-based web service discovery technique. Goals are representations of objectives for which fulfillment is sought through the execution of Web services. They can be descriptions of services that would potentially satisfy the user's desires [20]. WSMO could be used to allow UbiComp system designers to describe their desired functionality in terms of capabilities in the definition of WSMO goals at design-time. At runtime, any web service with a capability that matches the goal will be returned to the requesting system for further processing. Figure 2 depicts our proposed semantic mediation framework for architecting federated ubiquitous systems in terms of capabilities required to orchestrate context data and capability exchanges between heterogeneous UbiComp systems. The framework can be realized by implementing the Web Services Execution Environment[1] (WSMX) based on the conceptual models provided by WSMO and a UbiComp domain ontology. WSMX is a component-based execution environment for WSMO that aims to support the discovery composition, mediation, selection, and invocation of Web Services based on a set of user's requirements [25]. The main actors in Figure 4 include the following:

- *Originating System Architect/Developer* represents the architecture/development team of the system that supplies context data and capabilities (services) to the federation. The provided context data may be location, identity and state of people, groups or physical objects, or other kinds of context such as physical (e.g. location and time), environmental (e.g. weather and light) and personal information (e.g. mood and activity) [26]. Examples of provided services include music/video playing, information providing, event notification, and lighting services [17].

- *Receiving System Architect/Developer* represents the architecture/development team of the system that discovers and uses context data and capabilities from the federation. They interact with the semantic mediation framework mainly to post requirements for dynamic discovering of capabilities provided by other systems in the federation.

- *System End-user* represents the constituent system's end users as they move between smart spaces and interact with their corresponding intelligent systems.

[1] www.wsmx.org

A system can be originating, receiving, or both at the same time. In the following, we elaborate on the main tasks the system architects/developers will be performing at design-time.

Build and register system ontology: The system architect/developer constructs a formal representation of the constituent UbiComp system vocabulary (terms and definitions). If the ontology already exists, this activity is optional. Originating and Receiving System architects/developers respectively publish ontologies they will use for WSMO service and goal descriptions in the WSMX repository.

Map ontologies: In addition, the system architect/developer publishes mapping rules between their ontologies and the UbiComp foundational ontology. Every system that joins the federation will have to map its domain (system) ontology to the federation's foundational ontology. All terms in the ontology that were developed in the previous task are mapped to the terms of a "foundational" ontology serving as a basis for the federation-specific ontology. [15] define a foundational ontology or an upper ontology as an ontology that contains objects and concepts that transcend the boundaries of a single knowledge domain. Foundational ontology can simplify information exchange by creating an environment in which the terminologies of disparate UbiComp knowledge domains are all rooted in a common space.

Publish system capabilities: The system architect/developer exposes the originating system's services through a public interface based on WSMO goal-based service descriptions. A WSMO service describes a capability (non-functional and functional) and an interface provided by the originating system (service provider). The originating system architect/developer annotates the WSDL descriptions of the offered services with pointers to the semantic concepts defined in the domain ontologies using WSML allowing describing, semantically, all relevant aspects of Web services. This can make our federated approach possible by allowing the originating system architects/developers to describe the system capabilities they offer to the federation. Specifically, it allows describing the offered services at two levels of granularity: categorization and capability description [27]. Services can be described in terms of the functional category within which they fall, using the ontologies constructed earlier, and their functionality, using logical conditions that must hold before and after service invocation (preconditions and effects). The descriptions contain formal specification of functionality that the service can provide, which is the definition of conditions on service "inputs" and "outputs" which must hold before and after the service execution respectively. Functional descriptions of services are used for discovery purposes in order to find services which satisfy the receiving system's requirements. Non-functional descriptions of services, such as Quality of Services (QoS) can be used to describe user preference for service selection [19].

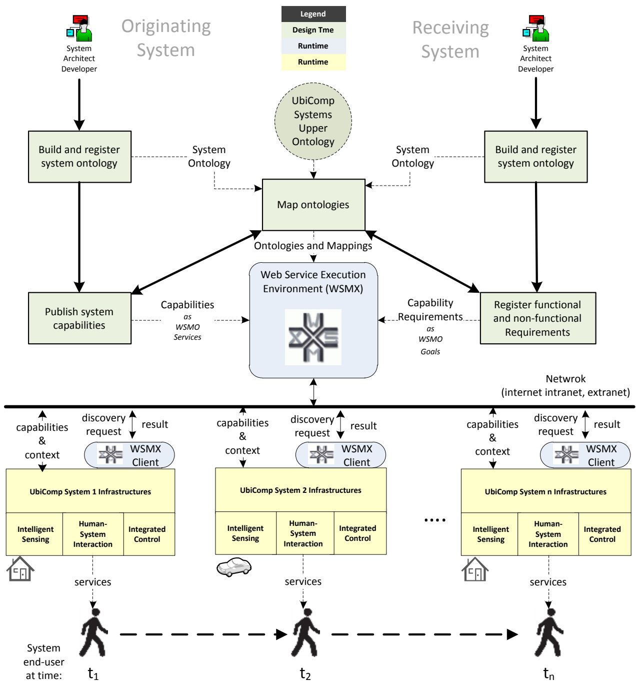


Figure 2. UbiComp FoS Semantic Mediation Framework.

Register functional and non-functional Requirements: The receiving system architect/developer creates WSMO goals that describe requests as well as interfaces through which the receiving system wishes to perform conversation with potential services. A goal is made up of the following elements [20]: *nonfunctional properties*, which are the aspects of the Web service that are not directly related to its functionality (e.g. accuracy, financial, network-related QoS, performance, reliability, scalability, and security), *imported ontologies*, used in the goal as a terminology to define the other elements that are part of the goal, *mediators*, which address the handling of heterogeneities occurring between elements that shall interoperate by resolving mismatches between different used terminologies (data level), communicative behavior between services (protocol level), and on the business process level, *requested capability*, describing the capability of the services that the receiving system would like to have, and, finally, *interface*, which describes the interface of the service that the receiving system would like to have and interact with.

At runtime, the receiving system sends a service request captured as a WSMO goal to WSMX, using a WSMX client, which on receipt of the goal it executes its internal discovery, selection and orchestration processes. The Discovery service matches the goal to potential services using several set-theoretical relationships (e.g. exact match, plug-in match, subsumption match, intersection match and disjointness). The Selection service selects the best service based on the preferences provided by the receiving system in the WSMO goal description. Preferences are non-functional properties, such as SLA and QoS, which capture constraints over functional and behavioral service descriptions. Finally, the orchestration service manages the run-time conversation between the receiving system (service requester) and the originating system (service provider) [19]. Adhering to our federated systems engineering approach, this proposed semantic mediation framework would not impose a particular standard that needs to be agreed by all stakeholders, thus allowing different system designers to select their own data formats and services, while still being able to use context and capabilities offered by other systems.

4 Conclusions and Future Work

The lack of a central authority or a centrally agreed purpose in virtual SoS environments creates a need for more complex technologies able to deal with interoperability and integration of capabilities among distributed and heterogeneous adaptive systems. Advances in semantic web technologies have paved the way for capturing and mediating those dynamic interactions. In this paper, we propose a semantic mediation framework for orchestrating the system engineering activities related to publishing constituent system capabilities during the design stage of the "federated" system engineering lifecycle, as well as automating capability discovery, selection, and composition at runtime. We have begun to: (1) develop a reference implementation of the proposed framework through a UbiComp scenario, (2) define the foundational ontology of the framework which will represent the knowledge model of the UbiComp federation and its service function, and (3) develop a service ontology for describing non-functional properties and requirements specific to the UbiComp domain. Because there are restrictions on system services, non-functional requirements are often of critical importance, and functional requirements may need to be sacrificed to meet them. Finally, we will validate our solution through a set of experiments to evaluate and compare the performance of the reference implementation and show the limitations of our approach.

References

[1] M. Weiser, "The computer for the twenty-first century," *Scientific Ame*rican, Vol. 265, No. 3, pp. 94-104, Sep. 1991.

[2] R. Caceres and A. Friday, "Ubicomp systems at 20: Progress, opportunities, and challenges," *IEEE Pervasive Computing*, Vol. 11, No. 1, pp. 14-21, Jan. 2012.

[3] T. Kindberg and A. Fox, "System software for ubiquitous computing," *IEEE Pervasive Computing*, Vol. 1, No. 1, pp. 70-81, Jan. 2002.

[4] A. Sage and C. Cuppan, “On the systems engineering and management of systems of systems and federations of systems,” *Information, Knowledge, and Systems Management Journal*, Vol. 2, No. 4, pp. 325-345, Dec. 2001.

[5] N. Duan, “Design principles of a federated service-oriented architecture model for net-centric data sharing,” *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, Vol. 6, No. 4, pp. 165-176, Oct. 2009.

[6] G. Lewis, E. Morris, S. Simanta, and D. Smith, “Service orientation and system of systems,” *IEEE Software*, Vol. 28, No. 1, pp. 58-63, Jan. 2011.

[7] M. Maier, “Architecting principles for system-of-systems,” *Systems Engineering,* Vol. 1, No. 4, pp. 267-284, 1998.

[8] N. Shadbolt, T. Berners-Lee, and W. Hall, “The semantic web revisited,” *IEEE Intelligent Systems*, Vol. 21, No. 3, pp. 96-101, May 2006.

[9] J. Tscumperlin, Model-driven Semantic Interoperability using Open Standards: A Case Study. (2007) [Onlne]. Available: http://www.minedu.govt.nz/~/media/MinEdu/Files/EducationSectors/PrimarySecondary/Initiatives/ITAdminSystems/ModelDrivenSemanticInteroperability.pdf

[10] A. Ojo, E. Estevez, and T. Janowski, “Semantic interoperability architecture for governance 2.0,” *Information Policy*, Vol. 15, No. 1-2, pp. 105-123, Apr. 2010.

[11] J. Waters, B.J. Powers, and M.G. Ceruti, “Global interoperability using semantics, science and technology (GIS3T),” *Computer Standards & Interfaces*, Vol. 31, No. 6, pp. 1158-1166, Nov. 2009.

[12] J. Ye, G. Stevenson, and S. Dobson, “A top-level ontology for smart environments,” *Pervasive and Mobile Computing*, Vol. 7, No. 3, pp. 359-378, June 2011.

[13] P. Moore, B. Hu, and J. Wan, “Smart-Context: A context ontology for pervasive mobile computing,” *The Computer Journal*, Vol. 53, No. 2, pp. 191-207, Feb. 2010.

[14] H. Chen, F. Perich, T. Finin, and A. Joshi, “SOUPA: Standard ontology for ubiquitous and pervasive applications,” Proc. First Annual Int. Conf. on Mobile and Ubiquitous Systems: Networking and Services, pp. 258-267, Aug. 2004.

[15] J. Hebeler, M. Fisher, R. Blace, and A. Perez-Lopez, *Semantic Web Programming*, Wiley Publishng, Inc., Indianapolis, IN, 2009.

[16] J. Graybeal, A.W. Isenor, and C. Rueda, “Semantic mediation of vocabularies for ocean observing systems,” *Computers & Geosciences*, Vol. 40, pp. 120-131, Mar. 2012.

[17] CL Wu and LC Fu, “Design and realization of a framework for human-system interaction in smart homes,” *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, Vol. 42, No. 1, pp. 15-31, Jan. 2012.

[18] M. Cai, W.Y. Zhang, and K. Shang, “ManuHub: A semantic web system for ontology-based service management in distributed manufacturing environments,” *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, Vol. 41, No. 3, pp. 574-582, May 2011.

[19] T. Vitvar et al., “Semantically-enabled service oriented architecture: Concepts, technology and application,” *Service Oriented Computing and Applications*, Vol. 2, No. 2, pp. 129-154, May 2007.

[20] D. Fensel, H. Lausen, A. Polleres, J. De Bruijn, M. Stollberg, D. Roman, and J. Domingue, *Enabling Semantic Web Services: The Web Service Modeling Ontology*, Springer, Berlin, 2007.

[21] D. Martin, M. Burstein, J. Hobbs, O. Lassila, D. Mcdermott, S. Mcilraith, M. Paolucci, B. Parsia, T. Payne, E. Sirin, N. Srinivasan, and K. Sycara, OWL-S: semantic markup for web services, W3C Member Submission. (2004) [Onlne]. Available: http://www.w3.org/Submission/OWL-S

[22] R. Akkiraju, J. Farrell, J. Miller, M. Nagarajan, M.T. Schmidt, A. Sheth, and K. Verma, Web service semantics – WSDL-S, W3C Member Submission. (2005) [Onlinie]. Available: http://www.w3.org/Submission/WSDL-S/

[23] J. Farrell and H. Lausen, Semantic annotations for WSDL and XML schema, W3C Recommendation. (2007) [Online]. Available: http://www.w3.org/TR/sawsdl/

[24] M. Stollberg and D. Fensel, “Semantics for service-oriented architectures,” *Agent-Based Service-Oriented Computing*, Chapter 5 (N. Griffiths, K.-M.Chao, Eds.), Springer-Verlag, London, 2010.

[25] C. Pedrinaci, J. Domingue, and A.P. Sheth, “Semantic web services,” *Handbook of Semantic Web Technologies*, Chapter 22 (J. Domingue, D. Fensel and J.A. Hendler, Eds.). Springer-Verlag, Berlin, 2011.

[26] E. Christopoulou, C. Goumopoulos, and A. Kameas, “An ontology-based context management and reasoning process for UbiComp applications,” Proc. of the 2005 joint conference on Smart objects and ambient intelligence: innovative context-aware services: usages and technologies, pp. 265-270, 2005.

[27] J. Kopecky, T. Vitvar, C. Bournez, and J. Farrell, “SAWSDL: Semantic annotations for WSDL and XML Schema,” *IEEE Internet Computing*, Vol. 11, No. 6, pp. 60-67, Nov. 2007.

An Application to Two-Hop Forwarding of a Model of Buffer Occupancy in ICNs

Marco Cello
DITEN
University of Genoa, Italy
marco.cello@unige.it

Giorgio Gnecco
DIBRIS
University of Genoa, Italy
giorgio.gnecco@dist.unige.it

Mario Marchese
DITEN
University of Genoa, Italy
mario.marchese@unige.it

Marcello Sanguineti
DIBRIS
University of Genoa, Italy
marcello.sanguineti@unige.it

Abstract – *An application of the model proposed in Cello at al., A Model of Buffer Occupancy in ICNs, IEEE Communications Letters, to appear is investigated. Such a model provides a relationship in the z-domain between the discrete probability densities of the buffer state occupancies of the nodes in the network and the sizes of the arriving bulks. Under a class of two-hop forwarding strategies, expressions are obtained for the average buffer occupancy and its standard deviation.*

Keywords: Intermittently-connected networks, congestion control, ad-hoc networks, Markov chains, epidemic routing.

1 Introduction

In the last years various applications emerged, where networks operate under conditions in which the assumptions of "universal connectivity" and "global information" do not hold. Examples are sensor networks [8], social networks or pocket switched networks [6], smart environments, and vehicular ad-hoc networks [18]. A common denomination of such contexts is *Intermittently Connected Networks* (ICNs). As in such contexts the networks may be disconnected most of the time or it may even happen that there is never an end-to-end path available between source and a destination, classical routing and data delivery-approaches (see, e.g., [1]) fail [14].

In [12] it was proved that the expected throughput of reactive protocols (which compute a route only when it is needed) is connected with the average path duration pd, the time to repair a broken path tr, and the source data rate r through the relationship: $throughput = \max(0, r(1 - \frac{tr}{pd}))$. However, node mobility leads to frequent disconnections, thus reducing the average path duration significantly. Consequently, in most cases tr is expected to be larger than the path duration, which implies that the expected throughput is close to zero. Other approaches to deal with routing in ICNs involve the use of additional communications resources (e.g., satellite, UAV, message ferries) forced to follow a given trajectory between disconnected parts of the network, in order to bridge the gap [10, 20] (DataMule, Message Ferries, etc.). In other cases, such as in inter-planetary networks [2], intermittent connectivity is predictable, so classical routing algorithms may be adapted to compute shortest delivery time paths by taking into account future connectivity [7].

Often, neither additional resources with controlled behavior nor predictable trajectories are available. In such cases, one of the most common approaches is *epidemic routing* [17], which is based on the replication and transmission of messages to newly-discovered contacts that do not already possess a copy of the message. In epidemic routing each node maintains a buffer, consisting of messages that it has originated and messages that it is buffering on behalf of other nodes. When two nodes meet each other, they decide how many and which stored messages are exchanged. In turn, each node requests copies of messages from the other. In the simplest case, epidemic routing is flooding: each time a contact happens, all messages that are not in common between the two nodes are replicated.

In general, however, message replication performed by epidemic routing paradigms imposes a high storage overhead on wireless nodes [19] and very likely node buffers run out of capacity. More sophisticated techniques can be used to limit the number of message transfers. Existing epidemic protocols try to avoid congestion by limiting, either in a deterministic [13] or in a non-deterministic way [11, 16], the number of copies of a message inside the network. So, an analytical framework for congestion control management is needed.

This is the subject of the present contribution, which extends our previous work [3] by applying the model proposed therein to the kind of epidemic routing known as *two-hop forwarding*. First, we describe the analytical framework developed in [3], based on bulk arrival and bulk service queues, to model ICN nodes behavior (Section 2). Then we discuss a relationship between the stationary discrete probability densities of the state occupancies of the ICN node buffers and the discrete probability densities of the sizes of the arriving bulks. As a further step, we investigate a class of two-hop forwarding strategies used by ICN nodes and sometimes exploited in epidemic routing (Section 3). For such a class of forwarding strategies, we derive an expression for the average buffer occupancy (Section 4). Finally, we discuss related literature (Section 5) and draw some conclusions (Section 6).

2 Model description

We consider the following ICN scenario. M nodes, deployed in an area, follow a certain mobility model, for which we impose the following property: for each couple of ICN nodes ($m, n \in \{1, \dots, M\}, m \neq n$) the number of encounters between them in any given interval of time is a Poisson random variable. This obviously means that the intermeeting time between two generic ICN nodes m and n is an exponentially-distributed random variable. Popular mobility models such as Random Waypoint and Random Direction [5] enjoy such a property.

We denote by $\mathcal{L} \subseteq \{1, \dots, M\}$ the set of destination nodes. We model each node as a battery of queues (Figure 1). Within a specific node $j \in \{1, \dots, M\}$, there are $|\mathcal{L}(j)|$ queues, where $\mathcal{L}(j) \triangleq \mathcal{L} \setminus \{j\}$. Each l-queue ($l \in \mathcal{L}(j)$) within the node j receives incoming data for the destination node l in two different modes. Either the data directed to l are internally generated by node j or they have been sent to j by other nodes during previous encounters with j, on the basis of a forwarding strategy. When node $j \in \{1, \dots, M\}, j \neq l$ encounters node l which is the destination of data it holds in its l-queue, it empties the l-queue completely sending all its data to l. To allow this operation, we assume that the maximum data exchange time between two nodes is much smaller than the average duration of the encounter.

More formally, node j encounters the destination node l with average rate $\mu^{j,l}$ [encounters/s] and sends to l all the packets buffered in its l-queue. Node j generates data in bulks, assigned to l, with average rate $r_s^{j,l}$ [generations/s] and, at each generation, it produces $I_s^{j,l}$ [bulks/generation], set to 1 in this letter. The average rate of bulk generation is $\lambda_s^{j,l} = r_s^{j,l} I_s^{j,l}$ [bulks/s]. We assume an exponentially-distributed time between two consecutive bulk generations. Node j meets any node different from the destination l with average rate $E_e^{j,l} = \sum_{h=1, h\neq j,l}^{M} \mu^{j,h}$ [encounters/s] and, at each encounter, receives $I_e^{j,l}$ [bulks/encounter], set to 1 in this letter. The corresponding bulk generation process has rate $\lambda_e^{j,l} = E_e^{j,l} I_e^{j,l}$ [bulks/s] and is a Poisson process, since it is the sum of independent Poisson processes. The two processes of bulk generation with associated average rates $\lambda_s^{j,l}$ and $\lambda_e^{j,l}$ are assumed to be independent. Due to the assumption on the mobility model and on the generation of bulks, the global process of bulk arrivals in the l-queue is Poisson process. We denote by $\lambda^{j,l} = \lambda_s^{j,l} + \lambda_e^{j,l}$ [bulks/s] its average rate.

The size of each bulk (i.e., the number of packets in the bulk) is also a random variable. We denote by $g_{k,e}^{j,l}$ the probability that the bulk assigned to l and received by j during an encounter is composed of k packets, and by $g_{k,s}^{j,l}$ the probability that the bulk generated by node j and assigned to node l is composed of k packets. The average arrival rate of bulks of k packets in the l-queue, measured in [packets/s], is $\lambda_s^{j,l} g_{k,s}^{j,l} + \lambda_e^{j,l} g_{k,e}^{j,l}$, as indicated in Figure 1. For $k \in \mathbb{N}_0$, we denote by $\{g_{k,s}^{j,l}\}$ and $\{g_{k,e}^{j,l}\}$ the sequences whose components are $g_{k,s}^{j,l}$ and $g_{k,e}^{j,l}$, respectively; $\{g_{k,s}^{j,l}\}$ and $\{g_{k,e}^{j,l}\}$ represent the discrete probability densities of the sizes of the two kinds of bulks. We denote by $p_k^{j,l}$ the stationary probability that the l-queue of node j has k packets and by $\{p_k^{j,l}\}$ the sequence that represents the discrete probability density of the size of the l-queue in node j.

The model introduced above allows us to model the evolution of each l-queue as a continuous-time Markov chain with bulk arrivals and bulk services. The transition rate from a generic state h to the state $h+k$ is $A_k^{j,l} = \lambda_s^{j,l} g_{k,s}^{j,l} + \lambda_e^{j,l} g_{k,e}^{j,l}$, which is the average arrival rate of bulks of length k. On the other hand, if we consider the overall bulk arrival process with average rate $\lambda^{j,l} = \lambda_s^{j,l} + \lambda_e^{j,l}$, then $A_k^{j,l}$ can be expressed as $A_k^{j,l} = \lambda^{j,l} g_k^{j,l}$ where $g_k^{j,l}$ is the probability of a k-length arrival. So, $g_k^{j,l}$ can be simply computed as in (1):

$$g_k^{j,l} = \frac{\lambda_s^{j,l}}{\lambda_s^{j,l} + \lambda_e^{j,l}} g_{k,s}^{j,l} + \frac{\lambda_e^{j,l}}{\lambda_s^{j,l} + \lambda_e^{j,l}} g_{k,e}^{j,l} . \quad (1)$$

The quantity $g_{k,s}^{j,l}$ can be interpreted as an endogenous component, since it is associated with the packets generated inside the currently considered node, and $g_{k,e}^{j,l}$ as an exogenous component, since, in general, it depends on the forwarding strategies of the other nodes (see Section 3 for a class of possible models for $\{g_{k,e}^{j,l}\}$). The two terms $\frac{\lambda_s^{j,l}}{\lambda_s^{j,l} + \lambda_e^{j,l}}$ and $\frac{\lambda_e^{j,l}}{\lambda_s^{j,l} + \lambda_e^{j,l}}$ in (1) play the role of weights. We denote by $\{g_k^{j,l}\}$ the sequence composed by the $g_k^{j,l}$ values, which represents the discrete probability density of the size of the bulk (independently from its origin). The complete model of the Markov chain with its transition rates is shown in Figure 2. The model is derived similarly to the ones discussed by Kleinrock in [9, pp.134-139] for different problems, in which only the arrivals or only the services are in bulk form.

To simplify the notation, supposing to refer to node j and queue l, in the remainder of the letter we omit the superscripts j and l. So, (1) becomes

$$g_k = \frac{\lambda_s}{\lambda_s + \lambda_e} g_{k,s} + \frac{\lambda_e}{\lambda_s + \lambda_e} g_{k,e} . \quad (2)$$

It can be proved easily that for any choice of the discrete probability density $\{g_k\}$, the continuous-time Markov chain in Figure 2 is irreducible and positive recurrent and that these properties are inherited by its associated embedded Markov chain. For this case, it is known that the stationary discrete probability density $\{p_k\}$ always exists and is unique [4, Theorems 3 and 8, Chapter 5]. The next proposition from [3] provides a relationship between $\{p_k\}$ and

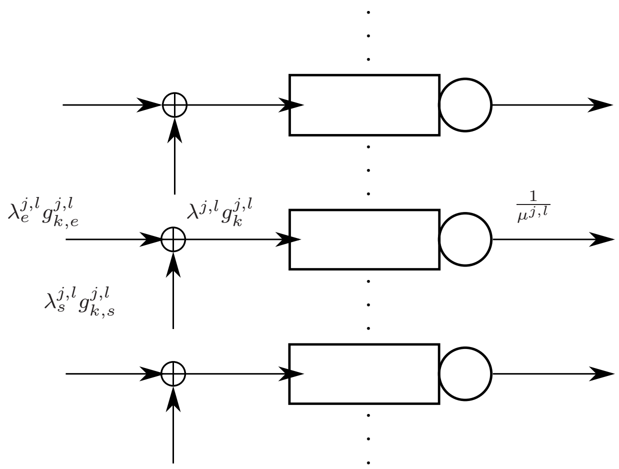


Figure 1: Model of the generic ICN node $j \in \{1, \ldots, M\}$ and its l-queue, where $l \in \mathcal{L}(j)$ ($k \in \mathbb{N}_0$).

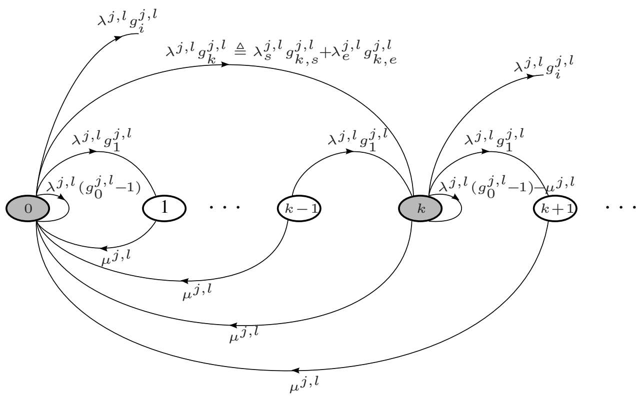


Figure 2: Transition rates for the continuous-time Markov chain related to l-queue inside node j.

$\{g_k\}$ by using their z-transforms $P(z) \triangleq \sum_{k=0}^{\infty} p_k z^{-k}$ and $G(z) \triangleq \sum_{k=0}^{\infty} g_k z^{-k}$.

Proposition 2.1 *$P(z)$ and $G(z)$ satisfy*

$$P(z) = \frac{\mu}{(\lambda + \mu) - \lambda G(z)} \,. \tag{3}$$

Note that such z-transforms have nonempty regions of convergence at least for $|z| > 1$. Indeed, considering for instance the case of the sequence $\{p_k\}$, one has

$$\sum_{k=0}^{\infty} |p_k z^{-k}| \leq \sum_{k=0}^{\infty} |z^{-k}| = \frac{|z|}{|z|-1} < \infty, \text{ for } |z| > 1.$$

3 A class of two-hop forwarding strategies

In two-hop forwarding, a packet can reach the destination only when one of these events occurs: 1) the source node meets the destination node; 2) the destination node meets another node that has previously received the packet from the source node itself. In this section and in the next one, we suppose that all the nodes of the network have the same traffic parameters and follow the same forwarding strategy. A possible forwarding strategy, which is a way to implement two-hop forwarding, is the following: inside each buffer, each node distinguishes between the packets generated by the node itself and the ones coming from the other nodes. When a node meets another one that is different from the destination, the latter sends all the packets generated by itself to the first node with probability q, otherwise no exchange is performed with probability $(1-q)$. In other words, suppose that the node analyzed j, which contains the l-queue under study, encounters the node $i \neq l$. Node i downloads to j all the packets generated by i and contained in its l-buffer with probability q. We denote by $\{\bar{p}_k\}$ the stationary probability that node i contains k packets generated by i itself in its l-buffer (such a probability does not depend on i, since by assumption all the nodes of the network have the same traffic parameters). The next corollary provides a relationship between the z-transforms $\bar{P}(z)$ and $G_s(z)$ of $\{\bar{p}_k\}$ and $\{g_{k,s}\}$, respectively. We denote by $\{\delta_{k,h}\}$ the Kronecker delta ($\delta_{k,h} \triangleq 1$ for $k = h$ and $\delta_{k,h} \triangleq 0$ otherwise).

Corollary 3.1 *$\bar{P}(z)$ and $G_s(z)$ satisfy*

$$\bar{P}(z) = \frac{\mu}{(\lambda + \mu) - (\lambda_s G_s(z) + \lambda_e)} \,. \tag{4}$$

Proof. The result is obtained by applying Proposition 2.1 neglecting the exogenous component (which does not influence the content of the portion of the buffer made up only of the internally generated packets), i.e., setting $\{g_{k,e}\} = \{\delta_{k,0}\}$, whose z-transform is $G_e(z) = 1$. Then one applies $G(z) = \frac{\lambda_s}{\lambda_s+\lambda_e} G_s(z) + \frac{\lambda_e}{\lambda_s+\lambda_e} G_e(z)$ (which follows by (2)). ■

So, turning back to the model of two-hop forwarding, the size of the portion of the l-buffer in the node i made only by the packets generated by i is k with probability $\bar{p}_k$, and i sends k packets to j with probability $\bar{p}_k q$. On the other hand, node i does not send anything to node j in two cases: the first one happens with probability $\bar{p}_0 q$, i.e., when the portion of the buffer in i used for l and composed only by the packets generated by i is empty; the second one happens with probability $(1-q)$ because of the forwarding strategy. More formally, $\{g_{k,e}\}$ is given by

$$\{g_{k,e}\} = (1-q)\{\delta_{k,0}\} + q\{\bar{p}_k\} \,, \tag{5}$$

for $q \in [0, 1]$. Corollary 3.2 provides an expression for the z-transform $P(z)$ of the sequence $\{p_k\}$ under the class of two-hop forwarding strategies (5).

Corollary 3.2 *If $\{g_{k,e}\}$ has the form (5), then*

$$P(z) = \frac{\mu}{(\lambda+\mu) - \left(\lambda_s G_s(z) + \lambda_e\left((1-q) + q\frac{\mu}{(\lambda+\mu)-(\lambda_s G_s(z)+\lambda_e)}\right)\right)} \,. \tag{6}$$

Proof. It is obtained by applying Proposition 2.1 with $\{g_{k,e}\}$ of the form (5) (whose z-transform is $G_s(z) = (1-q) +$

$q\bar{P}(z)$) and replacing $\bar{P}(z)$ by its expression provided by Corollary 3.1. ■

For the following analysis, the values of $P'(z)$ and $P''(z)$ (the first and the second complex derivatives of $P(z)$, respectively) computed at $z = 1$ are also needed. Starting from the expression of $P(z)$ in Corollary 3.2, simple computations provide the following corollary.

Corollary 3.3 *If $\{g_{k,e}\}$ has the form (5), then*

$$P'(1) = \frac{\lambda_s}{\mu} G'_s(1) + \frac{\lambda_e}{\mu} G'_e(1)\,, \tag{7}$$

$$\begin{aligned} P''(1) \;=\;& \frac{\lambda_s}{\mu} G''_s(1) + \frac{\lambda_e}{\mu} G''_e(1) + 2\frac{\lambda_s^2}{\mu^2}(G'_s(1))^2 \\ &+2\frac{\lambda_e^2}{\mu^2}(G'_e(1))^2 + 4\frac{\lambda_s\lambda_e}{\mu^2} G'_s(1)G'_e(1)\,, \end{aligned} \tag{8}$$

where

$$G'_e(1) = q\frac{\lambda_s}{\mu} G'_s(1)\,, \tag{9}$$

$$G''_e(1) = q\frac{\lambda_s}{\mu} G''_s(1) + 2q^2\frac{\lambda_s^2}{\mu^2}(G'_s(1))^2\,. \tag{10}$$

Note that the computations of formulas (7) and (8) do not require inverting z-transforms.

4 Buffer occupancy

The analysis detailed in the previous sections allows us to analyze the average buffer occupancy (11) and its standard deviation (12):

$$\sum_{i=0}^{\infty} i p_i = -P'(1)\,, \tag{11}$$

$$\sqrt{\sum_{i=0}^{\infty}\left(i - \sum_{k=0}^{\infty} k p_k\right)^2 p_i} \;=\; \sqrt{P''(1)(P''(1) - 2P'(1))} \tag{12}$$

Formulas (11) and (12) are checked by exchanging the order of differentiation and summation in the definitions of $P'(z)$ and $P''(z)$, then taking $z = 1$. Figure 3 shows the behaviors of the average buffer occupancy (11) and its standard deviation (12) for the class of forwarding strategies (5), by varying the parameter q and the values $\mu, \lambda_e, \lambda_s$. Note from the figures that both expressions are linear in q. For illustrative purposes, we consider for the endogenous component $\{g_{k,s}\}$ a model in which the conditions $G'_s(1) = -1$ and $G''_s(1) = 2$ hold. An example of such a model is

$$\{g_{k,s}\} = \{\delta_{k,1}\}\,, \tag{13}$$

whose z-transform is

$$G_s(z) = z^{-1}\,, \tag{14}$$

i.e. all bulks are composed of 1 packet.

Similar curves can be obtained for more complex models.

Figure 4 shows the behaviors of the average buffer occupancy (11) and its standard deviation (12) for a Poisson discrete probability density

$$\{g_{k,s}\} = \left\{\frac{a^k e^{-a}}{k!}\right\}\,, \tag{15}$$

($a > 0$ is a parameter), whose z-transform is

$$G_s(z) = e^{-a(1-z^{-1})}\,. \tag{16}$$

where $a = 3$, $G'_s(1) = -3$ and $G''_s(1) = 15$.

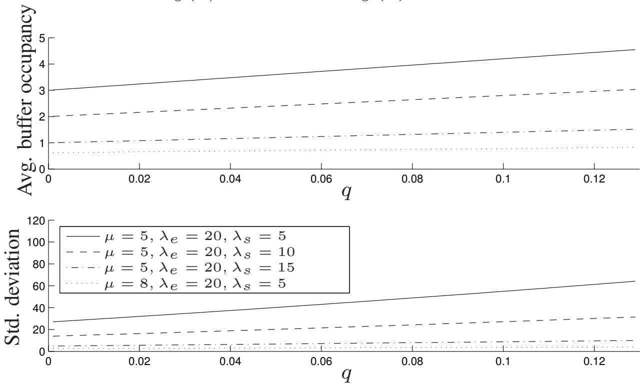


Figure 3: Average buffer occupancy and its standard deviation for the class of forwarding strategies (5), by varying the parameter q and considering a model for the endogenous component $\{g_{k,s}\}$ for which one has $G'_s(1) = -1$ and $G''_s(1) = 2$.

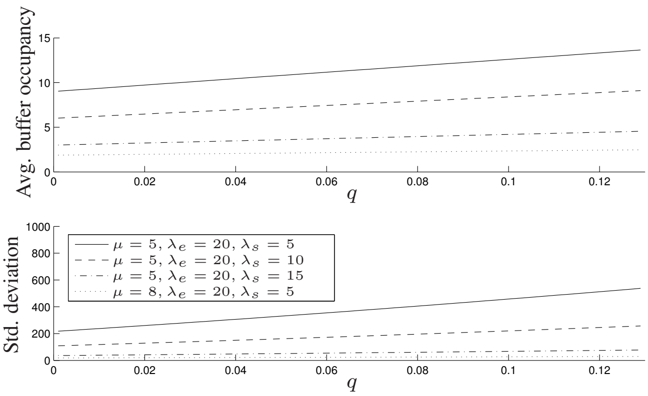


Figure 4: Average buffer occupancy and its standard deviation with a Poisson discrete probability density for the endogenous component $\{g_{k,s}\}$ for which $G'_s(1) = -3$ and $G''_s(1) = 15$.

5 Related Literature

An elegant model was proposed in [11] to analyze the delivery delay and its relative trade-offs with energy consumption and buffer requirements in the so-called (p, q)-*epidemic* *routing*. In [11] p and q represent, respectively, the probability that a node accepts a packet copy from another node when none of them is the source and the probability that a node accepts a packet copy from the packet source node. With a proper tuning of the values of p and q, (p, q)-epidemic routing models flooding, randomized flooding, or two-hops forwarding. The model is based on a continuous-time Markov chain, in which the state represents the number of copies of a specific packet in the system.

In [15], the authors developed a mathematical framework based on a Markov chain to get insights into the global congestion behavior. Their analysis is greatly simplified by replacing some random variables in the model with their expected values.

Differently from [11] and [15], in this paper we have focused on the behavior of a network single node, estimating both the discrete probability density of the size of its l-queue and the exogenous component $\{g_{k,e}\}$ of $\{g_k\}$ for a class of two-hop forwarding strategies. In [3], a similar analysis was done for a different class of forwarding strategies used in epidemic routing.

6 Conclusions

We have applied to a class of two-hop forwarding strategies a relationship between the discrete probability densities of bulk and queue sizes, which, under fixed traffic rates, depends only on the traffic generated by single nodes towards a specific destination. This allows to compute the average buffer occupancy and its standard deviation for a specific queue that contains traffic for a given destination within a node, knowing only the discrete probability density of the size of the bulks generated by that node towards the given destination. This has immediate practical advantages, e.g., in congestion control. Although the results are formulated for a single destination, they can be extended to the case of traffic directed to multiple destinations with possibly different generation rates.

References

[1] J. Broch, D. A. Maltz, D. B. Johnson, Y.-C. Hu, and J. Jetcheva, “A performance comparison of multi-hop wireless ad hoc network routing protocols,” Proc. 4th Annual ACM/IEEE Int. Conf. on Mobile Computing and Networking, pp. 85–97, 1998.

[2] S. Burleigh, A. Hooke, L. Torgerson, K. Fall, V. Cerf, B. Durst, K. Scott, and H. Weiss, “Delay-tolerant networking: An approach to interplanetary internet,” *IEEE Communications Magazine*, Vol. 41, No. 6, pp. 128 – 136, June 2003.

[3] M. Cello, G. Gnecco, M. Marchese, and M. Sanguineti, “A model of buffer occupancy in ICNs,” *IEEE Communications Letters*, to appear.

[4] R. G. Gallager, *Discrete Stochastic Processes*, Kluwer, 1996.

[5] R. Groenevelt, P. Nain, and G. Koole, “The message delay in mobile ad hoc networks,” *Performance Evaluation*, Vol. 62, No. 1-4, pp.210–228, 2005.

[6] P. Hui, J. Crowcroft, and E. Yoneki, “Bubble rap: Social-based forwarding in delay tolerant networks,” Proc. 9th ACM Int. Symp. on Mobile ad hoc Networking and Computing, pp. 241–250, 2008.

[7] S. Jain, K. Fall, and R. Patra, “Routing in a delay tolerant network,” Proc. 2004 Conf. on Applications, Technologies, Architectures, and Protocols for Computer Communications, pp. 145 – 158, 2004.

[8] P. Juang, H. Oki, Y. Wang, M. Martonosi, L. S. Peh, and D. Rubenstein, “Energy-efficient computing for wildlife tracking: Design tradeoffs and early experiences with zebranet,” *SIGPLAN Not.*, Vol. 37, No. 10, pp. 96–107, October 2002.

[9] L. Kleinrock, *Theory, Volume 1, Queueing Systems*, Wiley-Interscience, 1975.

[10] Q. Li and D. Rus, “Communication in disconnected ad hoc networks using message relay,” *J. Parallel Distrib. Comput.*, Vol. 63, No. 1, pp. 75–86, January 2003.

[11] T. Matsuda and T. Takine, “(p,q)-epidemic routing for sparsely populated mobile ad hoc networks,” *IEEE J. on Selected Areas in Communications*, Vol. 26, No. 5, pp. 783–793, June 2008.

[12] N. Sadagopan, F. Bai, B. Krishnamachari, and A. Helmy, “Paths: Analysis of path duration statistics and their impact on reactive manet routing protocols,” Proc. 4th ACM Int. Symp. on Mobile ad hoc Networking and Computing, pp. 245–256, 2003.

[13] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, “Spray and wait: An efficient routing scheme for intermittently connected mobile networks,” Proc. 2005 ACM SIGCOMM Work. on Delay-Tolerant Networking, pp. 252–259, 2005.

[14] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, “Efficient routing in intermittently connected mobile networks: The single-copy case,” *IEEE/ACM Trans. Netw.*, Vol. 16, No. 1, pp. 63–76, February 2008.

[15] N. Thompson, S.C. Nelson, M. Bakht, T. Abdelzaher, and R. Kravets, “Retiring replicants: Congestion control for intermittently-connected networks,” Proc. INFOCOM, pp. 1–9, March 2010.

[16] Y.-C. Tseng, S.-Y. Ni, Y.-S. Chen, and J.-P. Sheu, “The broadcast storm problem in a mobile ad hoc network,” *Wirel. Netw.*, Vol. 8, No. 2/3, pp. 153–167, March 2002.

[17] A. Vahdat and D. Becker, “Epidemic Routing for Partially-Connected ad hoc Networks,” Technical report, Department of Computer Science, Duke University, April 2000.

[18] H. Wu, R. Fujimoto, R. Guensler, and M. Hunter, "Mddv: A mobility-centric data dissemination algorithm for vehicular networks," Proc. 1st ACM Int. Work. on Vehicular ad hoc Networks, pp. 47–56, ACM, 2004.

[19] X. Zhang, G. Neglia, J. Kurose, and D. Towsley, "Performance modeling of epidemic routing," *Computer Networks*, Vol. 51, No. 10, pp. 2867–2891, July 2007.

[20] W. Zhao, M. Ammar, and E. Zegura, "A message ferrying approach for data delivery in sparse mobile ad hoc networks," Proc. 5th ACM Int. Symp. on Mobile ad hoc Networking and Computing, pp. 187–198, 2004.

Natural language processing based Services Composition for Environmental management

Armando Ordoñez, Juan C. Corrales
Telematics Engineering Group
University of Cauca
Popayán,Colombia
{jaordonez,jcoral}@unicauca.edu.co

Paolo Falcarin
ACE School
University of East London
London, UK.
falcarin@uel.uk

Abstract - *Automated services composition is an active research area nowadays. The application of this technology to specific domains, as well as the use of basic Telecommunications services such as call or SMS, offers new challenges. This paper presents the main components of an architecture for automated composition for Environmental Management. The proposed architecture uses Natural Language processing and AI Planners techniques domains. Although the present architecture is focused in a particular domain, most of the principles discussed here are applicable to similar scenarios.*

Keywords: AI planning, NLP, automated composition.

1 Introduction

As available services grow exponentially and are updated on the fly, it has become impossible for human capacity, to analyze all of them and generate manually a composition plan [1]. The use of new technologies in services composition field (like semantic annotations and AI planners) allows visualizing an automated composition.

Previous works from Academia [2][3][4] and Enterprise [5][6][7] have emerged around this topic. However, few academic approaches include implementations in production scenarios. On the other hand, enterprise works, are usually closed and hard to adapt to particular domains. This paper gathers important experience from previous works and presents an architecture for automated services composition in the Environmental Management. The architecture is based on AI Planning with user context information (that can include devices, network and user profile).

The Environmental management domain contains particular features: i) composed services can be formed by Web Services or by basic Telecommunication features like Call or send SMS, ii) the number of services can be limited due to the specialized field of action, iii) finally, in environmental management the procedures for emergencies handling and monitoring are standard. Therefore, these procedures and the associated services can be described using semantic annotations by experts. The latter is a very important issue, since few service providers have taken up the opportunity to mark-up their services, for the simple reason that they don't envisage the use of their services by an automated planning system [8]. The latter makes possible the application of semantic techniques in this domain. Although the present architecture is focused in a particular domain, most of the principles discussed here are applicable to similar scenarios.

This paper is organized as follows: section 2 presents the motivating scenario. Section 3 describes the proposed approach. Section 4 presents related work and Section 5 draws the conclusions.

2 Environmental management Domain

To illustrate how decision making takes place in the practice, we present a sketch of an environmental management system (see Fig. 1). The Environmental manager is on charge of decision making about environmental alarms and crops. In order to do so, the manager has information from sensor networks. The manager can also use Telecommunication and Web services to process basic data and send information to both farmers and sensors. Reuse of functionalities is a very important issue for some developing countries where budgets for technologies are limited.

Usually, environmental manager is an ecologist or biological expert. Therefore, his knowledge about underlying technologies is usually low. The preferred way to introduce information to the system is voice. Commonly the user expresses his request in an informal way; we will illustrate this with two examples:

- "I need calculate hydrological balance of zone one and receive the resulting map to my mobile"
- "the river flow of zone two is greater than 15% of average, emit an alarm to every farmer within a radius of 2 miles from the river"

The first request can be entered through voice, by a mobile device. Next, the system gather information of sensors of zone one (preconfigured in the system or specified by coordinates). The system uses hydrological services from Internet, sending sensor data and maps from Google maps. Finally the resulting image is sent by MMS to the user mobile.

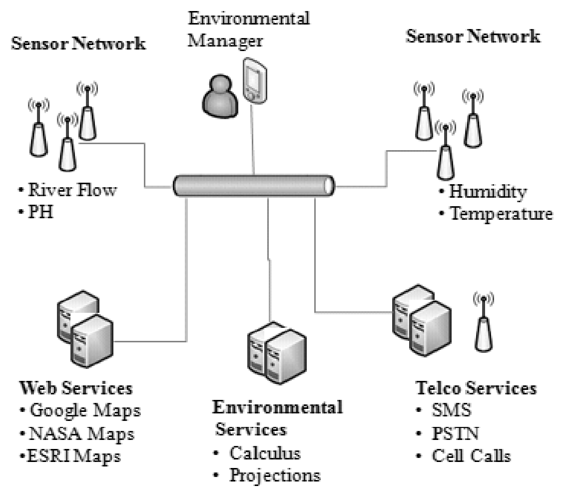


Figure 1. Telecommunication and Web services interaction in Environmental Management Systems

In the second request, sensor data are evaluated. If necessary, an emergency map is generated. This map is created drawing a radius of 2 miles from the sensor. To do so, the system uses SIG services and maps from internet. Finally, the system informs about the alarm to farmers inside the emergency area; the best way to send the information is selected: SMS, Cell Phone call, fixed telephone call, voice message. In both cases, services from Web and Telecommunications are used. These services work together and in coordination to save lives or help to make decisions about crops.

3 Our architecture

Next, we describe the architecture of our approach for automated composition (see Fig. 2.).

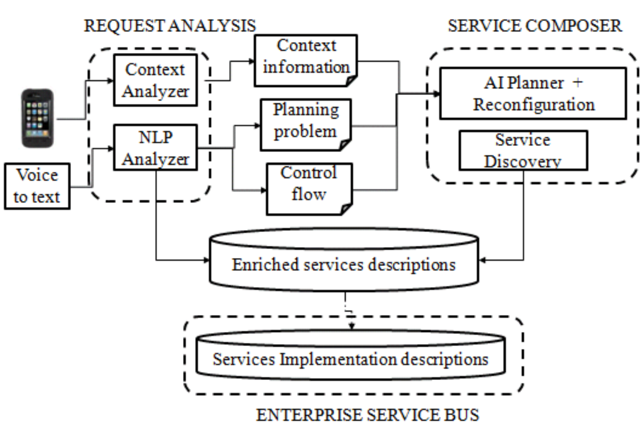


Figure 2. Architecture for NLP based automated planning

The first step of the architecture deals with Natural language processing (NLP). In the literature other alternatives have been found for user request treatment, such as Mashups [9] and Services creation environments [10]. However the natural language offers a better mechanism for end users without expertise to express their requests [11].

The *NLP Analyzer* extracts the *Planning problem* from user request. At the same time, it gathers information about control flow and context. For instance, if the request is: "*check sensors and inform farmers*". NLP Analyzer obtains: *(check (Sensors)) AND (inform (farmers)). "Check"* and *"Inform"* are translated to problems for the planner and *"AND"* describes a concurrent execution that will be considered latter in the *Service Composer*.

The *Context Analyzer* allows extract user context information and is mapped to services criteria. This information is used for *Services Discovery* in a later stage. The user context information analysis is performed in two steps: first the user profile is analyzed looking for preferences; Secondly device references are checked in capabilities repositories like the Wireless Universal Resource File (WURFL) and the Composite Capability/Preference Profiles (CC/PP) in order to analyze devices capabilities and features. Table 1 shows user context criteria identified from the request. Table 2 shows how these user criteria are mapped to services criteria.

Table 2. User context Criteria

	User Criteria	Values
User context	Network	GPRS/ WLAN/ GSM
	Device	Cell phone, Laptop
	Location	Outdoor, indoor
User preferences	Data subscription	Yes/No
	Only Free services	Yes/No
	Voice subscription	Yes/No
	Delivery quality	low, medium, high

Table 2. User-Service Criteria Mapping

User Criteria	Service criteria	Weight
Network	payload size	bytes
Device	payload size	bytes
Location	voice , text	integer
Data subscription	require subscription	Boolean
Only Free services	Cost	value
Voice subscription	voice, text	Boolean
Delivery quality	delivery warranty	integer

With the context information, the planning problem and the control flow, the *AI Planner* component obtains a plan and executes it, taking care of reconfiguration in case of to be necessary. This plan uses the *Services Discovery* component to select suitable services for each action.

Next, a description of each component of the architecture is presented.

3.1 Request Analysis

This component holds two sub components: *Context and NLP analyzer*. In this component, User Requirements are modelled with a query Q specified as a couple $<R;P>$, where R represents the request part of the query and P represents the user preferences of the query (device, network, position). In turn, R is specified as an n-uple $R= [\{s_1,s_2 \ldots s_n\}, F]$. Each s_n denotes one *Planning Problem* and F represents control flow information. For its part, each s is composed of 3-uples $<I;O;C>$; where I denotes the input data the user provides, O denotes required information to be provided as a result of the query and C denotes a functionality (associated with a service). For example:

Q: is a request made by an environmental manager from a cell phone. Where:

P: Cell phone reference and network capabilities.

R: "*I need calculate hydrological balance of zone one and receive the resulting map to my mobile*".

From R we can expand:

s_1: "*calculate hydrological balance of zone one*"

s_2: "*receive the resulting map to my mobile*"

F: *AND* (sequence of actions)

Analyzing s_1: "*calculate hydrological balance of zone one*"

I = Zone one. (A system variable, geo-coded location or a set of coordinates)

O = Hydrological balance map

C = Calculate hydrological balance service.

The previous processing is done by phases: the user expresses his need through a sentence: "*check sensors and inform farmers*". In order to retrieve the services required to satisfy the user need, the request goes through a linguistic processing module, responsible for: Text segmentation required to separate the words in the phrase; Removing stop words that are considered to be irrelevant (e.g. want, the, to, and); Stemming (e.g. checks becomes check); Spell-checking to correct the misspelled words and the words "damaged" during stemming. Some techniques used to implement the modules above are GateNLP, OpenNLP, Apache UIMA, among others. One of the most important is GateNLP [12], which offers an architecture that contains functionality for plugging in all kinds of NLP software: (taggers, sentence splitters, named entity recognizers) and all are Java based. The output text segments for the user request in the considered scenario are: *check, sensor, and, inform, farmers.*

Once completed these operations, the request is more consistent, but remains complex. Therefore, we established the *Named Entity Recognition*, which performs a classification between "*Control*" and "*Functional*" words according to its meaning, from which, control words are directed to *AI Planner* Module, whereas functional words are directed to *Semantic Analyzer*. The Functional Words are classified in *parameters* and *goals* for the planning problem (the request in AI planning language) in PDDL [13], some parameters are checked in the contact list, or other local repositories. For example "*Mark*" is searched and the number is replaced in the request, equally "*Zone 1*" is replaced by the coordinates of the zone.

3.2 Services composer Module

Recently, several authors have addressed the potentials and drawbacks of applying AI planning techniques to Web services composition. Between the issues of services composition there is the problem of changing context. Nowadays, this is more interesting since user can changes frequently of device and location. Therefore, devices capabilities can determine the type of service that could be provided. Equally, the problem of incomplete information may be considered [8]: Many planning problems in the Web service domain require the querying of information about the world to make decisions, for instance to check the value of one variable in a sensor in a given moment. Therefore, assumption of a complete initial state description is not valid. Further, there is the problem of nondeterministic behaviour of services: service execution can fail during or get unwanted results, for instance when a call cannot be established with a farmer, or the value of a sensor measure is maladjusted.

To deal with these issues, our approach for the planning module is based on the work of Peer [14]. In the present planning module (fig. 3) an AI planner is embedded in an execution monitoring engine with monitoring functions. This module translates Web service composition goals to planning problems. Next, the planning module executes the generated plans using the Businesses process engine. Finally, this framework monitors the progress of the problem solving process and to react according to predefined rules. Next, a deeper description of the three levels is presented:

Level1 provides access to AI planner, in this level the basic problem of generate a plan is performed. The level 1 can support different planning methods through the standardization of the interface and the language for problems and domain description PDDL. In this level, there exist the assumption of complete knowledge and fully deterministic operations.

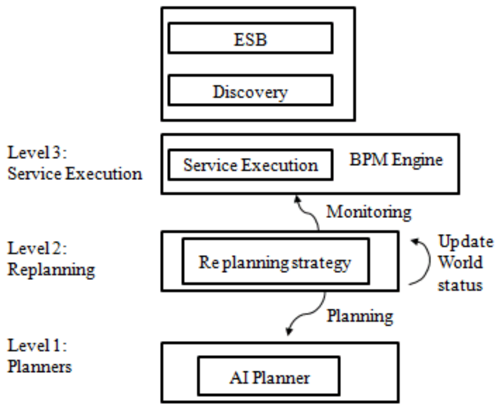


Figure 3. Internal Architecture of services composer module

Level 2 is on charge of monitoring continually the services execution. As long as the execution of the process is performed, the world status changes too. In case of a failure in execution, this level is on charge of request the level 1 to create another plan with the new world status corresponding to the status left by the older process.

Level 3 is on charge of execute the services, to do so this level uses the *control flow* information from *request analyzer* and creates a executable process that discover real services from the *Service Discovery* Module.). This plan is translated to BPMN and executed in a JBPM Engine.

The *Services* D*iscovery Module* is on charge of select services for implementing the functionalities that the planner can require. One service description can represent many implementations of the service. This is reached through the association of WSDL with different *Endpoints*. So, when a problem with a particular service appears, another implementation of the service can be selected. Besides, if a service fails during execution, a related service can be selected. For selection of services, a scoring function based in LSP [15] is used. To perform de services discovery the criteria from the context are used in the formula. This selection creates a ranking that is used for the re-planning module in case of the failure of one of the services during execution time. I.e. if the first selected service (preferred according to the user preferences) fails, the re-planning module uses the next in the ranking.

3.3 Enriched services description

For services description, our approach uses a semantic language. Effectiveness of semantic description is widely known, as well as some disadvantages related with difficulties in Ontology creation and high processing time. For our case study, we define a domain ontology [16] that describes the most important elements of the domain focusing our efforts in early warnings for disaster prevention and crops control. In this manner, the concepts and relation don't become son numerous.

SESMA is selected as the semantic language for services description in our approach. Nowadays, other proposals for mark-up exist. Among them are: OWL-S [17] and WSMO/WSML [18]. SESMA is easy to use and provides convenient support to mark up nondeterministic service operations. The syntax and semantics of the SESMA constructs are described in detail by Peer et al [14]. In the SESMA model, each service consists of a set of operations. Each operation can be described by four elements: a unique identifier of the operation, a set of preconditions, a set of effects, and a set of input and output variables.

SESMA can be used to annotate service descriptions based on WSDL and BPEL fragments. In our approach, this process is conducted by domain experts in using standard protocols for environmental management.

3.4 Services Implementation description

The *Service Implementation Description* module is based on an ESB. This module provides the connectivity layer between services. The description of a service is wide; it is not restricted by a protocol, which connects a service requestor to a service provider. In this manner, service description is completely isolated from real implementation. This way, the planner can use services regardless the implementation details.

On the other hand, telecommunications services have some intrinsic differences with Web services, these differences are clearly described by Bond et al. [19]. Nevertheless there is a new trend that has made possible count with Telecommunication functionalities in the Cloud. New business players like Twillio [20] and Skype [21] offers SMS, Calls and all services through Internet. In this vein, using SOA components like Enterprise services Bus (ESB) is perfectly acceptable for domains using basic Telecommunications services.

4 Related Work

Our approach is focused on automation of user centred service composition, and we intend to apply our framework to convergent environments. Previous works have proposed frameworks for automated services composition, like Kim et.al [22]. The authors present phases for automatic composition, focusing only on web domain without concern on execution phase.

Shia et al. [23] and da Silva et al. [24] present frameworks for automatic composition. These frameworks exploit natural language processing and semantic annotations for services matchmaking based on SPATEL language. They do not address the validation of the non-functional properties and are focused only on the request analysis and plan generation. Our approach deal with all the phases including execution and reconfiguration based on JSLEE environments.

Sirin et al. [25], provide an algorithm to translate OWL-S service descriptions to a SHOP2 domain and makes planning based on services. However, this works lacks of details of implementation in real environments and focus only on Web services. Other authors present approaches for Web service composition based on AI planning with preferences [26] [27]. However, they propose extensions to standard planning language and adaptation of planners, adding new levels of complexity to the automated composition process.

Zhu et al. [28] present Hybrid Service Creation and Execution Environment (HSCEE), a Template-based service creation platform with low latency service execution. HSCEE is based in on BPEL templates but most of the design tasks are manually. Equally natural language processing is not considered.

Table 2 outlines a comparison of related work, based in the following criteria: first, if the work deals with all the phases in the service composition process including reconfiguration. Second, if the approach for service composition is user-centred. Third, if the approach takes in account convergent considerations.

TABLE I. COMPARISON OF RELATED WORKS

works	Include all phases	User centred	Consider Convergence
[22]	No, just the request analysis and the service creation	No	No, they focus on web domain
[23] [24]	No, just the request analysis and the service creation	No	Yes
[25] [26][27]	No, just the OWL-S based planning	No	No, focused on Web domain
[28]	Yes, but not all of them are automatic or Natural language based.	Yes	Yes

Acknowledgements

The authors thank University of East London - Telcomp project for supporting the Ph.D. student Jose Armando Ordóñez.

5 Conclusions and future work

Automated composition is a very intensive research area; most of the approaches presented until now lacks of results in production environments. Equally some domains like environmental management require of specific adaptations. This paper describes the general components of architecture for automated services composition applied to environmental management. Specific elements of each module are detailed previously [29][30]. The present architecture performs an analysis of the request in Natural Language and translates it to PDDL. Likewise, the architecture includes mechanism for Services discovery and Re-plannig for fault handling.

Future works include the development of mechanisms for automation of deployment of services in Telecommunication environments. These environments will provide more reliability in order to support a wide amount of telecommunication services and provide a higher reliability. Equally, automated creation of planning domains will facilitate the use of the architecture in different domains with a higher number of services.

References

[1] S. Oh, D. Lee, S. Kumara, "A Comparative Illustration of AI Planning-based Web Services Composition" ACM SIGecom Exchanges. Vol. 5, pp. 1-10, 2005.

[2] P. Falcarin, "Service Composition Quality Evaluation in SPICE Platform," High Assurance Services Computing. pp. 89-102, 2009

[3] Yelmo, J.C et al., "A User-Centric service creation approach for Next Generation Networks," Proc. IEEE, ITU-T/IEEE Innovations in NGN: Future Network and Services (K-INGN 08), IEEE Press. May. 2008

[4] Omelette Project (2012) [Online]. Available: http://www.ict-omelette.eu/home

[5] SIRI Site. (2012) [Online]. Available: http://www.apple.com/iphone/features/siri.html

[6] Zypr Site (2012) [Online]. Available: http://www.zypr.net/

[7] Vlingo site (2012) [Online]. Available: http://www.vlingo.com/

[8] M. Carman , L. Serafini , P. Traverso, "Web service composition as planning," Proc. ICAPS 2003 Workshop on Planning for Web Services, Trento, July 2003.

[9] Z. Zhao, S. Bhattarai, J. Liu, Ji and N. Crespi, “Mashup services to daily activities: end-user perspective in designing a consumer mashups,” Proc. 13th International Conference on Information Integration and Web-based Applications and Services, pp.222-229, Ho Chi Mihn, December 2011.

[10] N. Laga, “Widgets and composition mechanism for service creation by ordinary users”, IEEE Communications magazine, Vol. 50, Issue 3, pp. 52-60, March 2012.

[11] K-H. LEE, J. Lim, “Constructing Composite Web Services from Natural Language Requests,” Journal of Web Semantics, Vol. 8, No. 1, 2010.

[12] H. Cunningham, V. Tablan, I. Roberts, M. A. Greenwood, and N. Aswani, “ Information Extraction and Semantic Annotation for Multi-Paradigm Information Management” M. Lupu, K. Mayer, J. Tait, and A. J. Trippe, editors, Current Challenges in Patent Information Retrieval, Vol. 29. Springer, 2011.

[13] C. Knoblock, A. Barrett, D. Christianson, M. Friedman, C. Kwok, K. Golden, S. Penberthy, D. E Smith, Y. Sun and D. Weld, “PDDL — The Planning Domain Definition Language,” Computer and Information Science, Vol. 78, pp. 1-27, 1998.

[14] J. Peer, Description and Automated Processing of Web Services. Switzerland. University of St. Gallen, 2006

[15] H.Q. Yu, and S. Reiff-Marganiec. “A Method for automated Web Service Selection,” Proc. IEEE Congress on Services - Part I, pp. 513-520, July 2008.

[16] W. Xu, S. Zlatanova, “Ontologies for Disaster Management Response,” Geomatics Solutions for Disaster Management, Lecture Notes in Geoinformation and Cartography, Springer-Berlin, pp. 185-200, 2007.

[17] W3C (2004). OWL-S W3C Member Submission [Online]. Available: http://www.w3.org/Submission/OWL-S/

[18] WSMO (2004). WSML Working Draft [Online]. Available: http://www.wsmo.org/2004/d7/d7.2/v1/

[19] G. Bond, E. Cheung, I. Fikouras, R. Levenshteyn, “Unified telecom and Web Services Composition: problem definition and future directions.” Proc. In: 3rd Principles, Systems and Applications of IP Telecommunications (IPTComm’09), ACM, July 2009.

[20] Twilio site (2012) [Online]. Available: http://www.twilio.com

[21] Skype API (2012) [Online]. Available: http://developer.skype.com/

[22] A. Kim, M. Kang, C. Meadows, E. Ioup, J. Sample, A Framework for Automatic Web Service Composition,”. Washington D.C.: US Naval research Lab, 2009.

[23] M. Shiaa, et al., “Towards the Automation of the Service Composition Process: Case Study and Prototype Implementations,” Proc. ICT Mobile and Wireless Communications Summit (ICT-MobileSummit 2008), June, 2008.

[24] Da Silva, E. Goncalvez, L. Ferreira, M. Van Sinderen, “Towards runtime discovery, selection and composition of semantic services,” Computer Communications, vol. 34, pp. 159-168, Feb., 2011.

[25] E. Sirin, B. Parsia , D. Wu, J. Hendler, D. Nau, D, “HTN Planning for Web Service Composition using SHOP2,” Proc. Web Semantics: Science, Services and Agents on the World Wide Web In International Semantic Web Conference 2003, Oct. 2004.

[26] N. Lin, U. Kuter, E. Sirin, “Web Service Composition with User Preferences,” Proc. European Semantic Web Conference (ESWC2008), Springer-Verlag, May 2008.

[27] S. Sohrabi, N. Prokoshyna, S. McIlraith, S, “Web Service Composition via Generic Procedures and Customizing User Preferences,” Proc. European Semantic Web Conference (ESWC2008), Springer-Verlag, May 2008.

[28] D. Zhu, Y. Zhang, B. Cheng, B. Wu, and J. Chen, “ HSCEE: A Highly Flexible Environment for Hybrid Service Creation and Execution in Converged Networks,” Journal of Convergence Information Technology, Volume 6, Number 3. March 2011.

[29] E. Pedraza, J. Zúñiga, L. Suarez, J.C. Corrales, “Automatic Service Retrieval in Converged Environments Based on Natural Language Request,” Proc. The Third International Conferences on Advanced Service Computing, pp. 52-60, Rome, Sep. 2011.

[30] A. Ordonez, J.C. Corrales, P. Falcarín, J. Guzman, “User Centred Automated Composition in Telco 2.0,” The Fourth International Conferences on Advanced Service Computing. Nice, July 2012 (in revision).

Proposal of Symbiosis ADS concept and Negotiation Support Methods for Cooperative Resource Allocation

Koichiro IIJIMA
Yokohama Laboratory
Hitachi, Ltd.
Japan
koichiro.iijima.kd@hitachi.com

Takashi FUKUMOTO
Yokohama Laboratory
Hitachi, Ltd.
Japan
takashi.fukumoto.kx@hitachi.com

Michiki NAKANO
Yokohama Laboratory
Hitachi, Ltd.
Japan
michiki.nakano.yt@hitachi.com

Abstract - *We propose a conceptual system model for public infrastructure systems, the Symbiosis-Autonomous Decentralized System (Symbiosis-ADS) and negotiation support methods. In this conceptual model, each agent system is owned by an individual stakeholder in a city, including individual residents or companies, and operates autonomously depending on the owner's decisions. These agent systems disclose information about their resource consumption and supply to other agents. An agent that has over consumed (resource shortage) and another agent that has under consumed (resource surplus) communicate and make an agreement, then one gives or lends resources to the other (accommodation). For this conceptual model, we developed negotiation and planning methods for supporting each agent's and owners' decision making by minimizing the negative effect of information incompleteness and environmental uncertainty. With a prototype system and simulation, we evaluated the effectiveness of these negotiation methods in a situation with information incompleteness.*

Keywords: Autonomous Decentralized System, Multi agent negotiation Simulation, Energy Management System, Factory EMS

1 Introduction

With increasing energy consumption and rising environmental awareness, public infrastructures and smart cities, where public resources are efficiently used, are needed. In a smart city, infrastructure systems must be able to adapt to the growing number of residents and changing environmental situations. Therefore, it is important that technology based on IT systems enable communication between several public infrastructure systems, for example, power, railway, transportation, and resident-side systems or equipment, to automatically control all the systems in the city to balance supply and demand and to support human activities.

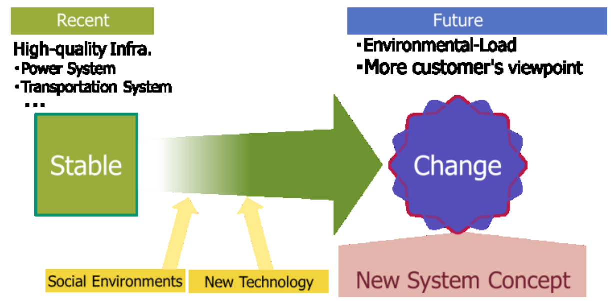


Figure 1. Aims of Technology Development for Public Infrastructure.

From the resource allocation view, future public infrastructure systems should operate not only in accordance with pre-defined scheduling but also should adapt to changes in residential, social, and environmental needs. For example, in electric power infrastructure systems, prediction and pre-planning of energy demand for several decades are not possible, and re-construction of public infrastructure systems greatly impacts the environment.

Therefore, future public infrastructure systems should result in dynamic optimal resource allocation depending on the expansion of or changes in a city and, in some cases, communicate and cooperate with residents to reduce demand and supply fewer resources. In electric power infrastructure systems, such as a community energy management system [5], one type of demand response (DR) system [6] enables optimal resource (energy) allocation by coordinating the demand and supply sides for preventing power shortages and reducing the CO_2 emissions.

2 Purpose

Our goal was to develop a cooperative system between a large variety of stakeholders in a city, including residents, factories, and infrastructure suppliers. With this system, a stakeholder can easily give or lend surplus resources to the other stakeholders, so (semi-) optimal resource allocation throughout the city is possible, minimizing the shortage of resources and the disadvantage to stakeholders.

The important issues in the cooperative system within multi-owner systems are information incompleteness (unobservability beyond the ownership boundary) and environmental uncertainty (uncontrollability beyond the system's ownership boundary).

Information incompleteness occurs when cooperating systems cannot always determine the other stakeholders systems' inner state or inner data due to private data hiding, unreliable communication paths, or forging data for the owner's specific reasons. Environmental uncertainty occurs when cooperating systems cannot always adjust to the other systems' resource usage or come to an agreement with other systems to reduce resource usage because of the other systems' judgments based on their preferences or purposes.

3 Proposed Conceptual Model

We propose a public infrastructure conceptual system model system called "Symbiosis - autonomous decentralized system (Symbiosis-ADS)". In the Symbiosis-ADS, all agent systems (owned by individual stakeholders, including individual residents or companies) have autonomy and can decide how to use their resources and control their subsystems and equipment that consume resources. For cooperation and optimal resource allocation, each agent discloses information to other agents about its resource consumption and supply. An agent that has over consumed (resource shortage) and another agent that has under consumed (resource surplus) communicate and make an agreement, then one gives or lends (accommodates) resources to the other.

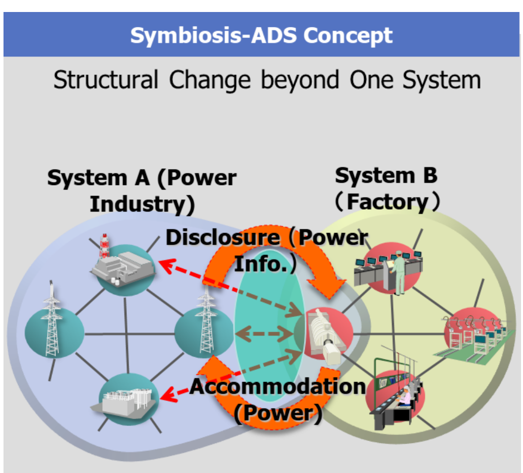


Figure 2. A Concept Model of Symbiosis-ADS

In our conceptual system model, there is no central planning or control system because public infrastructure systems must be highly reliable and extensible. If one agent system (or the central system) crashes, cooperation and resource allocation functionality should be active in the remaining agent systems. Residents and companies in a city may move away from the city and other residents and companies may move in from other cities, so the cooperative system must be highly and unpredictably extensible so that new participants and agent systems can participate in or be removed from the cooperative system. Therefore, the Symbiosis-ADS conceptual model has no central system or a system with limited central-management functionalities and distributed cooperation and resource allocation functionalities in each agent system. The distributed system model is natural because all stakeholders (and their agent systems) in a city make autonomous individual judgements and have their own preference and purposes (judgement criteria).

Each agent system has two groups of functionalities. One is pre-existing, resource consuming and supplying management/control components or equipment. The other is Symbiosis-ADS-specific components for communication and cooperation with other agent systems and negotiation planning of resource allocation. Therefore, each agent system has two layers of sub-system groups corresponding to the two functionalities; information service function layer (ISL) and information control layer (ICL).

A component in the ISL manages resource states in all cooperative systems. It makes resource demand and supply plans and collects usage and supply history on the system. It also receives or predicts resource demand and supply states on other agent systems. It plans how to use and supply resources. For example, if a resource shortage may occur, it plans the saving usage of resource or requests other agents to give or lend resources to prevent a resource shortage. The ICL contains resource-consuming and resource-measuring equipment and supervisory control and data acquisition (SCADA) systems. It communicates with the ISL components, measures resource usage, and sends resource usage and device states to the ISL components. It receives resource usage plans from the ISL components and executes these plans or rejects impracticable ones.

From the ISL view, the ICL and other agent systems are equivalent; these are system components that can supply (or save) resources and consume resources. However, other agent systems exhibit high information incompleteness and high environmental uncertainty, although the ICL components exhibit low information incompleteness and low environmental uncertainty. The Symbiosis-ADS conceptual model plans and executes the best mix of resource suppliers (saving resources or taking them from others) and minimizes the negative effect due to information incompleteness and environmental uncertainty; thus, enabling resource allocation with stakeholder cooperation.

The disclosure, accommodation, and negotiation processes are executed autonomously by each agent system's ISL components depending on the owner's (stakeholder's) individual decision making or planning methods. Therefore, Symbiosis-ADS is a System-of-Systems concept for cooperative resource allocation based on autonomous distributed decision making systems.

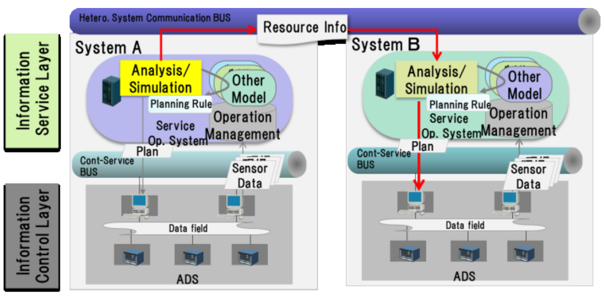


Figure 3. Abstract System Model of Symbiosis-ADS

4 Related Work/Problems

Many studies on multi-agent/multi-stakeholder cooperation have been conducted.

One group of studies was conducted in the context of distributed control [1][2]. Consensus protocols [1], for example, averaging consensus problems, are methods that calculate the same inner value of all agents with local information exchanging between agents without any central mechanism for information collection and calculations.

Another group of studies focused on game theory from which analysis or decision-making support methods were developed for negotiations between individual agents, especially concerning economics aspects. For example, studies on mechanism design [3][4] have resulted in negotiation mechanisms as resource allocation rules depending on the agent's preferences and utilities. Computational and automated mechanism design research [4] focused on a methodology of automated generation of these mechanisms or computational complexities.

Each study presupposed that all agents can disclose all inner states/data (low information incompleteness), and/or all agents can be controlled from the central controller or can accept pre-defined control rules (low environmental uncertainty). All agents having pre-defined inner state update rules or the same target values/functions is assumed with almost all distributed control methods. For example, with almost all game theory-based mechanisms, it is assumed that the central planner or markets can collect the agent's preference and determine the agent's utility functions.

For cooperation between real stakeholders (residents) in a city, a method in which making an agreement with selfish and autonomous agents based on incomplete information (locally collected or missing information) is needed.

5 Negotiation Methods

We developed a [basic negotiation process model and prototype system and the negotiation method to enable our Symbiosis-ADS conceptual model to be applied to resource accommodation in a city. The negotiation methods estimate surplus and consist of a save ability estimation process.

All agent systems disclose their resources information, and if one agent system has resource shortage and plans accommodation from other agents, it estimates the other agents' resource states (surplus) and ability to save resources. Depending on the estimation, it selects negotiation partners and determines the amount of resource accommodation requests. An agent to be requested accepts or rejects according to their resource demand or resource management ability.

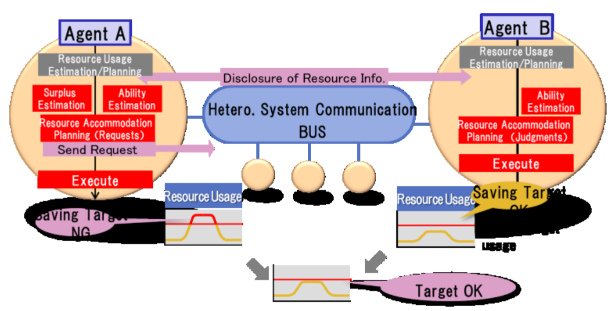


Figure 4. Abstract Negotiation Process

The surplus estimation process predicts resource surplus depending on other agents' resource consumption and negotiation/accommodation history and searches and recommends agents that have the most resource surplus and requires the least use of saved resources.

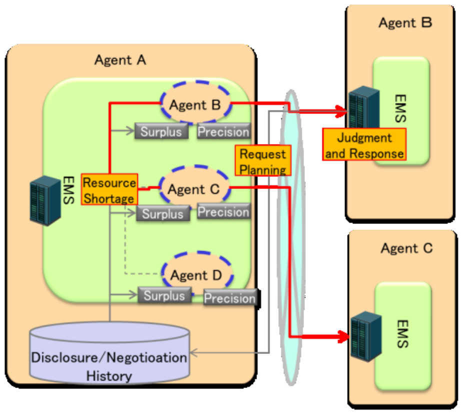


Figure 5. Surplus Estimation and Request Planning

If another agent system discloses its resource information, it might contain the same information incompleteness. For example, the precision of disclosed resource usage planning might be low, or for same reason other agents might hide their real resource usage plan. Formula (1) is a simple way of estimating other agents' resource surplus from disclosed resource information and resource usage history information.

$$P_i^{surplus}(surplus(t)) = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{supply_i(t)-surplus(t)} e^{-\frac{(x-(demand(t)+shift))^2}{2\sigma^2}}dx \quad (1)$$

$surplus_i(t)$, $supply_i(t)$, $demand_i(t)$ denote resource information disclosed by agent i and $shift, \sigma$ denote the average error and deviation of disclosed resource information and resource usage information calculated from the disclosed history, respectively.

The saving ability estimation process selects agents or inner ICL components to request resource saving based on the resource saving-method level. For inner ICL components, the level is estimated as the amount of the automation and visualization of [the/a?] resource saving method, automated equipment control, for example, auto switching of the energy consumption saving mode, or information about how to reduce power consumption to equipment operators. For other agents, the level is estimated as the rejection and failure rates from their negotiation/accommodation history.

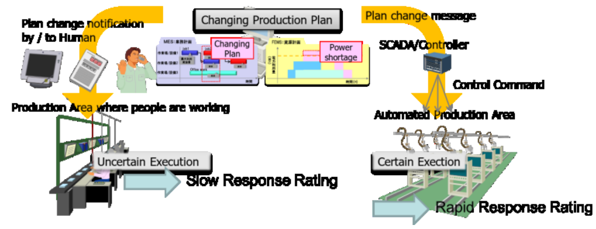


Figure 6. Overviews of saving ability (response time) estimation process

Formula (2) is inner ICL component i's resource saving ability, especially concerning minimum response time. If there is time remaining when resource saving requested is higher than pre-defined response time, the ICL component can execute the resource saving process. One agent system and its inner ICL components can define the automation level (Level) of energy management and calculate the response time of each automation. $Level(i)$ is a pre-define automation level, and ResponseTime(l) is the calculated minimum response time of Level l of the resource saving process.

$$P_{ICLi}^{response}(t) = \begin{cases} 1 & if\ t > ResponseTime\,(Level(l)) \\ 0 & otherwise \end{cases} \quad (2)$$

Formula (3) is another agent system's resource saving ability. Because another agent system's automation level generally cannot be obtained, saving ability is estimated from negotiation/disclosure history information. $Time_{reject}, Time_{accept}$ denote the average response times when the request is accepted or rejected, and $\sigma_{accept}, \sigma_{reject}$ denote deviation of response times. The right side of Formula (3) is the predicted minimum response time based on the Mahalanobis distance measurement.

$$P_i^{response}(t) = \begin{cases} 1 & if\ t > \frac{\sigma_{accept}\,Time_{reject}+\sigma_{reject}\,Time_{accept}}{\sigma_{accept}+\sigma_{reject}} \\ 0 & otherwise \end{cases} \quad (3)$$

The estimated surplus and saving ability is used as an approximation of other systems' accommodation costs. It is assumed that the cooperation systems determine whether other systems' judgments depend on these costs, support decision making for negotiation planning; select negotiation partners, and calculate an incentive to the accommodation.

Formula (4) is the optimization method for resource accommodation request planning, where $request_i(t)$ is a control parameter and is the amount of resource accommodation request for agent I , and $P_i(t, request_i)$ is the negotiation success (accept) rate of agent i when $request_i(t)$ is requested. We assume that the success rate will be high when agent i has enough resource surplus and resource saving ability, so $P_i(t, request_i)$ can be simply calculated as

$P_i^{surplus}(surplus(t)) * P_i^{response}(t)$. The objective function is minimizing the deviation of accommodated resources from other agents based on the binomial distribution portfolio optimization method. Constraint formulas (5) and (6) are the minimum expectation values of the amount of accommodated resource and the maximum amount of accommodation requests, respectively.

$$\text{minimize } \Sigma_i \left(request_i(t)^2 P_i(t) \left(1 - P_i(t)\right)\right) \quad (4)$$

$st.$

$$request(t) \leq \Sigma\left(request_i(t) P_i(t)\right) \quad (5)$$

$$request_i(t) \leq supply_i(t) - demand_i(t) \quad (6)$$

By estimating other agents' resource states (surplus) and the automated level of resource management from historical data, Symbiosis-ADS minimizes information incompleteness and environmental uncertainty.

6 Application/Experiments

We are implementing the Symbiosis-ADS conceptual model and the negotiation methods for a factory energy management system (FEMS). FEMSs are deployed at individual factories in a city. With these FEMSs, regional community energy management can be executed with power resource accommodation.

We developed a FEMS prototype system based on the Symbiosis-ADS conceptual model and a simulation environment. In the simulation, each FEMS estimates its own energy demand (and supply) and discloses the energy resource information to others.

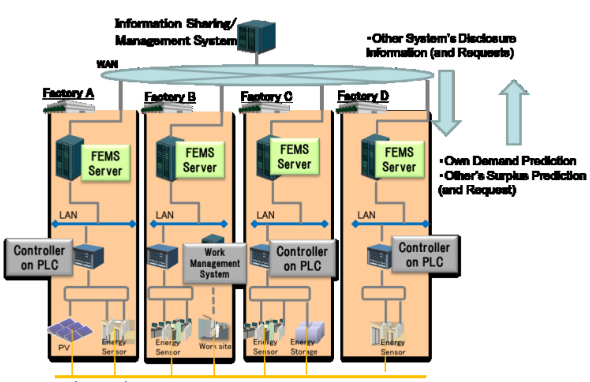


Figure 7. Overviews of simulation assumption

In the simulation, each FEMS does not imitate its disclosed resource information, but estimates low-precision resource demand. When a FEMS requests resource to others, it estimates the resource surplus of others and plans a mixture of requests to minimize fluctuation of resources to use. Our simulation environment properties are listed in Table 1.

Table1. Simulation settings

Number of FEMSs	10
Precision of each FEMS's demand estimation	Normal distribution error N(μ, σ^2), $(-10 < \mu < 10, 5 < \sigma^2 < 25)$ (Other FEMS cannot directly obtain error information.)
Judgement Rules of FEMS to be requested	･FEMS accepts request if amount requested is lower than resource surplus with margin of safety. ･margin of safety depends on error σ^2
Number of Negotiations	15epochs (days)*24 times (hours)

Figure 8 shows the results of the simulation. The Y-axis indicates the average amount of resource shortage. In a no-negotiation case (W/O Negotiation, blue line), shortage remained high. The negotiation process without any estimation (W/O Estimation, green line) made an improvement, and negotiation with surplus estimation (W/O Response Estimation and ALL, purple and red lines) made further improvement. On the other hand, the effectiveness of response time estimation (Formula (3)) was not clear. The negotiation process with response time estimation made little improvement; therefore, the response time estimation method requires more finely tuned modelling.

From these results, our negotiation methods reduce the shortage of each FEMS's energy supply by accommodating surplus energy resources between FEMSs. By estimating other agents' resource state (surplus), the negotiation methods enable inter-FEMS energy (re-)allocation by adapting to disclosed energy demand and supply information, especially in a low-precision energy demand estimation (incomplete information disclosed) situation.

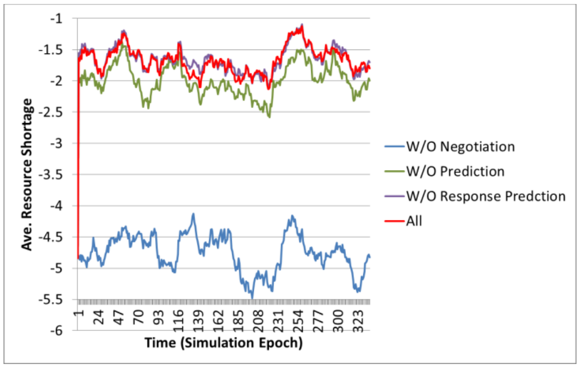


Figure 8. Results of simulation (average of energy shortage)

7 Conclusion

We proposed a conceptual system model called Symbiosis-ADS and simple negotiation methods. We are implementing this model and these methods for operating energy management systems based on multi-agent energy accommodation. With a prototype system and simulation, we evaluated the effectiveness of the negotiation methods by estimation and the ability of resource allocation within an independent multi-agent community.

References

[1] A. Nedic and A. Ozdaglar, "Distributed Subgradient Methods for Multi-agent Optimization," IEEE Trans. on Automatic Control, Vol. 54, No. 1, pp. 4861, 2009.Nobel Laureate, *His book*, Publisher, Location, Year.

[2] J. R. Marden, G. Arslan and J. S. Shamma, "Cooperative Control and Potential Games," IEEE Trans. on Systems, Man and Cybernetics, Part B: Cybernetics, Vol. 39, No. 6,pp. 1393–1407, 2009.

[3] Dash, R. K., Jennings, N. R., and Parks, D. C, "Computational-Mechanism Design: A Call to Arms",IEEE Intelligent Systems, Vol. 18,No. 6(2003),pp. 40–47.

[4] Conitzer, V. and Sandholm, T., "Automated Mechanism Design : Complexity Results Stemming from the Single-Agent Setting", In Proc. of the International Conf. on Electronic Commerce (ICEC), 2002.

[5] Y. Kamiyoshi, T. Nakabe, G. Mine, H. Nishi ,"Construction of energy measuring system in a University for Cluster Energy Management System", In proceeding of Annual Conference on IEEE Industrial Electronics Society,2010

[6] Open Automated Demand Response Communication Specification, http://openadr.lbl.gov/

System of Systems Complexity and Decision Making

Zhang Yingchao
Beijing Institute of System Engineering
Science and Technology on Complex Systems Simulation Laboratory
Beijing, China
yingchao_zhang@sohu.com

Abstract - *Complexity in System of Systems (SoS) is a key factor in SoS decision making. This paper analyzes the main characteristics of SoS, especially its complexity in terms of monolithic emergence, component systems adaption and uncertainty in SoS evolution. It further discusses the effects of SoS complexity on SoS decision making. In light of SoS complexity, the paper figures out the major problems that need to be addressed in four areas, namely SoS adaption, abnormity analysis in SoS evolution, exploratory analysis of SoS capability and SoS simulation. In conclusion the paper proposes a preliminary SoS decision making model based on the analysis of SoS adaption.*

Keywords: System of Systems (SoS), SoS Complexity, SoS Decision Making

1 Introduction

System of Systems (SoS) has become an important way for large scale application systems to exist in the information time. The development and consequently putting into use of a wide range of large network application systems that are highly informationized and integrated, such as the giant transporting network, national smart grid and the information equipment systems, contributed a great deal to the national economic growth and national defense development.

In the mean time, since SoS is a complex system, the factual existence of complexity in SoS needs to be considered in SoS decision making. As recognized by most SoS engineering researchers, SoS is a typical complex system and complexity in SoS will be more and more obvious as SoS further develops and more widely used. On the one hand, complexity is fundamental to achieving the integrative effect of SoS; on the other hand, it is the driving force that might cause the SoS to collapse. Since complexity is so relevant to the quality of SoS development and management and the methodology in SoS study, this paper discusses the issue of SoS decision making in terms of SoS complexity.

2 SoS and its major characteristics

2.1 SoS: an integrated whole

In the information time, with the fast development and application of high techs in information areas such as computer-mediated communication, internet of things, distributed controlling and cloud computing, individual systems that used to be interdependent and totally irrelevant to each other can now interconnect with each other through the “bridge” of “information” to form a larger system composed of “independent systems”, or a “system of systems”. For such a “system of systems”, although its components are widely distributed and loosely related and work independently, it can function as an integrated whole with larger capability, therefore is generally called a system of systems or “SoS”. It can be illustrated as below:

$SoS(n)$ means an integrated whole made up of interrelated component systems $e(1),...,e(i),...e(n)$, record as

$$SoS(n) = \{E(n), R_z^*\}$$

$$E(n) = \{e(i) \mid i = 1,2,...n; n \geq 2\}$$

R_z^* in the formula stands for the aggregation of interrelated component systems $e(1),...,e(i),...e(n)$.

Application of information technology in SoS lifts the speed of information transmission and level of information sharing among components and further enhances the capability of independent systems in environment cognition and response, thus makes it possible for “1＋1>2” in that it enables the various systems to fulfill missions unachievable to each individual system and shortens the physical distance between the components and the waiting time in their interactions through voluntary and collaborative interaction. Information technology indeed breaks isolation and independence among different systems, enabling separate components to aggregate into an integrated whole for larger SoS capability. In such a condensed spatial-temporal dimension, integrated function makes SoS a powerful mechanical body.

2.2 Major characteristics of SoS

SoS reflects the property of diversity in the category of system (such as simple system, complex system, ecosystem and military system). It demonstrates the diversification, interdependence and wholeness of systems as defined by Ludwig Von Bertalanffy in his general system theory in another way. Major characteristics of SoS are reflected in four aspects:

1)The component systems are independent to each other. Different component system are not only widely distributed spatially and mutually independent in their development temporally, but also workable autonomously.

2)The component systems are interdependent on each other. Different components may influence each other through the network that links them, in which they may work collaboratively within a SoS or counteractively between two SoSs, i.e. the confrontation of two rivalry battle SoSs.

3)There is obvious organizational factor involved. SoS is an integration of human, soft elements and hardware system, in which human and soft elements play a large role in making the SoS a flexible "machine".

4)They are to fulfill the shared mission. Guided by a common objective, various component systems interact with each other and gradually integrate into a whole, consequently wholeness effect emerges. The shared mission drives the component systems as well as the SoS itself to evolve continuously.

3 Complexity in SoS

3.1 Emergent behavior

SoS aims to let emergence happen to generate SoS growth. It is an ultimate goal for the SoS to achieve the wholeness effect through collaboration and interaction of various component systems and coordination of various soft elements. Occurring in the course of SoS dynamic evolution rather that its stationary state, emergence is the innovative quality of SoS displays in its structure, behavior and state, representing a promotion in SoS capability.

By emergence, we mean there are phase changes in the system, and there emerges new property and new function—a macro property resulted from the integration of abundant micro behaviors of the components in a system that can only be observed or perceived in the system as a whole instead of in parts of the component systems. Emergence can neither be predicted nor reduced. Neither can the cause and effect for such emergence be retrieved. For simple systems, no matter how large a size they have, no matter how often interaction occurs, they can never escape from the linear rule of "1+1=2" , for which there is only linear add-ups and behavior can be predicated, property can be reduced and the cause and effect can be retrieved. Therefore, emergence can only come from complex rather than simple systems. Since SoS aims to let emergence happen, SoS belongs to the category of complex system, and can be termed as "complex system of systems".

In the Spectrum of System, Mitre put SoS between ordinary system (simple system) and complex system, meaning that SoS can either be simple system or complex system. However, if "emergent behavior" is applied to define SoS or complex system of systems, the position of SoS in the spectrum of system shall be within the complex system, as illustrated in Figure 1:

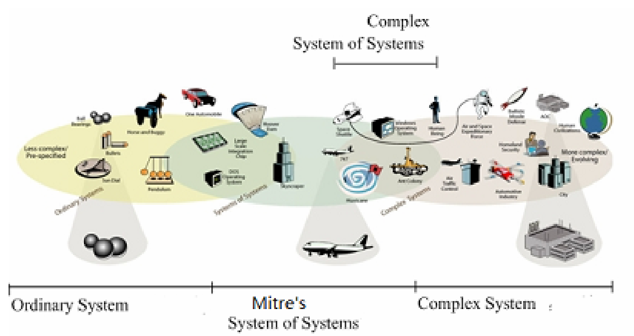


Figure 1. Position of complex system of systems in Mrtre's Spectrum of System

3.2 Adaption

As a natural property of SoS components, adaption is a critical mechanism for SoS emergence to happen. In his complex adaptive system theory, John Holland claimed that "adaption builds complexity", arguing that system are made up of agents, adaption of the agent is one of the mechanisms to have resulted in system evolution complexity, and thus drives the system to continuously evolve in its structure and level.

From the perspective of SoS, adaption of its components is the driving force for SoS emergent behavior and root for SoS complexity. Adaption of SoS components comes from not only autonomous perception and smart processing of the hardware system on the environment, but also adaption of “human” or “organization” that engage in the SoS operation. Pulled together by the SoS central mission, adaptive interaction in the SoS facilitates the components to coordinate and collaborate with each other and consequently drives the evolution of SoS as a whole and the macro emergence of SoS capability.

3.3 Uncertainty in SoS evolution

SoS evolution includes evolution in the process of SoS gradual forming and that in its application. The process of SoS gradual forming and its application in confrontational setting keeps SoS in the process of dynamic evolutionary development, in which changes in SoS mission, application setting (including adversary of the SoS), technical condition and soft elements such as human and doctrines makes SoS evolution quite uncertain. Such uncertainty influences SoS emergence outcome, either to achieve SoS integration and promote its integrative efficiency or to result in the failure and collapse of SoS architecting.

4 Issues of complexity that needs to be considered in SoS decision making

4.1 SoS engineering: methodology for SoS decision making

SoS engineering is the basic way for SoS decision making. While understanding of SoS engineering in different fields or among different scholars are quite different, different concepts of “SoS engineering” are presented in various fields and from different angles for SoS decision making. Here “SoS engineering” can be defined as system engineering method set to realize SoS designing, verification, architecting, management control and assessment under the guidance of system science. Scope of SoS research covers: SoS needs analysis that deals with acquisition, analysis, demonstration, verification and management of SoS needs; SoS research designing that studies rules and characteristics of model for SoS planning and architecting; SoS management control that probes into issues such as SoS decision making control, status monitoring and abnormal state handling; SoS evaluation that explores the compressive evaluation methods for SoS status and SoS effect and; SoS simulation experiment that explores ways to support modeling, simulation and experiment of SoS engineering with simulation tools.

Rather than negating traditional system engineering study methodology or abandoning system science, SoS engineering shall still be guided by system science to deal with issues of multi-system integration by streamlining interrelation between the whole and the parts from the angle of structure and evolution; in the meantime, traditional system engineering methodology shall be integrated into SoS development and used to guide existing system optimizing and new system designing and developing. From the perspective of SoS development and implementation, United States Department of Defense defined SoS engineering as “deals with planning, analyzing, organizing, and integrating the capabilities of a mix of existing and new systems into a SoS capability greater than the sum of the capabilities of the constituent parts”. While system engineering focuses on work related to needs analysis, system development, integration and testing of existing and new systems, SoS engineering deals with multiple systems integration of existing and new systems.

4.2 SoS and influence of its complexity on SoS decision making

Since emergent behavior, components adaption and uncertainty in evolution of SoS influence the whole process of SoS architecting and organizational management, traditional planning methodology for simple systems cannot be used to address SoS and its complexity, because of:

1)Fuzzy planning. SoS decision making cannot be designed “perfectly” in advance. Traditional system engineering methodology emphasized on top-layer precise designing at the early stage of engineering to clearly define functions of systems and their application context, and the system will not be put into use until it is completed. Forming of SoS is a temporal process of emergent behavior, a gradual process that can put the SoS into use while it is building, therefore no strict and precise architectural plan can be developed at the early stages. In this sense, SoS can only rely on “evolution” rather that “architecting”.

2)Limited objective. Objective of SoS decision making doesn’t emphasize on “the best”. While system engineering is based on “optimized method”, there is no best practice for SoS optimization since it is a process that witnesses the coordination and suiting of component systems with SoS overall mission and the environment, and it is the dynamic balance between performance of the components, the system and relevant soft elements that can only achieve comparative “satisfaction” among them.

3)Evaluation under uncertain condition. SoS evaluation shall focus on the overall exploration of possible space for SoS evolution, especially a comprehensive analysis of SoS overall capability for different missions, under different application environment and in different development stages. Multiple influense of random factors and hazards upon spatial and temporal dimension of the SoS shall be fully considered as well.

4.3 Several problems that needs to be solved in SoS decision making in the perspective of SoS complexity

In light of SoS complexity, we believe that the following problems need to be dealt with as priorities in the stages of SoS designing, verification, architecting, management control and evaluation.

1)Study on SoS adaption to facilitate top-layer designing of SoS integration. In the course of SoS integration, enough consideration shall be given to adaption of the component systems, study shall focus on how to architect a basic model for open SoS network interacting with SoS components on the base of information system and information network to enable the system platform autonomously and interactively shaping its integrated function as needed.

2)Abnormal state analysis in SoS evolution to support risk control in SoS evolution. Factors like SoS confrontation, technology development, changes in application environment and central mission and new system joining-in may cause abnormal state in the process of SoS evolution. Risk in SoS decision making shall be reduced by timely judging the critical state that may slide into an abnormal state, retrieving the complex cause and effect of abnormal state, predicting its consequential cascade reactions, developing auxiliary SoS evolution road-maps and setting up preventive mechanism against evolution failure.

3)Explorative analysis of SoS capability to support SoS comprehensive assessment. MOP and MOE of the SoS shall be explored in an overall point of view pertaining to all components of the SoS and the complete course of SoS evolutional life-circle so as to: develop an explorative analytical methodology on the SoS capability; fully verify consequence of various uncertain factor on SoS capability and; examine gaps between SoS capability and SoS central mission.

4) SoS simulation to set up research platforms for SoS engineering. Simulation is an important tool for system complexity study. SoS simulation that focuses on SoS behavior modeling on the base of information network modeling and aims to achieve the wholeness effect shall be used to provide an experimental study platform for SoS adaption study, evolution abnormal state study, SoS capability analysis as part of the overall SoS architectural planning, and decision making, and organizational management also.

5 A preliminary SoS decision making model based on adaption analysis

5.1 Analysis of SoS adaption

Analysis of SoS components adaption—a fundamental factor in macro emergence of SoS capability—is the key to SoS decision making, although it is hard manage.

While SoS adaption aims to let emergence happen to generate SoS capabilities, other complex system does not. SoS adaption on the one hand derives from adaption of individuals to the local environment, for which we term as local adaption; on the other hand it also derives from adaption of individuals to its mission, for which we terms as SoS adaption. These two kinds of adaption can be defined as follow:

Behavior capability of SoS component at time T $Entity_T(n)$ is a whole composed of the behavior of n possible entities $b(1),...,b(i),...b(n)$, record as:

$$Entity_T(n) = \{b(i) \mid i = 1,2,...,n; n \geq 1\}$$

Seth ϕ and φ as function for local and SoS adaption of the SoS components respectively, if:

$$\phi(b(m_1)) = \max\{\phi(Entity_T(n))\}$$

$$\varphi(b(m_2)) = \max\{\varphi(Entity_T(n))\}$$

When SoS components takes $b(m_1)$ as its behavior at time T, the framework that is used to determine SoS components adaptive behavior is local adaption of SoS components, record as: $Adaption_{local}$; when SoS components takes $b(m_2)$ as its behavior at time T, the framework that is used to determine SoS components adaptive behavior is SoS adaption of SoS components, record as: $Adaption_{SoS}$.

5.2 A SoS decision making model based on adaption analysis

Adaption of SoS is some kind of coupling of $Adaption_{local}$ and $Adaption_{SoS}$. Experiments demonstrated that, when components are apt to $Adaption_{local}$, there is certain property of cyclicity in the SoS adaption model caused by short range order of $Adaption_{local}$; when components are apt to $Adaption_{SoS}$, there is certain property of convergence in the SoS adaption model caused by long range order of $Adaption_{SoS}$.

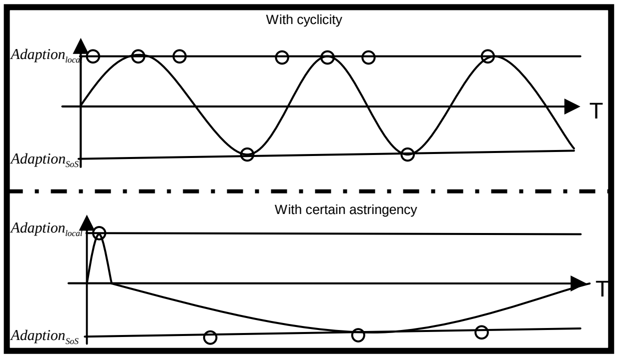


Figure 2. Properties in the two models of SoS adaption

Rules for SoS decision making based on analysis of adaption. When the SoS central mission is relatively clear, SoS planning shall take SoS adaption as a model, under which property of convergence resulting from long range order of SoS adaption can make SoS evolution meet the capability requirement for SoS more quickly to achieve SoS designing objective; On the contrary, when the SoS central mission is not clear yet, SoS planning shall take local adaption as a model, making full use of the autonomous mechanism of the components in adaption. SoS planning objective shall be further clarified in the course of SoS evolution to avoid SoS evolution quickly converges to the extent beyond expectation, or even generating SoS with less efficiency and capability that cannot meet the capability requirement for SoS.

6 Conclusion

John Horgan, a U.S. science writer mentioned in his work The End of Science that the science and technology development history of human kind is a course that witnesses the alterative emergence of science discovery and technology application peaks. The ongoing development of complex system study shows people the factual existence of complexity. How to make use of complexity to achieve prosperity in the information time? SoS engineering tries to develop an engineering methodology in the information time to meet the opportunities and challenges presented in it. In this process, problems of complexity posing by informatization and systematization on multiple system integration and its application must be addressed, for which our continuous efforts is needed to get an upper hand in large scale SoS architecting and management in this information time by improving the practicing and theory development of SoS decision making.

References

[1] Mo Jamshidi. System of Systems Engineering [C]. New Jersey: A John Wiley&Sons,INC.,2009.

[2] Zhang Weiming, Liu zhong et.al, SoS Engineering Theory and Methodology [M], Beijing, Science Press, 2010.

[3] Brain E.White. Engineering enterprises using Complex-System Engineering(CSE). [EB/OL]. [2011-7-28].http://www.sosece.org/pdfs/1stsosececonf/presentations/white.pdf.

[4] Arthur F.Krummenoehl.System of Systems Engineering[EB/OL].[2011-8-1].http://www.sosece.org/pdfs/1stsosececonf/presentations/Krummenoehl.pdf

[5] Kirk Dye System of Systems Complexity in Fielding a Joint Systems[EB/OL]. [2011-8-11] http://www.sosece.org/pdfs/1stsosececonf/presentations/dye.pdf.

[6] DoD.System of Systems System Engineering Guide:Consideration for Systems Engineering in a System of System Environment.[EB/OL] [2011-8-20].www.theriac.org/pdfs/SoS-SE-Guide.pdf

Modeling System of Systems Acquisition

Nil Kilicay-Ergin
Great Valley School of Graduate Professional Students
Penn State University
PA, USA
nhe2@psu.edu

Paulette Acheson
Engineering Management & Systems Engineering Department
Missouri University of Science & Technology,
MO, USA
pbatk5@mail.mst.edu

John Colombi
Air Force Center for Systems Engineering
Air Force Institute of Technology
OH, USA
john.colombi@afit.edu

Cihan H. Dagli
Engineering Management & Systems Engineering Department
Missouri University of Science & Technology,
MO, USA
dagli@mst.edu

Abstract – *System of systems (SoS) acquisition is a dynamic process of integrating independent systems. This paper describes modeling of the SoS acquisition environment based on the Wave Process Model. Agent-based modeling methodology is utilized to abstract behavioral aspects of the acquisition process.*

Keywords: system of systems, system of systems acquisition process models, agent-based modeling

1 Introduction

Today, increasing number of system acquisitions focus on integration of independent systems into a System of Systems (SoS). Traditional system engineering approach of scoping the system boundaries and optimally allocating requirements to system components is not suitable for SoS development anymore. This is mainly due to the fact that SoS component systems are independent and have their own functionality, development processes, funding and operational missions. In addition changes in external environment such as funding, national priorities can alter the dynamics of the acquisition process [1], [3]. Therefore, SoS engineering processes need to consider change as an important dynamic beyond technical considerations.

Evolutionary acquisition models are more suitable to SoS development as they emphasize stakeholder involvement, interim milestones, increased iteration and concurrent development [4]. The Systems Engineering of SoS (SoS SE) model is developed by the US Department of Defense to identify the core elements of the SoS acquisition [1]. The Incremental Commitment Model [3] is a risk driven framework that can be tailored for SoS development. The Wave Model [1] maps SoS SE model's core elements to a series of time-sequenced iterative process to guide implementation of the framework for practitioners of SoS development.

Regardless of the acquisition framework, in order to address SoS needs, SoS architects have to collaborate with individual system architects to leverage individual system functionalities. However, most system architects assume that SoS participants exhibit nominal behavior but deviation from nominal motivation leads to complications and disturbances in systems behavior. It is necessary to capture the behavioral dimension of SoS architecture to be able to represent the full problem space to guide SoS analysis and architecting phase [5].

This paper builds on the Wave Process Model to abstract behavioral aspects of the acquisition process. An agent-based model of the process is discussed to analyze the impact of individual system motivations on the overall SoS architecture evolution. It is envisioned that this type of model will help us in understanding the intricate dynamics of the SoS development and improve acquisition process. In the following sections, the Wave Process Model is provided as background information (Section 2), a conceptual agent-based model of the process is described (Section 3), and finally future directions for the research are discussed in the conclusions section.

2 SoS SE: The Wave Model

The Wave Process Model presents the core elements of SoS SE model in a series of six time-sequenced major steps. The wave model or bus-stop approach is a development approach that is similar to the effect of periodic waves crashing at the shore or a bus that periodically stops at a specific location. The SoS has specific places in the development where it can accept updates from the individual systems. Individual systems can plan their deliveries to coincide with the SoS 'bus-stops' or can evaluate the effect of missing a planned SoS wave. Figure 1 illustrates the major elements of the Wave Process Model. The steps in the model are briefly introduced below as background information. For further details refer to [1].

2.1 Initiate SoS

This step involves understanding the SoS objectives and operational concept (CONOPS) as well as gathering information on core systems to support desired capabilities.

2.2 Conduct SoS Analysis

This step establishes an initial SoS baseline architecture for SoS engineering based on SoS requirements space, performance measures, and relevant planning elements.

2.3 Develop and Evolve SoS Architecture

This step evolves the initial SoS baseline and develops the SoS architecture. The SoS architecture includes individual systems, key SoS functions and interdependencies among systems. The architecture identifies necessary changes in contributing systems in terms of interfaces and functionality in order to implement the SoS architecture.

2.4 Plan SoS Update

This step plans for the next SoS upgrade cycle based on the changes in external environment, SoS priorities, options and backlogs.

2.5 Implement SoS Update

This step establishes a new SoS baseline based on SoS level testing and system level implementation. This step is the end of wave cycle or 'bus-stop' where updates from individual systems can be integrated into the SoS.

2.6 Continue SoS Analysis

This step is the beginning of the next wave cycle and continuous to analyze the current SoS architecture for future SoS evolution.

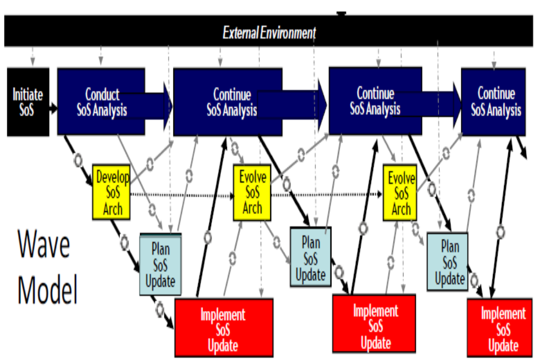


Figure 1. SoS Wave Process Model [1]

3 Agent-based SoS Acquisition Model

Agent based models (ABM) consist of a set of abstracted entities referred to as agents, and a framework for simulating agent decisions and interactions. Agents have their own goals and are capable of perceiving changes in the environment. System behavior (global behavior) emerges from the decisions and interactions of the agents. The approach provides insight into complex interdependent processes. Agent-based modeling methodology has several benefits over other modeling techniques; it captures emergent patterns of system behavior, provides a natural description of a system composed of behavioral entities and is flexible for tuning the complexity of the entities [6]. The methodology is used in a wide range of application domains such as financial markets [8], autonomous robots [9], and homeland security [10] to analyze and understand behavior of complex systems.

The agent-based methodology is suitable for analyzing the behavioral aspects of the acquisition process as agents can capture independent systems' behavior and SoS engineering activities. Figure 2 illustrates the agent-based SoS acquisition model. In the model, agents represent individual systems where agents embody the systems and the people responsible for them. The Wave process model applies to acknowledged SoS [7] thus there is also a specific agent responsible for the SoS engineering effort and for coordinating the individual system agents.

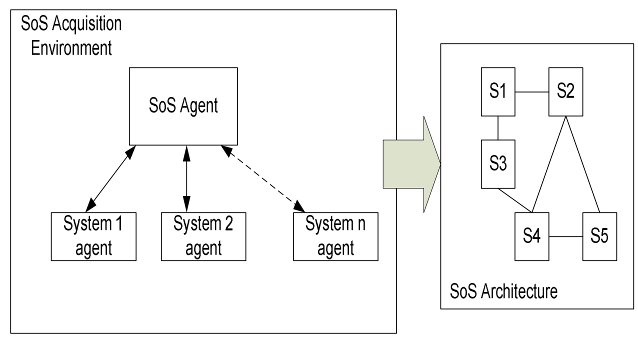


Figure 2. Agent-based SoS Acquisition Model

The following subsections outline the general structure of the agent-based model and how key elements of the Wave Process model are abstracted for analyzing the impact of individual system behavior on SoS architecture and acquisition process.

3.1 SoS Acquisition Environment

The SoS agent and the individual system agents are influenced by the changes in the SoS acquisition environment. Thus the environment model includes external factors/variables such as national priorities, threats and SoS funding. As the SoS acquisition progresses through wave cycles, these variables are updated to reflect acquisition environment changes. Table 1 summarizes the model elements in mathematical notation.

Table 1: SoS Acquisition Environment

External factors/variables:
$E_0 = f(National\ priorities, SoS\ funding, threats)$
Changes in external environment at wave time T: σ_T
External factors/variables at time T: $E_T = E_0 \sigma_T$

3.2 SoS Agent Behavior

SoS agent is responsible for the overall SoS engineering activity and coordinates with individual system agents to achieve the desired SoS mission. In the model, it is assumed that an initial SoS mission is already determined and an initial baseline SoS architecture is available. The SoS agent follows the six core SoS engineering activities outlined in the Wave Process Model in order to develop the SoS. The SoS architecture evolves based on the behavior of individual systems as well as changes in the external environment.

Initiate SoS – During the initialization phase, the wave interval, the time interval from one wave to next, is determined. At each wave interval time, the SoS agent identifies SoS target measures which comprises desired SoS capabilities and SoS performance parameters for these capabilities in order to meet mission objectives. Since some of the capabilities may have higher priority levels than others, weighted value of each capability is also identified at this phase. Table 2 summarizes the abstracted model elements in mathematical notation.

Table 2: Initiate SoS

Simulation time: t Wave interval: epic Wave time: T=epic. t At Wave time: T=0 Determine SoS desired capabilities: $So\bar{S}.C_i = (C_1, C_2, ..., C_n)$ Determine weighted value for each SoS capability: $SoS.w_i = (w_1, w_2, ..., w_n)$ Determine SoS desired performance parameters: $SoS.P_i = (P_1, P_2, ..., P_n)$ Identify initial SoS Target Measures: $SoS.M_0 = [a_{ij}]_{n\times 3}$ where $a_{i1} = SoS.C_i,\ a_{i2} = SoS.P_i, a_{i3} = SoS.w_i$

Conduct SoS Feasibility Analysis –The SoS agent allocates SoS capabilities to individual systems or group of systems. This allocation identifies interface and functionality requirements for individual systems. This allocation defines a baseline SoS architecture. The program management measures such as schedule, funding are also identified. The SoS baseline architecture and program measures information is sent to individual systems as a connectivity request to the SoS architecture. Individual systems should evaluate whether they can develop the requested interface with other systems and capabilities in the given deadline and funding. Table 3 summarizes the abstracted model elements in mathematical notation.

Table 3: Conduct SoS Analysis

Identify set of individual systems to satisfy the target SoS measures: $SoS.M_0 \rightarrow System.S_i = (S_1, S_2 ..., S_n)$ Allocate SoS capabilities to individual systems: For i=1..n   For j=1..n $SoS.Callocated_i = (C_i, S_{i,} S_j)$ where $S_i \rightarrow S_j$ and $S_i \neq S_j$ Determine deadline for each allocated SoS capability: $SoS.d_i = (d_1, d_2 ..., d_3)$ Determine funding for each allocated SoS capability: $SoS.f_i = (f_1, f_2 ..., f_3)$ Define initial baseline SoS Architecture: $SoS.A_0 = [a_{ij}]_{n\times n}$ where $a_{ij} = (SoS.Callocated_i)$ Send SoS Connectivity Request to individual systems: $SoS.R_i = f(SoS.A_0, SoS.f_i, SoS.d_i)$

Develop and Evolve SoS Architecture – The SoS agent updates the baseline SoS architecture based on information received from individual systems. Individual systems may decide to cooperate at the requested deadline, may decide to cooperate at a later time or may decide to not cooperate at all depending on their motivation. These decisions will affect the SoS architecture evolution. At this step, based on information received from individual systems, the expected SoS architecture at the end of the wave cycle is updated. Table 4 summarizes the abstracted model elements in mathematical notation.

Table 4: Develop and Evolve SoS Architecture

Receive information from individual systems (see Section 3.3, Table 8): $System.Information_i$ Architecture update factor: $Beta_T = f(System.Information_i)$ Expected SoS architecture at wave time T: $SoS.A_T = SoS.A_0 + Beta_T$

Plan SoS Update – At the end of the wave cycle, the SoS agent evaluates changes in the external environment. The SoS target measures and wave interval for the next cycle is updated based on environment changes and architecture gaps analysis. The gap analysis is also conducted at the end of the wave cycle during the SoS implementation step which is described in the following step. Table 5 summarizes the model elements in mathematical notation.

Table 5: Plan SoS update

At wave time T:
- Adjust/update SoS Target Measures:
Capability update factor
$SoS.\Delta C_i = (\Delta C_1, \Delta C_2, ..., \Delta C_n)$
$SoS.\Delta C_i = f(E_t, SoS.Gap_T)$
Performance update factor
$SoS.\Delta P_i = (\Delta P_1, \Delta P_2, ..., \Delta P_n)$
$SoS.\Delta P_i = f(E_t, SoS.Gap_T)$
SoS Target measures update factor
$SoS.Alpha_T = [a_{ij}]_{n\times 2}$ where
$a_{i1} = SoS.\Delta C_i$ and $a_{i2} = SoS.\Delta P_i$
at T=0
$SoS.Alpha_T = 0$
SoS Target measures at time T:
$SoS.M_T = SoS.M_0 + SoS.Alpha_T$
- Adjust wave rhythm interval
$Epic = f(E_T, SoS.Gap_T)$
- Adjust budget/schedule for allocated capabilities
$SoS.d_i = f(E_T, SoS.Gap_T)$
$SoS.f_i = f(E_T, SoS.Gap_T)$

Implement SoS – At the end of the wave cycle, the current SoS architecture is evaluated against initial SoS baseline architecture to identify the functionality and performance gaps. This step is an input to planning SoS update step. Table 6 summarizes model elements in mathematical notation.

Table 6: Implement SoS architecture

At wave time T:
Gap analysis:
$SoS.Gap_T = f(SoS.M_T - SoS.M_{T-1})$

Continue SoS analysis – The next wave cycle of the SoS development starts once the SoS target measures and wave interval time are updated.

3.3 Individual System Behavior

Individual systems receive request for connectivity to SoS architecture. Since each system is independent and has its own goals and motivations, the system has the option to cooperate or not cooperate with the SoS agent. The decision depends on several factors including system's willingness to cooperate which measures the degree of selfishness of the individual system to be part of the SoS, and system's ability to cooperate which depends on system's resources that will allow the system to be part of the SoS. If individual system decides to cooperate, it sends information to the SoS agent on the probability of meeting the requested capability at the given deadline. If individual system decides to not cooperate, it has the option of requesting a later deadline to provide the capability. Table 7 and Table 8 summarize the abstracted model elements in mathematical notation for individual systems.

Table 7: Evaluate SoS Connectivity Request

Individual system: $System.S_i$
System performance: $System.p_i$
System capability: $System.c_i$
Willingness to cooperate: $System.willingness_i$
Ability to cooperate: $System.ability_i$
-Receive Connectivity Request from SoS agent:
$SoS.R_i$
-Evaluate SoS request:
$System.coop_i = f(System.willingness_i,$
$System.ability_i, SoS.R_i)$
$System.coop_i = \begin{cases} 1 \text{ if cooperate} \\ 0 \text{ if not cooperate} \end{cases}$

Table 8: Reply back to SoS agent

If $System.coop_i = 1$
$System.Information_i = (System.c_i, System.p_i,$
$System.av_i)$
where
$System.av_i = P(SoS.R_i)$
else
Time to cooperate:
$System.cooptime_i = t$ where $t > SoS.d_i$

4 Conclusions

This paper outlined a generic agent-based model for analyzing behavioral dynamics of SoS acquisition based on Wave Process Model. Modeling key aspects of such an intricate process is challenging but necessary for understanding the behavioral dynamics. Future research will focus on generation of models for various SoS acquisition applications based on this generic model. It is envisioned that these models will provide insights into strategies for improving the SoS engineering process.

5 Acknowledgment

This material is based upon work supported, in whole or in part, by the U.S. Department of Defense through the Systems Engineering Research Center (SERC) under Contract H98230-08-D-0171. SERC is a federally funded University Affiliated Research Center managed by Stevens Institute of Technology.
Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Department of Defense.

References

[1] J. Dahmann, G. Rebovich, R. Lowry, J. Lane and K. Baldwin, "An implementers' view of systems engineering for systems of systems" Systems Conference (SysCon), 2011 IEEE International, pp. 212-217, April 2011.

[2] J. Dahmann, J. Lane, G. Rebovich, K. Baldwin, "A model of systems engineering in a system of systems context" Proceedings of the Conference on Systems Engineering Research, LA, CA, April 2008.

[3] J. Lane, and J. Dahmann, "Process evolution to support system of systems engineering" ULSSIS'08, Leipzip, Germany, May 2008.

[4] R. Creel, and B. Ellison, "System-of-systems influences on acquisition strategy development" https://buildsecurityin.us-cert.gov/bsi/articles/best-practices/acquisition/981-BSI.html, 2008, last accessed March 2012.

[5] J. P. Dauby and S. Upholzer, "Exploring Behavioral Dynamics in Systems of Systems", in Complex Adaptive Systems, Editor C. Dagli, Procedia Computer Science Volume 6, Page 34-39, Elsevier 2011.

[6] E. Bonabeau "Agent based Modeling: Methods and Techniques for Simulating Human Systems," Proceedings of the National Academy of Sciences, Vol. 99, pp. 7280-7287, 2002.

[7] Systems Engineering Guide for System of Systems, Version 1.0, 2008, http://www.acq.osd.mil/se/docs/SE-Guide-for-SoS.pdf, last accessed May 2012.

[8] N. Kilicay-Ergin, D. Enke, and C. Dagli, "Models of Trader Behavior and Analysis of Financial Market Dynamics" International Journal of Knowledge-based and Intelligent Engineering Systems, Vol. 16, pp. 99-116, 2012.

[9] D. D. Dudenhoeffer, and M. P. Jones, "A Formation Behavior for Large-scale Micro-robot Force Deployment" Proceedings of the Winter Simulation Conference, Orlando, FL, 2000.

[10] W. E. Weiss, "Dynamic security: An agent-based model for airport defense" Proceedings of the Winter Simulation Conference, Miami, FL, pp. 1320-1325, 2008.

The Originating Concept: a Foundation for System of Systems Architecting Decision Making

Vincenzo Arrichiello
SELEX-SI Academy
SELEX Sistemi Integrati SpA - FINMECCANICA
varrichiello@selex-si.com

Abstract - *The paper describes an approach enabling the development of an artefact that can support the analysis of needs and definition of requirements, and make available, to the Architectural Decision-Making process, a "measurement tool" for the evaluation, based on criteria and performance parameters, of the overall effectiveness of the alternative solutions. The approach encourages the active involvement of the main stakeholders enabling the development of a shared understanding and the building of consensus. The information content of the artefact facilitates the generation of diverse alternative solutions and provides also a sound foundation for all the following development activities of the System of Systems.*

Keywords: Systems Architecting, Decision-Making, Needs Analysis

1 Introduction

In the development process of a System of Systems, the term "systems architecting" is frequently used to refer to the activities of the upper left of the "Vee". At the highest level of synthesis, systems architecting can be seen as comprising three main tasks :

- understanding the problem and needs and translating these into clearly defined requirements, criteria and performance parameters
- generating a set of viable solutions, evolving a system concept of operation that satisfies the needs
- select the preferred solution and define its characteristics to enable its development.

It is well known that the fate of a project is strongly dependent on how well these tasks are performed. While the susbequent phases of the system's development (where the focus shift to requirements flow down, interfaces and configuration management, verification, integration and validation) can rely on well defined and proven processes and methods, those initial task are often regarded as to represent the "art" component of systems engineering.

A SoS, due to its scale and complexity, must be able to address a wide range of functional and non-functional needs, many of these conflicting each other. The identification, through extensive trade-offs, of the best value solution requires a sound decision making process. The intent of the present paper is to describe an approach to carry out the first of above tasks, having the purpose of providing a solid foundation for the decision process, and, at the same time, to the whole development process of the SoS.

Stated in more detail, the proposed approach has the following goals:

- to stimulate, facilitate, and support a pervasive and comprehensive analysis of needs, requirements and constraints,
- to produce a “workable” structured formalization of the high level requirements, facilitating the definition of alternative candidate solutions,
- to define a "Value Model" suitable to provide the reference for the evaluation of alternatives, enabling the selection of the best balanced solution,
- to facilitate the development of a common understanding and consensus with the stakeholders about the high level requirements.

1.1 Understanding Needs

One of the fundamental tasks of Systems Engineers, as identified by Parnell and Driscoll, is to "Convert customer needs to system functions and requirements." [1] To do that, one has to identify "the stakeholders involved with the system throughout its life cycle, and their needs, expectations, and desires" as the definition of the "Stakeholder Requirements Definition Process" of the ISO/IEC 15288 Standard states. [2] This is not an easy job; for starters, also the needs that are clearly stated must be analysed in depth and refined to be put in the form of proper requirements.

Expectations and desires are intrinsically difficult to discover since they are not well perceived, let alone articulated, by the stakeholders. All too often only when they start interacting with the real, delivered, system, users and the other stakeholders discover needs that they could not anticipate, but that now they perceive as severe deficiencies.

Unstated needs may originate either from really unknown needs or, somewhat surprisingly, from characteristics seen as too obvious to be stated. Unstated needs and requirements are a major cause of costly reworks and delays. Beside preventing these problems, the early discovery of unstated requirements can also inspire innovative ideas and products.

The "Excited Qualities" ("Delighters") defined by the well known Kano model, are, almost by definition, related to unexpressed, latent needs; the ability of the system's designer to anticipate the features that provide an added benefit to the customer, can result in a significant increase of the perceived value of the SoS.

Lastly, the thorough analysis of needs and requirements can help to limit the extent of the so called "requirements volatility", reducing the causes of confusion and dissatisfaction for customers and designers.

The Originating Concept approach is devised to help guarantee that the ensuing project activities are well rooted in a clear and exhaustive understanding of needs, expectations, desires, conditions and constraints.

1.2 Defining Evaluation Criteria

With reference, again, to Parnell and Driscoll, another fundamental task of the Systems Engineer is to "Define system performance measures to guide design synthesis, system validation, and successful system realization." The need to support and strengthen the design process with quantitative measures to make it successful ("how well") and, at the same time, effective ("good enough") is well known to engineers.

The high level needs and requirements often encountered when dealing with complex, large scale SoSs, does not lend themselves, for their very nature, to be easily related to measurable parameters; on the contrary the level of accomplishment of these needs can often be defined only in qualitative terms. The perceived value of a SoS is, to a large extent, related to its ability to address those high level concerns; to properly guide the development of an effective and successful solution to the customer's problem, the designer must have a "measurement tool" able to account also for there aspects.

2 The Originating Concept

The adjective "originating" was chosen to underline that the artefact produced with this approach is intended to be the true origin that underlies the whole system's development; this choice is also intended to differentiate it from the other "concepts" that are widely referenced in the literature (e.g. Operational Concept, Concept of Operations, Con-Ops, etc.). That does no mean that the originating concept approach is a total departure from the concepts at the basis of artefacts like the "Operational Concept Document" as defined by the AIAA "Guide to the Preparation of Operational Concept Documents", or by the IEEE "Guide for Information Technology-System Definition-Concept of Operations Document". On the contrary, it builds on the basis of these authoritative references with the addition of components able to support the decision process for the selection of the preferred architectural solution, and also to provide a guiding reference for the generation of alternative solutions. In fact, the IEEE documents states that "Classifying the desired changes and new features into essential, desirable, and optional categories is important to guide the decision making process during development of the proposed system."[3] The AIAA document suggest to include a subsection to "describe the system's goals and the objectives and expectations for it, quantified where possible, and the key performance attributes for the system." [4] The method to be applied to define and prioritize goals and objective is not, however, dealt in detail by either of the documents.

The Origination Concept approach provides a way to integrate the analysis of mission needs, goals and objectives on one side, with the development of a Value Model that defines their relation and relative importance on the other. This way, not only a better coherence of the end result can be achieved, but also a better overall effectiveness of the process, exploiting the synergies and overlapping existing between the two activities.

2.1 Methods and constructs

The Originating Concept makes use ad integrates some well known methods and constructs. A brief description of these is provided in the following, mainly to highlight the aspects that the approach take advantage of.

2.1.1 Mission Analysis method

The "Mission Analysis" method has been developed by the military "to provide an audit trail from the broadest national objectives down to operational activities at the tactical engagement level".[5] Through a series of subsequent steps, the method leads to the development a hierarchy of objectives linking the high level to the low level ones. The hierarchy structure provides an explicit description of the relations between the objectives. The lower level of the hierarchy contains the "Operational Tasks"; these are defined as: what "force elements are to accomplish in order to achieve an Operational Objective"[5] Operational tasks have associated measures and criteria; measures provide for the quantification of levels of performance, and criteria define acceptable levels of performance.

2.1.2 Value Focused Thinking

Value Focused Thinking (VFT) is a decision making methodology developed by Ralph L. Keeney; he defines the methodology as "a philosophical approach and methodological help to understand and articulate values and to use them to identify decision opportunities and to create alternatives." [6] The methodology is contrasted with the more usual alternative-based approach, that starts with a set of alternatives and then defines the selection criteria. The VFT methodology involves the development of a Value Model comprising two elements:

- The Qualitative Value Model: a structured objectives hierarchy built following a top-down recursive approach.
- The Qualitative Value Model: a means for the evaluation of how well each of the alternative solutions achieves the fundamental objective

The Qualitative Value Model comprises two elements:

- the Objectives Hierarchy
- a set of Value Measures / Measures of Effectiveness

The starting point for the development of the Objectives Hierarchy is the "Overall Fundamental Objective" that defines the highest level goal. It is decomposed into a first layer of objectives; each of these are then decomposed in further layers, until a level is reached where for each objective can be defined one or more metrics (Value Measures/Measures of Effectiveness); these enable the assessment of how well the objective is attained by a specific solution. Building the objectives hierarchy promotes an in-depth analysis of values, since defining lower level objectives add to the definition and understanding of the higher level ones. As the hierarchy defines the relations between objectives at different layers, the attainment of each higher level objective can be evaluated on the basis of the attainment of corresponding lower level ones. This can be carried on down to the lowest level objectives whose attainment can be assessed on the basis of measurable criteria. Sources of goals and objectives can be: approved documents, the actual stakeholders or Subject Matter Experts acting as "surrogates".

To result in a proper Qualitative Value Model, the methodology requires that the decomposition process results in an objective hierarchy that is:

- complete (the objectives and value measures must cover all the aspects needed for the evaluation of the overall objective),
- non-redundant (objectives and value measures on the same layer should not overlap),
- decomposable (the grade attained in a specific value measure must be independent from the grades of all the other measures at the same level),
- operable (understood in the same way by all the concerned parties),
- small in size (easier evaluation of alternatives, focus on most important aspects).

The development process of the Qualitative Value Model helps the "hard thinking" needed to discover and analyse not just needs, but also undesired effects and consequences. The ordered hierarchy building approach greatly facilitates the discovery of hidden objectives and unstated needs. An additional benefit provided by the adoption of the values-first VFT approach is to broaden the range of alternatives considered, thanks to the reduction of the anchoring effect of already identified alternatives usually affecting the alternative-based approach.

The Quantitative Value Model makes the Value Model actionable, providing a means for the evaluation of how well each of the alternative solutions achieves the fundamental objective, elaborated on the basis of the degree to which the alternative solution meets the lowest level objectives of the hierarchy. To this purpose, the Quantitative Value Model complements the objectives hierarchy and Measures of Effectiveness of the Qualitative model with some additional features: Utility Functions, Weights of relative importance, and Mathematical Expression.

The Measures of Effectiveness (MoEs) give a quantitative evaluation of how well each alternative solution attains the lowest level objectives. The performance data of MoEs are measured each with specific units and ranges, so they can not be directly combined.

The Utility Functions first purpose is to transform these data into normalized and dimensionless "utility" figures. Since the utility refers to the value perceived by the stakeholders, a further feature is the ability to take into account their "preferences"; these are usually based not only on objective criteria, but also on subjective, judgmental ones.

The Weights provide the means to prioritize objectives, on the basis of their relative importance for the stakeholders, or their ability to concur to higher level objectives.

Finally, the Mathematical Expression enable the calculation of one overall evaluation figure (Overall Figure of Merit) for each of the alternative solutions.

The method guarantees that all the solution are evaluated in a consistent "fair" way, reducing the effects of the human innate judgmental bias.

A wide choice of methods can be applied to define the elements of the Quantitative Value Model; a large body of literature is available on the topic.

The attainment of a higher level objective is computed combining with the mathematical expression the results of its lower level elements; this is repeated layer by layer from the lowest to the highest.

2.1.3 A triad of constructs: "Rosetta stone"

The main groups of stakeholders involved in the development of a SoS are the customers, the users and the developers. While they are well entitled to hold different perspectives on the SoS, to enable a fruitful cooperation these must be reconciled to provide a common ground of understanding. To that purpose, the Originating Concept approach exploits the affinity between three constructs: Capability, Operational Thread and Use Case.

Capability is a construct coined in the last years by the military to support the Capability Based Planning approach. It is defined by the "Joint Capabilities Integration and Development System" as: "the ability to achieve a desired effect under specified standards and conditions through combinations of means and ways across the doctrine, organization, training, materiel, leadership and education, personnel, and facilities (DOTMLPF) to perform a set of tasks to execute a specified course of action." [7]

The Operational Thread construct is widely used by the military. It is defined by the "NATO Architecture Framework" as: "a set of operational activities, with sequence and timing attributes of the activities, and includes the information needed to accomplish the activities." The same document further elaborates: "A particular operational thread may be used to depict a military capability. In this manner, a capability is defined in terms of the attributes required to accomplish a given mission objective by modelling the set of activities and their attributes. The sequence of activities forms the basis for defining and understanding the many factors that impact on the overall military capability." [8]

The Use Case construct is mostly used for the specification of functional requirements in Software Engineering, but it is also widely applied to the same purpose in Systems Engineering, and to the description of business processes (Business Use Case). One of the most cited definition of the Use Case is the one from Ivar Jacobson: "a sequence of transactions performed by a system, which yields an observable result of value for a particular actor". [9] One of the reason of the effectiveness of the Use Case construct is the combination of the a description using the natural language, with a strong structuration and formalization; the use of natural language allows "non-technical" stakeholders to fully understand and get effectively involved in the definition of the functionalities. On the other hand, just thanks to its structuration and formalization, the resulting information is in a format well suited to be used by the developers as the basis of their activity.

If the three constructs are analysed together, it becomes apparent that they show a strong affinity; more specifically, it can be pointed out that all of them make reference to two main features:

- a desired effect (result, objective)
- a set of tasks (activities,transactions)

On the other hand, the constructs were originated, and are mainly used, in different contexts:

- The Capability belongs mostly to the field of high level planning and acquisition ("customer")
- The Operational Thread is mainly used by operatives: users and Subject Matter Experts.
- The use case one is one of the main components of the "toolbox" of developers.

The conjoined use of the three constructs also helps in "bridging" the divide between the mindsets and languages of the groups, providing for a powerful means of information exchange, knowledge sharing, and consensus building. To state this in a more suggestive way, one can see the three constructs as forming together a sort of "Rosetta Stone" enabling effective communication between users and developers.

It must be noted that, even though two of the constructs are part of the military lexicon, they are suitable to be applied in any operational context, where processes (sequences of actions) are used to produce results.

2.2 Expanded Operational Effectiveness Model

The basic definition of Operational Effectiveness from the DAU Systems Engineering Fundamentals Guide reads: "Operational effectiveness is the overall degree of a system's capability to achieve mission success considering the total operational environment." [10] In a more articulated way, the Defense Acquisition Guidebook [11] defines the "Affordable System Operational Effectiveness" as the combination of Mission Effectiveness with Life Cycle Cost/Total Ownership Cost. Mission Effectiveness is, defined as the combination of Design Effectiveness and Process Efficiency; Design Effectiveness is, finally, defined as the combination of Technical Performance and Supportability. Process Efficiency includes: production maintenance, logistics and operations. Supportability includes: reliability, maintainability, and support features Technical Performance includes functions and performances. This last term refers to the functional needs; all the other needs can be classified as non-functional needs.

The "Expanded" Operational Effectiveness further broadens the scope of the concept to include additional non-functional elements that play a role in the SoS ability to achieve customers' satisfaction, like: modularity, scalability, flexibility, adaptability and robustness. This includes also new concepts, like sustainability (the use of resources without depletion or damage) and socio-economic return. These last require to take into account the interaction of the SoS with a wider context (not just technical, but also environmental and socio-economical).

The Expanded Operational Effectiveness Model intent is to provide a holistic description of the complete set of high level requirements that the candidate SoS solutions must address. The analysis of the functional needs requires a focus on the operational aspects, processes, services and functions that the SoS must perform; on the other hand, non-functional needs are related to qualities, constrains and limitation that are global, in the sense that apply to all the constituents of the SoS. Therefore it seems preferable, when proceeding to develop the Expanded Operational Effectiveness Model, to address separately the functional and non-functional needs, obtaining two distinct models, to be later integrated in an overall model.

2.2.1 Functional Needs Model

The needs and expectations for complex, large scale SoSs are, usually, expressed by high level objectives and strategies. The level of abstraction of this kind of reference makes it not suited to directly support the development of the SoS. For the guidance of the generation of alternative solutions and of their evaluation, are required concrete, low level objectives and measurable criteria. These can be obtained by means of the analysis and decomposition of the high level needs.

The Originating Concept method for the development of the Functional Needs Model combines the Mission Analysis method with the synthesis of the Capability, Thread and Use Case constructs. The top-down decomposition procedure of the Mission Analysis starts from the high level objectives and develops a hierarchical structure. The clear and easily understandable form of the objectives hierarchy facilitates communication and consensus building between the customers', users' and developers' communities. The decomposition is carried out until a level of resolution is reached where two conditions are met:

- a clear effect resulting from the achievement of the objective can be identified,
- one, or a set of, measures of performance that define how well the effect is attained, can be identified.

It is apparent how he lowest level elements of the hierarchy of objectives, defined this way, satisfy the formal definition of Capability seen before. Each of the identified Capabilities, can then be analysed in further detail in order to define an associated Thread. This can be done with the support of operatives (users or Subject Matter Experts) that define a suitable Course of Action to attain the desired effect. This activity provides an excellent opportunity to actively involve the users. The definition of threads and Courses of Action is preferably supported by the definition of scenarios; an extensive coverage of this specific topic is provided by [4].

In a last step, the Capabilities and Threads descriptions are refined to comply with the Use Case format. The adoption of this more formal description goes in the direction of the developers' perspective, making it ready to be applied for the generation of alternative solution and for the further development activities, greatly reducing the risk of misunderstanding and wrong assumptions. Nonetheless, as seen, the Use Case construct allows the operatives to continue to be involved, thus they can validate the end result.

The Originating Concept model of the functional needs produced following the above approach, comprises:

- a hierarchy of objectives defining the interrelations between the high level objectives and lowest level ones (Capabilities)
- a set of Use Cases, one for each of the Capabilities, that defines the high level requirements for the SoS and is completed with the reference for the evaluation of their fulfilment.

The affinity of the composition of the above model with the Qualitative Value Model of VFT seen before is quite apparent. Indeed the VFT method has been demonstrated suitable for the analysis of functional needs of systems [1]. However, specially when dealing with SoSs where the dominant perspective is more often operational than functional, the Mission Analysis approach is preferable as it is already familiar to many stakeholders, and it is well aligned with their perspective of the SoS. In view of a later integration with the non-functional needs model, when developing the objective hierarchy with the Mission Analysis, one has to take care to make it also compliant with the characteristics required for a proper Qualitative Value Model. The limited additional effort is rewarded also by the better quality of the end result and by the facilitated detection of missing, or redundant, elements.

The set of Use Cases, notably when related to suitable scenarios, provides also a good reference for the Validation and Verification processes.

2.2.2 Non-functional Needs Model

Probably due to the fact that the functional needs are related to what the SoS is expected to provide, in term of functionalities and services, all stakeholders tend to focus on them and to give a low priority to the analysis of the non-functional needs. This is a major shortcoming, since failing to properly address non-functional needs can typically lead to an unsatisfied customer or, worse, cause the product to be not accepted.

Typically non-functional needs are defined as general goals, often high level ones. As a remarkable example, let just be cited the definition of sustainability of the National Environmental Policy Act: "To create and maintain conditions under which humans and nature can exist in productive harmony and that permit fulfilling social, economic, and other requirements of present and future generations." Also the concerns about the potential interactions of the SoS with its operating context, are usually defined as general goals (e.g. the need, for security systems, to "minimize the interference with the operation of other systems"). The non-functional needs model includes also the cost goals (e.g. Life Cycle Cost, Total Ownership Cost), thus providing a complete reference.

To be useful in guiding the generation and evaluation of alternatives, these high level goals must be articulated into lower level objectives with the final intent of identifying measurable characteristics. The development approach of the VFT Qualitative Value Model is optimally suited to define the objectives hierarchy of the non-functional needs. The Non-functional Needs Model that is developed this way, has the same characteristics and is fully compatible and integrable with the Functional Needs Model seen before.

The product of the integration of the two models is the Expanded Operational Effectiveness Model that provides a comprehensive view of all the needs, expectation and desires of the stakeholders and, also, of the constraints and expected qualities.

2.3 Expanded Operational Effectiveness Evaluation

The Expanded Operational Effectiveness Model is a Qualitative Value Model; to support the evaluation of the effectiveness of the candidate solutions, it must be complemented by a corresponding Quantitative value Model. The end result is the Originating Concept. The Qualitative Value Model comprises the information about the relative importance of the objectives, in terms of their contribution to the achievement of the highest level goal. This information is quantified and recorded by the Utility Functions and the Weights. Some of these parameters can be defined using "hard" techniques, based on quantitative models and mathematical formulations (sometimes just "back of the envelope" calculations). In many cases this approach proves not practical or feasible; an alternative approach is to revert to "soft" techniques, that rely on judgments, based on experience, of knowledgeable individuals. The best option is to involve in this activity the stakeholders, to elicit and capture their knowledge about the problem and its operational context. If the stakeholder are not available, Subject Matter Experts can act as "surrogates"; in this case, the end result must be reviewed and agreed upon by the stakeholders. While the results obtained with hard techniques are more repeatable and unbiased, the soft techniques have a superior capability to accurately represent the stakeholders' appraisal and perception of the value of the objectives in the hierarchy.

To support and facilitate the elicitation of the information from the stakeholders, many methods are available. One of the main advantages of these methods is to make explicit and document also the subjective component of the judgments. This allows for the review, validation and refinement of the model, thus making it more dependable and trustworthy. Since the quantitative model is associated with the complete effectiveness model, it is assured that the priorities are assigned to all the objectives in a uniform and consistent way; this is an essential condition to be able to identify the best balanced solution.

The ultimate result is a "measurement tool" that provides, for each candidate solution, an overall evaluation of the effectiveness that is the best approximation of the customer own evaluation.

To evaluate the effectiveness of a candidate solution, it must be defined and modeled to a level that enables to estimate its score in each of the Measure of Effectiveness / Value Measures of the Expanded Operational Effectiveness Model; these are then elaborated with the quantitative model relations and parameters.

3 Conclusions

The Originating Concept approach shows promise to enable the development of an artefact that can provide the needed support to the Architectural Decision-Making process for a SoS. The set of models that form the artefact provide also the reference for the generation of alternative solutions and contains most of the information required to support the development of the preferred solution, its validation and verification. All the components of the Originating Concept are defined in formats that enable all stakeholders (customer, users, developers) to get an in-depth understanding and to effectively participate in their development, refinement and validation.

References

[1] Decision Making in Systems Engineering and Management, G.S.Parnell, P.J.Driscoll, D.L.Henderson, John Wiley & Sons, 2008

[2] International Standard ISO/IEC 15288 Systems and software engineering — System life cycle processes

[3] IEEE Guide for Information Technology-System Definition-Concept of Operations (ConOps) Document - IEEE Std 1362-1998

[4] Guide to the Preparation of Operational Concept Documents - BSR/AIAA G-043A (DRAFT-2011)

[5] Strategies to Tasks: A Framework for Linking Means and Ends, David E. Thaler, RAND, 1993

[6] Value-Focused Thinking: A Path to Creative Decision making, Ralph L. Keeney, Harvard University Press, 1996

[7] CJCSI 3170.01G, JOINT CAPABILITIES INTEGRATION AND DEVELOPMENT SYSTEM, CHAIRMAN OF THE JOINT CHIEFS OF STAFF INSTRUCTION, 1 March 2009

[8] NATO Architecture Framework v3, CHAPTER 4

[9] I. Jacobson, M. Griss and P. Jonsson. Software Reuse.' Architecture Process and Organization for Business Success. Addison-Wesley 1997,

[10] Systems Engineering Fundamentals, Defense Acquisition University Press, 2001

[11] Defense Acquisition Guidebook, January 10, 2012

An Interoperable Reconstruction and Recovery Decision Support Tool for Complex Crises Situations

Francesca Matarese, Dario Di Crescenzo, Antonio Strano
SESM, c/o SELEX Sistemi Integrati
Via Circumvallazione Esterna, loc. Pontericcio
80014 Giugliano in Campania, Italy
fmatarese@sesm.it, ddicrescenzo@sesm.it, astrano@sesm.it

Florence Aligne, Juliette Mattioli
Decision Technologies and Maths. Lab., Thales
1, rue A. Fresnel, 91767 Palaiseau Cedex, France
florence.aligne@thalesgroup.com
juliette.mattioli@thalesgroup.com

Abstract - *Today, more people than ever are threatened by crisis situations, both from natural origin or caused by malicious acts. These situations lead to both economic and humanitarian tragedies. Reconstruction and recovery operations are increasingly longer-lasting, costly and complex, especially when CBRN decontamination is necessary. Coordinated needs assessment, reconstruction and recovery planning can contribute significantly to mitigate the economic and humanitarian impact of such disasters; yet, no advanced software platform is available to support this process as a joint operation in which information is continuously updated and shared between the organisations, progress monitored and accountability facilitated. FP7 DESTRIERO project aims at developing an advanced net-centric information management tool, which structures and presents information to collaborative groups of (international) stakeholder organisations, supports damage and needs assessment and recovery planning. Based on international standards, procedures and methodologies, the tool will offer an integrated framework with innovative functionalities as described in this paper.*

Keywords: Post-Crisis Needs Assessment, Reconstruction and Recovery Planning, International Cooperation, Interoperability, Common Operational Picture.

1 Introduction

Today, more people than ever are threatened by crisis situations, both from natural origin, such as earthquakes and floods, or caused by malicious acts, such as terrorist attacks. Besides dramatic **structural damages** also **CBRN contamination risks** can occur as a consequence of these events, leading to both economic and humanitarian tragedies. Ever wider geographic areas are affected, sometimes crossing national borders, while **reconstruction and recovery operations are increasingly longer-lasting, costly and complex,** especially when decontamination is necessary.

In such situations, emergency management, but also **Post-Crisis Damage and Needs Assessment (PDNA)** and **Reconstruction and Recovery Planning (RRP),** is usually coordinated by local authorities or dedicated civil protection organisations, with the support of a variety of different national and international relief organisations **acting relatively autonomously.** The damage assessment needs analysis, recovery, and reconstruction planning process is typically coordinated through periodic **physical meetings** of the involved organisations, in which information is shared about the situation, priorities set and responsibilities allocated. Follow-up and execution of tasks is managed by each individual relief organisation, supported by a range of more or less proprietary not interoperable tools. No advanced software platform or tools are available to support this process as a joint operation in which information is continuously updated and shared between the organisations, progress monitored and accountability facilitated.

FP7 DESTRIERO project[1] aims at developing an advanced **net-centric information management tool**, which structures and presents information to collaborative groups of (international) stakeholder organisations and supports damage and needs assessment as well as recovery planning. Based on international standards, procedures and methodologies (e.g. Damage and Loss Assessment (DaLA) Methodology), the tool will offer an integrated framework with innovative functionalities.

This paper presents the approach that has been chosen for the DESTRIERO project. Section 2 details the concept and objectives targeted in DESTRIERO. These objectives will be fulfilled by a set of interoperable tools with innovative functionalities that are described in Section 3.

[1] DESTRIERO, A **DE**cision **S**upport **T**ool for **R**econstruction and recovery and for the **I**ntEroperability of international **R**elief units in case **O**f complex crises situations, including CBRN contamination risks. The proposal is entering negotiation phase, in the frame of FP7 2012 Security Call.

2 Concepts and objectives

DESTRIERO aims at developing a next generation post-crisis needs assessment tool for reconstruction and recovery planning. It will include structural damage assessment through advanced remote sensing enriched by in-field data collection (buildings, bridges, dams) and related data integration and analysis. The approach is based on international standards, novel (automated) data and **information interoperability across organisations** and systems, in combination with an advanced **multi-criteria decision analysis tool** and methodology for multi-stakeholder information analyses, priority setting, decision making and recovery planning.

2.1 Context and needs

Disasters have a major impact on the living conditions, economic performance and environmental assets and services of affected countries or regions. Consequences may be long term and may even irreversibly affect economic and social structures and the environment. Globally, statistics show that disasters cause more socially significant and irreversible damage in developing countries, where the poorest and most vulnerable population groups feel the most severe impact. In the developed world, on the other hand, an increasing and significant degree of protection against disasters has been achieved over the years thanks to the availability of resources and technology for the introduction of effective prevention, mitigation and planning measures, together with vulnerability reduction schemes. However, even in these countries damages have risen significantly as a result of the greater concentration and value of societal activities. Moreover, in industrialised countries, disasters cause massive damage to the large stock of accumulated capital such that recovery and reconstruction may extend over longer periods of time [1].

Adequately coordinated needs assessment, reconstruction and recovery planning can contribute significantly to mitigate the economic and humanitarian impact of such disasters[2], and speed-up the required actions to restore the situation in "**a state close to the original**".

The needs assessment and recovery planning process

When a disaster occurs, national-emergency bodies are generally in charge of assessing **inflicted damages** and **humanitarian needs** during the first emergency stage. Immediately **after the first emergency (humanitarian) stage**, a further assessment must be made of the direct and indirect effects of the event and their consequences on the social well-being and economic performance of the affected country or area. Assessments can be defined as "the set of activities necessary to understand a given situation". They include "the collection, up-dating and analysis of data pertaining to the population of concern (needs, capacities, resources, etc.), as well as the state of infrastructure and general socio-economic conditions in a given location/area"[3]. Needs assessment in post-disaster situations, which are already unfavourable conditions, is obviously complicated in the case of CBRN contamination.

With assessment results in hand, it is possible to **determine the rough extent of the recovery and reconstruction requirements**, which is an urgent task since those affected cannot wait long under the conditions prevailing after a disaster occurs. In the case of **CBRN** contamination, it has to be urgently established what the extent of contamination is, and what will be required to restore the situation to a safe environment. Such an exercise is mandatory for identifying and undertaking reconstruction programme and projects, many of which will require the international community's financial and technical cooperation. The recovery process may last many years, depending on the type of damage and specific geographic area. **In case of CBRN contamination, the recovery period may even last up to 10 or 20 years** [1].

A key responsibility at the country level is to ensure that humanitarian actors build on local capacities and maintain appropriate links with government and local authorities, state institutions, civil society and other stakeholders. The nature of these links will depend on the situation in each country and the willingness and capacity of each of these actors to lead or participate in humanitarian activities. Whatever the situation, needs assessment requires coordination, joint planning and monitoring of the assessment activities of multiple independent organisations. This is a relatively complex situation that is even amplified if international support mechanisms are involved. In international humanitarian crises situations, needs assessments are often carried out with support from the United Nations, European Union and other public and private international organisations and NGO's, using international standards and procedures such as:

- The **Inter-Agency Standing Committee (IASC) Operational Guidance for Coordinated Assessments in Humanitarian Crises**[4] to

[2] "Experience has shown that there are significant benefits to coordinating needs assessments and that doing so can help save more lives and restore more people's livelihoods" OCHA – United Nations Office for the Coordination of Human Affairs.

[3] Glossary of key terms (and concepts), World Health Organisation

[4] Developed by the Needs Assessment Task Force (October 2011) of the Inter-Agency Standing Committee (IASC), to facilitate coordination between International Organisations

facilitate collaboration between international organisations within the UN established <u>Inter Cluster Coordination Mechanism</u>;

- The **Damage and Loss Assessment (DaLA) Methodology**[5] that provides guidelines to evaluate the damage caused by a disaster;

- The **Post Disaster Needs Assessment (PDNA)**[6] that aims at increasing national capacity to lead efforts to determine recovery requirements and priorities from early to full recovery.

The <u>IASC Inter Cluster Coordination Mechanism</u> which is normally chaired by OCHA requires complex coordination and reporting lines between involved stakeholder organisations as schematically indicated in Figure 1 that presents the IASC Humanitarian Coordination Architecture[2].

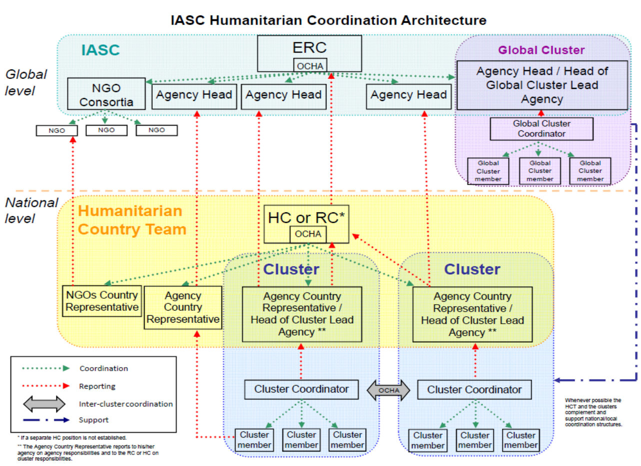


Figure 1: IASC Humanitarian Coordination Architecture

Whereas significant efforts are undertaken by International Organisations such as EC, UN and World Bank to align such international collaboration at the higher political level (i.e. International Cluster Organisation) and to coordinate stakeholders in the early recovery management stages (e.g. DaLa and PDNA), no supporting software tool is available and <u>important gaps remain in optimising coordination at the **operational** level</u>, towards longer-term recovery planning in a joint multi-stakeholder set-up. This has to do with the fact that different organisations use different procedures, language, tools and reporting standards.

[5] Developed by the Economic Commission for Latin America and the Caribbean (ECLAC), 1970

[6] Under development by the United Nations (led by UNDP), the World Bank, the European Commission, as per the Joint Declaration on Post Crisis Needs Assessment and Recovery Planning, 25 September 2008

"An issue in many emergencies is not always an absence of assessment information but rather the capacity to quickly validate and analyze the information necessary to determine priorities and guide planning of the humanitarian response. All too often, assessment data is difficult to access (i.e. spread over different systems and organisations), insufficiently shared or used, and data sets from different assessments are not comparable. Also there is insufficient time to aggregate data from multiple assessments, information needs are not sufficiently prioritised and data collection processes are cumbersome" [6]. This all leads to a sub-optimal needs assessment and recovery operation.

Hence solutions in supporting the operational joint needs assessment, recovery and reconstruction planning process have to be found towards:

- Coordinated assessments of different recovery actors to avoid incongruent data, redundancy of effort and unnecessary delays;

- Sharing of assessment data from different stakeholder organisations across different systems (data and information interoperability) and coherent analyses of this information towards establishing joint conclusions and priorities in recovery needs;

- Identifying cross-cutting needs related to recovery capacity, information management and communication;

- Coordinating and (re)planning recovery efforts of involved relief organisations based on established priorities and progressive results;

- Managing accountability of (humanitarian) aid contributions based on continuous update information on progress regarding a multitude of the recovery projects financed by multiple donors.

2.2 Objectives

DESTRIERO will involve a representative group of international end-users to provide exact user requirements, analyse state-of-the-art assessment tools and define the functional requirements for a joint PDNA and RRP process, as well as supporting tool-set. The main overall objective of the project is to develop a new net-centred collaborative framework and tool-set, to support, improve and speed-up the operational recovery process over longer periods of time, with the following four specific objectives:

Objective #1: Faster and better damage assessment for planning and monitoring of progress of recovery.

DESTRIERO will improve damage assessment by integrating satellite data, aerial photos and data from the field (e.g. from mobile devices) into a coherent information management tool. The tool will offer remote sensing functionalities to create a first and fast territorial picture of the scope of the damage, establishing at the same time the fundament for a GIS based visual interface (see objective #2) and enable continued monitoring during the recovery process.

Objective # 2: Facilitate fast and intuitive access for distributed users to visualise the dynamic "common operational picture", during the planning and reconstruction period.

DESTRIERO will integrate state of the art visualisation techniques based on I-NAV (Integrated NAVigation Services) developed in the FP6 LIMES project (FP6 Programme – Aeronautics&Space/GMES Security). The LIMES and the other GMES projects foreground will be integrated to present PDNA and RRP information in a multilayer GIS type user interface, using satellite damage assessment and monitoring maps as a basis. Tagging of geographical areas (e.g. on contamination) and damaged buildings (e.g. on state of reconstruction) on satellite images will enable distributed users in control rooms or on the field to drill down to different layers of damage reports, pictures, remote control data, needs assessment and priority reports as well as progress information on-going the recovery phase.

Objective #3: Better collaborative decision making during the planning and reconstruction phase.

DESTRIERO will support standardisation of assessment data, interoperability between different information systems from stakeholders and capabilities to compare damage and recovery requirements towards prioritisation and joint decision making. DESTRIERO tool will offer standard damage assessment templates, built on internationally agreed procedures and standards (e.g. DaLa), to collect data in homogeneous formats. A novel decision support tool, based on a recently developed methodology and a prototype to support multi-criteria consensus building to determine priorities in the crisis reconstruction phase and strategic decision-making, will be one of the building blocks of DESTRIERO.

Objective #4: Improve management information in relation to PDNA and RRP.

DESTRIERO will provide a single access point and on-line library with PDNA and RRP Frameworks, high level overviews of planned and running recovery projects including multi-donor funding information and the possibility to trace related surveillance data for progress monitoring. As such DESTRIERO will overcome insufficient accountability among humanitarian agencies and inconsistencies in donor policies by enhancing humanitarian response capacity, predictability, accountability and partnership.

The innovative set-up of the DESTRIERO system has been visualised in Figure 2, highlighting:

- the underlying PDNA and RRP process, starting at the early recovery phase, through to initial and continued damage and loss assessment, joint priority setting and decision making for recovery planning and continuous monitoring and updating of assessment data and information;
- the supporting building blocks and tool-set (system interfaces, surveillance techniques, Joint Decisions Support and Management information, GIS based user interface) to be integrated in DESTRIERO.

The involved (inter) national stakeholder organisation, their different involvement levels and the international standards and procedures to be taken into consideration in DESTRIERO.

3 DESTRIERO tools set

The DESTRIERO platform will be modelled on the key **internationally agreed methodologies, procedures and standards** for disaster recovery. To support the information needs as well as the decision making and planning processes, DESTRIERO will furthermore incorporate **state of the art ICT and Decision Support System technologies**.

3.1 International standards and procedures

According to existing guidance material on needs assessment in post-crisis situations, some standards for information presentation and sharing have been defined, in order to obtain homogeneous information and to apply objective prioritisation criteria. Unfortunately these standards are not yet fully shared among relief organisations and mostly used to inform higher management levels of relief organisations. DESTRIERO will strengthen the use of these standards by integrating them within the system.

Use of the ECLAC Damage and Loss Assessment (DaLA) Methodology: DESTRIERO will integrate the DaLa approach; assessment standards and reporting templates to facilitate seamless alignment with (inter)national recovery operations.

Use of operational Guidance for Coordinated Assessments in Humanitarian Crises: DESTRIERO will map the Operational Guidance procedures and standards within the platform to support common and coordinated assessment planning and adoption of common reporting standards to improve the production of comparable assessment data.

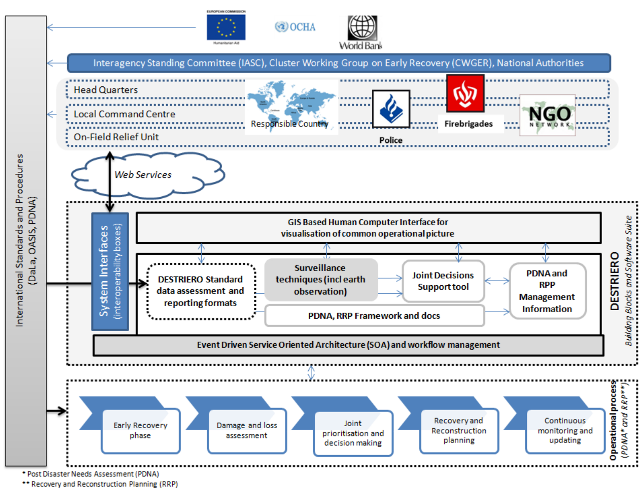


Figure 2: DESTRIERO context and tool set

Use of Post Disaster Needs Assessment (PDNA): DESTRIERO will bridge this gap, integrate these international standards, methodologies and frameworks and provide support tools for sharing of key-documents (e.g. PDNA frameworks), carry out the PDNA (e.g. with remote sensing, mobile terminals and common reporting using agreed standards), facilitate interoperability and sharing of resulting assessment data and information across and between organisations (web-services and interfaces), joint decision making based on common operational picture to develop and implement Recovery Frameworks, including continuous monitoring and management over longer periods of time. As such the DESTRIERO will provide tools that underpin (not replace) project and program management frameworks and methodologies that are used by coordinating and relief organisations.

3.2 State of the art technologies at the base of DESTRIERO

DESTRIERO integrates four state of the art technological solutions, including several novel prototype tools:

1. Harmonising, coordinating and aligning **data collection** processes, offering state of the art surveillance technologies within an integrated information management system for PDNA and RRP.

2. Improving and harmonising presentation of information integrating a prototype **multi-layer GIS** type of user interface for dedicated PDNA and RRP **information from various sources.**

3. Facilitating **joint-decision making** for PDNA and RPP, integrating a recently developed state of the art methodology and prototype ICT Decision Support software Tool (**MYRIAD**) [8][9]. This prototype derives from Qualitative Decision Theory [3][7], in which users enter simple preferences, and uses an elaborate aggregation model that enable to represent complex decision strategies in a flexible way [5].

4. **Making services interoperable** by using newly emerging **Event-Driven Service Oriented Architecture** approach to manage large scale data collection and analyses through a network-centric platform (**SWIM-BOX**)[10], based on the FP6 SWIM-SUIT (System-Wide Information Management SUpported by Innovative Technologies) project results. This approach is used to support the management of large scale data collection and analyses through a network-centric platform.

This leads to the DESTRIERO tool-set, as schematically indicated in Figure 3.

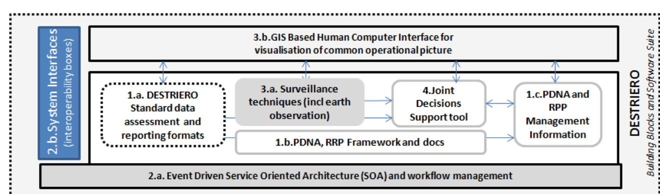


Figure 3: DESTRIERO Tool-set

4 Conclusion

DESTRIERO software will, therefore, be characterised by the following features:

1. Interfacing different distributed and heterogeneous information systems, within a joint interoperable information platform based on Event Driven Service Oriented Architecture (SOA),

2. Integrating international data assessment and presentation of standards and procedures within a joint PDNA and RRP workflow,

3. Integrating state of the art GIS based multilayer visualisation techniques and software for aggregated data from distributed heterogeneous systems,

4. Integrating novel decision support methodologies and support tool to implement joint interpretation of assessment data and strategic decision making as well as priority setting for recovery and reconstruction planning,

5. Facilitating monitoring and reporting based on reporting procedures of international donor organisations.

With the integration of interoperable tools and innovative functionalities, DESTRIERO will contribute directly to the next generation damage and post-crisis needs assessment tool for reconstruction and recovery planning.

References

[1] Joseph A. Bishop, “Guidelines for Environmental Assessment Following Chemical Emergencies”, Joint UNEP/OCHA Environment Unit, 1999

[2] “Civil Protection and Humanitarian Aid - Facing disasters and crisis together”, International Conference, Rome, 10-11 November 2011

[3] D. Dubois, H. Fargier, P. Perny, “Qualitative decision theory with preference relations and comparison uncertainty: an axiomatic approach”, Artificial Intelligence 148 (2003) 219-260.

[4] Economic Comission for Latin America and the Caribbean. “Handbook for Estimating the Socio-economic and Environmental Effects of Disasters”, United Nations, ECLAC, 2003

[5] M. Grabisch and Ch. Labreuche, “A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid”, Annals of Operation Research, n° 175, pp. 247-286, 2010

[6] IASC, “IASC Operational Guidance for Coordinated Assessments in Humanitarian Crises”, February 2011

[7] J. Doyle, R. Thomason, “Background to qualitative decision theory”, The AI Magazine 20 (1999) 55-68

[8] Ch. Labreuche, F. Le Huédé, “MYRIAD: a tool suite for MCDA », Int. Conf. of the Euro Society for Fuzzy Logic and Technology (EUSFLAT), Barcelona, Spain, pp. 204-209, September 7-9 2005

[9] Ch. Labreuche, “A general framework for explaining the results of a multi-attribute preference model”, Artificial Intelligence, n° 175, pp. 1410-1448, 2011

[10] D. Di Crescenzo, A. Strano, et alii, “SWIM-SUIT Prototype System Segment Design Document”, SWIM-SUIT Project, 2010

A systems-of-systems approach to the development of flexible, cost-effective training environments

Luminita Ciocoiu
School of Electronic, Electrical and Systems Engineering
Loughborough University, UK
l.ciocoiu@lboro.ac.uk

Michael J de C Henshaw
School of Electronic, Electrical and Systems Engineering
Loughborough University, UK
m.j.d.henshaw@lboro.ac.uk

Ella-Mae Hubbard
School of Electronic, Electrical and Systems Engineering
Loughborough University, UK
e.hubbard@lboro.ac.uk

Abstract - *In today's aircrew training context, although there is an abundance of training systems that can enhance training and reduce costs, the challenge for the military training organizations to select the most cost-effective training systems to address their immediate and future needs is unresolved. The urgency of this dilemma is exacerbated by shrinking defense budgets. This paper shows how the systems engineering perspective can help the decision-making process for selecting the training media equipment to construct a cost-effective training media environment. A multidisciplinary approach and systems engineering techniques were used to develop a theoretical model of the Mission Training Environment arrangement . Implications of the approach, such as that the training environment can be viewed as a system of systems and that the choice is based on combination of equipment, will be discussed.*

Keywords: Training System, Training Environment, training media, LVC, decision-making, systems of systems.

1 Introduction

In the aircrew training domain, research has intensified in an effort to provide solutions that will ensure an increase in pilot performance (the new operational equipment and environment is more challenging) while there is also a reduction in costs (make the best of existing systems) without compromising safety (for aircrew and civilians).

Lower safety risks, reduced costs and increased operational readiness are benefits offered by the virtual environments. However, virtual environments come in many forms (from virtual simulation to live simulation) and, furthermore, the importance of training in a live environment cannot be underestimated. Therefore, an obvious direction in research is to explore the Mixed Media Training Environments benefits for aircrew training.

Such Mixed Media environments are rarely used (and even more rarely designed); they are constructed on an ad hoc basis as a bottom-up development for a specific exercise, at a specific location and then torn down when the exercise has finished [1]. This makes them difficult to research.

1.1 Cost and opportunity

Asymmetric warfare and shrinking budgets are demands that influence how the military prepares for its activities. Furthermore, these shape the defense industry in terms of development of products and services that aid the process of preparedness of the military, which has resulted in an abundance of "off the shelf" products ready to be used in training programs. Recent developments in the simulation domain have also resulted in high quality products that offer new possibilities to achieve cost-effective training [2], [3].

However, the lack of measurement techniques to assess the benefit of using particular systems in particular ways pose difficulties when it comes to deciding which is the optimum mix of products and services to be used to deliver a cost-effective training exercise.

1.2 Media and aircrew training

The matter of choosing the right mix of training media equipment to deliver cost-effective aircrew training is a question that, in one form or another, has been researched for some time within various domains and, despite the progress made, there are still some issues that need to be resolved [2].

Besides the lack of measurement techniques highlighted earlier, another issue is that although there is a common understanding of the meaning of Live, Virtual and Constructive (LVC) concepts, such that Live means real people operate real equipment; Virtual means real people operate simulating systems; and Constructive means simulated people are operating simulated systems [4], there is no commonly accepted classification and concomitant definitions of media encompassed within the training systems [2]. As a result, terms such as "blurred boundaries" and "blended technology" [1], [2] are more often used.

There is also the problem of capturing and integrating different types of data, such as qualitative data and tacit knowledge, into a rigorous, objective analysis that can aid the process of selecting the training media (equipment) to create an optimum training environment to deliver a cost-effective training exercise.

1.3 The question

With all this in mind, there is an unresolved question of how to create an optimum training media environment to deliver a cost-effective training exercise. This is the question to be answered within this research.

2 Approach

In trying to address as many issues as possible, in an integrative way for the benefit of the overall solution, a multidisciplinary approach was taken to define the problem space and to search for solutions. Therefore, various views from disciplines, such as, Human Factors, Operational Research and Systems Engineering have been taken into account.

2.1 Systems Engineering perspective

A Systems Engineering approach [5] is usually recommended when the problem has a high degree of complexity and there are systems integration challenges. The approach allows the engineer to deal with the complexity by decomposition of concepts and analysis of smaller problems, whilst maintaining focus on the potential interactions between such problems. Furthermore, it helps to define the environment and the boundaries of a problem [6].

The standards and guidelines for System Engineering are usually directed more towards development of new systems, rather that optimization of extant systems, although in practice they are applied to both new and extant systems. For development of extant systems, other approaches that are more specific to Operational Research domain are recommended.

Nevertheless, the inherent holistic thinking and multidisciplinary characteristics of the systems engineering approach makes it ideal to be followed in the present case, as it allows and encourages consideration and integration of multiple perspectives.

2.2 Operational Research

Finding the balance between LVC looks like a straightforward problem to be solved through application of an optimization technique, as such numerical techniques are often used to balance costs against effectiveness [7], [8]. But to be able to apply an optimization technique, certain steps have to be followed and certain criteria have to be satisfied, such as that the problem, the desired solution, the variables, and the dependencies between variables require strict definition [8].

The mathematical optimization models are designed to optimize a specific objective criterion which is subject to a set of constraints and the solution of the model is feasible only if satisfies all the constraints [7]. Therefore, the quality of the solution depends on the completeness of the model through which real-world parameters are reduced or lumped together into assumed real-world parameters. If the abstracted model is incomplete, the solution may not be optimal for the real world system and this raises concerns regarding the adequacy of the mathematical model. This may raise some conflict between the traditional parsimonious modeling approach of finding the simplest model which represents the situation and the SE approach, which focuses on a holistic and integrative view.

Researchers have drawn attention to the fact that human behaviour must also be taken into account when constructing these models to ensure that the solution is adequate and there is no possibility to even fail [8] and that means that human factors data need also to be incorporated and express in these models.

Furthermore, when the context of a system varies greatly, optimization can provide only a short-term advantage and may not be the best solution to make the system more efficient. Fisher [9] also points out that, although optimization is a good technique to increase the efficiency of traditional systems, optimization may undermine adaptability and can become inefficient as the circumstances on which the systems are operating are changing (e.g. increased variability of context = changing training requirements). Users may also be reluctant to repeat the optimization when circumstances change, leading them to rely on inaccurate information.

The application of optimization techniques to solve the problem of finding the balance between LVC for construction of the training media environment for a given training exercise should not be disregarded, but more work is required to fully accommodate all the necessary criteria within the optimization technique for it to be adequate.

2.3 Human Factors and Training Needs Analysis

The purpose of a training exercise, whether it takes place in a live, virtual or mixed media environment is to teach the trainee new knowledge and skills (or develop exiting knowledge or skills). The environment and the method chosen to train have not only to ensure the acquisition and development of skills and knowledge but also to ensure that these skills and knowledge are transferable to real, live situations.

Therefore, two additional variables must be taken into account to decide on the most appropriate arrangement for a training environment. These are degree of transferability of skill and knowledge learned in the training environment, and individual cognitive particularities of trainee.

Distinctions can be drawn between different types of training exercises based on learning stages. Meador [10] makes a distinction between acquisition and retention (or reacquisition), and Frank et. al [1] distinguish between Familiarization, Acquisition, Practice and Validation in their FAPV model. These distinctions have a significant impact on establishing the context of a training exercise and defining the training requirements.

Training requirements are usually derived from the analysis of training needs. Figure 1 shows the TNA (Training Needs Analysis) process. The diagram is an adaptation of the UK MoD TNA Process Diagram depicted in JSP822 report [11], [12].

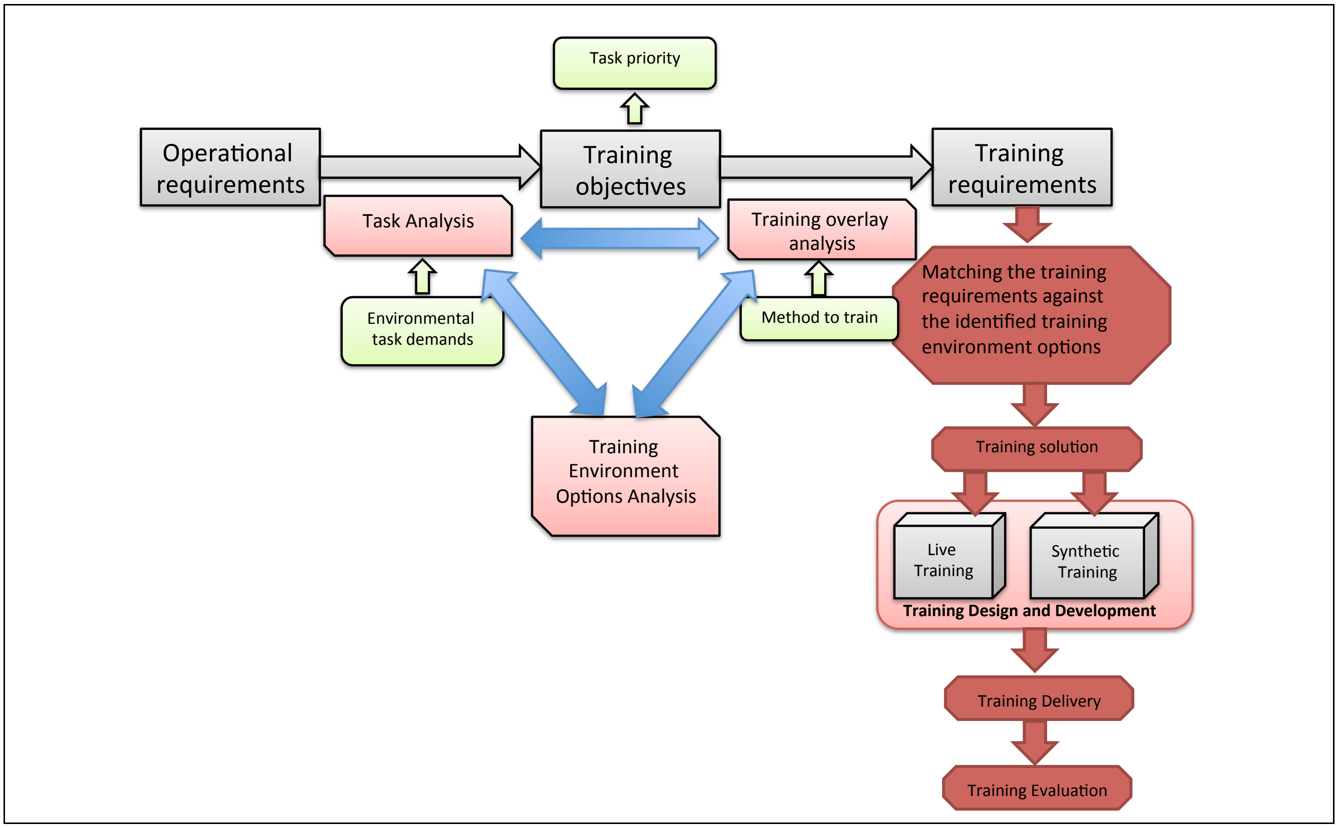


Figure 1. TNA Process Diagram

As can be observed from Figure 1, human factors particularities are captured at the training needs analysis stage that precedes the design and development phase of a training exercise. Furthermore, the training media environment system is decided based on the training requirements resulting from previous analyses.

However, this process is very restrictive. It is a major deficiency that such a process only allows the selection of one training environment system per exercise; it does not allow the possibility of choosing a combination of training systems to create a training environment. A combination may prove to be more cost-effective because it will maximize the usefulness of the available resources. Furthermore, the process of Figure 1 does not take into consideration factors such as schedule, maintenance, cost and other variables that impact the cost-effectiveness of a given training environment.

3 Method

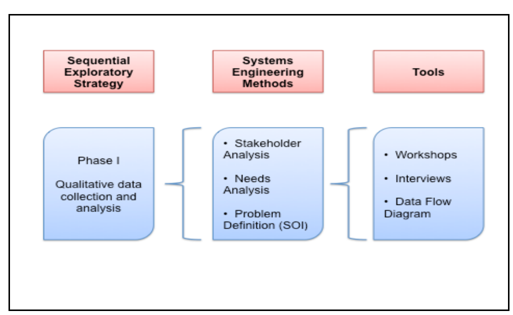


Figure 2. Applied methodology, methods and tools

Because of the high importance of defining and representing the problem through as accurate a model as possible, and because of the issues with this, such as the need to integrate quantitative and qualitative information into the model, a *mixed methods methodology (sequential exploratory strategy)* [13] was used in parallel with *systems engineering methods* to develop a Theoretical Model of Media Environment arrangement for the Mission Training Scenario. The methodology, and the process of methods and tools that was followed are presented in Figure 2.

4 Results

A Theoretical Model of Training Environment arrangement that is presented in Figure 3, has been developed based on the analysis of the information captured from Subject Matter Experts (SME's). The theoretical model is represented by a data flow diagram.

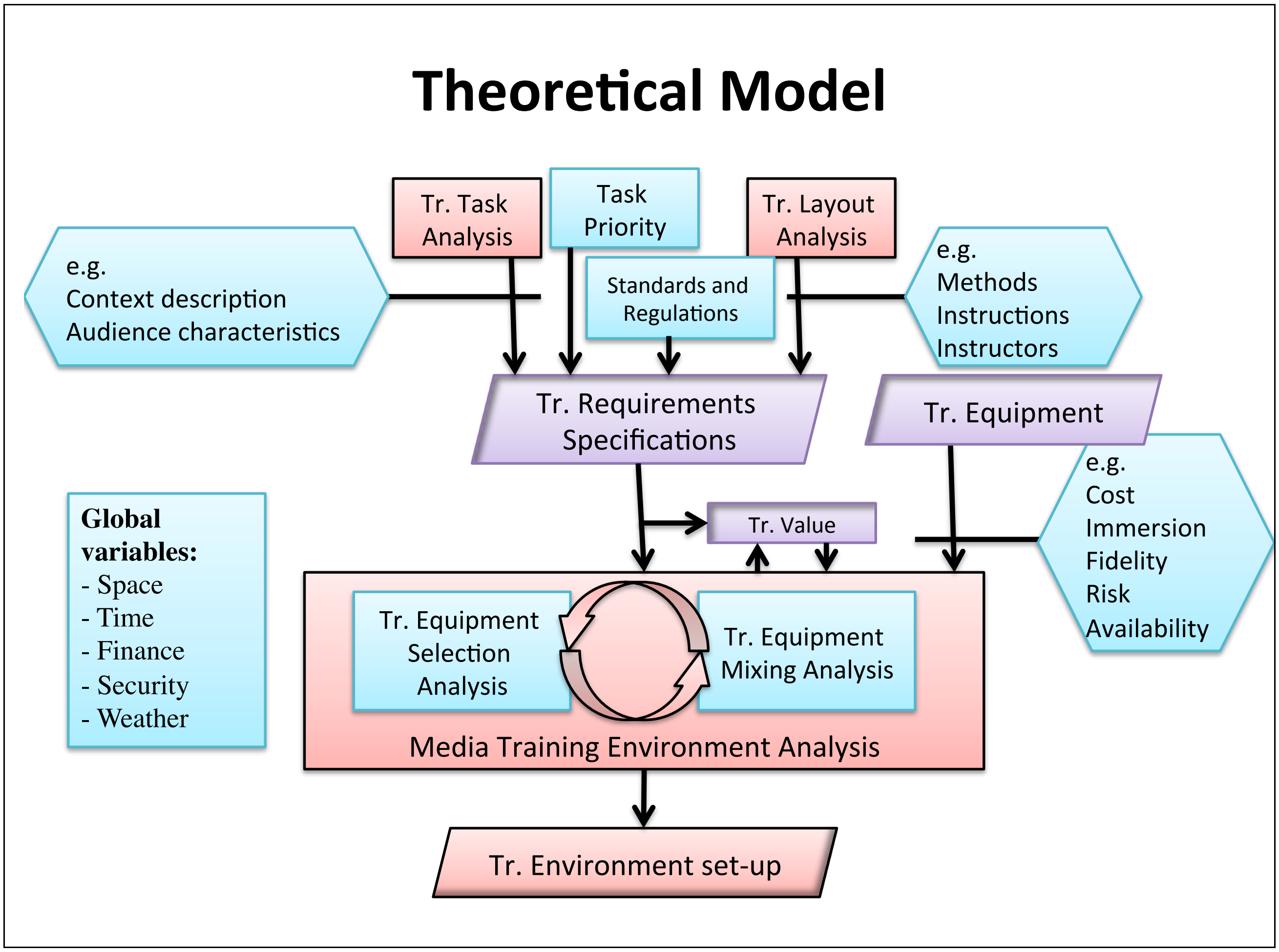


Figure 3. Theoretical Model of Training Environment arrangement

The model presents the System of Interest (SoI), which is the "Training Media Environment Analysis", and the Wider System of Interest (WSoI), which is represented by the factors that influence the behaviour of the SoI.

The model shows the input data necessary for the system, the needed transformation functions and the required output. Furthermore, some specific and global variables upon which the decision making of selection relies are highlighted.

5 Discussions

The idea put forward by the model is that, if the properties exhibited by the relationship between training equipment and LVC technology, Media, Training Equipment and Training Environment are investigated, the following may be concluded:

Media ⊃ Training Equipment

And,

Training Equipment = Training Environment

And as,

Training Equipment = Training System

Then,

Training Environment = Training System

However, this relationship is true only in the case when (after the decision that was made on the selection of the training equipment) the result is to use only one training equipment to deliver the training exercise (for example, a ground base simulator).

If the decision is to use more than one training equipment to deliver a training exercise, then the following can be concluded:

Media ⊃ Training Equipment ⊃ Training Environment

Then if,

Training Equipment = Training System

Training Environment = Training System

Means that,

Training Environment Sub-system = Training Equipment Systems

Furthermore, if we look at the developments made in the synthetic training domain it can be observed that there is an abundance of off the shelf products that are cheaper than bespoke ones and highly efficient in delivering cost-effective training. But these training systems have not necessarily been designed to be used alongside other training systems.

Because of this interoperability particularity, we propose that the Training Media Environment should be considered to be a Training System of Systems rather than a Training System. Therefore, it can be considered that the setting up of a training media environment is not only a matter of identifying and selecting a cost-effective training system but rather a matter of constructing and managing a System of Systems Training Environment that comprises a mix of LVC technologies.

Furthermore, as the emergent behavior of a system depends on the interactive behavior of its components, the decision of selecting the components of a system is, or should be, directly influenced by the effect resulting from the combination of different components. This means that, the decision making process of selection of the training media equipment to construct a training media environment should be tightly coupled with the training systems mixing analysis.

The developed Theoretical Model of Training Environment Set-up that is proposed in this paper is the first step in a research project the aim of which is to develop a tool to help decision makers in selecting the most appropriate blend of training media to construct a cost-effective training environment for aircrew training. By bringing together, data resulted from training needs analyses and training equipment analyses, coupled with the overall context variables, a more comprehensive tool to aid the decision making process can be built.

This theoretical framework will help the development of a tool that will integrate quantitative as well as qualitative data in its analyses. This will also be beneficial in capturing tacit knowledge that is usually lost when the experts that are making the decisions retire. Furthermore, this model will contribute towards making cost-effective decisions, because it promotes the idea of making the most out of the available resources.

6 Limitations

Although, the proposed theoretical model has been validated at the conceptual level, with the help of military aviation domain SME's, the verification process has not been carried out at this stage. The scope of model applicability is limited to the particular training application associated with mission training scenarios. Although, it is possible that it could be extended to other training applications and domains, there has been no attempt, so far, to validate it more widely.

Further development of the theoretical model may yield additional main variables that have not so far been captured; this will be tested during the next development phase.

7 Conclusions

In answering the question of how to create an optimum training media environment to deliver a cost effective exercise, this research proposes a novel, multidisciplinary approach to be taken forward.

The theoretical model that was developed at this stage represents a first step into integrating multiple types of data into an analysis that will help decision makers, in their process of building an optimum media training environment, to deliver a cost-effective training. The model comprises variables linked with human characteristics as well as with equipment characteristics. Furthermore, it incorporates some global variables that have usually been missed so far. The scope of the model is to address the more complex training needs of the future, and takes a wider perspective of the solution; hence may also generate more cost-effective solutions of greater flexibility.

Furthermore, significantly and explicitly the model includes consideration of human issues and because of this characteristic it could be applied to complex civilian roles as well (e.g. emergency response).

The approach that is put forward in this paper has its limitations, however, it offers an alternative, integrative way to explore the phenomenon of constructing Mixed Media Environments for the benefit of the next generation of aircrew training.

Acknowledgments

The authors would like to thank the Abbey Wood De&S, Flight Simulation and Synthetic Trainers team, Sqn Ldr Ian Thompson RAF and Chris Pilfold for their valuable inputs into this research. Furthermore, we are very grateful to Wg Cdr Mike Saunders, BAE Systems and ABTC RAF Waddington for their cooperation and helpful comments. Without their help the ideas put forward in this paper would not have been developed.

We would also like to acknowledge BAE Systems MAS and the Engineering and Physical Sciences Research Council (EPSRC) who are funding this research.

References

[1] G.A. Frank, R.F. Helms II and D. Voor, “Determining The Right Mix of Live, Virtual, and Constructive Training,” Proc. Inservice/Industry Training, Simulation and Education Conference, Arlington, pp. 1268-1277, 2000.

[2] B. Kirby, G. Fletcher and H. Dudfield, “Live Virtual Constructive Training Blend Optimisation Study,” NATO Modeling and Simulation Group Symposium, Bern, pp. 18, 1-10, October, 2011.

[3] J.H. Prost, B.T. Schreiber, W. Bennett Jr. and K. Kleinlein, “Providing Effective and Efficient Training: A Model for Comparing Simulator Improvements,” Proc. Simulation Interoperability Workshop Conference, Orlando, pp. 11-17, September, 2008.

[4] J. van der Pal, M. Keuning and A. Lemmers, “A Comprehensive Perspective on Training: Live, Virtual and Constructive,” NATO Modeling and Simulation Group Symposium, Bern, pp. 13, 1-10, October, 2011.

[5] A.T. Bahill and B. Gissing, “Re-evaluatingSystems Engineering Concepts Using System Thinking,” IEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews, Vol 28, No. 4, pp. 516-527, November, 1998.

[6] L.L. Bordelon, “Systems Engineering – An Everyday Process,” In Mission Systems Engineering Educational Notes RTO-SCI-176, pp. 1-1 – 1-20, Neuilly-sur-Seine, France: RTO, 2006. Available from: http://www.rto.nato.int.

[7] R. Baldick, *Applied Optimization: Formulations and Algorithms for Engineering Systems*, University Press, Cambridge, 2006.

[8] H.A. Taha, *Operations Research: An Introduction (9th Ed.),* Pearson Education, 2010.

[9] D.A. Fisher, *An Emergent Perspective on Interoperation in Systems of Systems*, Carnegie Mellon, Pittsburg, 2006.

[10] D.P. Meador, *Modeling Training Effects on Task Performance Using a Human Performance Taxonomy*, Wright State University, November, 2008.

[11] J. Pike and J. Huddleston, “Training Needs Analysis for Team and Collective Training”, HFIDTCPIII Report, April, 2011.

[12] JSP 822, “Training Needs Analysis,” Defence tRAINING support Manual, Part 5, Ch 3, 2007.

[13] J.W. Cresswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 3rd Ed, Sage, 2009.

System of Systems Architecture Framework (SoSAF) for Production Industries

Asif Mahmood, Francesca Montagna
Dip. di Ingegneria Gestionale e della Produzione
Politecnico di Torino, Italy
Email: firstname.lastname@polito.it

Abstract – *Streamlining of System of Systems (SoS) concepts by contemporary researchers have marked an epoch in the development of cross domain postmodern technologies. This demands us to reflect for a moment, and refine and mature SoS concepts and principles on universal grounds independent of any field (i.e., military). This paper spotlights the necessity of a generic System of Systems Architecture Framework (SoSAF). It also applies the SoS concepts in the Production system field just using the SoSAF paradigm for introducing a unique architecture for production SoS. In this regard, the designing process for Production SoS has been elaborated in three phases. The first phase identifies the production scenarios, the second phase explains analysis and selection of production strategies, and the third phase constructs an innovative Production SoSAF.*

Keywords: Production, System of Systems, Executable, Architecture.

1 Introduction

The dynamics of modern markets require the development of new product and production processes to be carried out side by side. Moreover, in the production systems the interactions and their relationships among many variables are not known [1] due to increasing complexity. In the past, these connections remained unnoticed due to huge buffers of inventories [2]. But now thanks to adopting just-in-time and concurrent strategies, even small perturbations can cause domino effect exacerbated by the enhanced dynamicity and connectivity involved along the processes [2].

It is irony that most of the production companies still try to manage their complex issues through informal information channels without any well established architecture [3]. A more complex organization requires more formal procedures like architecture to comprehend complexities and to have an effective control over production flows and capacities. Without an appropriate architecture, the systems are just piled together so they may conflict with each other. Besides, to resolve complex issues, systematic mechanisms based on formal procedures and protocols must be designed for strong control and coordination. It demands for integration of technology and processes at multiple levels of systems through concrete frameworks and methodologies. This ubiquitous and dynamic connectivity among systems has spawned a new class of systems of systems (SoS) [4]. In such cases point solution may appear to be a quick-fix approach but it often plaques the overall strategy of the system. Even the flawless designing of subsystems or systems alone cannot secure a high global progress of the system of systems.

Architecture formation is an important part of SoS design process. The design of an SoS requires that analysis methods are appropriate to the type of entities that constitute the system of systems [5]. Production SoS [6] design process comprises different phases starting from idea generation, strategy formulation, selection and architecture modeling. In order to map such an architecture, the designing and execution of a Production System architecture based on System of Systems Approach - we name it Production SoSAF (System of Systems Architecture Framework) - is being introduced. Having such architecture is a prerequisite rather than a luxury for well-informed decisions because it provides a comprehensive view of the processes and acts like a regulator. It should avoid, for example, a maintenance system to be developed in isolation from manufacturing, quality and scheduling; rather it would facilitate to evolve in an interactive and collaborative environment with other systems. In other words, in SoSAF, the net centric environment would establish communications across collaborating systems and operate with the network instead of point to point communications.

SoSAF is a generic architectural framework yet to be constructed that can have inspirations from various military, governmental and non-governmental architectures. Since currently in the literature there is not a generic dedicated architecture (SoSAF) available to address System of Systems issues, so the conventional Department of Defense Architecture Framework (DoDAF) has been analyzed and adapted to construct the production system architecture.

Section 2 initiates the designing process of Production System of Systems. It encompasses the phase-wise description of the designing process. Phase-I elaborates how production scenarios are identified while Phase-II utilizes the Analytical hierarchical Process technique for the explanation and selection of strategic objectives and strategies. Phase-III models and explains the innovative Production SoSAF. Finally, section 3 ends with some conclusions highlighting the future research areas linked to this study.

1.1 Designing of production System of Systems

Literally, SoS is not designed but brought together, organized and assimilated to secure SoS level mission capabilities. In this way, the relationship of systems with its subsystems and with other systems must be clearly defined. SoS design thus provides the technical framework for assessing changes needed in systems or options able to address the requirements [7]. Compared to the design of simple systems, the complexity of designing SoS solution is daunting [8]. Diverse spectrum of missions and operations must be taken into account architecting SoSs [9]. The diversity in production SoS is increased by the multi engineering-and-non-engineering disciplines involved that can be manageably reduced through holistic modeling of components, systems and context together where they can be analyzed.

SoS design process consists of three phases [10] as shown in Fig. 1. In phases 1 and 2, in order to reduce uncertainty and complexity, all the relevant information elements have to be considered and the results must be reached with a context understanding. Tools concerning problem structuring and scenario simulation can be adopted. Hence, different strategic options can be obtained defining the number of systems in the SoS, technologies, complexity and control levels in the architecture. These strategic options are transformed into a set of possible strategies in the next phase. In fact, the first two phases provide the raw material for the construction of architecture in the third phase.

The relationship and communication among the phases highlight the underlying process of connectivity. These three phases interoperate in a loop showing a never ending improvement process and to react to emerging events. This reiteration ensures the alignment of shop floor strategy and corporate strategy in due course. Phase 1 and Phase 2 have already been validated [11], while Phase 3 described here is in the process of validation these days.

1.2 Phase I: Scenario Identification

In this phase the critical contextual factors of systems, operations and organization with their uncertainties and complexities are identified. Scenario identification is not about prediction but it does forewarn about

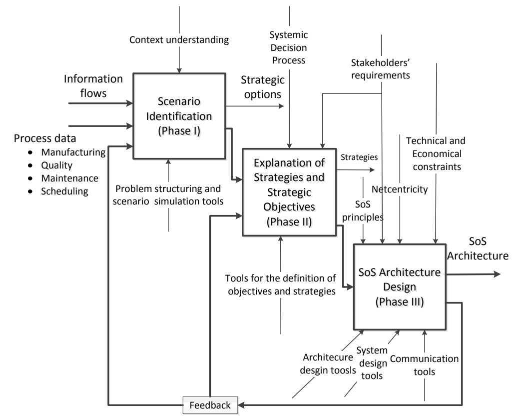


Figure 1: Production System of Systems Design Process

diverse environmental condition the organization could face with. Production scenarios generate a high level view to be helpful to make explicit production strategies for the next phase. Some models are created and used to assist in the definition and resolution of scenarios that arise in the planning, operation and coordination of activities during production operations [12]. Besides them, here a map of uncertainty is drawn to make better strategic decisions based on the interaction and evolution of factors such as economic conditions, technological developments and other organizational dynamics. There have been different methods developed in Problem Structuring Methods (PSMs) along with the soft Operational Research (OR). Strategic Choice Approach [13] is a method, supported also by a software tool STRategic ADvisor (STRAD), to structure the 'problem space' and generate alternatives in uncertainty conditions; it is hence very useful for the definition of more strategic scenarios. It lists the current issues of concern in an "overview window" and classifies each of them as either a decision area, an uncertainty area, or a comparison area (that is, a criterion for choice). To develop a further picture of the "problem space", it maps the interconnections among decision areas (a decision graph) in the "focus window" and combine the strategic options for each decision areas to generate the possible alternatives (i.e. scenarios).Then it examines them, taking one pair at a time, in order to check the compatibility of the strategic alternatives proposed (in term of validity of the combinations generated). Each time a new incompatibility result, the number of alternatives available is likely to be reduced. Further comparing and choosing phases requires the definition of the criterion for choice (Comparison areas), and the alternatives are evaluated using both value judgments and clear scales of measurement. Figure 2, sketches the complete scenario generation process.

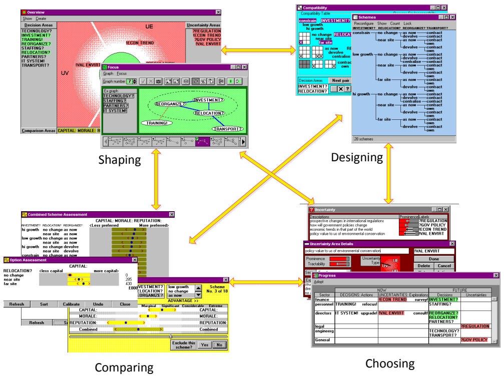


Figure 2: STRAD application for Scenario Identification

1.3 Phase II: Explanation of strategic objectives and strategies

In this phase the production strategies and the objectives applied to achieve these strategies are well defined and explained in the context of scenarios identified in the previous phase. These different strategies describe the production factors related to production capacity, technology advances, vertical integration and other organizational issues. Many methods and models are present in the MCDA literature that focus on synthesis and organization of the information to guide decision makers in identifying a "preferred course of action" [14]. Among those that can used to define possible strategies from feasible scenario alternatives, AHP (Analytical Hierarchy Process) [15], and ANP (Analytical Network Process) [16] are very effective.

The idea is comparing alternatives through the definition of decisional criteria, organized in a hierarchy in relation to a specific goal. Weights among criteria are determined through the matrix of pairwise comparisons with respect to the overall goal. The criteria should be related to the main problem dimensions (economical dimension, organizational-managerial dimension, risk dimension) and the LM strategies should not consider the manufacturing system only, but should involve the revision of the internal logistic system, the quality management system, and also the production management system. More specifically, not all the problems in production SoS can be structured hierarchically due to their interactions. ANP, that is the mathematical extension and generalization of AHP, transfigures hierarchies into networks thus building the model close to a real world connectivity of systems and their components and it includes feedback loops and interdependencies of nodes within and among different clusters (inner and outer dependencies). Thus ANP comes out with more precise results by accommodating complex multidimensional systems of relationships.

1.4 Phase III: Design of SoS architecture and of the component systems

In this phase the ideas and strategies developed in previous phases are materialized. It includes the selection and number of systems, and their compositions to link the component systems. These systems are heterogeneous systems because they perform different functional activities in the production SoS, employ different resources of hardware and software, interact differently abroad the organization (suppliers, customers, etc.). The communication, cooperation and collaboration among these heterogeneous systems are the essential purposes of production architecture. To achieve this, multiple systems are required to be simultaneously connected [17] and there is a need to delineate the overall behavior of the SoS as a whole [18].

For designing of Architecture and its components two major things are needed; a reference Framework and a modeling Language. For Production SoS, there are various architectures available that could potentially act as reference, for example, some military architecture are MODAF (UK), NAF (NATO), DNDAF (Canada), MDAF (Italy), AGATE (France) and ADOAF (Australia), and the list is very long for other nondefense architectures. DoDAF (defense) can be adapted to production (nondefense) architectures because it can distinctively be used to design network of systems where communication is the basic element for the successful completion of mission (the strategic goal to face a scenario in the production case case) and because various multiple perspectives can be extracted as views of the whole production architecture.

DoDAF V1.5 2003 introduced the concept of net centricity to establish more focus on data. In DoDAF V2.0, the views are referred to as viewpoints, and the products as models to emphasize data centricity. DoDAF V2.0 changed the structure of the framework and more viewpoints have been added: Capability Viewpoint (CV); Project Viewpoint (PV); Data and Information Viewpoint (DIV); Services Viewpoint (SvcV); Standards Viewpoint (StdV) (Renamed from Technical Standards View-TV). We extend DoDAF V1.5 [19] but use the terminology of DoDAF 2.0 [20] to create Production System of Systems Architecture Framework.

Different problems in architectural frameworks may require a single or multi-viewpoints with different levels of details. Each representation caters specific objectives and engenders different effect on the overall system operation. Based on this, the production SoSAF can be arranged into five viewpoints with three perspectives for each viewpoint as shown in Fig. 3.

Manufacturing Viewpoint (MV) is the viewpoint around which other viewpoints are grounded. It adds the value by converting raw materials and components into end products. *Quality Viewpoint* (QV): QV is not just the description of procedures, policies and stan-

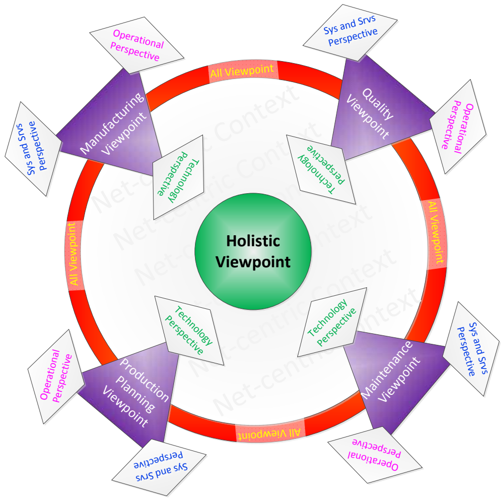


Figure 3: Production System of Systems Architecture Framework (Production SoSAF-Static Aspects)

dards but a total quality culture; it ensures the quality program as an integral part of the strategic objectives of the company, extending their values to the other viewpoints. *Technical Maintenance Viewpoint* (TV): TV describes the functions and readiness of the battle field (shop floor). It keeps the maintenance data and past solutions in historical data servers to act as input data for future applications. It is simpler for a technician/maintainer to decode the problem when corrective maintenance action is connected with errors. The datum flows at different terminals and stores in a common database in the form of history of equipment condition, cause and frequency of disruptions. *Production Planning & Scheduling Viewpoint* (PV): Production planning and scheduling are correlated and two sequential integrated processes, and therefore, should be dealt likewise. To ensure the optimum utilization of resources and shortest production lead time, short, medium and alternative plans are required. It issues operational orders to the machines, supervise and coordinate the execution process, direct routing and equipment selection. It is the complex and dynamic activity due to myriad variables involved and unanticipated events. All these viewpoints are surrounded by *All Viewpoint* (AV) to describe the context, scope, plan, scenarios, constraints, assumptions and information of the entire architecture.

There have to be clear relationships and communications among these viewpoints. On one hand, these viewpoints provide insight of objectives, capabilities, constituents' description and enable to slice the problem in a specialized fashion. On the other hand, they coalesce together to constitute a strategic framework that presents a holistic view of Production SoSAF. *Holistic Viewpoint* (HV) is not a separate viewpoint but integrates these heterogeneous viewpoints to bring complementary capabilities together.

Among four fundamental viewpoints, QV and MV are Core Viewpoints (CV), while the other two TV and PV are Service or Auxiliary Viewpoints (SV). Each viewpoint has three perspectives that are: *Systems and Services Perspective* (SP) supports operational activities by providing system functions and service resources. It provides services in the form of a series of processes to keep production process an ongoing phenomenon; *Operational Perspective* (OP) describes the activities, operational nodes and the information exchanges to complete SoS missions; *Technical Standard Perspective* (TP) provides the set of rules, criteria, technical standards for operational activities. Every viewpoint is described by these three perspectives, for example, MV is described: by SOPs related to the machineries, equipment in practice (technical standard perspective): by managerial practices (operational perspective), and: by flow charts of machines and material, automation and other means to secure smooth operation (systems and services perspective).

Viewpoints can be defined also with respect to their models, for example, QV-1 (due to raw material), QV-2 (due to process), QV-3 (due to manhandling), etc. Similarly, QV-2a (error prevention), QV-2b (inspection upgrading), and so forth can also be defined. These models coordinate and regulate the material handling and processing events to add value from raw material to finished goods. They are documents or graphics to act as templates. These templates are stuffed with the appropriate data and structured together to describe the viewpoints for analysis and decision making process. Through these models, systems become manageably less complex with the flexibility of increasing the number of models upon deeper understanding. These viewpoints in production SoSAF are autonomous but not independent having central command and control over individual viewpoints and distribute complex decision making process over all viewpoints. This setup enjoys full local autonomy, rapid information flow, and simplicity of redesigning and adaptability and it requires strong netcentric environment. This system works in a heterarchical structures to manage shop floor operations. The heterarchical set up has been described well in [21], where systems act like negotiating partners rather than under a strict supervisory control.

Two viewpoints can find themselves beneficial in different positions. For example, PV can be favored by intermediate planning of production quantities responding to market, while QV may be interested in long term planning to produce goods with consistent qualities. In such and other cases, production SoSAF exercises its prerogative to assign several decision making functions including shop floor activities such as scheduling, channeling the WIP and resource distribution through particular viewpoint(s) and model(s). Additionally, it also defines interdependencies and correlations among the viewpoints, thus developing a structure to synchronize the implementation of numerous choices and plans.

For designing architectures a language is needed. Many possibilities exist [22], SysML can be used here because it is a general purpose graphical modeling language for specifying, analyzing, designing and verifying complex systems used for systems engineering. SysML is a semi-formal user friendly language whose semantics can be extended with other applications. SysML has total nine diagrams including seven diagrams (UML4SysML) of UML 2.0. SysML thus has the capability to model views of a wide range of disciplines across Production SoSAF such as design, production, operation, distribution, etc.

2 Conclusion and Further Research

In industrial perspective, the integration problem is usually bottom-up. The interacting role of both the active resources (that perform directly) and passive resources (such as material, machine, information, data bases, etc.) is important. Information interacts and flows among these resources of the organization. The system to system communication patterns are not deterministic in such complex systems because the systems interact with each other to give rise to a peculiar behavior. In order to make this information and communication valuable, an architecture (platform) has been proposed in this research work.

In this regards, first of all the SoS designing process for production System has been described. Then an architectural framework for production system - Production SoSAF - has also been developed. With such a platform, the loosely coupled systems would establish an environment of interdependency while maintaining their autonomous status in terms of management and operations. That will surely help improve communications among stakeholders, better integration of hardware and software, and efficient root-cause analysis of the problems, and hence early risk identification. Further research associated with this work is to define more models (products) for Production SoSAF.

References

[1] David W. Coit And Alice E. Smith, "Using Designed Experiments To Produce Robust Neural Network Models Of Manufacturing Processes", Proceedings of the Fourth Industrial Engineering Research Conference (IERC), pp. 229-238, May 1995.

[2] Brynjolfsson, E., A. Renshaw and M. V. Alstyne., "The Matrix of Change", Sloan Management Review 38, pp. 22-40, Winter, 1997.

[3] Joe Barkai Benjamin Friedman, "Collaboration in Product Development: The Sum of Its Parts Drives Success", Manufacturing Insights # MI220385, October 2009.

[4] Robert J. Cloutier, Michael J. Dimario and Hans W. Polzer, "Net Centricity and System of Systems, System of Systems Engineering- Innovations for the 21st Century", (M. Jamshidi, Ed.), John Wiley Series on Systems Engineering, New York, 2008.

[5] D. A. DeLaurentis and William A. Crossley, "A Taxonomy-based Perspective for Systems of Systems Design Methods", IEEE International Conference on Systems, Man and Cybernetics, pp. 86-91, 2005.

[6] F. Montagna, "The SoS approach for Lean Manufacturing Systems", 9 convegno AITEM, Torino, Italy, September, 2009.

[7] Dahmann, J. S., "Systems Engineering for Department of Defense Systems of Systems, System of Systems Engineering", (ed M. Jamshidi), John Wiley & Sons, Inc., Hoboken, NJ, USA. doi: 10.1002/9780470403501.ch9, 2008.

[8] Reggie Cole, "SoS Architecture, Systems of Systems Engineering Principles and Applications", (ed by M. Jamshidi), CRC Press 2009 Print ISBN: 978-1-4200-6588-6 eBook ISBN: 978-1-4200-6589-3 DOI: 10.1201/9781420065893.ch2, 2009.

[9] Dagli, C. H. and Kilicay-Ergin, N., "System of Systems Architecting, System of Systems Engineering", (ed M. Jamshidi), John Wiley & Sons, Inc., Hoboken, NJ, USA. doi: 10.1002/9780470403501.ch4, 2008.

[10] A. Alfieri, M. Cantamessa, F. Montagna, "The SoS approach for lean manufacturing systems", forthcoming on IJTM, Special Issue on Lean Principles and Information Management, Vol. 57, No. 1-2, 2012.

[11] M. Cantamessa, F. Montagna, "Beyond Lean Manufacturing: developing an integrated methodology to design effective manufacturing systems", ASME Engineering Systems Design and Analysis Conference, Haifa, Israel, July 7-9, 2008.

[12] Charu Chandra and Sameer Kumar, "Enhancing manufacturing operations effectiveness through knowledge based design", Integrated Manufacturing Systems 14/3, pp#278-292, MCB UP Limited, 2003.

[13] Friend, J., "The strategic choice approach, Rational analysis for a problematic world: problem structuring methods for complexity, uncertainty and conflict", J. Rosenhead, ed., Chichester: Wiley, pp. 121-157, 1989.

[14] Belton, V., Steward, T.J., "Multiple Criteria Decision Analysis, an Integration Approach", Kluwer Academic Publishers, Norwell, MA., 2002.

[15] Saaty, T.L., "The Analytic Hierarchy Process", McGraw-Hill, New York, NY, USA, 1980.

[16] Saaty, T.L., "Decision Making with Dependence and Feedback: The Analytic Network Process", (Second ed.), Pittsburgh, USA: RWS Publications, 4922 Ellsworth Avenue, Pittsburgh, PA 15213 USA, 2001.

[17] Bowen, R.M., and Sahin, F., "A net-centric XML based system of systems architecture for human tracking", System of Systems Engineering (SoSE), 2010 5th International Conference on , vol., no., pp.1-6, 22-24 June 2010.

[18] Col Kenneth Flowers and Cyrus Azani, "Open Systems Policies and Enforcement Challenges", Proceedings of the National Defense Industrial Association Systems Engineering Conference, Dallas, Texas, October 25-28, 2004.

[19] DoD Architecture Framework Version 1.5, Volume I, 23 April 2007.

[20] The Department of Defense Architecture Framework (DoDAF), version 2.0, Volume 01: Introduction, Overview & Concepts, Manager's guide, May 2009.

[21] Vieira, G. E. and Veiga, C. P., "Hierarchical and centralized architectures for distributed production planning, scheduling and control activities", Proceedings of IEEE International Conference on Service Operations, Logistics and Informatics, Chicago, USA, pp. 341-346, July 22 - 24, 2009.

[22] Friedenthal, Sanford; Moore, Alan; Steiner, Rick; "Practical Guide to SysML - The Systems Modeling Language", (2nd Edition), Elsevier, 2012.

System-of-system Approaches and Challenges for Multi-Site Manufacturing

Simon Ford
University of Cambridge
sjf39@cam.ac.uk

Ursula Rauschecker
Fraunhofer IPA
Ursula.Rauschecker@ipa.fraunhofer.de

Nikoletta Athanassopoulou
IfM-ECS
naa14@cam.ac.uk

Abstract - *In the multi-site manufacturing domain, systems-of-systems (SoS) are rarely called so. However, there exist a number of collaborative manufacturing paradigms which closely relate to system-of-system principles. These include distributed manufacturing, dispersed network manufacturing, virtual enterprises and cloud manufacturing/manufacturing-as-a-service. This paper provides an overview of these terms and paradigms, exploring their characteristics, overlaps and differences. These manufacturing paradigms are then considered in relation to five key system-of-systems characteristics: autonomy, belonging, connectivity, diversity and emergence. Data collected from two surveys of academic and industry experts is presented and discussed, with key challenges and barriers to multi-site manufacturing SoS identified.*

Keywords: Integrated multi-site manufacturing, system-of-systems, SoS, collaborative manufacturing, distributed manufacturing, dispersed network manufacturing, virtual enterprise, cloud manufacturing, manufacturing-as-a-service.

1 Introduction

Competitive pressures mean that manufacturers need to deliver more product variants within shorter lead times, often in low and fluctuating volumes, and at lower prices. This has given rise to a number of manufacturing paradigms, including lean production, agile manufacturing, bionic manufacturing and holonic manufacturing, which attempt to improve the manufacturer's capacity to meet these shifting requirements. At the same time, there is an increasing trend towards more geographically distributed manufacturing at both the intra- and inter-firm level, brought about by costs, tax policies and specialisation on core businesses. Few vertically integrated manufacturing enterprises exist today with the vast majority of manufacturers embedded within globally distributed supply networks.

The geographic distribution of manufacturing can be seen at both the intra- and inter-firm level. Intra-firm manufacturing networks typically involve large global manufacturers in industries such as automotive, aerospace and electronics, which have multiple manufacturing sites distributed across the world. At the inter-firm level, geographically distributed manufacturing occurs as individual manufacturers self-organise and collaborate in the delivery of specific products to customers.

In parallel to manufacturing paradigms in general, a number of similar collaborative manufacturing paradigms have emerged. These include distributed manufacturing, dispersed network manufacturing, virtual enterprises and cloud manufacturing/ manufacturing-as-a-service. A common feature of these new paradigms is the temporary nature of the collaborative manufacturing effort. Manufacturers self-organise into collaborative networks in response to customers' needs, dissolving once these needs have been satisfied. This principle behaviour is highly equivalent to SoS characteristics. For this reason, the future of system-of-systems in integrated multi-site manufacturing is currently being explored in Road2SoS, an EU FP7 project. The objective of this project is the development of SoS roadmaps in 4 domains: (1) integrated multi-site manufacturing, (2) multi-modal traffic control, (3) smart grid and distributed energy generation, and (4) emergency and crisis management. This paper provides a basis for understanding the underlying characteristics of SoS in the multi-site manufacturing domain. It goes on to present and analyse data from two surveys with domain experts, identifying current challenges and potential future trends.

2 Multi-site manufacturing paradigms

In the production domain, systems-of-systems are rarely called so. However, there exist various terms which closely relate to system-of-system principles. This section gives an overview on the most important terms and paradigms, their characteristics, overlaps and differences.

2.1 Distributed manufacturing

Distributed manufacturing is "a new pattern of interfirm relationships evolving network-wide integration by creating different forms of interentity processes" [1]. It is a temporary alliance of manufacturers, suppliers and customers, which combine their knowledge and competencies in order to meet clearly defined market opportunities. The organisations within it are nearly autonomous as they execute their tasks. While individual operations can either be coordinated through a central organisation or decentralised, there is a need for some form of coordination and synchronisation through information exchange [2]. Drivers of distributed manufacturing include the availability of reconfigurable manufacturing systems, technologies for decision-making support, enhanced human-machine interfaces, and collaboration software [1].

2.2 Dispersed network manufacturing

The concept of dispersed network manufacturing originates in industrial networks theory and is defined as "an organisational manifestation for collaboration between and coordination across loosely connected agents" [3]. These agents are typically small and medium sized enterprises (SMEs), which are geographically distributed, often on a regional scale rather than a global one. Their objective is to pool their capabilities in order to provide the necessary production capacity that a larger customer requires [4]. To meet this objective, the membership of the network may change over time. The degree of autonomy possessed by each member within the network varies and the interactions between members lead to the emergence of collective behaviours that go beyond the control of any single firm [5].

Related to dispersed network manufacturing is the concept of a 'dispersed manufacturing network', the "particular combination of SMEs that has formed in accordance with the principles of DNM [dispersed network manufacturing]" [4]. In this topology, SMEs form dynamic networks in order to satisfy customer demand, dissolving following the satisfaction of those demands. While there is no fixed membership or ownership of the network in both dispersed network manufacturing and dispersed manufacturing networks, a key difference lies in the strength of relationships between network members. In dispersed manufacturing networks, these relationships are much more significant and maintaining the relationship is a key enabler for this concept [4].

2.3 Virtual enterprises and organisations

A virtual enterprise is "a temporary alliance of enterprises that come together to share skills or core competencies and resources in order to better respond to business opportunities" [6]. It is created when organisations from a production network collaborate to form a supply chain for a single order and thereafter appear to the end customer as a single organisation. Following the fulfilment of this order, the virtual enterprise is dissolved. For the next customer order, collaboration can be continued through the original virtual organization or a different virtual organization may be established in the production network.

A related concept is the virtual organisation. The distinction between a virtual enterprise and a virtual organisation is that the former involves only profit seeking enterprises within its member, while the latter includes both for profit and not for profit organisations (e.g. universities, government agencies, charities). Three types of virtual organisations can be identified. The first type is a short-term virtual enterprise that delivers highly specialised, customised products through non-hierarchical control. The second type is the consortium virtual enterprise, in which semi-standardised products are made, with the partnership of a medium to long-term horizon. The control structure is non-hierarchical and co-operative. The final type of virtual organisation is an extended enterprise, in which a dominant enterprise uses hierarchical control to coordinate some or all of its suppliers. The extended enterprise typically delivers standardised products, with the collaboration of a long-term nature [7].

2.4 Cloud Manufacturing / Manufacturing-as-a-Service

Cloud manufacturing is an attempt to transfer key characteristics such as scalability (flexible increase or decrease of capacity and/or capability), location transparency (the user has not to care about resources), or costs depending on really occurred usage from cloud computing to the manufacturing domain. As a result, according to cloud-computing principles like Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) or Software-as-a-Service (SaaS) provided by computing clouds, Manufacturing-as-a-Service (MaaS) is thought to provide the ability to manage production networks by means of service orchestration and execution. Cloud manufacturing is very closely related to the term grid manufacturing. It has more or less occurred simultaneously with the transition from grid to cloud computing in the IT domain.

3 Why is multi-site manufacturing a type of system-of-systems?

In considering what constitutes a system-of-systems, Sauser and Boardman define five characteristics: autonomy, belonging, connectivity, diversity and emergence [8]. Examining the new manufacturing paradigms indicates that forms of each can be considered as systems-of-systems.

3.1 Autonomy

An autonomous system is "situated within and a part of an environment that senses its local environment and acts upon that environment in pursuit of its own agenda" [9]. Each of the firms in the collaborative manufacturing networks pursues its own agenda. When these networks are solely SMEs, they are both managerially and operationally autonomous. However, when the networks involve subsidiaries of larger firms, the entities are only partially autonomous.

3.2 Belonging

Constituent systems within an SoS choose to be part of the larger system because of their needs, beliefs or fulfilment [10]. The collaborative nature of the inter-firm manufacturing networks is a defining characteristic. Firms enter the network of their own accord and set the terms of their involvement upon entering.

3.3 Connectivity

System-of-systems feature interoperability and a communication capability between the constituents of the SoS so that social functionality is enabled [11]. This interoperability is essential for operations in integrated multi-site manufacturing. Firms use ICT to transmit and share information across the network, distributing the total production load.

3.4 Diversity

Another attribute of system-of-systems is that they feature visible heterogeneity. That is, they include “distinct or unlike elements or qualities in a group” [8]. Whether inter-firm or intra-firm, each of the collaborative manufacturing networks is an amalgam of such heterogeneous entities. Each firm has distinctive capabilities and competencies and participates in the network so that it can obtain access to those complementary capabilities and competencies that it does not possess.

3.5 Emergence

The final core concept in Sauser and Boardman’s model is that system-of-systems exhibit emergent attributes, including unexpected structures and behaviours [8]. The multi-site manufacturing network paradigms described in this paper are each expected to be transitory in nature as they dissolve following the delivery of the customer’s requirements. However, there is the potential for collaborations to endure and to take on new forms beyond the completion of the initial network’s objectives.

4 Challenges and future trends

It is apparent that manufacturing across multiple sites is no longer determined solely by the supply network but has begun to operate in collaborative networks that possess system-of-systems attributes. In an attempt to understand how integrated multi-site manufacturing is evolving and developing system-of-system characteristics, two surveys have been conducted at a European level as part of the Road2SoS project. These surveys sought the views of experts in the field, bringing together perspectives from industry and academia.

The aim of the first survey was to capture perspectives on the future direction of technological developments in multi-site manufacturing. The survey was conducted with 17 experts in the multi-site manufacturing domain. 10 surveys were completed through interview, with the other 7 conducted through an online questionnaire.

The objective of the second survey was to capture insights into the market factors that are enabling or inhibiting the adoption of multi-site manufacturing. The survey was conducted as an online questionnaire and was completed by 19 industrial experts in multi-site manufacturing.

4.1 SoS concepts in multi-site manufacturing

Respondents to the first survey provided a wide range of definitions as to what constitutes a system-of-systems in multi-site manufacturing. Example definitions included:

- A complex system which is initiated by integrating autonomous systems by means of interfaces and infrastructure in order to reach a common goal.
- Multi-site production running autonomous production sites. Each site is focusing on specific markets strategically and to gain competitive advantage these sites are controlled centrally.
- The application of ICT to the distributed business processes linking diverse manufacturing enterprises in supply chains, value networks and business ecosystems.
- A system-of-systems consists out of heterogeneous components which can work in their specific domain independently but can be orchestrated working together in a way to have a larger system.

This data indicates that there remains no clear definitional consensus of SoS in multi-site manufacturing among academics and practitioners who work in the area. The novelty of the SoS field is also highlighted by respondents (Figure 1), as the maturity of SoS architectures and tools lags well behind that of SoS concepts.

While over half the respondents believe that SoS concepts are mature or well advanced, around 70% of respondents believed that SoS architectures and SoS tools are only elementarily developed or underdeveloped. Overall, two-thirds of respondents considered the multi-site manufacturing SoS to be elementarily developed.

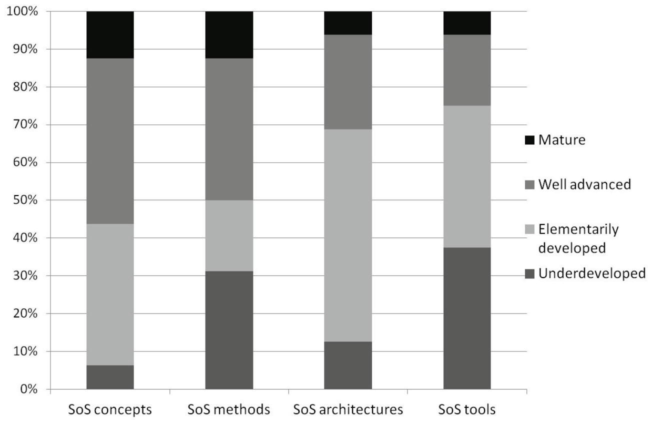


Figure 1. The maturity of SoS concepts, methods, architectures and tools in multi-site manufacturing *(Survey 1, n=16)*

There are also a variety of terms used to describe complex systems, systems-of-systems and collaborative networks. Figure 2 shows the responses to the concepts and principles of SoS relevant to multi-site manufacturing. The most relevant concepts and principles are shown on the left of the figure, with the six most relevant judged to be: (1) flexibility, (2) adaptability, (3) autonomy of the subsystems, (4) network, (5) modularity, and (6) evolutionary.

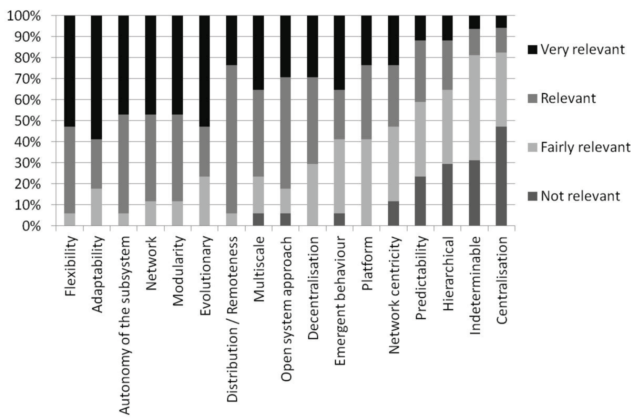


Figure 2. Main principles and concepts in multi-site manufacturing SoS *(Survey 1, n=17)*

4.2 Drivers of multi-site manufacturing SoS

The drivers for the adoption of SoS approaches in multi-site manufacturing are numerous. Table 1 summarises the survey responses collected on this topic, which were scored on a five-point Likert scale. Each of the drivers described are contribute to improving the firm's competitiveness. The most highly rated are concerned with how SoS approaches can help improve the firm's ability to respond to customer demands while improving manufacturing efficiency and reducing costs.

Table 1. Domain drivers in multi-site manufacturing SoS *(Survey 2, n=18)*

Domain driver	Mean
Reduction of inventories and lead times	4.44
Increased responsiveness to customer demands	4.39
More economical use of resources, cost reduction	4.33
Maximized manufacturing efficiency	4.28
Increased flexibility	4.28
Increased ability for on-demand production	4.06
Reduced equipment integration and production ramp-up times	3.94
Enabling product and/or process innovations	3.89
Consistent/integrated product tracking and tracing	3.83
Reduced need for maintenance	3.82
Better control of process parameters	3.78
Integrated, continuous scheduling	3.72
Product customization	3.61
Reduction of environmental impact	3.61
Increased scalability	3.59
Condition monitoring	3.35
More diversified production	3.22
Decreased dependency on strong players in the market	3.22
Increased degree of automation	3.17
Greater number of product variants	3.17

4.3 Challenges of multi-site manufacturing SoS

Survey respondents provided insights into the key challenges for the implementation of system-of-systems in integrated multi-site manufacturing in the next five years. These challenges include the following:

- Reference implementations and SoS showcases to increase acceptance
- Development of (quasi) standards for integration of systems and description of systems functionality
- Understand how to link production with the broader value chain
- Data capture and interpretation
- Being able to access the information (avoid human information overload), and improve visualisation
- Predictive software
- Understanding how to create resilience and adaptability to change
- Resolve legal issues (also supported by implementation of the systems)
- Skills shortages

Focusing on the ICT and technological challenges, survey respondents highlighted the importance of interoperability. As Figure 3 illustrates, the seamless integration of systems and components, between both new and legacy systems, in multi-organisation networks are key technical challenges that need to be overcome for SoS practices to be more widely adopted.

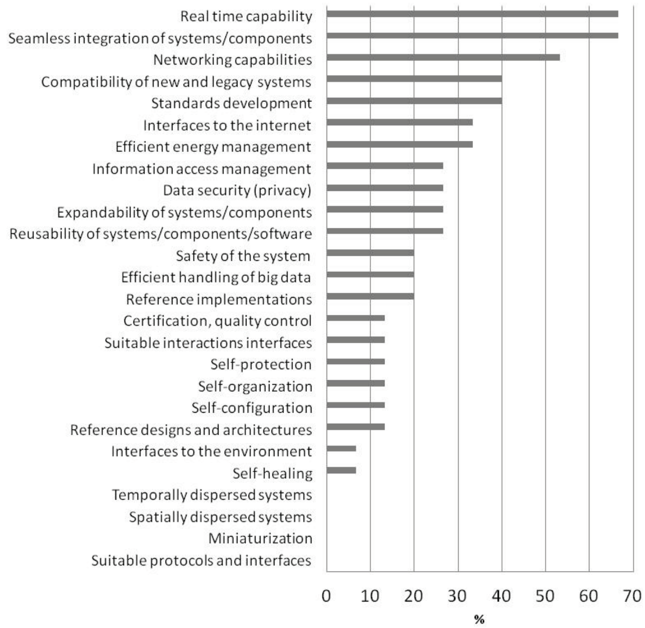


Figure 3. Technology and ICT challenges in multi-site manufacturing SoS *(Survey 2, n=17)*

To overcome these challenges, survey respondents identified a number of technologies that would impact the future development of system-of-systems in multi-site manufacturing. The most identified technologies were the following:

- Cloud computing
- Communications standards for information exchange (e.g. SAP)
- Manufacturing service descriptions
- Plug-and-produce seamless reconfigurable production environments
- Product tracking throughout the supply chain
- Integration of design tools in order to specify/customise product parameters throughout the whole supply chain
- Order management software in non-hierarchical networks

4.4 Barriers to multi-site manufacturing SoS

Respondents also identified a number of key socio-economic barriers to the implementation of systems-of-systems in multi-site manufacturing. These barriers include the following:

- Distribution of knowledge about SoS in industry
- Short term focus, e.g. efficiency-only focus
- Economic viability and liability
- Political interests, power, and incentive structures
- Economic interdependencies
- Protection of intellectual property
- World trade barriers
- Uncertainty in global macro-economic trends
- Migration of production to low cost countries
- Lack of finance to enable the implementation of IT infrastructure in emerging economies

Focusing on the social, economic, political and legal barriers to the implementation of SoS in multi-site manufacturing highlighted that these barriers are as significant as the technical challenges. Table 2 summarises the survey responses collected on this topic, which were scored on a five-point Likert scale. The most highly rated barriers concern how organisations work together and finding ways to capture economic value from SoS approaches.

Table 2. Barriers to implementing SoS in multi-site manufacturing *(Survey 2, n=15)*

Issue	Mean
Problems related to multiple ownership	4.27
Lack of availability of skilled personnel	4.00
Intellectual property issues	4.00
Lack of appropriate business models	3.93
Concerns about security and privacy	3.87
High initial investment	3.87
Lingual and cultural differences in global networks	3.80
Antitrust policies hindering cooperation of companies establishing System of Systems implementations	3.80
Risk-benefit ratio unclear	3.67
Time to market too long or unclear	3.64
Absence of demonstration / technology and approach insufficiently tested	3.50
Uncertain demand	3.47
Lack of organizational acceptance	3.40
Concerns about system stability and failures	3.40
Software licensing	3.38
Concerns about false information conveyed by the system	3.27
Individual action is highly risk fraught	3.23
Regulatory issues	3.07
Certification	3.00
General preference for centralized, hierarchical systems	2.93
Lack of public acceptance	2.80
Lack of public funding	2.73

5 Conclusions

This paper has focused on the emergence of SoS approaches in the domain of integrated multi-site manufacturing. It is apparent that SoS are becoming more widespread given the technological advances that have been made in ICT and the potential for greater responsiveness to customer demand. However, significant technical challenges and socio-economic barriers remain.

The Road2SoS project aims to build of these survey insights. Future work will bring together expert practitioners at a roadmapping workshop. The survey data will be synthesised with desk research to create a pre-populated roadmapping template. This preliminary roadmap will be used at the workshop, with participants working together to develop a clearer vision of the future challenges, enablers and barriers that remain in multi-site manufacturing SoS. Possible future applications scenarios for the domain in the next 15 years will be explored and specific actions for the materialisation will be developed. It is expected that this work will shape the EU's Horizon 2020 research and funding landscape.

This work is co-financed by the European Commission under the 7th Framework Programme. Grant Agreement No. 288274.

References

[1] H. Kühnle, "Distributed Manufacturing: Paradigm, Concepts, Solutions and Examples," in *Distributed Manufacturing: Paradigm, Concepts, Solutions and Examples*, H. Kühnle, Ed. London: Springer-Verlag, 2010, pp. 1-9.

[2] T. Sauter and A. Treytl, "Communication Systems as an Integral Part of Distributed Automation Systems," in *Distributed Manufacturing: Paradigm, Concepts, Solutions and Examples*, H. Kühnle, Ed. London: Springer-Verlag, 2010, pp. 93-111.

[3] R. Dekkers, "Introduction," in *Dispersed Manufacturing Networks: Challenges for Research and Practice*, R. Dekkers, Ed. London: Springer-Verlag, 2009, pp. 3-12.

[4] H. Noori and W. B. Lee, "Dispersed Network Manufacturing: An Emerging Form of Collaboration Networks," in *Dispersed Manufacturing Networks: Challenges for Research and Practice*, R. Dekkers, Ed. London: Springer-Verlag, 2009, pp. 39-58.

[5] H. Kühnle, "Self-Similarity and Criticality in Dispersed Manufacturing: A Contribution to Production Networks Control," in *Dispersed Manufacturing Networks: Challenges for Research and Practice*, R. Dekkers, Ed. London: Springer-Verlag, 2009, pp. 59-76.

[6] L. M. Camarinha-Matos and H. Afsarmanesh, "Collaborative networks: a new scientific discipline," *Journal of Intelligent Manufacturing*, vol. 16, no. 4-5, pp. 439-452, 2005.

[7] M. T. Martinez, P. Fouletier, K. H. Park, and J. Favrel, "Virtual enterprise – organisation, evolution and control," *International Journal of Production Economics*, vol. 74, no. 1-3, pp. 225-238, 2001.

[8] B. Sauser and J. Boardman, "Taking Hold of System of Systems Management," *Engineering Management Journal*, vol. 20, no. 4, pp. 3-9, 2008.

[9] R. K. Garrett, S. Anderson, N. T. Baron and J. D. Moreland, "Managing the Interstitials , a System of Systems Framework Suited for the Ballistic Missile Defense System," *Systems Engineering*, vol. 14, no. 1, pp. 87-109, 2010.

[10] A. Gorod, B. Sauser, and J. Boardman, "System-of-Systems Engineering Management : A Review of Modern History and a Path Forward," *IEEE Systems Journal*, vol. 2, no. 4, pp. 484-499, 2008.

[11] M. J. Dimario, J. T. Boardman, and B. J. Sauser, "System of Systems Collaborative Formation," *IEEE Systems Journal*, vol. 3, no. 3, pp. 360-368, 2009.

System of Systems Thinking in Product Development: A System Dynamic Approach

Alemu Moges Belay*
Department of Production
University of Vaasa
Vaasa, Finland
albel@uwasa.fi

Petri Helo
Department of Production
University of Vaasa
Vaasa, Finland
phelo@uwasa.fi

Torgeir Welo
Dept. of Engineering Design and Mat.
Norwegian University of Science & Technology, Norway
torgeir.welo@ntnu.no

Abstract - *Companies ceaselessly strive to have a systemic approach to survive. One of those approaches is system of system engineering (SoSE) that enables the decision-makers to understand the interaction of different systems in terms of costs, lead time, and quality. This paper considers product development (PD) in light of SoSE. Different PD stages treated as a system and analyzed to see the effects of concurrent engineering on total cost and lead time. We also analyze the marketing strategy with its components to have sound decision based on SoSE. We use a system dynamics approach and the study found that, the more cost allocated in early stage, the more reduced the total cost (by half) and the lead time reduced by a quarter. From marketing system analysis with interdependent variables, firms can control to a certain level. Some variables could be unpredictable and difficult to control as the system grows more complex.*

Keywords: System of systems engineering (SoSE), System, System dynamics, Concurrent engineering, System engineering, Product development, Design, Marketing.

1 Introduction

Companies want to sell their products as fast as possible to have a bigger market share. But, time-to-market is not an easy task with several complicated systems, processes and activities especially in high technology industries that require a huge amount of investment (for example power plant and aerospace industries). Decisions that are made at early design stage during product development have some quantifiable impacts of direct and indirect product development related activities and costs. This is also extended to the project time and cost overrun problem that mainly due to design changes (S. Chritamara et.al, 2002)

According to Department of Defense (DoD, 2008) and the National Defense Industry Association (NDIA), modeling and simulation is considered as a technical tool-set and is applied throughout the system development life cycle (from concept to delivery). In this study, we use one of the modeling and simulation methods i.e system dynamics, by considering the complex product development in SoSE level. J. W. Forrester in 1961 developed idea of System Dynamics (SD) and (Sterman 2000), developed a theory to approach complex systems, non-linearity, and with several feedback loops of information in a system. One of the tools that help to understand the tradeoffs in the product development stage is dynamic simulation models. Several researchers applied system dynamics in different areas of study. Marujo, LG (2009) applied in the rework impact evaluation in the overlapped product development schedule and its aim was to reduce the lead-time of activities. He provided the general model to estimate the extended design time, strictly related to the necessary rework fraction, considering over-lapped activities using system dynamics. Sterman also indicated the dynamic behavior of product development with an involvement of multiple feedbacks, complex framework and has several interdependent activities. The whole process of product development should be seen as a system including analyzing of time to market. However, the previous studies treat product development by focusing on certain part of the development. This paper wants to see the overall effect of the CE concept in terms of cost and lead time. In addition to that, efforts have been made to analyze marketing section as a separate system in product development processes but as a part of the whole. This is to show the challenges of controlling even a single system let alone several systems. This encourages and shows the significance of system of systems engineering to have an overall insight on the outputs from the interaction of several systems.

The paper is structured as follows: It begins with a brief introduction and provides some definitions of the key terms in SoSE. The second section explains a system of systems in concurrent product development with the effects of design changes in cost. The third and fourth section covers the concurrent product development and parallel computing in CE respectively. The fifth and sixth section deals with the system dynamic models in the product development stages and in the marketing strategy as a part of the system. Finally the findings and conclusion are presented.

1.1 Terms in system of systems engineering

System: system is an integrated set of elements that accomplish a defined objective (INCOSE, 2000).

System Engineering: Systems engineering is an interdisciplinary approach or a structured, disciplined, and documented technical effort to simultaneously design and develop systems products and processes to satisfy the needs of the customer (DoD, Version 1. 2006).

System of systems: According to (Maier, 1999) SoS is "An assemblage of components which individually may be regarded as systems and which possesses two additional properties: Operational Independence of the Components: If the system-of-systems is disassembled into its component systems the component systems must be able to usefully operate independently. That is, the components fulfill customer-operator purposes of their own. Managerial Independence of the Components: The component systems not only 'can' operate independently, they 'do' operate independently. The component systems are separately acquired and integrated but maintain a continuing operational existence independent of the system-of-systems." And another definition by (DoD, Version 1. 2006) is a set or arrangement of systems that results when independent and useful systems are integrated into a larger system that delivers unique capabilities.

2 SoS in concurrent PD

Systems engineering (SE) is an approach to convert the required capabilities of operations into an integrated system design by pondering simultaneous or concurrent executions of activities throughout the product life cycle. However, the challenge becomes more severe as systems to go larger and more complex. This is typical in the process of lean product development in which firms strive to achieve optimal solutions in different dimensions, such as, cost, quality and lead time. Each of the product development stages (concept architecture, design, test, and production) has enormous activities and processes that need to be integrated to obtain an integrated system design. In such situation the idea of a system of systems is a prominent approach to integrate and coordinate these distinct stages and related processes. While applying SoS, it is significant to know and understand well the systems that take a part, their interdependencies and relationships, and the influence of some external factors. Regardless of some debates on similarities and differences between system, system engineering and system of systems, we follow accepted view of system encompasses different parts, interactions, relationships and a whole that is greater than the sum of the individual parts. In this paper, system thinking is applied at all stages of product development because it enables to model complex processes and helps to look for better alternatives and see the effects on the quantifiable factors like lead time, cost and quality.

The paradigm shift of approaching SE through concurrent engineering plays a significant role in lean product development processes. This is because concurrent engineering opens an improvement window through communication, reduce uncertainty and solve a problem throughout the networks of all stakeholders and functions.

According to Siemens PLM Software, 2011 and department of defense (DoD, 2008), give an insight and showed how SoSE teams changed and enhance the traditional SE into system of system level. At the same time, it is noted that some methodologies are found to be similar to that of basic lean philosophy. For instance, application of concurrent engineering is one of the common engineering methods that are mentioned in both systems of system engineering and lean product development in which we would like to link with system dynamics in this study. Understanding and using lean principle rewards remarkable benefits and helps to manage and control the large and complex systems. Lane and Valerdi (2010) showed the manifestation of lean principles in the SoSE and give their insight that SoSE teams are using lean concepts whether or not they are aware of it.

2.1 Effects of Design change on cost

During the product development process, change may happen in any stages and the ultimate cost required may vary accordingly. As the change goes later, the total cost will go higher and that affects the profit and time to market (see Table 1.) Table 1 Typical cost of design changes in major electronics (Port et.al, 1990),

When design changes are made	Costs
During design	$1000
During design testing	$10,000
During process planning	$100,000
During test production	$1,000,000
During final production	$10,000,000

Andersen (2008), presented that by the time a product is designed 80% of the cost, by the time a product goes into production 95% of its cost is determined, so it will be unmanageable to remove cost at that late a date. Similarly, Siemens product lifecycle management (Siemens PLM Software, 2011), recognizes that decisions made at an early stage in the product lifecycle account for 90 percent of a product's costs. Siemens has implemented SE methodologies for example using CE based software

(Team-center) to push all key decision making to the front of the PD process. However, in our model of system dynamics we consider 80 % of the cost commited before the design stage.

3 Concurrent product development

In product development, existences of industrial wastes are inevitable and it would be wise to minimize as much as possible in order to obtain sound profit. Some of the most common wastes are over processing, transportation, motion, inventory, waiting time, defect, overproduction and etc. Most of them are time and process dependent and that are expressed in terms of additional costs, time, rework, defect rate and etc. Enormous approaches (like lean, JIT, TQM, concurrent engineering, six sigma and etc.) have been proposed for the last few decades in order to tackle such challenge. For example when we consider in the light of concurrent engineering, there are activities that could be executed in parallel while developing a new product. The first major step is identifying the processes or activities that can be done sequentially and in parallel then the Amdahl principle could help in optimizing the time and determine speed up factor accordingly.

Individual systems, activities or steps can be designed as the overall processes have been designed. For instance, we can take one of the personnel involved in engineering activities in the product development process and optimize the overall system process design in order to achieve the economic objectives. It's clear that one has to be critical and know how much it will reduce on the product cost. This paper focuses on the influence of concurrency from the perspective of its impact on product life cycle and its costs at different stages of development. But, it is important to realize that improving cycle time or doing things in concurrent way is not a universal solution since there are situations that fail to achieve a significant result.

4 Parallel computing in CE

Recently all types of businesses including NPD have become under extreme pressure to provide products to the client/customer quicker than ever before. With such an extreme pressure being applied to respond, companies that provide products/services can no longer afford or survive to perform work in a sequential manner. According to Martin et al. (1998), one way of breaking down these barriers and improving communication and teamwork among functional groups is by using Concurrent Engineering (CE) methodologies. Doing tasks in parallel rather than sequentially is the basic principle of system engineering and CE. The competitive edge gained by utilizing this methodology is the ability to deliver products/services more expeditiously and at a lower cost while still meeting the customer's expectations. To demonstrate the basic principle of doing activities in parallel, Amdahl principle is used to see the effect of the same. There are three types of executing different activities, sequential, overlap and parallel (See figure 1).

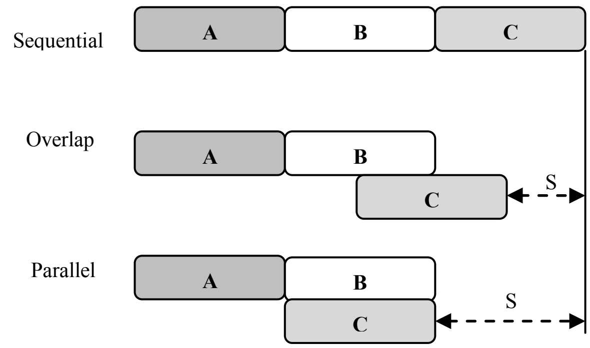


Figure 1 Different execution of activities in PD process

Amdahl's law for parallel computing in determining the speed up factor is important to see the percentage of improvement in product development process. Barry Wilkinson and Michael Allen (1999) represent this principle in the following diagram (Figure 2).

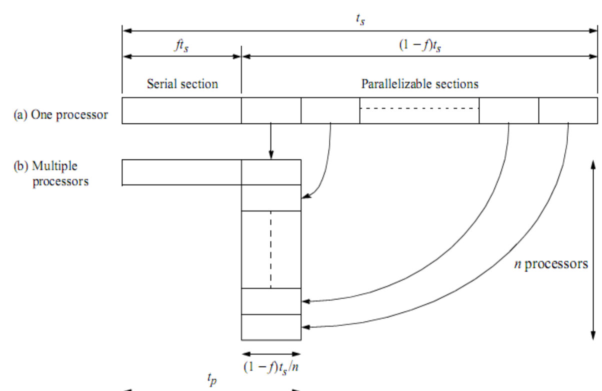


Figure 2 Amdahl principle representations.

$$S_{(n)} = \frac{\text{Execution time using one processor (single processor system)}}{\text{Execution time using multi processor with n processors}} = \frac{t_s}{t_p}$$

Speedup factor is given by

$$S(n) = \frac{t_s}{ft_s + (1-f)t_s/n} = \frac{n}{1+(n-1)f}$$

Where ts time to do in sequential way in one processer and tp execution including parallel ways of doing (multiprocessor) and fraction is represented by f.

Managers can use the Amdahl law as an indicator in which way the optimum combination of different activities can be done for the better decision making in PD. This only considers the sequential and parallel way of execution that doesn't comprise the overlapping processes. This can be one of the pitfalls of this method and the future work is developing a model that incorporates the overlapping processes.

Prasad (1997) considers the trend of revenue follows the S-curve and from an area of triangle, total revenue for early (on-time market introduction) is calculated as: Revenue loss term (R_{loss}) due to delay in introducing the new product is calculated as:-

$$R_{loss} = \frac{R_{early} - R_{delayed}}{R_{early}}$$

By taking the analogy of Prasad, delay loss can be calculated from speed up factor. That is if there is no any speeding up activities, the speedup factor is 1.

$D_{loss} = \frac{S_s - S}{S_s}$ where D_{loss} is delay loss, S_s is speed up with speeding factor and S without speed up factor that is 1.

5 System dynamic model

The mathematical representation of the model is based on the assumption of cost that follows growth function and the product development cost (cost at concept, design, test and production) is split into cost before and cost after design (test) for simplicity of the analysis

$T_C=C_D+C_T$, where C_D is cost before and including design whereas C_T cost after design (testing). Adding and subtracting some percentage of cost fraction ‘f’ and considering

$$C(T) = C_0 \cdot e^{\alpha T}$$

Adding and subtracting some percentage of cost fraction ‘f’ from each component of the cost and Anderson (2008) assumption of early cost allocation (80% of the total cost that is C_D=0. 8T_C and C_T=0. 2T_C), we rewrite the growing formula as:

$$C(T) = C_0 e^{\alpha T} = 0.8 T_C e^{\alpha T} + 0.2 T_C e^{\alpha T}$$

$$C(T) = C_0 e^{\alpha T} = (0.8 - f) \cdot C_D e^{\alpha T} + (0.2 + f) \cdot C_T e^{\alpha T}$$

And the effects may extend to the total time that may arise from reworks of each stage of the product development processes that is proportional to the ratio of the two costs. The interactions of some parameters are shown below in Figure 3.

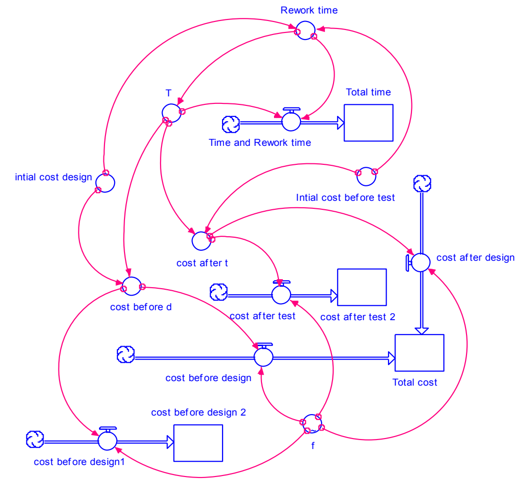


Figure 3 Model for cost distribution

6 Marketing as a system in SoSE

System dynamics model based on the basic marketing theory is developed and simulated. Due to space limitation we omit to present all mathematical models. To clarify the basic concept, we consider some parameters. Our assumption is a certain product has potential customers and the available customers will increase through advertisements (number of talkers per product). So the demand will increase. As the demand increase it has two effects i.e. on one hand it will increase the number of customers but on the other hand the number of potential customers decrease. This process will continue as we keep advertising and until the market saturate but it will be difficult to predict after this point since several additional parameters are incorporated and naturally the demand decline through time (See Figure 5, 6, and 7). Our idea is, if we consider marketing as a system that takes part in actual product development processes (in a concurrent engineering environment), the whole system becomes very complex and difficult to analyze and therefore, system of system thinking is eminent. Here we want to represent in the system of system approach in the following figure 4

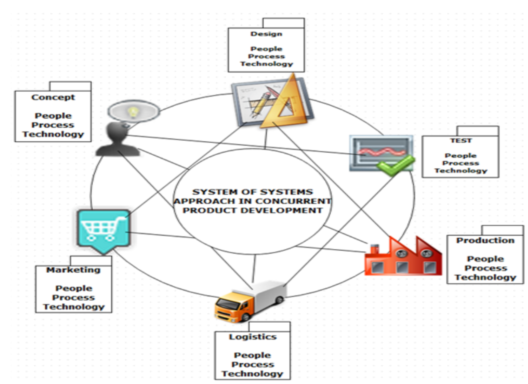


Figure 4 SoS incorporating marketing and related functions

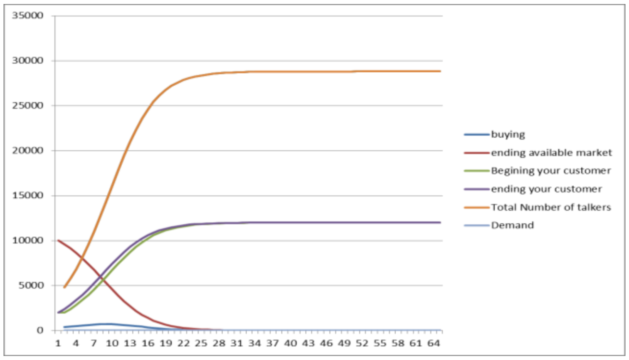


Figure 5 market situation before the market is saturated and advertise or talkers per product (=4)

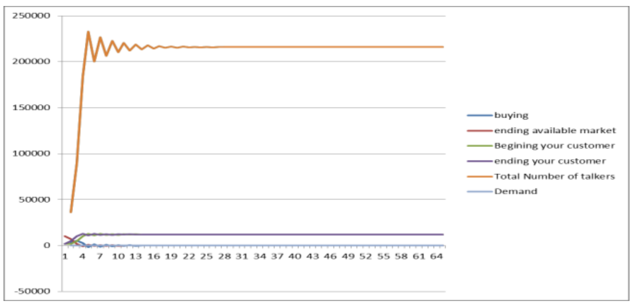


Figure 6 market situation after the market is saturated and advertise or talkers per product (=30)

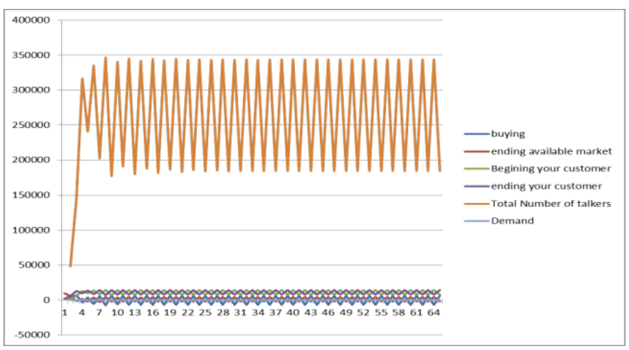


Figure 7 market situation after the market is saturated and advertise or talkers per product (=50)

7 Findings and implication

Based on the model developed (figure 3), the research found out the following results (figure 8, 9): A product development process with more Cost/budget allocated or invested before design come up with a reduction of total cost by almost half. In addition to that, the total time for developing a product is improved by 17%. When we invest more of the costs after design there and there will be a proportional delay loss. That means, time to market or delivering product to the customer on the schedule is affected by not doing so. The general trends of the cost curves somehow follow a similar trend as the literature and previous studies depicts (Andersen 2008).

From system of systems perspective, marketing is considered as one of the systems in lean PD like design and production. In this paper, the product is produced to be sold and understanding the marketing behavior has an impact on the decision of what and when and to whom. We found out marketing should be a part of the whole system in the product development process and it will help in minimizing unnecessary wastes by avoiding unnecessary investment on advertisement and reduce on rework by understanding the customer needs and know the future market situation.

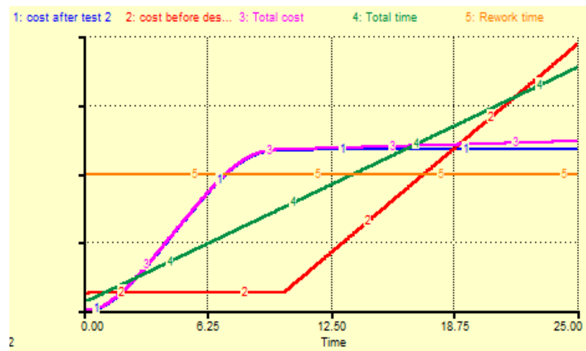


Figure 8 the total cost diagram when considering cost before design is higher than cost after design

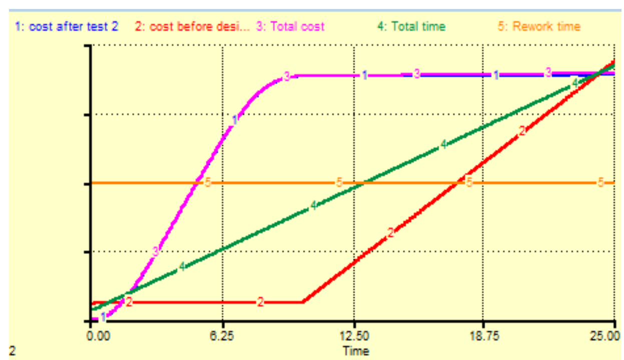


Figure 9 the total cost diagram considering when cost before design is less than cost after design (test)

8 Conclusion and summary

In product development, speed/time, cost and quality have been considered as the major goals during product development processes. That is why firms put their effort relentlessly to have satisfied customers by providing quality products with affordable price and keeping the delivery schedule as per the customers' requirements. However, due to challenging nature of development processes, managers need to know which system or activities should be improved more.

The paper studies the dynamic behavior of the complex system of the product development process while applying CE in SoSE approach. As we believe and the paper shows, the effect of concurrent engineering in companies that runs huge investment is eminent and significant savings can be gained by optimizing and systematically analyzing this engineering method. System approach is used as it helps to visualize and based on its capability to incorporate several subsystems and parameters. From the model and simulation, valuable findings are noted. When we change the cost fraction before and after design with the same amount, the total cost of the system varies significantly at different stages of the development processes. This variation indicates that there are some hidden costs that need to be optimized systematically, for instance doing some activities in parallel or overlap each other.

In the most concise terms, we found that applying a system approach using concurrent engineering (more invest before design stage) cuts about half of the total system cost, reduce time from 650 months to 540 months that is approximately 17% total reduction in time or improve time to market that rewards higher market share. The paper insights and try to clarify Amdahl's principle for future use to determine the speed up factor on complex systems. The marketing strategy is treated as a system and the challenge is explained in the model that invites SoSE approach in complex product development processes. Knowing the amount of customers will be an input for the PD processes. The future work will be combining each system of PD and see the impact of CE in the whole system.

9 References

[1] B. Prasad Analysis of pricing strategies for new product introduction, *Journal of Product & Brand Management*, Vol. 5, No. 4, pp.132–141, 1997.

[2] B. Wilkinson and M. Allen. *Parallel Programming: Techniques and Applications Using Networked Workstations and Parallel Computers*, Prentice Hall, USA, 1999

[3] David. M. Anderson, *how to design for low cost, design in high quality, design for lean manufacture, and design quickly for fast production, Design for Manufacturability & Concurrent Engineering*, CIM Press, Cambria, California, 2008.

[4] Department of defense (DoD), *Guide for Integrating Systems Engineering into DoD Acquisition Contracts* http://www.acq.osd.mil/se/docs/Integrating-SE Acquisition-Contracts_guide_121106.pdf. Version 1. 2006 accessed on 31.06.2012.

[5] Department of defense (DoD), *Guide for Integrating Systems Engineering into DoD Acquisition Contracts* http://www.acq.osd.mil/se/docs/SE-Guide-for-SoS.pdf version 1. 2008 accessed on 31.06.2012.

[6] International Council on Systems Engineering (INCOSE), *Systems Engineering Handbook* http://g2sebok.incose.org/documents/assets/MSS//Final/sh %20hdbk%202.2.pdf.2000. Accessed on 01.06.2012

[7] J. O. Clark, ''System of Systems Engineering and Family of Systems Engineering from a Standards, V-Model, and Dual-V Model Perspective,'' Proceedings of 3rd IEEE International Systems Conference (SysCon 2009), Mar., pp.381-387. 2009

[8] J.A. Lane, and R. Valerdi, ''Accelerating system of systems engineering understanding and optimization through lean enterprise principles,'' Systems Conference, 2010 4th Annual IEEE San Diego, CA, 5-8 April 2010, pp.196 – 201.

[9] LG. Marujo, (2009). Rework Impacts Evaluation through System Dynamics Approach in overlapped Product Development Schedule. *Journal of Technology Management & Innovation*, Volume 4, Issue 2

[10] M.G., Martin, M.R., Weeks, and K.J. ,Davis, "Delivering Successful Service and Outsourcing Projects Using Concurrent Engineering Methodologies", Proceedings of the 29th Annual Project Management Institute, Seminars & Symposium Long Beach, California, USA: Papers Presented October 9 to 15. 1998

[11] M.W. Maier. Architecting principles for systems of-systems', *Systems Engineering*, Vol. 1, No. 4, pp.267–284. 1999

[12] Port, Otis, Zachary Schiller, and Resa King. A Smarter Way to Manufacture: How 'concurrent engineering' can reinvigorate American industry, Business Week, April 30, 1990, pp. 110-117. 1990

[13] S. Chritamara, S.O. Ogunlana, N.L. Bach. System dynamics modeling of design and build construction projects. *Construction Innovation: Information, Process, Management*, Vol. 2 Iss: 4, pp.269 – 295. 2002

[14] Siemens PLM Software. *Enabling innovation through integrated systems engineering, Create, capture and deliver a systems perspective through integrated lifecycle processes and cross-discipline synchronization.* A White paper.http://www.plm.automation.siemens.com/de_ch/Images/9009_tcm782-4883.pdf,2011 Accessed on 02.06.2012

[15] Sterman, J.D., 2000. *Business Dynamics: systems thinking and modelling for a complex world*, New York: Irwin McGraw-Hill.

[16] Sterman, J.D., 2002. All models are wrong: reflections on becoming a systems scientist. *System Dynamics Review*, Vol. 18 No.4, pp. 501-531.

A Mathematical Framework for the Planning and Control of Complex Systems

Davide Giglio, Simona Sacone, Silvia Siri
Department of Communication, Computer and System Science (DIST)
University of Genova, Italy
davide.giglio@unige.it, simona.sacone@unige.it, silvia.siri@unige.it

Abstract – *A two-level framework is proposed in this paper for a class of complex systems in order to optimize, at a tactical planning level, the activities over a medium-term horizon, and to control, at an operational level, the system resource functioning over a short-term horizon or in real time. At the tactical planning level, an aggregate model of the system is defined and an optimization problem is stated and solved; the optimal solution of such a problem provides a reference plan to be tracked. At the operational level, both optimal control strategies are employed to face perturbations that may affect the system, and a detailed discrete-event or hybrid model is defined to verify some properties and performance indexes of the system. The application of the proposed framework to a logistic intermodal node is illustrated.*

Keywords: Modeling, tactical planning, optimization and control, discrete-event systems.

1 Introduction

A mathematical framework for the planning and control of complex systems, such as manufacturing systems and logistic intermodal nodes, is proposed in this paper. These systems are very complex since they are formed by many subsystems, in some cases interacting, in other cases competing for some shared resources. Moreover, the overall systems are generally able to perform a function that cannot be performed by each subsystem alone. These characteristics make the systems considered in the present work possibly defined as "Systems of Systems" [1, 2, 3].

The proposed framework acts at two levels, respectively the tactical planning level, in which an aggregate plan over a medium time horizon is defined, and the operational or control level, in which more detailed decisions must be taken (even in real time) and stochastic events affecting the systems are taken into account. More specifically, the former level is devoted to the definition of the optimal way in which the system must work, according to static and expected data. The result of this phase is an optimal plan that will be applied to the system. Though, the system operation is generally influenced by stochastic aspects, in particular in complex systems. These stochastic elements can be viewed generally as perturbations of the expected system behavior that has been considered in defining the optimal plan. For this reason, such a plan, when applied in real time, can result no more effective, or even no more feasible. Then, the system is periodically monitored and, if necessary, a re-planning procedure must be applied. Moreover, the latter level is formalized through a discrete-event model which represents the system resources at a higher level of detail and whose dynamics has the optimal plan, produced by the tactical planning level, as the reference plan to be tracked. Such a detailed model can be used to analyze the system performances and to check that the system avoids deadlock or forbidden states.

The aggregate models defined for the system at the tactical level are generally hybrid models [4], combining an event-driven dynamics (for instance modeling arrivals and departure processes) with a time-driven dynamics (characterizing for instance the production or the inventory process). In particular, among the modeling frameworks or paradigms which have been proposed for hybrid systems, the considered class of systems can be related to the "continuation paradigm" [5], in which discrete actions or events act as disturbances of a continuous dynamics. As regards instead the operational level, more detailed models are generally needed to correctly represent the system dynamics; then, discrete-event models are often adopted, mainly using the petri net formalism (refer for example to [6] and [7] for a manufacturing and a logistic application, respectively) or discrete-event simulation frameworks (see for instance [8, 9] for seaport terminals or [10] for an example of application to a manufacturing system).

Some specific cases of hybrid models applied to supply chain systems or inventory-production systems have been proposed in [11] and [12], respectively, by the same authors; in those cases, some asynchronous processes are considered (i.e. pulse processes or piecewise constant processes) and a finite-horizon optimization problem is stated. For the solution of the problem, in [11] dynamic programming techniques are applied and a feedback control strategy is found, whereas in [12] an algorithm providing the optimal solution in polynomial times is defined.

The paper is organized as follows. In Sections 2 and 3 the tactical and the operational levels are described. Section 4 reports an example referred to a real intermodal terminal, whereas some concluding remarks can be found in Section 5.

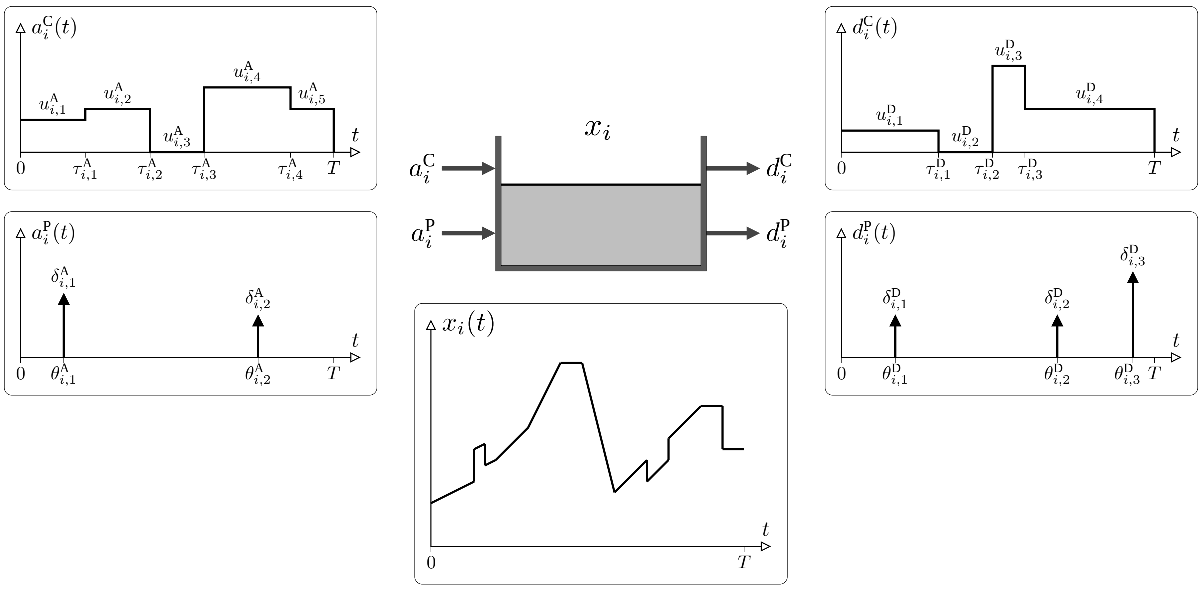


Figure 1: Sketch of the considered model.

2 The tactical level

At the tactical planning level, the system is modeled in an aggregate way, by means of a set of state variables representing the presence of different types of items in the system (they can be for instance finite products in the inventory of a manufacturing system or containers in the yard of a seaport). The dynamics of these variables depends on arrival processes, which feed the system with items, and departure processes, which withdraw items from the system. Arrival and departure processes can be, in the most general case, of two types:

- "continuous" processes, defined as piecewise constant patterns, changing rate at asynchronous time instants;
- "pulse" processes, defined as sequences of events asynchronous in time.

Referring to real systems, the arrival process can model for instance the arrival of finite products in the inventory of a manufacturing node (according to the production rate) or the arrival of containers by ship in a port. If this process is modeled as a piecewise constant function, it represents items entering the system in a continuous way, with a rate (e.g., parts/hour, containers/day) that changes over time. The pulse arrival can model for instance the arrival of a means of transport (a train or a ship at an intermodal node, or a truck at a manufacturing system) carrying some items that instantaneously enter the system. Analogously, the departure process can model items exiting the considered system in a continuous or pulse way.

In the considered model, let $A(t)$ denote the arrival process and $D(t)$ denote the departure process. Different types of items can be generally considered (products of different types, with different destinations, and so on); let index i denote each type of item, with $i = 1, \dots, n$. A simple sketch of the considered system is in Figure 1. For such system a finite-horizon optimization problem will be stated, over the time horizon T; hence, it is assumed that both the arrival process and the transportation offer refer to the time interval $(0, T]$.

Consider the arrival process $A(t)$. Let $A^{\mathrm{C}}(t)$ be the "continuous" arrival process. It consists of several functions, namely $a_i^{\mathrm{C}}(t)$, $i = 1, \dots, n$; $a_i^{\mathrm{C}}(t)$ is assumed to be a piecewise constant function, in the interval $(0, T]$, changing value at asynchronous time instants. In this connection, let $\tau_{i,q}^{\mathrm{A}}$, $q = 1, \dots, Q_i - 1$, be the generic time instant at which the arrival rate changes, and let $u_{i,q}^{\mathrm{A}}$, $q = 1, \dots, Q_i$, be the rate in the interval $(\tau_{i,q-1}^{\mathrm{A}}, \tau_{i,q}^{\mathrm{A}}]$. It is assumed $\tau_{i,0}^{\mathrm{A}} = 0$ and $\tau_{i,Q_i}^{\mathrm{A}} = T$, $\forall i$. An example of $a_i^{\mathrm{C}}(t)$ is at the top-left of Fig. 1. As regards the "pulse" arrival process $A^{\mathrm{P}}(t)$, it consists of several functions $a_i^{\mathrm{P}}(t)$, $i = 1, \dots, n$, which correspond to discrete-event processes. In particular, $a_i^{\mathrm{P}}(t)$ is a sequence of arrival events, asynchronous over time (that is, not equally spaced). In this connection, let $\theta_{i,p}^{\mathrm{A}}$, $p = 1, \dots, P_i$, be the generic time instant at which goods arrive, and let $\delta_{i,p}^{\mathrm{A}}$, $p = 1, \dots, P_i$, be the quantity of arriving goods. It is assumed $\theta_{i,0}^{\mathrm{A}} = 0$ and $\theta_{i,P_i}^{\mathrm{A}} < T$, $\forall i$. An example of $d_i^{\mathrm{C}}(t)$ is at the middle-left of Fig. 1. The departure process is absolutely analogous, with some "continuous" and some "pulse" processes, as shown as example at the right of Fig. 1. With these types of arrival and departure processes, the inventory levels in the system present the pattern shown at the bottom of Fig. 1, that is a piecewise linear function with "jump" discontinuities (related to the pulse processes).

According to the considered system, different decision variables are taken into account. If for instance the system state represents the inventory of finite products, the decision variables are the production rates and the delivery time instants in order to satisfy an external demand. Instead, if the system under study is an intermodal logistic node, since the arrival processes are generally fixed, then the decisions regard the delivery of containers from the node. On the basis of such decision variables, a finite-horizon optimization problem is generally stated. The objective function to be minimized may include different cost terms, according to the specific system considered, such as holding costs, costs related to the demand satisfaction (earliness, tardiness, backlog, and so on) and production costs. The constraints of the optimization problem include the state equations and some physical limits of the system (for instance the maximum production effort in a manufacturing system or the maximum available space in the yard of a terminal).

One of the main features of the proposed models consists in the fact that the decisions concern the timing of the operations. In other words, if the decisions regard a continuous piecewise constant process, the decision variables are not only the values of the rates over time, but also the time instants at which these values change. When the decision deals with pulse processes, the decision variables are both the quantities associated with the pulses and the related asynchronous time instants when these events occur. It is worth noting that the decisions on timing make the considered planning problem more complicated than in the case of discrete-time horizon with fixed and known time intervals and, in fact, this planning problem cannot be formulated and solved via mixed-integer programming. On the contrary, the optimization problem defined within the proposed framework has generally a parametric structure and includes nonlinearities and combinatorial aspects. In fact, since some asynchronous time instants are matter of decision in the optimization problem, the occurrence of some events is not known at the beginning; this aspect is very crucial in the statement of the optimization problem because it is necessary to introduce some sets of binary variables indicating, for each pair of events to be decided, whether a given event occurs before or after the other one. The defined optimization problem can be solved in some cases analytically (e.g., in [12]) or in all cases with commercial solvers, leading to an optimal (or suboptimal) solution of the tactical planning phase.

3 The operational level

The second phase of the proposed framework deals with the operational level or with a real-time operating mode. In order to face such phase, two different approaches are proposed.

First of all, when it is possible, optimal (closed-loop) control strategies as functions of the system state are determined. Closed-loop strategies are useful when the system is affected by perturbations (for the considered systems, the perturbations can be of many types, such as late arrivals or departures, or machines working with an effort that is lower than the nominal one, a demand different from the estimated one, and so on). Generally speaking, these closed-loop strategies can be found by applying optimal control techniques. Optimal control strategies provide in real time the new optimal reference plan to be tracked without the need of re-executing the original optimization problem.

Optimal control strategies has been determined by the authors, up to now, for a reduced model with respect to the one described in the previous section. In such a reduced model, which refers to a class of inventory-production systems, only the continuous arrival process is taken into account and the delivery process is assumed given a-priori and, thus, it is not matter of decision. In this case, the optimal control strategies can be determined analytically, on the basis of the solution provided by the tactical level. More specifically, let $a^\circ(t)$ and $x^\circ(t)$, $0 \leq t \leq T$, be, respectively, the "optimal path" of the continuous arrival process (piece-wise constant function of t) and the "optimal path" of the state of the system (piece-wise linear function of t), which are obtained by solving the optimization problem stated and solved at the tactical level. $x^\circ(t)$ represents the state when no perturbation affects the system. However, when some perturbations occur, the actual state may differ from $x^\circ(t)$. In this connection, consider a generic time instant $t^\star \in [0, T)$ at which the system state is measured, and let $x^\star$ be the measurement. Moreover, assume that a perturbation has occurred somewhere in $[0, t^\star)$, such that it results $x^\star \neq x^\circ(t^\star)$. The optimal control strategy, which is determined at this level, provides the new optimal path of the arrival process, namely $\tilde{a}(t, t^\star, x^\star)$, which has to be applied from $t^\star$ onward, in order to minimize the costs of the system. $\tilde{a}(t, t^\star, x^\star)$ is a closed-loop strategy, being a function of the actual system state at $t^\star$, namely $x^\star$. The function $\tilde{x}(t, t^\star, x^\star) = x^\star + \int_{t^\star}^{t} \tilde{a}(\eta, t^\star, x^\star)\,\mathrm{d}\eta - \int_{t^\star}^{t} d(\eta)\,\mathrm{d}\eta$ is the "optimal perturbed path" of the system state (piece-wise linear function of t), which results from the application of $\tilde{a}(t, t^\star, x^\star)$ (instead of $a^\circ(t)$).

The second approach proposed at the operational level regards the modeling of the system, at a higher level of detail, with a discrete-event or hybrid model (for instance, by using the Petri net formalism). As a matter of fact, the optimization at the tactical level is made without taking into account the resources of the system. Then, the detailed discrete-event or hybrid model is employed with the aim of allocating resources (eg., cranes, trailers, reach stackers, etc., of an intermodal node, or machines, robots, etc., of a manufacturing system) to the various operations (handling, transportation, production, etc.) and of sequencing operations so that the reference plans provided by the tactical level are fulfilled. Moreover, the detailed model is also adopted to verify some important properties and performance indexes for the system.

4 Example – Planning and control of an intermodal logistic system

Let us consider a simple example in order to better clarify the planning and control framework described in the previous sections. The data of this example regard a multimodal transport operator in a seaport terminal of the North of Italy and, in particular, the import flows, i.e. containers arriving by sea and destined to the hinterland, via road or rail. As it usually happens in such terminals, the operator knows the schedules of vessels about one year in advance, but of course he acquires the information about specific containers with a lower advance. Moreover, as regards the transportation offer, and in particular the availability of trains, the operator reserves in advance (with a yearly plan) some time slots for the departure of trains, depending on the estimation on transportation demand. Of course, he is able to make the final plan, i.e the decision about the modal split and the departure patterns for these containers, over a horizon that is certainly shorter, generally corresponding to a weekly or at least monthly plan. Therefore, at the tactical level an optimal plan must be defined over a horizon T equal to 1 week, given the pattern of arrivals and the transportation offer (based on realistic data provided by logistic operators). In particular, 6 days are considered, from Monday to Saturday, since on Sundays neither freight trains nor trucks are allowed to circulate.

In the considered case, the transportation demand to be served is relevant to 2 destinations (i.e., two inland intermodal terminals) and 3 owners; in particular, the first and third owner must deliver goods to both destinations, whereas the goods of the second owner must reach only the first destination. Moreover, let us define all the time instants in hours, hence $T = 144$. As already mentioned, in the present case import flows are considered, then the arrivals at the terminal are only relevant to ships (pulse arrival process) as follows:

Owner	Destination	Arrival time	TEUs
1	1	$t = 10$	120
1	2	$t = 10$ $t = 115$	80 150
2	1	$t = 38$ $t = 80$	60 110
3	1	$t = 10$	100
3	2	$t = 115$	100

The deliveries from the intermodal node can be realized according to the available transportation offer, that in this case corresponds to two different modes, i.e. rail and road. As regards rail transportation (pulse transportation offer), there are three possible trains for each destination with the following schedules and capacities:

Destination	Departure time	Capacity (in TEUs)
1	$t = 8$ $t = 66$ $t = 78$	75 60 80
2	$t = 44$ $t = 118$ $t = 136$	60 75 60

The transportation offer provided by trucks (continuous offer) is defined as a piecewise constant pattern as follows: it is equal to 10 TEUs per hour between 6 a.m. and 10 p.m form Monday to Friday, 6 TEUs per hour between 6 a.m. and 10 p.m on Saturday, 2 TEUs per hour between 10 p.m. and 6 a.m. every night. Moreover, a traffic penalty function is defined in order to discourage deliveries by truck in the peak hours; specifically, this cost function is higher on Monday morning (from 6 a.m. to 2 p.m.) and on Friday afternoon (from 2 p.m. to 10 p.m.). Finally, as regards the initial conditions, at the beginning of the planning horizon there are 10 TEUs for shipments with destination 1 (for both owners), and 20 TEUs for shipment of owner 3 and destination 2; no containers of the other shipments are present on the yard at the beginning of the week.

At the tactical level, an optimization problem can be stated and solved in order to find the optimal delivery processes for containers to be sent to destinations 1 and 2, taking into account the arrival process and the available transportation offer. The resulting problem is a nonlinear mixed-integer mathematical programming problem that has been solved by means of the commercial solver Lingo.

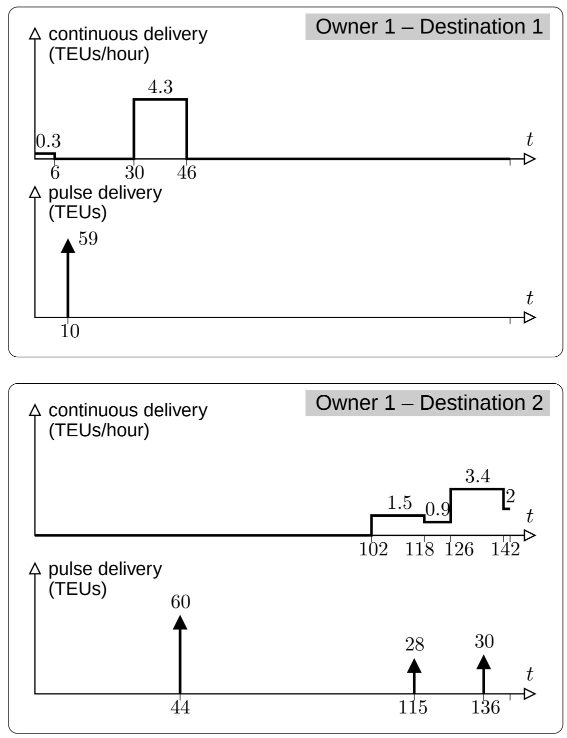


Figure 2: Optimal delivery processes for Owner 1.

In the determined optimal solution, trains travel fully loaded. The details of the compositions of the six trains departing from the seaport terminal can be derived from Figures 2÷4 (as an example, the first train to destination 1, departing at $t = 10$, will transport 59 TEUs of owner 1 and 16 TEUs of owner 3). It is worth highlighting that two trains (the first of the three trains to destination 1 and the second of the three trains to destination 2) are delayed, with respect to their scheduled time, of 2 and 5 hours, respectively; such delays allow waiting the containers arriving at the terminal by sea (at $t = 10$ and at $t = 115$). Of course, such delays are compatible with the maximum delay allowed in this system. Also trucks are used to deliver containers from the terminal. Analysing again Figures 2÷4, it is possible to note that trucks are mainly used in the second part of the considered time period, to deliver containers that cannot be placed

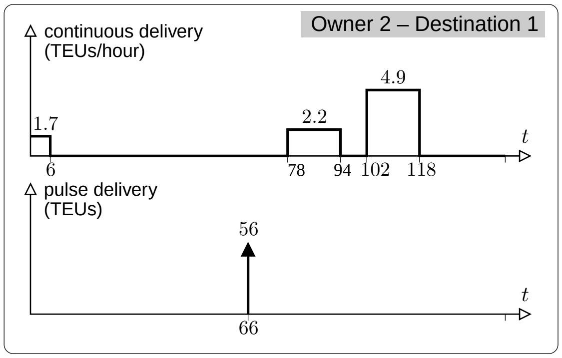


Figure 3: Optimal delivery processes for Owner 2.

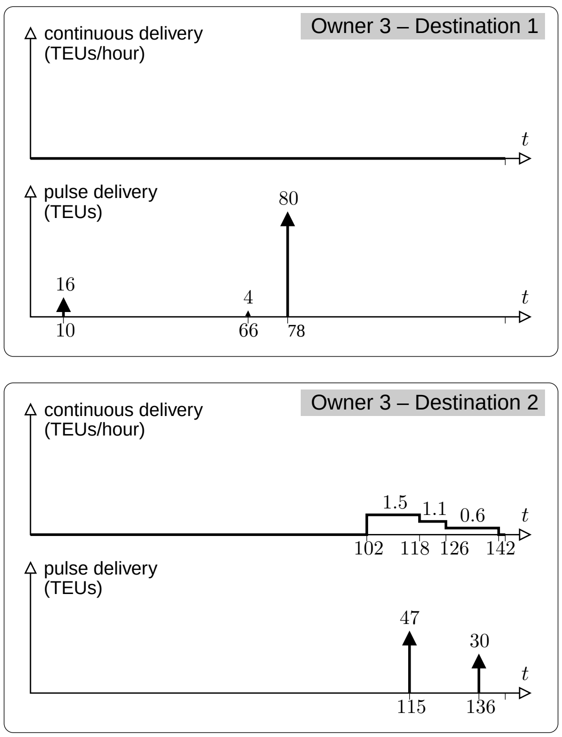


Figure 4: Optimal delivery processes for Owner 3.

on trains because of their limited capacity. The modal split in the optimal solution turns out to be: 410 TEUs delivered by train (55.5%) and 328 TEUs delivered by trucks (44.5%). Only a small number of containers remain in the terminal at the end of the considered time interval, corresponding to 26 TEUs.

At the operational level, a Petri net is defined to represent detailed activities in the intermodal logistic system. Such a net is actually used to test different allocation strategies which allow fulfilling the delivery plans provided by the tactical level. The part of the net referred to the quay at which ships arrive is illustrated in Figure 5.

5 Conclusions

In the paper, a mathematical framework for planning and controlling complex systems has been described. The considered systems are generally composed of different subsystems (represented as different inventory levels) interacting with each other, for instance because they share some resources. Such subsystems cannot be planned separately but it is necessary to optimize the system as a whole. The considered systems can be then classified as "Systems of systems". An example of an intermodal logistic node has been presented in the paper, describing the approach prosed both for the tactical planning level and for the operational or real-time level.

References

[1] N. Karcanias and A.G. Hessami, "Complexity and the notion of System of Systems: part (I): general systems and complexity", *Proc. World Automation Congress*, 2010.

[2] N. Karcanias and A.G. Hessami, "Complexity and the notion of System of Systems: part (II): defining the notion of System of Systems", *Proc. World Automation Congress*, 2010.

[3] A. Mostafavi, D.M. Abraham, D. DeLaurentis and J. Sinfield, "Exploring the dimensions of Systems of Innovation analysis: a System of Systems framework", *IEEE Systems Journal*, Vol 5, No. 2, pp. 256–265, 2011.

[4] C.G. Cassandras, D.L. Pepyne and Y. Wardi, "Optimal control of a class of hybrid systems," *IEEE Transactions on Automatic Control*, Vol 46, No. 3, pp. 398–415, 2001.

[5] M.S. Branicky, V.S. Borkar and S.K. Mitter, "A unified framework for hybrid control: model and optimal control theory,"*IEEE Transactions on Automatic Control*, Vol 43, No. 1, pp. 31–45, 1998.

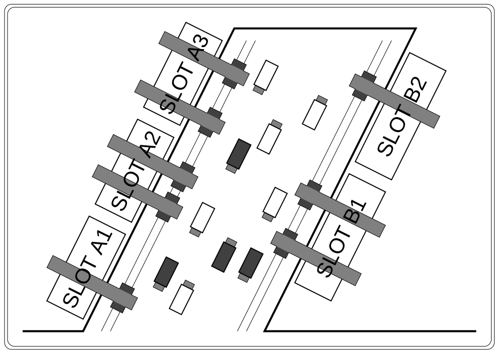
[6] H. Liu, Z. Jiang and R.Y.K. Fung, "Performance modeling, real-time dispatching and simulation of wafer fabrication systems using timed extended object-oriented Petri nets", *Computers & Industrial Engineering*, Vol 56, pp. 121-137, 2009.

[7] M.P. Fanti, A. Giua, C. Seatzu, "Monitor design for colored Petri nets: An application to deadlock prevention in railway networks", *Control Engineering Practice*, Vol 14, pp. 1231-1247, 2006.

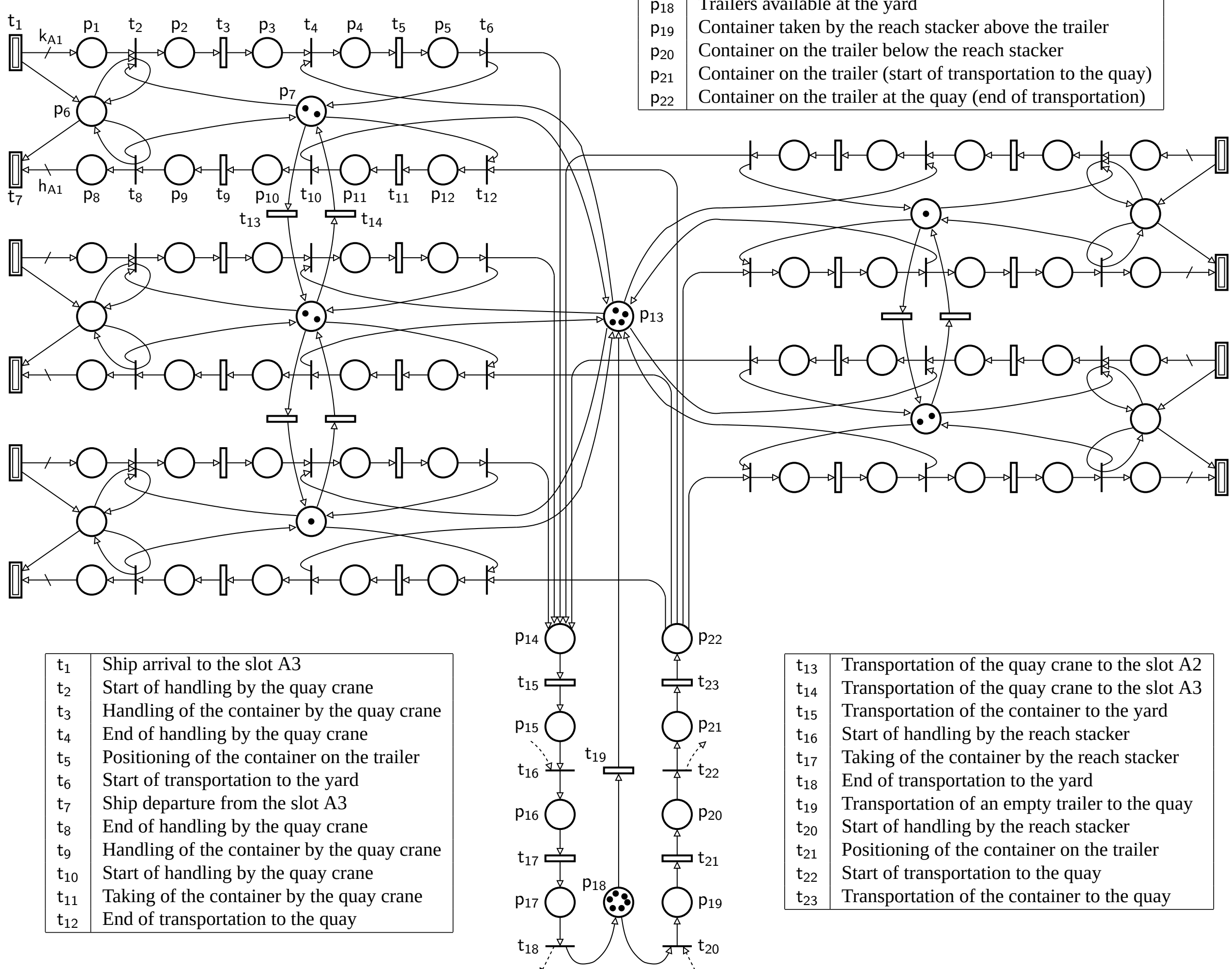
[8] A. Cartení and S. de Luca, "Tactical and strategic planning for a container terminal: Modelling issues within a discrete event simulation approach", *Simulation Modelling Practice and Theory*, Vol 21, pp. 123–145, 2012.

[9] M. Kia, E. Shayan F. and F. Ghotb, "Investigation of port capacity under a new approach by computer simulation", *Computers and Industrial Engineering*, Vol 42, pp. 533–540, 2002.

[10] J.-J. Bang and Y.-D. Kim, "Hierarchical Production Planning for Semiconductor Wafer Fabrication Based on Linear Programming and Discrete-Event Simulation",



p_1	Containers to be unloaded from the ship at the slot A3
p_2	Container taken by the quay crane (start of handling)
p_3	Container taken by the quay crane (end of handling)
p_4	Container taken by the quay crane above the trailer
p_5	Container on the trailer below the quay crane
p_6	Ship at the slot A3
p_7	Quay cranes available at the slot A3
p_8	Containers loaded on the ship at the slot A3
p_9	Container taken by the quay crane (end of handling)
p_{10}	Container taken by the quay crane (start of handling)
p_{11}	Container on the trailer below the quay crane
p_{12}	Container taken by the quay crane above the trailer
p_{13}	Trailers available at the quay
p_{14}	Container on the trailer (start of transportation to the yard)
p_{15}	Container on the trailer at the yard (end of transportation)
p_{16}	Container on the trailer below the reach stacker
p_{17}	Container taken by the reach stacker above the trailer
p_{18}	Trailers available at the yard
p_{19}	Container taken by the reach stacker above the trailer
p_{20}	Container on the trailer below the reach stacker
p_{21}	Container on the trailer (start of transportation to the quay)
p_{22}	Container on the trailer at the quay (end of transportation)



t_1	Ship arrival to the slot A3
t_2	Start of handling by the quay crane
t_3	Handling of the container by the quay crane
t_4	End of handling by the quay crane
t_5	Positioning of the container on the trailer
t_6	Start of transportation to the yard
t_7	Ship departure from the slot A3
t_8	End of handling by the quay crane
t_9	Handling of the container by the quay crane
t_{10}	Start of handling by the quay crane
t_{11}	Taking of the container by the quay crane
t_{12}	End of transportation to the quay

t_{13}	Transportation of the quay crane to the slot A2
t_{14}	Transportation of the quay crane to the slot A3
t_{15}	Transportation of the container to the yard
t_{16}	Start of handling by the reach stacker
t_{17}	Taking of the container by the reach stacker
t_{18}	End of transportation to the yard
t_{19}	Transportation of an empty trailer to the quay
t_{20}	Start of handling by the reach stacker
t_{21}	Positioning of the container on the trailer
t_{22}	Start of transportation to the quay
t_{23}	Transportation of the container to the quay

Figure 5: Petri net representing the activities of resources (yard cranes and trailers) at the quay at which ships arrive (illustrated in the top-left part of the figure) and the transportation of containers to the yard.

IEEE Transactions on Automation Science And Engineering, Vol 7, No. 2, pp. 326–336, 2010.

[11] D. Giglio, R. Minciardi, S. Sacone and S. Siri, “On optimizing production nodes in supply chain systems”, in *Innovations in Distribution Logistics*, L. Bertazzi, M.G. Speranza, J. van Nunen (Eds.), vol. 619, Lecture Notes in Economics and Mathematical Systems, Springer Verlag, pp. 149-174, 2009.

[12] D. Giglio, S. Sacone, and S. Siri, “Asynchronous regulation of service speed in inventory-production systems with time-varying positive demand”, in *Proc. of the 50th IEEE Conference on Decision and Control and European Control Conference*, 2011, pp. 5874-5880.

Model-based requirements engineering for system of systems

Jon Holt, Simon Perry, Mike Brownsword & Daniela Cancila
Atego Systems
Cheltenham,
GL50 1TA, UK
jon.holt@atego.com

Stefan Hallerstede and Finn Overgaard Hansen
Engineering College of Aarhus
Denmark
sha@iha.dk

Abstract - *Model-based systems engineering is a discipline of systems engineering where a model forms the heart of all the systems engineering activities and is the basis of many of the project artefacts. Systems modelling is no longer viewed as simply a 'good idea' but is, however, increasingly become an essential part of any systems engineering project. Indeed, this is the start point for the EU FP7 project known as COMPASS (Comprehensive Modelling for Advanced Systems of Systems) that has funded this research. The application of model-based systems engineering is becoming well understood at the systems level, however, there is a lack of research and subsequent industrial application at the system of systems (SoS) level. The main aim of the COMPASS project is to provide an approach for SoS including: processes, tools and new modelling techniques. This paper discusses how a model-based systems engineering approach was developed for requirements engineering that could be applied to both system of systems (SoS) and its constituent system (CS) level. This approach would meet the requirements of current best practice in system of systems in terms of established standards and research.*

Keywords: Model-based systems engineering, requirements engineering, systems modelling, system of systems

1 Introduction

Model-Based Systems Engineering (MBSE) describes an approach to systems engineering where the model forms the heart of all the systems engineering activities and also many of the artfefacts [1]. There are many benefits that can be realized when applying a model-based approach compared to a more document-centric approach. These include:

- -Reduced development time, due to increased automation, consistency checking and traceability analysis.
- -Enhanced analysis, due to the ability to automate trade studies, impact analyses and to perform simulations.
- -Increased re-uses, when used in the same context, then elements of a MBSE system may be re-used.

In order to realize these benefits, however, there are a number of areas that must be addressed [2].

1.1 Realizing MBSE

In order to realize these benefits and gain full advantage of MBSE, there are three basic areas that must be addressed:

- -People – there must be properly-educated, trained, experienced people available who hold the appropriate competence for their roles.
- -Process – in order to realize MBSE capability, there must be an effective set of processes in place, properly deployed and available to all people.
- -Tools – 'sharp' tools are required, particularly for automation, but not just in the form of CASE tools. Other tools will include notations, architectural frameworks, and so on.

One area of systems engineering that is essential for realizing successful systems is that of requirements engineering and requirements management. This paper focusses on the application of MBSE for requirements engineering.

1.2 Model-based requirements engineering (MBRE) for systems of systems (SoS)

There are many techniques available for carrying out requirements activities for systems engineering at a systems level, but are they equally applicable for systems engineering at a system of systems level? This paper looks at one established approach to model-based requirements engineering, known as ACRE [2], and assesses it suitability for being applied at a system of systems level.

In order to apply an MBSE approach at a systems level, there are a number of standards and best practice models that exist that may be used as a starting point for requirements engineering and management. Two such example are:

- -CMMI – capability maturity model integration [3, 4 & 5]. The CMMI explicitly covers requirements engineering (stated as 'requirements development' in the standard) and requirements management.

- -ISO 15288 – software and systems engineering life cycle processes [4]. This standard explicitly covers requirements engineering and requirements management. These processes are also expanded upon in the best-practice guidelines [1]

In order to be applied at a system of systems level, any technique must be able to be applied to the four basic types of SoS [6], which are:

- -Virtual SoS, that lack a central management authority and a centrally agreed upon purpose for the system-of-systems
- -Collaborative SoS, where constituent systems interact more or less voluntarily to fulfil agreed upon central purposes
- -Acknowledged SoS, have recognized objectives, a designated manager, and resources for the SoS; however, the constituent systems retain their independent ownership, objectives, funding, and development and sustainment approaches
- -Directed SoS, the integrated system-of-systems is built and managed to fulfil specific purposes. It is centrally managed during long-term operation to continue to fulfil those purposes as well as any new ones the system owners might wish to address

As well as being applied to these four types of SoS, there are a number of features that are unique to SoS that must be addressed [9]:

- -Consideration of perspective of the SoS and the constituent systems (CS). This includes considering the identification of capabilities for the SoS and also the needs of each individual CS.
- -Consideration of SoS current and future needs against individual capabilities. This includes understanding how best to engineer individual CS, understanding how the capabilities provided by these CS can be combined to meet the goals of the SoS and understanding the needs on the SoS environment.

Any approach to MBRE for SoS must meet all of the requirements listed above. On top of these requirements, it should not be forgotten that there was also a set of requirements for the COMPASS project itself, as described in the description of work [10].

Based on all of these source requirements, a number of use cases were drawn up that could be traced back to the source requirements. The use cases were used to put the source requirements into the context of the COMPASS project. The proposed approach for MBRE for SoS could be demonstrated to satisfy the use cases then, through traceability views, it is also possible show which of the source requirements they meet.

The approach that was to be developed would comprise: an ontology, a set of processes and framework – each of these will be expanded upon in subsequent sections of this paper.

2 An approach for context-based requirements engineering (ACRE)

As a start point for developing an approach for MBRE for SoS part of the COMPASS project, a number of approaches for requirements engineering were considered [11]. The approach that was decided upon was the ACRE approach, for a number of reasons:

- -The ACRE follows a true MBSE approach that describes a requirements ontology that is then used as a basis for a number of views that can be used to visualize a complete set of requirements.
- -The ACRE approach does not dictate any specific process and, hence it may be used with any process of methodology.
- -The ACRE approach may also be used at different levels of project on terms of scale (from a small, one-week project to a large multiyear project) and in terms of rigor (from non-critical to mission-critical systems).
- -The ACRE may also be visualized using any notation, or combination of appropriate notations.

The primary reason for selecting ACRE was its flexibility.

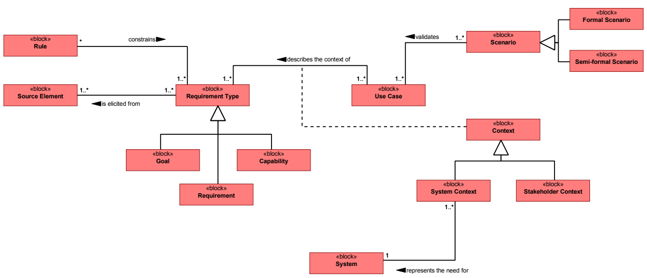
2.1 The ACRE ontology

In order to capture and describe the concepts and terminology, an 'ontology' was introduced. An ontology, in the context of this paper, provides a visualisation of all the key concepts, the terminology used to describe them and the inter-relationships between said concepts. The ontology, however, is much more than just a data dictionary and plays a pivotal role in the definition and use of any rigorous framework.

The use of ontologies for defining frameworks for architectures, such as enterprise architectures, process architectures, system architectures and so on, is one that is well-established and used extensively throughout industry. For examples of the use of ontologies, see [12], [13] and [14]. Whenever any framework is defined in terms of a set of views, then an ontology is essential. It is the ontology that enforces the consistency and rigour demanded by such frameworks.

The ontology introduced here will cover all of the concepts pertinent to model-based requirements engineering and a number of views will be defined based on this ontology. Each view will focus on, and expand upon, a subset of the ontology and instantiate, or realise, specific concepts in the context of a real system or project.

Based on the results of a survey of modelling techniques, SysML was decided upon for the modelling noatation for the project [15]. The following diagram shows the ontology using a SysML block definition diagram.



.Figure 1. The ACRE ontology

The diagram here shows that there is an abstract concept of a 'Need' that has three types: 'Requirement', 'Capability' and 'Goal'. One or more 'Need' is elicited from one or more 'Source Element'. One or more 'Rule' constrains one or more 'Need'.

One or more 'Use Case' describes the context of each 'Need' via the 'Context'. There are two types of context shown here: the 'System Context' and the 'Stakeholder Context', although this list is incomplete. Note that the original ACRE approach is aimed at system (or constituent system) level and, therefore, there is no reference to a system of systems here.

One or more 'Scenario' validates one or more 'Use Case' and there are two types of 'Scenario' – the 'Semi-formal Scenario' and the 'Formal Scenario'.

2.2 The ACRE framework

The ACRE approach comprises a number of pre-defined views that are based on the ACRE ontology. Figure 2 shows the set of views in the ACRE framework.

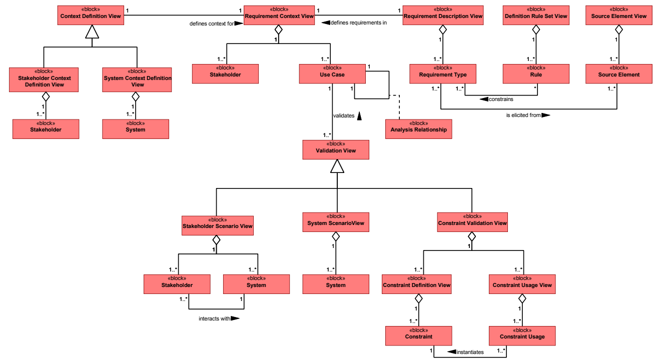


Figure 2. The ACRE framework

These views are described as follows:

- The Source Element View contains all relevant source information that is required to get the requirements right. It is essential that the origin of all requirements is known and this is what this view allows this definition. This view is used primarily as a mechanism to establish traceability and provide links between the requirements and any other aspect of the system.
- The Requirement Description View contains structured descriptions of each of the needs, where they are requirements, goals or capabilities. The main purpose of this view is to describe each individual requirement according to a pre-defined set of attributes. These attributes will vary depending on a number of factors. This view is primarily used for managing the requirements of a system and is often the basis of implementation for many of the commercial requirements management tools that are on the market today. Each requirement description provides a non-contextual description of the requirement.
- The Rule Set Definition View contains rules that may have to be applied to each requirement definition. For example, these may be complexity rules in the form of equations of more general text-based rules.
- The Requirement Context View takes the requirements and gives them meaning by looking at them from a specific point of view. This is known as putting the requirements into context and forms the basis of the approach presented in ACRE. The problem arises that these requirement descriptions may be interpreted in different ways depending on the viewpoint of the reader of the requirement description. It is essential then that each requirement is looked at from different points of view, or in different contexts. When a requirement is put into context it is known as a 'use case' and by considering these uses case and

the relationships between them and other use cases or stakeholders, it is possible to generate a complete point of view, or context. These contexts may be based on a number of elements, such as stakeholders or levels of hierarchy in a system.

- The Context Definition View identifies the points of view that are explored in the Requirement Context View. These points of view, or contexts, may take many forms including stakeholders and levels of hierarchy in a system.
- The Validation Views provide the basis for demonstrating that the requirements can be met or complied with in some way. These views can be informal scenarios, such as those based on sequence diagrams at various levels of abstraction or may be formal, such as mathematical-based scenarios.
- The Traceability Views, a key part of any requirements engineering endeavour is to provide traceability both to and from the original requirements. There are two levels of traceability relationships that exist in the model, the 'implicit' and 'explicit' relationships. The implicit relationships are the ones that are inherent in the modelling language itself. The explicit relationships are those that are not inherent in the modelling notation but that are dependent on the application of the modelling. These relationships can be identified directly from the ontology and the framework. It is often necessary, therefore, to define exactly where the traceability relationships exist. Indeed, it is possible to trace between almost any system element and any element in the framework.

3 Applying ACRE to SoS

The next step in the research was to define a set of processes that could be used to describe the approach. It was decided to use the ACRE ontology and framework as a start point and to define processes based around them. This would enable the first draft of the processes to be defined and then applied to a test model in order to assess the suitability of the process.

3.1 Defining the processes

The processes were to be defined using a best-practice model-based approach known as the 'seven views' approach [12]. The processes were identified based on the use cases and then defined according to the approach.

These processes were then applied to a test model in order to assess their suitability for SoS. This test model is an SoS that represents emergency service producers [16]. Based on the experience of applying these processes, it was then possible to re-define the processes based on the lessons learned and to re-visit the original ACRE ontology and framework.

One of the features of ACRE is that it is based on a context-based approach to requirements engineering. This context-based approach for systems engineering is particularly interesting from the point of view of SoS, bearing in mind the following points:

- A context represents a system from a specific point of view.
- A system of systems may be thought of as a different, higher-level point of view of a set of systems. Therefore, a context exists at both the system and SoS level [17].

Bearing these points in mind, it was anticipated that most of the knowledge and experience of applying ACRE at a systems level could be re-used, to a certain extent, at the SoS level as exactly the same modeling techniques could be applied.

4 The COMPASS SoS approach

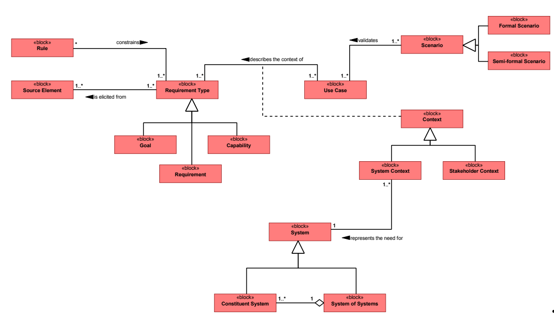
The COMPASS approach for MBRE for SoS comprises the following elements:

- The ontology, where all the key concepts and terminology are defined.
- The framework, where all the necessary views are defined
- The process set, where all the processes necessary to generate the views in the framework were defined.

Each of these is expanded upon on the following sections.

4.1 The COMPASS ontology

The COMPASS ontology was based on the ACRE ontology and can be seen below.



. Figure 3. The COMPASS ontology

The first point to notice here is that the ontology here is very similar indeed to the original ACRE ontology. In fact, the only area where the ontology has changed is in the definition of the systems that relate to the context. In the updated ontology it can be seen that the 'System Context' represents the need of two types of 'System': the 'Constituent System' and the 'System of System' (that itself is made up of a number of 'Constituent System').

4.2 The COMPASS framework

The COMPASS framework took as its starting point the ACRE framework. The COMPASS framework for SoS is shown in the diagram below.

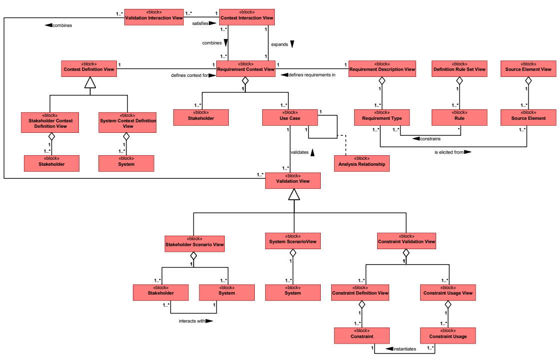


Figure 4. The COMPASS ontology

The framework shown here is very similar to the original ACRE framework. It was found that the basic ACRE approach could be applied at both the system and SoS level. However, this still left some gaps as, although a context could be defined at both the CS and SoS level, there was a need to understand the relationships between them. This resulted in the need to define two new views that are necessary when modeling requirements for SoS:

- The 'Context Interaction View'. A context can be defined at both the CS and SoS level, but these contexts need to be related together. For example, the capabilities of a SoS need to be mapped to the requirements of the underlying CS that realize them. The purpose of this view, therefore, is to allow the interactions between the SoS context and its CS contexts to be visualized. The Context Interaction view expands upon the SoS context and also combines a number of CS contexts.

- The 'Validation Interaction View'. It is possible to generate a number of scenarios for any context that can be used for validation purposes. Therefore, if a set of validation views exists at the SoS level and a set also exists for each CS, then there must be some relationship between them. The Validation Interaction View satisfies the SoS context by combining the set of CS validation views.

By enhancing the original ACRE framework with these new views it was now possible to use this framework as a basis for the COMPASS processes.

4.3 The COMPASS processes

The COMPASS processes describe the approach taken to generate the views in the framework, according to the ontology. The processes were defined using the seven-views approach to process modeling, which resulted in a set of eight processes being defined, as shown in the diagram below.

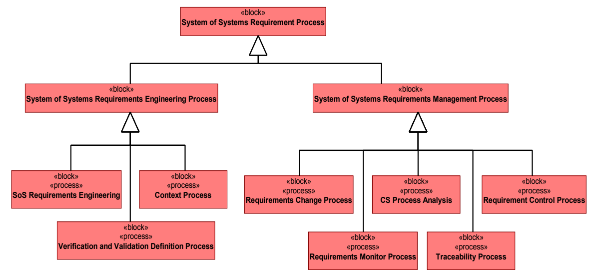


Figure 5. The COMPASS SOS requirements processes

The diagram here shows the library of processes (known as the 'process content view' in the seven views approach) that were defined for SoS requirements. The 'SoS Requirements Engineering Process' has three processes defined that are described briefly as:

- 'SoS Requirements Engineering' The overarching process that controls the requirements engineering.
- 'Context Process'. This process defines a context at either CS or SoS level
- 'Verification and Validation Definition Process' This process defines the verification and validation criteria for the SoS

The 'SoS Requirements Management Process' has three processes defined that are described briefly as:

- 'Requirements Change Process' that controls changes to the CS or SoS requirements
- 'CS Process Analysis' that allows the management processes of a CS to be understood
- 'Requirement Control Process' that ensures that all requirement are agreed to commitment is obtained
- 'Requirements Monitor Process' that allows changes in requirements at both CS and SoS level to be identified
- 'Traceability Process' that allows traceability views to be set up

The processes were described in far more detail, but this low level description (using the other six of the seven views) is not within the scope of this paper.

These processes can now be executed in different sequences to satisfy the COMPASS use cases.

4.4 Satisfying the original requirements

As was stated previously, the approach taken to define the processes was based on the ACRE approach. A single model was produced that contained the requirements for COMPASS (based on ACRE) and the process definitions (based on the seven views approach). As both of these approaches, ACRE and seven views, are true model-based systems engineering approaches, then it is a simple matter to combine the artifacts of both into a single SysML model.

The model was also fully traceable as follows:

- The COMPASS processes were executed in sequences as scenarios. These scenarios were then traced back to the COMPASS use cases. ('Validation View' to 'Requirement Context View')
- The use cases represent the original requirements in the context of the COMPASS project. These use cases were traced back to the original requirements. ('Requirement Context View' to 'Requirement Description View')
- The original requirements were derived from a number of requirements sources, such as standards, best-practice guides and project documentation. (requirement Description View' to 'Requirement Source View')

All of this traceability was built into the COMPASS model and, hence all of the traceability views can now be automatically produced.

5 Conclusions

This paper has presented an approach to carrying out model-based requirements engineering that is suitable for systems of systems. This work has been carried out as part of the EU FP7 project known as COMPASS [17].

The original requirements for the work were taken from a number of sources, including standards, best-practice guides, papers and project documentation. Not only did the work result in an MBSE approach for requirements for SoS, but the approach taken to carry out the work was also an MBSE approach.

The resultant approach comprises and ontology, a framework and a set of processes.

The approach has been tested on a test model and will now be applied to multiple case studies as part of the overall COMPASS project.

References

[1] INCOSE, 'INCOSE Systems Engineering Handbook', *INCOSE Publishing*, 2007

[2] JD Holt, SA Perry and MJ Brownsword, 'Model-based Requirements Engineering', *IET Publishing*, 2011

[3] 'CMMI for Development. Version 1.3' *CMMI-DEV (Version 1.3, November 2010)*. Carnegie Mellon University Software Engineering Institute. 2010

[4] 'CMMI for Acquisition. Version 1.3' *CMMI-ACQ (Version 1.3, November 2010)*. Carnegie Mellon University Software Engineering Institute. 2010.

[5] 'CMMI for Services. Version 1.3' *CMMI-SVC (Version 1.3, November 2010)*. Carnegie Mellon University Software Engineering Institute. 2010.

[4] 'ISO 15288 – Software and Systems Engineering Life Cycle Processes', *ISO Publishing*, 2009

[6] 'Systems Engineering Guide for Systems of Systems' *Office of the Under Secretary of Defense*, USA DoD, August 2008

[9] GA Lewis, E Morris, P Place, S Simanta, DB Smith, 'Requirements Engineering for Systems of Systems', *IEEE SySCon 2009 – 3rd Annual IEEE International Systems Conference*, March 2009

[10] 'Comprehensive Modelling for Advanced Systems of Systems' Annex 1 – Description of Work' Project Number: 287829, 2011

[11] JA Estefan, 'Survey of Model-Based Systems Engineering (MBSE) Methodologies', *INCOSE MBSE Focus Group*, 2007

[12] JD Holt, 'A Pragmatic Guide to Business Process Modelling – Second Edition', BCS Publishing, 2009.

[13] CE Dickerson & DN Mavris, 'Architecture and Principles of Systems Engineering', *CRC Press*, 2009

[14] JD Holt & SA Perry, 'Modelling Enterprise Architectures' *IET Publishing*, 2010.

[15] 'Initial report on SoS architectural models' COMPASS (287829) , document D22.1, March 2012

[15] 'LESLP major Incident Procedure Manual – seventh edition', *The Stationery Office*, 2007

[16] JD Holt, SA Perry & MJ Brownsword, 'Context-based Systems Engineering', SOSE 2010, 5th International IEEE Conference on Systems of Systems, June 2010

[17] COMPASS project web site: http://www.compass-research.eu/ - accessed 19th March 2012

Interface Specification for System-of-Systems Architectures

Richard Payne, Jeremy Bryans, John Fitzgerald, Steve Riddle
School of Computer Science
Newcastle University
Newcastle upon Tyne
{richard.payne, jeremy.bryans, john.fitzgerald, steve.riddle}@ncl.ac.uk

***Abstract*—Establishing that a system-of-systems (SoS) architecture respects global SoS-level properties is complex. Recording explicit technical interfaces at the boundaries of constituent systems would facilitate this, but support for the description of such interfaces is limited in current widely-used architectural notations. This paper identifies research challenges that arise from using the combination of SysML and the formal notation VDM to describe the interface specifications recorded at the boundaries of the constituent systems. The approach is illustrated with a case study based on an emergency services SoS.**

I. Introduction

The characteristics of systems-of-systems (SoSs) identified by Maier [15] make SoS engineering a significant challenge. At the same time, the complexity of SoSs and the reliance placed on them is increasing. Methods and tools for SoS engineering are therefore required in order to permit the SoS engineer to validate global SoS properties during design and evolution. An important area of potential benefit is the description and analysis of SoS architectures.

Although SoS engineering offers unique challenges, we believe it to be beneficial to adapt current methods and tools as starting points [6] rather than supplanting current best practice. In our view, a promising method for SoS architectural definition lies in interface specification, because it allows the internal definition of constituent systems to change, as long as it continues to respect the contractual interface specification. In developing techniques for modelling SoS architectures we therefore start from the established notation SysML [19]. While designed for systems engineering, several features of SysML mean that it can be used for description of SoS architectures. However, like many architectural description languages, support for the formal specification of the interfaces between the constituent systems is limited, and this in turn limits the extent to which automated or partially automated tools can be used to analyse semantically significant properties of models.

A possible approach to improving the level of formality is to use an existing formal specification language in combination with SysML, allowing tools and methods developed for the formalism to be used alongside those of SysML. Formal specification languages have a mathematically well-founded and precisely defined semantics. They have associated techniques that allow desirable properties of a system to be specified and demonstrated to a high degree of rigour. A suitable language is VDM [9], a formal model-based specification language. The analysis techniques for VDM include static analysis by syntax-checking, type-checking and proof, and dynamic analysis through testing and simulation. These analyses can increase the confidence that a SoS engineer has on the correctness of global SoS properties.

This paper identifies research challenges that arise from using SysML in combination with VDM to describe the interfaces of the constituent systems of SoSs. We illustrate these using a case study from the emergency response domain. We begin by summarising the current state of the art in interface contract specification in Section II, then introduce a case study based on an emergency response scenario in Section III with the SoS architecture defined in SysML, identifying some of the interfaces in the study and the specification of interfaces in VDM. Section IV considers the design and analysis of these interfaces using VDM. Section V concludes.

II. Interface Specification for Architectural Modelling

A. State of the Art in Interface Specification in Architectural Modelling

The state of the art in interface specification in architectural description notations is poor [23]. The most widely used notations (UML [18] and SysML [19]) allow basic signatures to be defined and pre- and postconditions to be specified textually, but these are rarely used. In AADL [8] models are defined in terms of component types and implementations which include subprograms (similar to operations) with basic signatures, though pre- and postconditions are not available. Formal architectural notations such as Darwin [14] and Wright [2] allow software components to have ports defined, and (in the case of Wright) have their message exchange protocols defined. However, these notations do not include the ability to specify other details such as operation pre/postconditions, and they do not contain architectural abstractions suitable for the definition of SoSs.

B. Interface Contracts and Design by Contract

Interface contracts are descriptions of the constituent systems of a SoS described *contractually* in terms of their expectations and the obligations placed on their behaviour.

They have much in common with the idea of Design by Contract, a software engineering technique introduced by Meyer [16] in which contracts make explicit the relationships between systems in terms of preconditions and postconditions on operations and invariants on states. In Meyer's approach, contracts mainly specify functionality. The interaction between operations can be described using notations such as UML sequence diagrams or in process algebraic notation such as CSP [11] or CCS [17].

The use of contracts in service selection and subscription is an active research field in service-oriented computing, in particular the use of contracts for the specification of non-functional properties. Beugnard et al. [3] expand the notion of a contract to architectures in which components provide services. A four-level structure for contracts is proposed, adding scheduling of component interaction and message passing as well as non-functional aspects of operations. Contracts are subscribed to prior to service invocation, after a period of negotiation.

C. *Potential Benefits of Interface Contracts*

The incorporation of interface contracts in architectural specifications may provide two main benefits:

- Interface contracts defined for the constituent systems of a SoS architectural design permit the analysis of SoS-level properties. These analyses give SoS designers the ability to experiment with consequences of different architectural designs.
- SoS designers can define the expected interfaces of the constituent systems, and these definitions may be provided to the system developers. This provides greater confidence to SoS designers that constituent systems will adhere to the expected properties on interfaces.

D. *An Approach to Interface Contract Specification*

In this paper we consider interface specification defined contractually for SoS architectural modelling. Based on the existing state of the art in architectural modelling, identified in Section II-A, current notations are limited for specifying interface contracts. As highlighted in Section I, however, we believe it to be beneficial to adapt existing notations from systems engineering practice as starting points. As such, we propose the use of the SysML, widely-used in industry, to define SoS architectures, along with the use of VDM to formally define interface contracts identified in SysML models. The contribution of the paper is to identify the research challenges that arise from using the combination of SysML with VDM to describe the interfaces of existing constituent systems of SoSs.

III. LESLP Case Study

In order to explore the consequences of formal interface specification in a contractual style, we use a case study based on the system formed by emergency services (fire, police, ambulance etc.) in response to a major incident. We refer to this system as the Major Incident Response (MIR). The MIR is a SoS in Maier's terms [15] and may be considered an *acknowledged* SoS [7]. The emergency services are operationally and managerially independent. Each service is itself geographically distributed and may evolve, for example as personnel come on and off duty during the course of a long-running incident. Emergent behaviour is also present – the comprehensive approach to management of the incident relies on voluntary and collaborative interaction. We give more details on the case study in Section III-A, and an architectural description in Section III-B. Section III-C supplements this description with a formal definition of the interfaces.

A. *Informal Description of the Case Study*

We base the study on the procedures for the coordination of the MIR in London, as outlined in the Major Incident Procedure Manual [20] published by the London Emergency Services Liaison Panel (LESLP). This documents the process for identifying a major incident, the initial information to be passed to the appropriate services and the roles and responsibilities of the service members at the scene.

The response to all major incidents follows a broadly similar structure. The members from each service attending the scene form Bronze command. For more severe incidents, a Silver command will be formed containing representatives of all the involved services. For long-running incidents, a Gold command may be formed at a geographically distant point. The Bronze, Silver, Gold hierarchy corresponds to the operational, tactical and strategic levels of command.

We pay particular attention to the rules outlined in [20] for communication of casualty information with the media, and the requirements these place on the interfaces between the emergency services. In the early stages of a major incident, confusion can arise if the media aggregate casualty figures from various sources. This can lead to "double-counting" and overestimation of the severity of incident. To avoid this, all casualty details must be given to Gold command, which is then responsible for coming to a more reliable estimate and communicating this to the media.

In [5] the case study was explored using the Event-B [1] formalism, but the model developed there does not provide an accessible representation of the SoS architecture, and does not consider the interface specification between the constituent systems. In this paper we focus on these aspects.

B. *Architectural Description of LESLP*

SysML [19] is a profile for UML 2.0, developed for system engineering, but also supporting the modelling of SoS architecture. It enjoys wide industrial support and a sound tool base. SysML provides several diagram types, with a "precise natural language" semantics, to support the description of the SoS structure, behaviour and requirements.

The MIR structure of the case study is given in Figure 1 as a SysML Block Definition Diagram (BDD). The MIR contains up to three emergency services and these are the constituent systems of the MIR SoS. The emergency services (ES) may be a police force, a fire brigade or ambulance service. All

ES contain one or more person and each person has a role (Bronze, Silver or Gold).

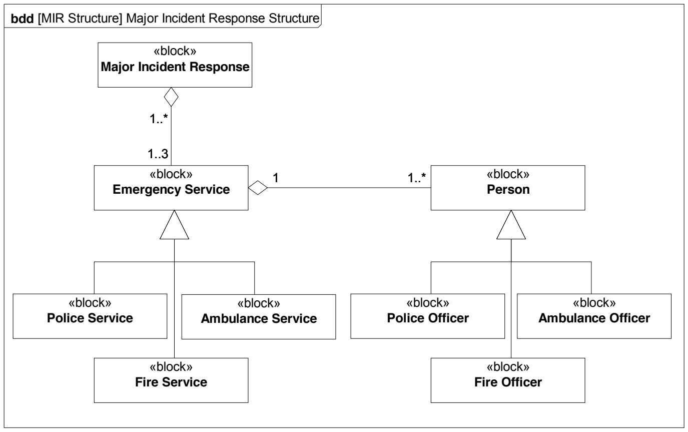


Fig. 1. Block definition diagram depicting Major Incident Response structure.

We consider one requirement of the MIR SoS, that only accurate casualty information should be released to the media, and show in detail how the constituent systems communicate[1]. These communications between instances of the constituent systems with a given role are described in an Internal Block Diagram (IBD) given in Figure 2. The IBD details the provided and required interfaces of the systems of the MIR SoS and depends upon the roles undertaken by the constituents. For example, in the IBD of Figure 2, an interface *info_to_silver* exists between Officers with a Bronze role and their respective emergency service Officers with Silver role.

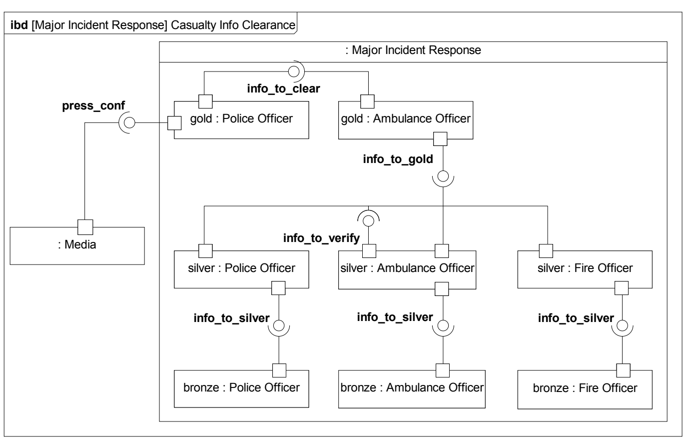


Fig. 2. Internal Block Diagram showing relationships of response constituents when releasing casualty figures.

The SysML interface definitions in Figure 3 relate to points of interaction for those operations made public by the relevant constituent systems. The operations are defined in terms of operation signatures detailing data input to and output from an operation call. For example, the *info_to_silver* interface, relating to the transferring of casualty information from Bronze officers to Silver Ambulance officers, consists of the operation *verifyCasualtyDetails(CasualtyDetails)* which requires some unverified casualty information of the *CasualtyDetails* data type (defined elsewhere in the model).

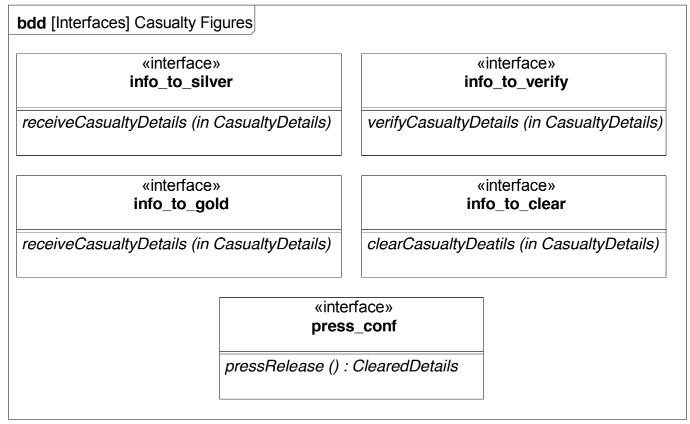


Fig. 3. Block Definition Diagram showing interface definitions when releasing casualty figures.

The SysML specification allows pre- and postconditions to be specified for operations, however this is optional and no analyses are available to ensure their correctness. SysML also omits the Protocol State Machine of UML 2.0 which dictates the response of an interface to specified sequences of events, constraining the order of operations. It is our opinion that this construct would add additional rigour to an interface specification and increase the range of analyses available.

C. Formal Definition of LESLP Interface Contracts

Given a SysML architectural specification, we strengthen the interface specifications of the constituent systems as described above. Using a formal model-based specification language allows increased confidence using a range of analysis techniques. In this paper, VDM is used to give formal definitions of the architectural interfaces. VDM is a model-oriented notation supporting descriptions of data and functionality. VDM has industry-strength[2] and open-source[3] tool support.

The SysML interfaces in Figure 3 may be modelled as VDM classes. VDM supports inheritance, so we define the operations of interfaces with pre- and postconditions, and provide no operation body at the interface. The operation definition, therefore, does not state *what* is done, the implementation of the algorithm must be provided by the constituent systems implementing the interface. This is denoted using the ***is subclass responsibility*** key phrase in the operation body. The system designer must give a specification for each constituent system which provides the interface, and ensure that the implementation of the constituent system contains

[1]We consider only the communications and interactions that are necessary to meet this requirement.

[2]VDMTools: http://www.vdmtools.jp/en

[3]Overture: http://www.overturetool.org

the interface operations, and that they are consistent with the interface specification.

Figures 4a) and 4b) define interface definitions of the *info_to_clear* and *press_conf* interfaces initially given in Figure 3. The ***operations*** keyword in each class denotes the available operations of the interfaces with signatures corresponding to those defined in Figure 3. Each operation is strengthened using pre- and postconditions. The *clearCasualtyDetails* operation of the *info_to_clear* interface, shown in Figure 4a), requires a parameter *cd*, of type *CasualtyDetails*. The operation does not return any result. As mentioned above, no operation body is given. The operation precondition, denoted by the keyword ***pre***, states that there must be 0 or greater reported casualties (*cd.number* >= *0*), they must have been verified (*cd.verified*), and are not already cleared for release (***not*** *cd.cleared*). No postcondition is provided.

The *press_conf* interface in Figure 4b) contains one operation, *pressRelease*. The operation has no parameters and returns a result of type *ClearedDetails*. This interface operation also has no body. The operation has no precondition, but has a postcondition, which is given after the ***post*** keyword. We may refer to the return variable using the ***RESULT*** keyword, The postcondition requires that the result has been verified and cleared (***RESULT***.*verified and* ***RESULT***.*cleared*).

```
class info_to_clear

  operations
  public clearCasualtyDetails: CasualtyDetails ==> ()
  clearCasualtyDetails(cd) ==
    is subclass responsibility
  pre cd.number >= 0 and cd.verified and
      not cd.cleared

end info_to_clear
```

(a) *info_to_clear* interface.

```
class press_conf

  operations
  public pressRelease: () ==> ClearedDetails
  pressRelease () ==
    is subclass responsibility
  post RESULT.verified and RESULT.cleared;

end press_conf
```

(b) *press_conf* interface.

Fig. 4. VDM representations of interface definitions.

Given these interface specifications, VDM classes are defined which implement the various interface definitions to correspond with the IBD in Figure 2. These classes are implemented using the ***is subclass of*** key phrase followed by the interface class name. The classes (for example Gold Police) also implement the *Police* abstract class detailing emergency service-specific variables.

The Gold Police system is defined as a VDM class below in Figures 5, 6 and 7. Figure 5 shows the Gold Police class which implements the *info_to_clear* and *press_conf* interfaces, defined on the first two lines. The class has a single instance variable, *cas_cleared*, a set of type *CasualtyDetails*, initially set to the empty set {}.

```
class Gold_Police is subclass of Police,
               info_to_clear, press_conf

 instance variables
 private cas_cleared : set of CasualtyDetails := {};
 ...
end Gold_Police
```

Fig. 5. VDM class representing *Gold Police* constituent system: preamble and instance variables.

In Figure 6, the Gold Police class implements the *clearCasualtyDetails* operation, as defined in the *info_to_clear* interface. The operation strengthens the precondition to ensure that the police officer has the rank Gold (*role* = *<Gold>*) and the remainder of the precondition is unchanged. The operation postcondition ensures that the number of *CasualtyDetails* items in the *cas_cleared* instance variable does not decrease (***card*** *cas_cleared* >= ***card*** *cas_cleared˜*)[4]. The postcondition also states that, for all *CasualtyDetails* items in the *cas_cleared* set which have the same location as the parameter *cd*, those details should be verified and cleared (***forall*** *c* ***in set*** *cas_cleared* & *c.loc* = *cd.loc* => *c.verified and c.cleared*).

```
class Gold_Police
 ...
 operations
 public clearCasualtyDetails : CasualtyDetails ==> ()
 clearCasualtyDetails(cd) ==
 (
  if not exists c in set cas_cleared & c.loc = cd.loc
  then cas_cleared := cas_cleared union
        {mk_CasualtyDetails(cd.number, true,
                             true, cd.type, cd.loc)}
 )
 pre role = <Gold> and cd.number >= 0 and
     cd.verified and not cd.cleared
 post card cas_cleared >= card cas_cleared~ and
      forall c in set cas_cleared & c.loc = cd.loc =>
                             c.verified and c.cleared;
 ...
end Gold_Police
```

Fig. 6. VDM class representing *Gold Police* constituent system: clearCasualtyDetails operation.

The operation body of the *clearCasualtyDetails* operation, given after the operation signature, has an *if* statement as the main structure. If there does not exist any casualty details in the *cas_cleared* set with the same location as the parameter *cd* (***if not exists*** *c* ***in set*** *cas_cleared* & *c.loc* = *cd.loc* ***then***), then a new *CasualtyDetails* item is added to the *cas_cleared* set

[4]In VDM postconditions, a variable name succeeded by the ˜ symbol refers to the initial value of that variable.

with the *verified* and *cleared* fields set to ***true***, all other fields are given as those supplied by the parameter, *cd* (*cas_cleared := cas_cleared* ***union*** {*mk_CasualtyDetails(cd.number,* ***true***, ***true***, *cd.type, cd.loc)*}).

The Gold Police class also implements the *press_conf* interface and so, in Figure 7, the *pressRelease* operation is given. The *pressRelease* operation in Figure 7 has a strengthened precondition, as with the *clearCasualtyDetails* operation, to ensure that police officers carrying out the operation have the rank <*Gold*>. The postcondition is also strengthened to ensure that the number of casualty details released is less than, or equal to, the number of cleared details the Gold Police know about. The operation uses a private function, *totalCleared*, to calculate the total cleared figure (***RESULT****.number <= totalCleared(cas_cleared)*). This is an important property of the casualty clearance scenario – that the Major Incident Response command do not release casualty figures exceeding the number of casualties that have been verified and cleared for release. The Gold Police are able to use their discretion in releasing a lower figure than is known. The *pressRelease* operation body passes the *cas_cleared* instance variable to a private *clearForPress* function, the result of which is returned (***return*** *clearForPress(cas_cleared)*). The private *clearForPress* and *totalCleared* functions are given in Figure 7. The *clearForPress* operation body is undefined using the ***is not yet specified*** key phrase – a policy decision for the Gold Police, not given here. Finally, the *totalCleared* function is a simple recursive function to count the number of casualties given a *CasualtyDetails* set.

```
class Gold_Police
 ...
 operations
 public pressRelease : () ==> ClearedDetails
 pressRelease () ==
 (
  return clearForPress(cas_cleared);
 )
 pre role = <Gold>
 post RESULT.cleared and RESULT.verified and
      RESULT.number <= totalCleared(cas_cleared);

 functions
 private clearForPress : set of CasualtyDetails
                                   ==> ClearedDetails
 clearForPress(cds) ==
  is not yet specified;

 private totalCleared : set of CasualtyDetails -> nat
 totalCleared(cds) ==
   cases cds:
     {} -> 0,
     others -> let cd in set cds in
                  cd.number + totalCleared(cds\{cd})
               end;

end Gold_Police
```

Fig. 7. VDM class representing *Gold Police* constituent system: pressRelease operation and auxiliary private functions.

The remainder of the VDM model of the Major Incident Response is omitted from this paper: it can be found in full in [22].

IV. Analysis of Interface Contracts

The purpose of a formal analysis is to confirm or refute specified properties that are required of the model under consideration. The analysis techniques available for VDM include static and dynamic techniques, with varying degrees of machine support. VDM's particular history of use in industry settings requiring extensive test-based validation of models and model-based testing of implementations mean that its tool sets have highly developed interpreters allowing rapid testing of models on high volumes of test cases [13]. Simulation is closely linked to testing. Rather than executing a single well-defined test, the model execution is driven by a *scenario* containing multiple decision points that may resolved by a user interacting with the model. This allows those not experienced in the notation or involved in the development to gain familiarity with the formal model, and provides a valuable way of exposing application domain experts to the model at an early stage.

VDM has a well-defined formal semantics, and therefore VDM models are amenable to logical proof [4]. Proof obligations arise naturally within a model, for example the obligation to prove that the specification of each operation is *satisfiable*, i.e., for all valid pre-states and inputs there is always a state of the model that satisfies the postcondition of the operation. Overture generates proof obligations automatically, and manages their manual "sign off" by the user, but currently little help is given to the user in automatically discharging them.

Testing, simulation and proof both have a contribution to make in increasing our assurance of the design of the MIR SoS. For example, both proof and unit testing of the operations would be valuable to ensure that explicit definitions for operations meet the implicit (pre- and postcondition) specifications. Proof is an expensive technique, and best applied only to a small number of key properties, but it provides a higher degree of confidence in a system. The MIR SoS was designed to ensure that only verified casualty figures were released to the media. Some properties relating to this purpose are suggested in [21], for example that *only Police Gold is authorised to release casualty figures*, or that *Unverified casualty figures must never be released to the general public.* Demonstrating properties such as these would be an appropriate application of proof technology.

V. Conclusion

From the work documented in this paper, we conclude certain requirements on the notation used for the specification of interfaces between constituent systems. Significant requirements include *the necessity of using a formal notation which includes architectural abstractions* and *the ability to describe the accepted orderings of events at the interface*. We also observe *the necessity to provide strong ties to an accepted* *industrial strength architectural description language* and *the ability to deal with different levels of abstraction*.

Whilst SysML enables SoS engineers to model complex SoS architectures, the facilities for interface definition are less satisfactory and little analysis is possible. Using a formal notation enables these analyses. However, existing formal notations with architectural abstractions are poor. We propose the use of formal specification notations, such as VDM, to define interfaces and concrete system specifications and to reason about their properties. In the paper, we demonstrate the use of VDM by defining interfaces corresponding to the SysML architectural model, and further specify systems implementing the interfaces. Whilst not performed in this paper, VDM allows strong static analysis support and dynamic support in the form of simulation, testing and proof obligation generation, in particular allowing the specification of interfaces at different levels of abstraction.

Current VDM tools lack model checking and proof support. Further, although it supports data-based specification of functionality, VDM does not contain abstractions to support description of event *orderings* at interfaces. Existing notations which could fill these gaps in analyses include the family of process algebras, e.g. CSP. This would allow the definition of protocols on interfaces and provide dynamic analysis through model checking, which could be used to avoid protocol mismatch [10]. An optimal approach however, would be to use a notation that provides data-based modelling (as in VDM) *and* event ordering (such as CSP) and which also contains the abstractions necessary to model SoS architectures and state and reason over global SoS properties.

The development of such a notation is the goal of our current work in the COMPASS project[5]. This project aims to improve the state of the art in SoS engineering by provision of modelling tools and analysis techniques based on an underlying modelling language (the COMPASS Modelling Langage, CML). CML provides VDM-style data modelling and CSP-style event ordering as outlined above, for representing SoS architectures and interface contracts. This language will have a formal semantic definition to support description of behaviour and composition of subsystem properties, based on Hoare and He's Unifying Theories of Programming (UTP) [12] and will integrate with SysML to support modelling using either textual CML, graphical SysML or a combination of the two.

Acknowledgments

The authors' work is supported by the EU FP7 Project COMPASS (http://www.compass-research.eu), the UK EPSRC Project Trusting Dynamic Coalitions, and the EPSRC Platform Grant on Trustworthy Ambient Systems (TrAmS-2).

References

[1] Jean-Raymond Abrial. *Modeling in Event-B: System and Software Engineering*. Cambridge University Press, 2010.

[2] Robert J. Allen. *A Formal Approach to Software Architecture*. PhD thesis, School of Computer Science, Carnegie Mellon University, May 1997.

[3] Antoine Beugnard, Jean-Marc Jézéquel, Noel Plouzeau, and Damien Watkins. Making components contract aware. *Computer*, 32(7):38–45, July 1999.

[4] Juan Bicarregui, John Fitzgerald, Peter Lindsay, Richard Moore, and Brian Ritchie. *Proof in VDM: A Practitioner's Guide*. FACIT. Springer-Verlag, 1994. ISBN 3-540-19813-X.

[5] Jeremy W. Bryans, John S. Fitzgerald, and Tom McCutcheon. Refinement-based techniques in the analysis of information flow policies for dynamic virtual organisations. In *Adaptation and Value Creating Collaborative Networks - 12th IFIP WG 5.5 Working Conference on Virtual Enterprises, PRO-VE 2011. Proceedings*, number 362 in IFIP Advances in Information and Communication Technology, pages 314–321, 2011.

[6] Radu Calinescu and Marta Kwiatkowska. Software Engineering Techniques for the Development of Systems of Systems. In *Foundations of Computer Software. Future Trends and Techniques for Development*, volume 6028 of *Lecture Notes in Computer Science*, pages 59–82. Springer Berlin/Heidelberg, 2010.

[7] Judith S. Dahmann, George Rebovich Jr., and Jo Ann Lane. Systems engineering for capabilities. *CrossTalk Journal (The Journal of Defense Software Engineering)*, 21(11):4–9, November 2008.

[8] Peter Feiler, David Gluch, and John Hudak. The architecture analysis and design language (AADL): An introduction. Technical Report CMU/SEI-2006-TN-011, Software Engineering Institute, 2006.

[9] John Fitzgerald, Peter Gorm Larsen, Paul Mukherjee, Nico Plat, and Marcel Verhoef. *Validated Designs For Object-oriented Systems*. Springer-Verlag TELOS, Santa Clara, CA, USA, 2005.

[10] Carl Gamble. *Design Time Detection Of Architectural Mismatches In Service Oriented Architectures*. PhD thesis, School of Computing Science, Newcastle University, UK, 2011.

[11] Tony Hoare. *Communication Sequential Processes*. Prentice-Hall International, Englewood Cliffs, New Jersey 07632, 1985.

[12] Tony Hoare and Hi Jifeng. *Unifying Theories of Programming*. Prentice Hall, April 1998.

[13] T. Kurita and Y. Nakatsugawa. The Application of VDM++ to the Development of Firmware for a Smart Card IC Chip. *Intl. Journal of Software and Informatics*, 3(2-3), October 2009.

[14] Jeff Magee, Naranker Dulay, Susan Eisenbach, and Jeff Kramer. Specifying distributed software architectures. In *Proceedings of the 5th European Software Engineering Conference*, pages 137–153. Springer-Verlag, 1995.

[15] Mark W. Maier. Architecting Principles for Systems-of-Systems. *Systems Engineering*, 1(4), 1998.

[16] Bertrand Meyer. *Object-Oriented Software Construction (2nd ed.)*. Prentice-Hall, 1988.

[17] R. Milner. *Communication and Concurrency*. Prentice Hall International (U.K.) Limited, 1989.

[18] Object Management Group OMG. Unified Modelling Language infrastructure version 2.2. http://www.omg.org/spec/UML/2.2/Infrastructure, February 2009.

[19] Object Management Group OMG. Systems Modelling Language version 1.2. http://www.omg.org/spec/SysML/1.2, June 2010.

[20] London Emergency Services Liaison Panel. *Major Incident Procedure Manual*, seven edition, 2007. Available at www.met.police.uk/leslp.

[21] Richard J. Payne and Jeremy Bryans. Modelling the Major Incident Procedure Manual: A System of Systems Case Study. Technical Report CS-TR-1320, School of Computing Science, Newcastle University, March 2012.

[22] Richard J. Payne, Jeremy Bryans, John Fitzgerald, and Steve Riddle. Interface Specification for System-of-Systems Architectures. Technical Report CS-TR-1335, School of Computing Science, Newcastle University, May 2012.

[23] Richard J. Payne and John S. Fitzgerald. Evaluation of Architectural Frameworks Supporting Contract-based Specification. Technical Report CS-TR-1233, School of Computing Science, Newcastle University, December 2010.

[5] www.compass-research.eu

Technical Challenges of SoS Requirements Engineering

Stefan Hallerstede[1], Finn Overgaard Hansen[1], Jon Holt[2], Rasmus Lauritsen[3],
Lasse Lorenzen[2] and Jan Peleska[4]
[1] Aarhus University School of Engineering, Denmark, [2] Atego, UK
[3] Bang & Olufsen, Denmark, [4] University of Bremen, Germany

Abstract – *Taken by themselves separate aspects of systems of systems (SoS) can be addressed by conventional system engineering techniques. That is, at least to a large degree, we know how to address the problems of distribution, emergence and evolution. The specific challenges posed by SoS arise from their combination. Additionally, we have to deal with independence of constituent systems (CS) of SoS, in particular, managerial independence. In this article we focus on technical challenges of mastering SoS requirements. Based on techniques for systems engineering we sketch problems that appear specifically in SoS engineering if we want to be able to use conventional engineering techniques as much as possible. The ultimate aim of our work is to develop tools that can support SoS requirements engineering.*

1 Introduction

The main objective of our work is an integrated approach to requirements engineering of SoS that comprises modeling and validation of requirements as well as verification of design and implementation relative to requirements models. Within our current research we do not address the problem of requirements elicitation of SoS. This research is carried as part of the EU FP7 project COMPASS, addressing the modeling of SoS by formal and informal means. Aspects such as requirements elicitation are investigated separately within the COMPASS project.

The purpose of this article is to provide an overview of the technical challenges we face and their intimate relationships. For example, dealing with requirements validation and verification in the sense above while allowing an SoS to evolve continually demands strong support for requirements tracing.

We consider the following widely recognized characteristics of SoS engineering [1] the main source of the technical challenges we face:

i. *Independence.* The component systems are able to operate and are managed separately.
ii. *Distribution.* The component systems are dispersed and communicate over larger distances.
iii. *Emergence.* The behavior of an SoS exceeds the behavior of its constituent systems (CS).
iv. *Evolution.* The SoS is in continual development and can never be considered fully completed.

Some of the characteristics of SoS engineering pertain generally to systems engineering but are more pressing in the context of SoS, e.g., evolution.

These characteristics are present to varying degrees in the different types of SoS [2]. For our current work we assume that an SoS falls into one of the two following categories:

i. *Acknowledged.* It is under control of one authority.
ii. *Collaborative.* It is *not* under control of one authority.

The industrial case studies described in Sections 2 and 3 present a case of each type. In Section 4 we sketch a requirements engineering process for SoS. Specific technical challenges are discussed in subsequent sections: validation in Section 5, tracing in Section 6 and verification in Section 7. Finally, section 8 draws some conclusions. Sections 2 to 7 have been structured into subsections dealing with *independence*, *distribution*, *emergence* and *evolution* in turn.

2 Acknowledged Case

A case study of an acknowledged SoS is brought forward by the Italian company Insiel. After describing this system in general terms we analyze the case study with respect to the SoS characteristics.

This case study deals with a new *unified emergency call-center* SoS that is to operate in the Fruili Venezia Giulia region of North Italy. One key service of the center to the general public is responding and coordinating efforts to handle emergencies. A similar SoS for the services of the London Emergency Services Liaison Panel (LESLP) has been studied in [3]. The Insiel system offers two views of the current status of the SoS to the call operators. One view permits call operators to get an overview of incoming calls. The other view provides the call operator with a map of the region that overlay on-going emergencies with deployed emergency response units. The system has a room where several call operators are situated. Each operator uses computer equipment and radio to communicate with the system. The system allows an operator to handle and group calls and to dispatch emergency response vehicles to an event. In a major crisis situation many incoming calls regarding the same crisis will appear and operators will have means to group such calls together. An additional set of wall mounted monitors visible to all operators in the call-center room are available to provide a common operational view in such situations.

The principal stakeholders for the system include:

- The Italian government.
- Health care services.
- Police and Fire brigades.
- Insiel.
- The general public at large in case of emergencies.

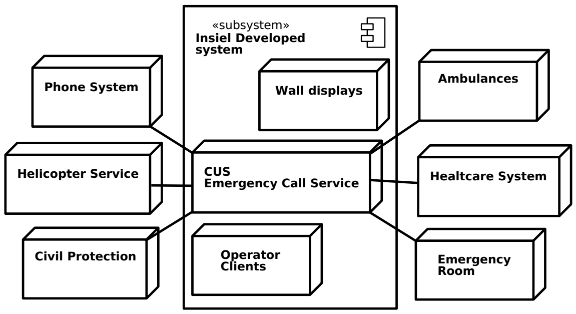


Figure 1: Overview of the Insiel CUS System of systems.

The CS developed by Insiel is communicating with several other CS as depicted in Fig. 1. The most prominent ones are:

- The Phone System.
- The Helicopter Service for mountain and forest rescues.
- The Civil Protection authorities.
- The Healthcare System.
- The Ambulances.

Based on the preceding general description of the case study, we analyse its SoS characteristics. The SoS is ordered by the Italian government, which is considered its single responsible authority. However, on the managerial level stakeholders owning the constituent systems clearly maintain their independence.

2.1 Independence

Accounting for independence of the constituent system is straightforward. As an example consider the phone system. It is operationally and managerially independent. A private phone company independent of the Italian government and Insiel is responsible for it.

2.2 Emergence

Emergent behavior is clearly identifiable. For instance, the overview of the region provided to the operators along with a view of all allocated emergency response resources cannot be provided by the constituent systems alone.

2.3 Evolution

The emergency response service is intended to have a long lifespan. Therefore this system will inevitably evolve on the basis of technological progress alone. One also expects that within its lifetime laws and regulations concerning its operation are going to change. As an example of technological change take the advent of the GPS that made navigational systems omnipresent. Similar technological advances in the future will place new demands on the capabilities of SoS in operation.

3 Collaborative Case study

A collaborative SoS case study is represented by B&O in form of the *connected Audio-Video (AV) products challenge*.

The AV range in the 1980ies was not as diverse as it is today. At that time B&O could deliver a complete range of AV equipment using proprietary technology to connect the equipment. Offering a complete range of products, B&O did not consider to make the proprietary technology publicly available but kept it private to be used only with B&O products and a few select Home-Automation (HA) installers.

Today the trend is towards more open technologies with interoperability as a major concern. The range of products in the AV domain today is much more diverse and complex than in the 1980ies making it expensive and difficult to deliver quality products for the entire range. It is no longer feasible to provide all connected AV products in a home installation from one company.

The modern AV system is no longer limited to just one supplier. Furthermore the border between the AV system and the HA is disappearing, being replaced by the concept of intelligent homes. In future B&O will no longer be able to rely solely on proprietary technology but instead have to follow open standards. Despite of this, even closer integration with HA installers will be necessary. To fulfill the goal of interoperability B&O needs an organisation do drive and specify an interoperability protocol for AV products, so that B&O, the HA installers and other consumer electronics providers have a common baseline for their products. Today the *Digital Living Network Alliance (DLNA)* is starting to become such an organisation, and it provides a protocol [4] that specifies basic interoperability standards.

3.1 Independence

The focus of modern AV systems is to connect stand-alone products into a combined system experience. The products themselves are considered as independent systems and so is their development which takes place at competing electronics providers. In addition, the DLNA organisation—which counts many electronics providers as its members—is developing the protocol and standard for product interoperability. DLNA itself is an independent organisation.

In general B&O has no influence or control over how other products are developed and how the DLNA organisation works with the interoperability protocol.

3.2 Distribution

An AV system is not necessarily situated in the same room or even house, but can be dispersed over several different geographic locations, e.g., summer house and car. However the more common scenario would be different AV products located in one house.

3.3 Emergence

The characteristic features on an AV systems are emergent. A simple example of such a feature is the cooperation of separate audio and video devices that can be controlled by one remote control. The user expects that the devices and the remote control work together without any additional intervention just by placing them close enough to each other.

A more intricate example is dealing with Digital Rights Management (DRM). Today each constituent system will have to comply with DRM regulations. The difficulty with DRM appears when many products are connected. It is necessary to ensure that the AV system still complies with all the constituent systems' DRM regulations. One might even end up in a situation where the constituent systems should be able to form an AV system, but due to legal constraints in the DRM regulations it is not allowed to do this.

3.4 Evolution

AV systems evolve per installation and by way of the underlying technology. AV Systems may simply evolve by adding new products, some of which may result in new emergent properties. If an existing product is added it may also lead to emergent properties such as audio streaming to multiple target devices.

As the AV market is very competitive and in constant evolution, is is evident that the AV system has to follow that evolution. For each new technology the SoS needs to adopt this, but still ensure that the original SoS is not destructed.

3.5 Collaboration

A separate body, the DLNA, drives the interoperability protocol for AV systems. The manufacturers of AV products try to attain collaboratively the minimum interoperability level specified by the DLNA.

4 Requirements Process

4.1 Traditional versus SoS Requirement Engineering

Traditional system engineering requirements processes are defined to support development of complete new systems, where all important requirements are defined up-front, before the system is architected and implemented. Another typical characteristic is that these systems have a single authority or customer, who controls the system development. These processes are not adequate for the development of System of Systems as described in [5] and [6]. As shown in

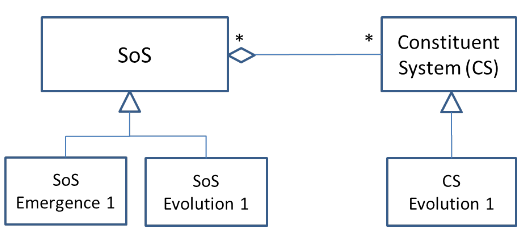


Figure 2: SoS System Model - at an early life-cycle stage

Figure 2, a System of Systems is composed of a number of constituent systems, some of which already exist and have a purpose of their own and are managed by their own authority. Another important aspect to consider is, that a given constituent system in principle can belong to more than one SoS, which could lead to conflicting requirements for the CS.

Figure 2 shows how the concepts of evolution and emergence are related to the SoS and its CS, illustrating that evolution can occur both at the SoS- and the CS-level, whereas emergent behaviour is defined to be only at the SoS-level.

4.2 Independence

In the SoS literature high level system goals are called *system capabilities* which are broken down to more specific requirements for either the SoS as a whole or allocated as more detailed requirements for the constituent systems participating in a given SoS.

Independence means operational as well as managerial independence of the constituent systems. This implies that new system capabilities, requirements and changes shall be dealt with at two levels, the SoS- and the CS-level. In relation to a process for SoS requirement engineering this process should account for this situation, where a given system capability has to be broken down to requirements that belong to either the SoS or to one or more of the constituent systems.

Handling an SoS requirement can be regarded as a top-down process whereas handling a requirement for a constituent system can be regarded as a bottom-up process.

An SoS requirement process has to account for these two different scenarios and offer a different approach for each. When a new capability, a specific requirement or a change request is introduced it should, as one of the first steps, be analyzed and characterized as either belonging to the SoS- or to the CS-level and be handled by the corresponding authority.

4.3 Distribution

An SoS will normally be a distributed system, where the constituent systems are geographically distributed and exchange information based on commonly agreed upon communication protocols. This implies that the stakeholders require a high coordination effort with respect to the development process as well as other engineering aspects such as verification of the complete SoS. Consequently, handling of a distributed stakeholder and user community should be included in a SoS requirement engineering process.

4.4 Emergence

As illustrated in Figure 2, emergence is a characteristic which belongs to the SoS-level and describes behaviour obtained at the SoS-level based on the collaboration of a number of CS to obtain this new behaviour. Somewhat simplified, an emergent SoS requirement is shown as a subtype but can also have other relations to the original SoS. One example could be an emergent SoS which both enhances and changes some of the original SoS behaviour.

4.5 Evolution

As shown in Figure 3 evolution can be introduced at the SoS-level or at the CS-level. Evolution is natural for these long living SoS, where changes can be caused by technological changes, by new or changed user capabilities, or by new legal regulations, introduced, for example, by the government.

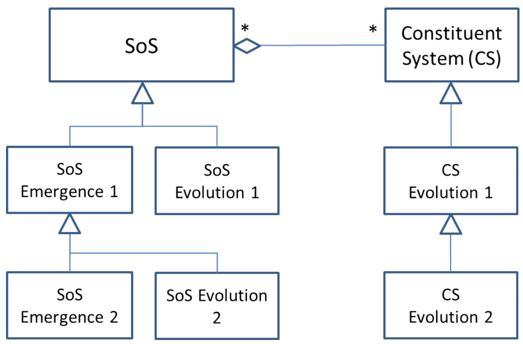


Figure 3: SoS System Model - at a later life-cycle stage

In addition to this an SoS and its constituent systems have long system life times, where each CS can be in a different stage of its individual system life-cycle.

An SoS requirement process should therefore cover a continual development life-cycle scenario, where new capabilities, requirements and changes to existing requirements are to be handled by the process at the SoS- and the CS-level.

4.6 An SoS requirement engineering process proposal

In further work we will fully define the *COMPASS SoS requirements engineering process* proposal based on the process requirements described above and plan to include capability engineering aspects as described in [7], where the authors propose the following two parallel steps: (i) develop SoS capability objectives and (ii) develop a concept of operations to improve the quality of the requirements. Following these two steps a set of high level SoS requirements can be formulated.

Another source to be incorporated in the COMPASS SoS requirements process is the SoS requirement process proposal described in [5] which consists of the following steps:

1. Identify SoS context
2. Identify SoS and individual system goals
3. Understand SoS interactions
4. Identify individual system capabilities and
5. Analyse the gap

We apply the initial proposal for the COMPASS SoS requirement process to the two industrial cases described in Sections 2 and 3.

5 Requirements Modeling

We consider semi-formal approaches, e.g. SysML [8] as well as formal ones, e.g. VDM [9] to requirements modeling and validation. In our experience semi-formal and formal methods can be complementary and sometimes they can only be deployed jointly [10].

The modeling approaches above (and similar ones) have been proved useful for systems modeling and have been used in industry to varying degrees. They assume that the development and maintenance is under control of one authority. Independence is not provided for. They may be able to deal with evolving requirements but there is no explicit support for this. The approaches have not been developed for this purpose. A similar statement can be made about emergence.

5.1 Independence

Dealing with independence requires agreement among the constituent systems about their functionality. Contract-based modeling [11] could be used to enforce agreement. Note that SoS requirements are naturally linked with architectural concerns. It is not possible to delay architectural design that concerns the composition of the CS themselves. This constraint introduces the partitioning of the requirements with respect to emergence. The contracts describe non-emergent properties of CS. Emergent properties are discussed below.

5.2 Distribution

Constituent systems of an SoS often are in *distant locations*. We take distant location to mean that the CS can only communicate by means of some communication network. Because of the resulting architectural concerns we must also deal with communication facilities during requirements modeling. Specification techniques for high-level modeling of communicating systems are known, e.g. [12]. However, specification of communication details –even abstract details– can make requirement models overly complex. Traditionally, one tries to avoid mentioning communication details in requirements for that reason. The partitioning of the requirements into emergent and non-emergent ones could help to identify those requirements that can be modeled on the level of the CS.

5.3 Emergence

We aim at an engineering method for SoS. As such we have a limited view of emergence as properties that exceed those of the CS but are attainable by engineering.[1] This is also referred to as *weak emergence* [13]. A formal engineering-oriented view of weak emergence based on refinement is discussed in [14]. Refinement [15] is used to bridge the abstraction gap between the emergent properties to be verified and the model consisting solely of the CS. Refinement has also been used in [16, 17] to model requirements, and validation of such models has been discussed in [18]. We believe, that some form of refinement will be necessary in order to address the problem of emergence.

5.4 Evolution

System evolution is an issue for any system intended for long-term use. What changes in the context of SoS is that

[1] As opposed to far-reaching philosophical definitions concerning, e.g., consciousness.

evolution is considered the normal case. Any method not taking evolution into account will be of no use. For our purposes we understand by evolution any continued development of an SoS and its CS. This may happen by adding, changing or removing constituent systems, as well as, changing emergent requirements. We observe that we have only little control over what is inside the CS. We know the contracts and can change them or may be forced to change them. As a consequence, emergent properties may fail to hold or new emergent properties may hold. The most interesting problem of evolution of SoS appears to be related to emergence and high-level requirements on CS. In [19] a method and tool for incremental modeling is discussed. It is incremental in two respects: It supports frequent changes to models and it supports a notion of refinement that permits to add detail to a model gradually. We believe, a similar approach could also benefit the modeling of SoS requirements.

6 Requirements Tracing

6.1 Independence

Independence has two different aspects [1]: operational independence and managerial independence. Operational independence imposes architectural constraints on the SoS. In principle, it is known how to deal with this in conventional requirements engineering for systems. Managerial independence imposes that requirements may have to be distributed among the stakeholders that supply constituent systems. As a consequence, tracing requirements to stakeholders becomes a dominant problem. Of course, conventional techniques of tracing to designs and programs still apply. But special focus needs to be placed on impact analysis, in particular, to constraining impact: If requirement changes affect too many suppliers of constituent systems at once, they may be difficult and expensive to implement.

6.2 Distribution

Distribution will not pose new problems that do not already appear in conventional systems engineering. One would expect, however, that there is some overlap between independence and distribution. The architecture of an SoS will usually match with managerial independence. Hence, the SoS architecture may serve as the basis for tracing requirements to different stakeholders. This is more evidence in favor of including architectural concerns early in the requirements modeling as opposed to the practice in traditional requirements engineering.

6.3 Emergence

Emergent properties cannot be expressed on the level of the CS. So we cannot expect to identify emergent requirements with CS. We have argued above that refinement could address emergence in SoS. The employed notion of refinement will have to provide the means to trace requirements. The abstraction gap closed by refinement with respect to property validation must also be closed with respect to requirements tracing. Such a method of refinement with requirements tracing is outlined in [17] for conventional systems engineering. It is based on the WRSPM model [20] for reasoning about system requirements.

6.4 Evolution

Evolution can be split into the two broad categories: CS evolution and SoS evolution. Potentially any change in the configuration of an SoS concerns all stakeholders. In practice, often more than one will be concerned. Changes in the contracts of some CS often affect correctness assumptions made in other constituent systems. Because regular changes are considered to be the normal case, predicting the effect of the changes and determining who is involved is of central importance.

When emergent requirements change, it will be important to recognise as soon as possible who will be involved in addressing the change and in what way. It is possible that, in order to implement the new emergent requirement, contracts have to be changed. It will be necessary to see which parts of the contracts are affected and how. Such information can feed into the effort prediction of the different stakeholders.

7 Requirements Verification

Verification needs to take into account the heterogenous characteristics of an SoS. Neither formal verification nor dynamic testing by themselves are sufficient for verifying SoS on a realistic scale. In particular, the SoS size will nearly always defeat any approach to comprehensively model software and hardware in a formal way. As a consequence, the proper integration of software and hardware has to be investigated by means of testing.

7.1 Independence

Verification has to be coordinated between the different stakeholders. Contracts will form the basis for this. In COMPASS we favor an approach where algorithmic properties of constituent systems should be formally verified, while hardware/software integration is verified by means of testing. On SoS level, constituent system properties are represented by contract abstractions, so that formal verification can be performed again for checking component cooperation logic. Again, system integration aspects have to be verified by means of dynamic testing.

7.2 Distribution

SoS have open architectures where CS may join or leave. On the one hand, this makes testing of the SoS difficult because only specific SoS configurations can be checked. On the other hand, it makes formal verification difficult because the assumptions that can be made about the CS are weak. Combining the two approaches, we expect that the most critical properties can be formally verified while the other behaviour requirements can be checked by dynamic testing. In the former case completeness of the verification is crucial, whereas in the latter we only verify the existence of a suitable implementation.

7.3 Emergence

Emergent properties cannot be verified on the various CS separately. Whether formal verification is used for specific properties or dynamic testing, the effort needs to be coordinated. Both methods should support a notion of refinement that could be used to bridge the abstraction gap between SoS and CS.

7.4 Evolution

Regular changes in the requirements and their models should be automatically taken into account as much as possible. A record about the amount of manually developed tests connected to requirements should be available so that the actual effort of a change can be predicted. To deal with frequent changes tracing of requirements to the tests that verify them is required: failing tests should be related to the requirements that they invalidate.

8 Conclusion

We have described technical aspects of requirements engineering in the context of SoS. In the coming years we intend to fully define and evaluate the development process sketched in Section 4. The evaluation will be based on the two case studies that we have described in Sections 2 and 3.

We have related the main technical activities of the requirements process, validation, tracing and verification to the core characteristics of SoS, namely, independence, distribution, emergence and evolution. This permits us to identify the main technical challenges of SoS requirements engineering. Ultimately, defining the process and addressing these challenges should clarify what kind of tool is needed in order to support SoS requirements engineering. We believe, the task of SoS requirements engineering will benefit from tools for modelling, verifying and simulating scenarios of SoS as does standard requirements engineering.

Acknowledgment

This research is funded by the EU FP7 COMPASS project under grant agreement no.287829.

References

[1] M. W. Maier, “Architecting Principles for Systems-of-Systems,” *Systems Engineering*, vol. 1, no. 4, 1998.

[2] J. S. Dahmann, G. R. Jr., and J. A. Lane., “Systems engineering for capabilities,” *CrossTalk Journal (The Journal of Defense Software Engineering)*, vol. 21, no. 11, pp. 4–9, 2008.

[3] J. W. Bryans, J. S. Fitzgerald, and T. McCutcheon, “Refinement-based techniques in the analysis of information flow policies for dynamic virtual organisations,” in *PRO-VE*, 2011, pp. 314–321.

[4] ipTVnews., “Installed base of dlna devices exceeds 440mn,” January 2011. [Online]. Available: http://archive.iptv-news.com/iptv_news/january_2011_2/installed_base_of_dlna_devices_exceeds_440mn

[5] G. A. Lewis, E. Morris, P. Place, S. Simanta, and D. B. Smith, “Requirements engineering for systems of systems,” in *IEEE SysCon*, 2009, pp. 1–6.

[6] C. Ncube, “On the engineering of systems of systems: Key challenges for the requirements engineering community,” in *RESS*, 2011, pp. 1–4.

[7] J. S. Dahmann, G. R. Jr., J. A. Lane, and R. Lowry., “Systems engineering artifacts for sos,” *”IEEE A&E Systems Magazine”*, pp. 22–28, January 2011.

[8] J. Holt and S. Perry, *SysML for Systems Engineering*. IET, 2008.

[9] J. Fitzgerald and P. G. Larsen, *Modelling Systems – Practical Tools and Techniques in Software Development*, 2nd ed. Cambridge University Press, 2009.

[10] R. Gmehlich, K. Grau, S. Hallerstede, M. Leuschel, F. Lösch, and D. Plagge, “On fitting a formal method into practice,” in *ICFEM’2011*, ser. LNCS, S. Qin and Z. Qiu, Eds., vol. 6991. Springer, 2011, pp. 195–210.

[11] R. Payne and J. Fitzgerald, “Evaluation of architectural frameworks supporting contract-based specification,” University of Newcastle upon Tyne, Computing Science, Tech. Rep., 2010.

[12] J. C. P. Woodcock and A. L. C. Cavalcanti, “A concurrent language for refinement,” in *IWFM’01: 5th Irish Workshop in Formal Methods*, ser. BCS Electronic Workshops in Computing, A. Butterfield and C. Pahl, Eds., 2001.

[13] M. A. Bedau, “Weak emergence,” in *Philosophical Perspectives 11: Mind, Causation, and World*, J. E. Tomberlin, Ed. Blackwell, 1997, pp. 375–399.

[14] J. Sanders and G. Smith, “Emergence and refinement,” *Form. Asp. Comput*, vol. 24, no. 1, pp. 45–65, 2012.

[15] C. C. Morgan, *Programming from Specifications*, 2nd ed. Prentice Hall, 1994.

[16] J.-R. Abrial, *Modeling in Event-B: System and Software Engineering*. Cambridge University Press, 2010.

[17] M. Jastram, S. Hallerstede, and L. Ladenberger, “Mixing formal and informal model elements for tracing requirements,” in *AVOCS 2011*, 2011.

[18] S. Hallerstede, M. Leuschel, and D. Plagge, “Validation of formal models by refinement animation,” *Sci. Comput. Program.*, 2011.

[19] J.-R. Abrial, M. J. Butler, S. Hallerstede, T. S. Hoang, F. Mehta, and L. Voisin, “Rodin: an open toolset for modelling and reasoning in event-b,” *STTT*, vol. 12, no. 6, pp. 447–466, 2010.

[20] C. A. Gunter, E. L. Gunter, M. Jackson, and P. Zave, “A reference model for requirements and specifications,” *IEEE Softw.*, vol. 17, pp. 37–43, 2000.