6th INTERNATIONAL INDUSTRIAL SIMULATION CONFERENCE

2008

ISC'2008

EDITED BY
Joel Colloc
Claude Petit
and
Claude Dussart

JUNE 9-11, 2008
LYON, FRANCE

A Publication of EUROISIS-ETI
6th Industrial Simulation Conference 2008

LYON, FRANCE

JUNE 9-11, 2008

Organised by
ETI- The European Technology Institute
Sponsored by
EUROSIS - The European Simulation Society

Co-Sponsored by

ADICTEL

AEKI

CREAX

ENSAIT

Delft University of Technology

Ghent University

IPK Fraunhofer

P4

Ville de Villeurbanne

UPV

Hosted by

Université de Lyon - CESH

Lyon, France
EXECUTIVE EDITOR

PHILIPPE GERIL
(BELGIUM)

EDITORS

General Conference Chairs
Joel Colloc, CESH, Université de Lyon, Lyon, France
Claude Petit, Université de Lyon, Lyon, France

General Program Chair
Khaled Ghedira, Campus Universitaire, La Manouba, Tunisia

Track Chair Workshop on Modelling and Simulation in the Textile Industry
Vladan Koncar
GEMTEX-ENSAIT
Roubaix, France

INTERNATIONAL PROGRAMME COMMITTEE

Discrete Simulation Methodology, Languages and Tools
Matthias Becker, University Hannover, Hannover, Germany
Carmen Bobeanu, Ghent University, Ghent, Belgium
Juan de Lara, Universidad Autonoma de Madrid, Madrid, Spain
Helge Hagenauer, Universitaet Salzburg, Salzburg, Austria
Sophie Hennequin, ENIM, Metz Cedex, France
Mauro Iacono, SUN II, Universita di Napoli, Naples, Italy
Bjorn Johansson, Chalmers University of Technology, Gothenburg, Sweden
Panagiotis Katsaros, Aristotle University of Thessaloniki, Thessaloniki, Greece
Pierre L’Ecuyer, Universite de Montreal, Montreal, Canada
Jan Lemeire, VUB, Brussels, Belgium
Lenin Lemus, UPV, Valencia, Spain
Goreti Marreiros, Polytechnic of Porto, Porto, Portugal
Maria do Rosário Moreira, University of Porto, Porto, Portugal
Aziz Naamane, ISIS-EPUM, Marseille, France
Jose Neves, Universidade do Minho, Braga, Portugal
Paulo Novais, Universidade do Minho, Braga, Portugal
Werner Pohlmann, Universitaet Salzburg, Salzburg, Austria
Jiri Safarik, University of West Bohemia, Pizen, Czech Republic
Werner Sandmann, University of Bamberg, Bamberg, Germany
Frederic Serin, IUT du Havre, Le Havre, France
Jaroslav Sklenar, University of Malta, Msida, Malta
Giorgio Tonella, Univ. de los Andes, Mérida, Venezuela
Antonella Zanzi, Universita’ degli Studi dell’Insubria, Varese, Italy
INTERNATIONAL PROGRAMME COMMITTEE

Simulation in Manufacturing
Reza Azadegan, Urmia University, Urmia, Iran
Pascal Berret, Université Bretagne Sud, Lorient, France
Peter Byrne, Dublin City University, Dublin, Ireland
Ana Camacho, UNED, Madrid, Spain
Luis Camarinha-Matos, Universidade Nova de Lisboa, Lisbon, Portugal
Eduardo Castellano, IKERLAN Technological Research Centre, Mondragón-Arrasate, Spain
Remy Dupas, Universite Bordeaux 1, Bordeaux, France
Baltazár Frankovic, Slovak Academy of Sciences, Bratislava, Slovakia
Alexander Felfernig, University of Klagenfurt, Klagenfurt, Austria
Michel Gourgand, Universite Blaise Pascal, Clermont-Ferrand, France
Bernard Grabot, LGP-ENIT, TARBES, France
Pouria Homayonifar, Darmstadt University of Technology, Darmstadt, Germany
Christophe Laroque, Universitaet Paderborn, Paderborn, Germany
Habtom Mebrahtu, APU, Chelmsford, United Kingdom
Carlo Meloni, Politecnico di Bari - DEE, Bari, Italy
Pascal Meyer, Forschungszentrum Karlsruhe, Karlsruhe, Germany
Birgit Oberer, TK, Austria, Villach, Austria
Henri Pierreval, IFMA, Aubiere Cedex, France
Marcos A. Rodrigues, Sheffield Hallam University, Sheffield, United Kingdom
Flavio Tonelli, University of Genova, Genova, Italy
Markus Vorderwinkler, PROFACTOR GmbH, Steyr, Austria

Simulation in Steel Manufacturing
Brian Hollocks, Bournemouth University, Bournemouth, United Kingdom

Simulation in Automotive Systems
Naoufel Cheikhrouhou, EPFL, Lausanne, Switzerland
Jesus Felez, UPM, Madrid, Spain
Morteza Montazer, Iran University of Science and Technology, Tehran, Iran
Rob Walker, Anglia Ruskin University, Chelmsford, United Kingdom

Simulation in Robotics
Leopoldo Arnesto Angel, UPV, Valencia, Spain
A. Chatzinkolaou, Athens, Greece
Mohsen Bahrami, Amirkabir University of Technology, Tehran, Iran
Andrzej Dzielinski, Warsaw University of Technology, Warsaw, Poland
Markus Koch, University of Paderborn, Paderborn, Germany
George L. Kovacs, Hungarian Academy of Sciences, Budapest, Hungary
Terje Kristensen, NTNU, Trondheim, Norway
Alexandru Marin, Polytechnical Inst.of Bucharest, Bucharest, Romania
Martin Mellado, UPV, Valencia, Spain
Bogdan Raducanu, Computer Vision Centre, UAB, Barcelona, Spain
Krzysztof Skrzypczak, Silesian University of Technology, Gliwice, Poland
Angel Valera, UPV, Valencia, Spain

Simulation in Electronics, Computer and Telecommunications
Kamel Barkaoui, CEDRIC-CNAM, Paris, France
Dietmar Boenke, Hochschule Reutlingen, Reutlingen, Germany
Christos Bouras, University of Patras, Patras, Greece
Boguslaw Butrylo, Bialystok Technical University, Bialystok, Poland
Christophe Claramunt, Naval Academy Research Institute, Brest, France
Tom Dhaene, University of Antwerp, Antwerp, Belgium
Domenico Giunta, ESA-ESTEC, Noordwijk, The Netherlands
Paulo Maciel, University of Pernambuco, Recife, Brazil
Maurizio Palesi, Universita di Catania, Catania, Italy
Marco Roccetti, University of Bologna, Bologna, Italy
Fernando Ronorat Segui, UPV, Valencia, Spain
Renate Sitte, Griffith University, Gold Coast, Australia
INTERNATIONAL PROGRAMME COMMITTEE

Simulation of Complex Multiprocessor Systems
Orhan Gemikonakli, Middlesex University, London, United Kingdom

Simulation in Logistics, Transport and Harbour Simulation
El-Houssaine Aghezzaf, Ghent University, Ghent, Belgium
Isabel Garcia Gutierrez, Univ. Carlos III de Madrid, Madrid, Spain
Martin Gronow, Berlin University of Technology, Berlin, Germany
Thomas Hanne, University of Applied Sciences Northwestern Switzerland, Olten, Switzerland
Daniel Krajzewicz, German Aerospace Centre, Berlin, Germany
Peter Lawrence, Monash University, Clayton, Australia
Marie-Ange Manier, UTBM, Belfort, France
Roberto Montemanni, IDSIA, Manno-Lugano, Switzerland
Jaap Ottjes, TU Delft, Delft, The Netherlands
Selwyn Piramuthu, University of Florida, Gainesville, USA
Rosaldo Rossetti, University of Porto, Porto, Portugal
Rik van Landeghem, Ghent University, Ghent, Belgium
Hans Veeke, TU Delft, Delft, The Netherlands
Pengjun Zheng, University of Southampton, Southampton, United Kingdom

Hospital Logistics Simulation
Track Chair: Christine Verdier, IMAG, Grenoble, France
Co-Chair: Giorgio Romain-Jacur, University of Padova, Vicenza, Italy
Francois Chapuis, SIS, University Lyon 1, Lyon, France
Claude Dussart, CESH, University Lyon 1, Lyon, France
Jose Machado, University of Minho, Braga, Portugal
Claude Petit, CESH, University Lyon 1, Lyon, France
Peter Summons, University of Newcastle, Australia

Complex Systems Modelling
Ivan Bruha, McMaster University, Canada
Frantisek Capkovic, Slovak Academy of Sciences, Bratislava, Slovak Republic
Gabriella Dellino, University of Bari, Bari, Italy
Werner Dubitzky, University of Ulster, Coleraine, United Kingdom
Alessandro Genco, Universita di Poloromo, Polormo, Italy
Krzysztof Fujarewicz, Silesian Technical University, Gliwice, Poland
Eugene Kindler, Ostrava University, Ostrava, Czech Republic
Marielle Mokhtari, Defence R&D Canada, Valcanter, Canada
Alexandre Nketsa, LAAS, Toulouse, France
Marc Parenthoen, CERV, Plouzane, France
Michal Pechoucek, Czech Technical University, Prague, Czech Republic
Francesco Quaglia, University of Rome I, Rome, Italy
Filippo Queirolo, Comotec, Morez, France
Miguel Rocha, University do Minho, Braga, Portugal
Michael Sobolewski, Texas Tech University, Lubbock, USA
Alfonso Urquia, UNED, Madrid, Spain

Simulation in Aerospace
Wolfgang Kuehn, University of Wuppertal, Wuppertal, Germany
Martin Spieck, DLR, Goettingen, Germany
Zdravko Terze, University of Zagreb, Zagreb, Croatia
Tao Wan, University of Bradford, Bradford, United Kingdom

Marine Simulation
Cyril Ray, Ecole Navale, Brest, France

Simulation in Industrial Design and Product Design
Chiara Catalano, IMATI-CNR, Genoa, Italy
Yan Luo, NIST, Gaithersburg, USA
Jorge Perez Mateos, Industria de Turbopropulsores SA, Madrid, Spain
Catarina Rizzi, University of Bergamo, Bergamo, Italy
INTERNATIONAL PROGRAMME COMMITTEE

Simulation in Engineering Processes
Chribsanti Angeli, Technological Institute of Piraeus, Athens, Greece
Blazej Balasz, Technical University of Koszalin, Koszalin, Poland
Alejandra Gomez Padilla, University of Guadalajara, Mexico
Fouad Riane, CREGI-FUCAM, Mons, Belgium
Christopher Schlick, RWTH Aachen, Aachen, Germany
Jan Studzinski, Polish Academy of Sciences, Warsaw, Poland
Joao Tavares, University of Porto, Porto, Portugal

Simulation in Energy and Power Systems
Saleh Al-Jufout, Tafila Technical University, Tafila, Jordan
Manuel Cordeiro, CETAV/UTAD, Vila Real, Portugal
Janos-Sebestyen Janosy, KFKI Atomic Energy Research Institute, Budapest, Hungary
Mohammad Montakhab, University of West England, Bristol, United Kingdom

Simulation in Multibody Systems
Track Chair: Bernd Schaefer, DLR, Wessling, Germany
Ignacio Garcia-Fernández, University of Valencia, Valencia, Spain
Jose Diaz Lopez, Dynasim AB, Lund, Sweden

Simulation in Chemical, Petroleum and Mining Engineering
Diganta Das, Loughborough University, Loughborough, United Kingdom
Klaus Hellgardt, Imperial College, London, United Kingdom
Vahid Nassehi, Loughborough University, Loughborough, United Kingdom
Mohamad R. Riazi, Kuwait University, Kuwait

Simulation in Military and Defense
Roberto de Beauclair Seixas, IMPA, Rio de Janeiro, Brazil
Rinardo Michelini, University of Genova, Genova, Italy
Carlos Palau, UPV, Valencia, Spain
Matthias Reuter, CUTC Gmbh, TU-Clausthal, Clausthal, Germany

Virtual Reality and Graphical Simulation in Industrial Simulation
Track Chair: Guodong Shao, NIST, Gaithersburg, USA
Emilio Camahort, Universidad Politecnica de Valencia, Valencia, Spain
Anders Hast, University of Gavle, Gavle, Sweden
Pieter Jorissen, University of Hasselt, Hasselt, Belgium
Fabrizio Lamberti, Politecnico di Torino, Turin, Italy
Sudhir Mudur, Concordia University, Montreal, Canada

Verification, Validation and Accreditation
Mokhtar Beldjhem, École Polytechnique de Montréal, Canada
Roberto Revetria, University of Genoa, Genoa, Italy
Agustin Yague, Technical University of Madrid, Madrid, Spain
Dmitri Zaitsev, Odessa National Telecom University, Odessa, Ukraine

Simulation and Training
Manuel Alfonseca, Universidad Autonoma de Madrid, Madrid, Spain
Wenji Mao, Chinese Academy of Sciences, Beijing, P.R. China
Dorin-Mircea Popovici, Ovidius University, Constanta, Romania
Eсин Rajabally, Loughborough University, Loughborough, United Kingdom
Gerhard Schreck, Fraunhofer IPK, Berlin, Germany
Ahmed Bin Subaih, University of Sheffield, Sheffield, United Kingdom

Workshops
Workshop on Modelling and Simulation in the Textile Industry
Track Chair: Vladan Koncar, ENSAIT, Roubaix, France
Antonela Curteza, "Gh. Asachi" Technical University of Iasi, Iasi, Romania
INTERNATIONAL PROGRAMME COMMITTEE

Peter Farber, Hochschule Niederrhein, Krefeld, Germany
Jiri Miltky, Technical University Liberec, Liberec, Czech Republic
Ali Shams-Nateri, Guilan University, Rasht, Iran

Workshop on Simulation in Sports
Track Chair: Raphaël Massarelli CRIS University Lyon 1, Lyon, France
Paolo Grigolini, University of Texas, Denton, USA (tbc)
Arne Sieber, Profactor Solutions GmbH, Seibersdorf, Austria (tbc)
Thierry Terret, CRIS, Lyon, France

RFID Track
Yo-Ping Huang, Tatung University, Taipei, Taiwan
Jaap Ottjes, Delft University of Technology, Delft, The Netherlands
Selwyn Piramuthu, University of Florida, Gainesville, USA
Hans Veeke, Delft University of Technology, Delft, The Netherlands

Workshop on Intelligent Transport Systems
Paul Davidsson, Blekinge Institute of Technology, Ronneby, Sweden
Petr Hanacek, Brno University of Technology, Brno, Czech Republic
Jairo Montoya Torres, Universidad de la Sabana, Chia, Columbia

NANOSIM
Clemens Heitzinger, Purdue University, West Lafayette, USA
Javier Marin, ETSI, University of Malaga, Malaga, Spain

Workshop Augusted Reality and Pervasive Systems in Simulation
Alessandro Genco, University of Palermo, Palermo, Italy

Workshop on Simulation in Lean Manufacturing
El-Houssaine Aghezzaf, Ghent University, Ghent, Belgium
Abdelaziz Bouras, University of Lyon II, Bron Cedex, France
Hendrik Van Landeghem, Ghent University, Ghent, Belgium
INDUSTRIAL SIMULATION 2008
Complex system modeling remains a very challenging task, even if a lot of simulation approaches are now available and provide quite appropriate solutions to unsolved problems. The analytic methods always fail to accurately and efficiently describe the behavior of complex systems where a small change of one object will provide a complete bifurcation of the whole system state. In fact analytic approaches, mainly top-down are complementary to simulation models that are bottom-up and provide a marvelous property, which is called "emergence".

These two classes of methods are the two faces of the same coin. We trust that it is necessary to make them cooperate together in order to cope with the complexity of our world and environment. One aspect becomes the main scope of many researches: "what is life?"

The understanding of the behavior of complex systems needs a multidisciplinary cooperation of specialists coming from different disciplines like mathematics, computer-sciences, cognitive psychology, sociology, management… which are currently exchanging many paradigms. The application fields are very numerous and varied in all the industrial and scientific domains.

Many questions remain unsolved at the moment: What could be a system, which is “alive”? Can we develop thinking machines provided with consciousness, inventiveness and behavior autonomy?

This problem can be defined as making computers shift from the status of pre-programmed servant machines doing repetitive tasks to the status of actual autonomous entities owning an artificial psychic system, providing thinking and auto-modifying capabilities. This new generation of computers could then be an actual interlocutor to human beings. However, this will necessitate the need to set up other ethical and epistemological questions. Is it wise to develop such types of systems?

The conference site namely, the Pôle Universitaire de Lyon PUL also called « Université de Lyon » welcomes approximately 130.000 students. The PUL is the regrouping of three Universities: Claude Bernard Lyon 1 (Sciences and medicine, Villeurbanne & Lyon), Lumière Lyon 2 (Human sciences, Bron & Lyon), Jean Moulin Lyon 3(law, management, Lyon) and four engineering schools: Ecole Centrale de Lyon (ECL, Ecully), Ecole Normale Supérieure de Lyon (ENS, Lyon), Institut National des Sciences Appliquées de Lyon (INSA, Villeurbanne), Ecole Normale Supérieure Lettres et Sciences Humaines (ENSLSH, Lyon). The University of Lyon is built on a long and strong scientific tradition.

On behalf of the International Scientific Committee of ISC-2008, we are very happy and proud to welcome you to the CESH (Centre Européen de Santé Humanitaire, Lyon Gerland) and the University of Lyon.

We hope you will enjoy your stay in Lyon and Villeurbanne and we trust that the conference will be the opportunity of new paradigm exchanges between industries and universities to develop new complex system and simulation methods.

Joël Colloc, Claude Petit
Conference chairs
CONTENTS

Preface ........................................................................................................................................... XI
Scientific Programme ......................................................................................................................1
Author Listing ...............................................................................................................................431

TUTORIAL

Humanitarian Demining: Efficiency by Intelligent Planning and Low-Cost Robotics
Rinaldo C. Michelini, Roberto P. Razzoli and Udayanga M. Hemapala ..................5

SIMULATION METHODOLOGY

SIMULATION AND AI

A Genetic Algorithm method for generating efficiency intervals in DEA with Imprecise data
Said Al-Gattoufi ............................................................................................................................17

Of the Stability of Closed Self-Organizing Maps (gSOMs) for Predictive Control
Matthias Reuter ..........................................................................................................................25

Integrating Immunity features in a Multi Agent Model
Saber Darmoul, Sonia Hajri-Gabouj and Henri Pierreval .......................................................31

Project Plan formalization and modeling via typed CPN
Sarka Kvetonova and Dusan Kolar ..............................................................................................34

Identification of Multivariable NARMA Models Using Artificial Neural Networks
Brahim Tilili, Faouzi Bouani and Mekki Ksouri ........................................................................40

DISTRIBUTED SIMULATION

Pipe-lined Parallelism Scheduling in homogeneous and heterogeneous environments
Carmen Odubasteanu and Călin Munteanu ..............................................................................47

An efficient method for Distributed Constraint Satisfaction Problems resolution
Kais Ben Salah and Khaled Ghedira ..........................................................................................52
CONTENTS

PROCESS SIMULATION

Implementation of the Process Interaction Approach in a General Purpose Language
Hans P.M. Veeke, Jaap A. Ottjes and Gabriël Lodewijks ........................................65

Multiple-model Design and Switching Solution for Nonlinear Processes Control
Ciprian Lupu, Dumitru Popescu, Catalin Petrescu, Alexandru Ticlea,
Bogdan Irinia, Catalin Dimon and Andreea Udrea ..............................................71

SIMULATION IN MANUFACTURING

LARGE SCALE PRODUCTION SYSTEMS

The Operational Strategies in focused Flexible Manufacturing Systems
F. Nucci and A. Grieco ..................................................................................................81

Validation of the design of a new steelmaking factory
Nicolas Pipard and Steve Thornton .............................................................................89

Evaluating the Impact of the Lean Approach on a Food Processing Line
Flavio Tonelli, Roberto Revetria and Ernesto Revello ..............................................93

MAINTENANCE PERFORMANCE SIMULATION

Availability Modelling for Optimizing The Complex Systems Maintenance
Radhouane Djeridi, Mathieu Glade, Samir Ghelam, Frédéric Matheron
and Aline Cauvin .......................................................................................................101

Evaluation of Maintenance Schedule of a Light Metal Industry Plant
using Petri Nets
Vrassidas Leopoulos, Charalampos Tsogkas, Evaggelos Champiomatis
and Leonidas Tsikonis .................................................................................................104

AUTOMOTIVE ENGINEERING

Sheet Hydroforming of Automotive B Pillar and Coefficient of Friction
Effects Analysis
Javad Marzbanrad and Mir Javad Peyman ..............................................................111

A new Approach for Modeling of Spot Welds for Numerical Analysis
Abolfazl Masoumi and Amir Najibi ........................................................................117
## CONTENTS

### Cutting Simulation Software for Mold Industry
Su-Jin Kim ........................................................................................................................................... 122

### Simulation and Performance Improvement Process at Ford Powertrain Operations
Habtom Mebrahtu and John Ladbrook ................................................................................................. 126

### Hybrid Powertrain Simulation
Valerian Croiorescu, Andrei Maciac, Mircea Oprean, Alexandru Marin and Cristian Andreescu ................................................................................................................................. 133

## SIMULATION IN ROBOTICS

### COMPONENTS SIMULATION, NAVIGATION SYSTEMS AND MOVEMENT

**A Simulation-based Method for Fault Diagnosis in Hydraulic Components**
C. Angeli and A. Chatzinikolaou .................................................................................................................. 143

**Formation Preserving Navigation of Agent Teams in 3-D Terrains**
Ali Galip Bayrak and Faruk Polat .................................................................................................................. 148

**Fusion of INS with GNSS and its application to Mobile Robot Navigation**
L. Domenech, L. Armesto and J. Tornero ...................................................................................................... 156

**Trajectory Planning Based on Robot Dynamics with Artificial Potential Fields**
José A. Bernabé, Marta C. Mora, Ana Ruescas and Josep Tornero .............................................................. 161

**Implementation of the Symbolic Simplification for the Calculation of Accelerations of Multibodies**
Dmitry Vlasenko and Roland Kasper ........................................................................................................... 166

## MEDICAL ROBOTICS

**Virtual Prototyping and Multi-Modal Interfaces to test the Control of an Orthosis**
Aron Pujana-Arrrese, Joseba Landaluze, Jesús Gimeno and Marcos Fernández .......................................... 173

**Movement Reconstruction in industrial and medical domains based on topological Petri Nets**
Markus Koch, Robert Beckelbons and Jürgen Schrage .............................................................................. 181
CONTENTS

COMMUNICATIONS SYSTEMS

Expansion of OPNET Modeler’s SIP Model for Performance Evaluation of Hierarchical Call Routing in NGN
Hajar Atrianfar, Zohreh Ayatollahi and Mohsen Sepasi ........................................189

Implementation of Decision Support Systems for Communication Systems
Calin Munteanu, Simona Caramihai, Carmen Odubasteanu and
Florian Antonescu ........................................................................................................194

Modeling and Simulation of Traffic with Compression at Media Gateways for Next Generation Networks
Dimitar Radev and Izabella Lokshina ........................................................................199

SIMULATION IN ENERGY

Protection System Refurbishment at Paks Nuclear Power Plant using Full-scope Simulator
János Sebestyén Jánosy ................................................................................................209

New application fields of nuclear power plant simulator
MyeongSoo Lee, JinHyuk Hong, SeungHo Lee, JeongKwan Suh and
DoHyun Hwang ........................................................................................................215

Optimisation System for a Cogenerative Power Plant
Florin Hartescu and Ion Miciu ....................................................................................219

Modelling and Simulation of Fluid Power Systems in an Intelligent Programming Environment
Gunnar Grossschmidt and Mait Harf .........................................................................224

LOGISTICS AND SUPPLY CHAIN SIMULATION

CUSTOMER DEMAND

Fuzzy Simulation Optimization Approach for the Transshipment Problem with Uncertain Customer Demands
Mohamed Hmiden, Lamjed Ben Said and Khaled Ghedira ........................................235

Proposition of a Simulation Tool for a Cooperative Effective ATP Function
Mariem Siala, Khaled Ghédira and Jean-Pierre Campagne ....................................241
# CONTENTS

## MULTI-PRODUCT SIMULATION

An Overview of Multilevel and Hierarchical Methods for Discrete Event Simulation of Complex Systems  
Esra E. Aleisa ................................................................. 249

Modelling and Simulation Features for Products Services  
Rinaldo C. Michelin and Roberto P. Razzoli .......................... 256

## SUPPLY CHAIN SIMULATION

Using Simulation to support sustainable Logistical Decision Making in Complex Supply Chains  
P.J. Byrne, Paul Ryan and Cathal Heavey ............................. 265

A Conceptual Model for SME Mechatronics Supply Chain  
Jihene Tounsi, Julien Boissière and Georges Habchi .................. 273

## LOGISTICS SIMULATION

TRANSSIM Node- A Simulation Tool for Logistics Nodes  
Larissa Neumann and Simon Deymann ................................. 283

A Design Approach for Asset Supply Logistics  
Sebastiaan A. Röben, Arnold van den Dool, Jaap A. Ottjes and Gabriël Lodewijks .................................................. 288

Using Container Call Time Information for Restacking Reduction  
Wing Sum Lee, Jaap A. Ottjes, Hans P.M. Veeke and Joan C. Rijsenbrij ........ 293

## LIFTING AND STACKING

Lift System Optimization by Simulation  
Jaroslav Sklenar, Michal Touš and Pavel Popela ...................... 301

Dynamic Control of Animation Schemes for the Efficient 3D Visualization of Material Flow Simulations  
Wilhelm Dangelmaier, Benjamin Eikel, Matthias Fischer and Christoph Laroque .................................................. 306
CONTENTS

RFID SIMULATION

Decision Support System for RFID
Gabriël Lodewijks, Adriana M. López De La Cruz and Hans P.M. Veeke ........313

Performance Measurement of RFID
Adriana M. López De La Cruz, Hans P.M. Veeke and Gabriel Lodewijks ........317

HOSPITAL LOGISTICS

Emergency Room Simulation Prototypes for Incident Management Training
Guodong Shao and Charles R. McLean .........................................................323

Modelling Causality in Medical Scenarios
Peter F. Summons .........................................................................................328

A step of modeling for the design simulation models in the hospital systems: Application to the New Hospital of Estaing
Julie Chauvet, Michel Gourgand and Sophie Rodier ........................................333

COMPLEX SYSTEMS SIMULATION

MEDICAL SIMULATIONS

Case-based Reasoning: A Perspective for Pharmaceutical Decision Making
Claude Dussart, François-Xavier Magaud, Joel Colloc and Claude Petit ...........343

Pregnancy-birth course in Veneto Region: development of a flexible simulation model for description and forecasting
Alessia Barella, Paola Facchin and Giorgio Romanin-Jacur .................................348

Kinematic Modelisation of Joint Displacement: Validation in Human Pointing Tasks
Clément Villars, Jérôme Bastien, Karine Monteil and Pierre Legreneur ...............353

Using a GA to Determine Genotype and Phenotype Relationships
Clyde Meli ........................................................................................................356

COMPLEX ENVIRONMENTS SIMULATION

Finite element modelling of Underground Water Flow in low permeability Lands
Abhijeet Kulkarni, Bahareh Kaveh-Baghbaderani and Vahid Nassehi ...............365
Modeling and Simulation of Feedback Seismic Structural Control
Svetla Radeva and Izabella Lokshina ...................................................... 371

FORWAVER – A Wavelet-Based Predictor for Non-Stationary Signals
Dan Stefanoiu, Janetta Culita and Florin Ionescu ...................................... 377

TEXTILE SIMULATION

MODELLING AND SIMULATION OF TEXTILE PROCESSES

Modelling and simulation of an active anti-ulcer mattress system
Jalloul Elfehri, François Boussu and Vladan Koncar ................................. 387

Modeling the Fiber Yarn Interface for Clothing Design
Ionut Neagu, Pascal Bruniaux, Xavier Legrand and Dorin Avram ............... 393

APPAREL INDUSTRY AND CLOTHING MODELLING AND SIMULATION

Towards a coupled simulation of cloth and soft tissue
Bastian Linneweber and Arnulph Fuhrmann ............................................. 401

Strategic Analysis of Speed and Flexibility in sourcing Textile Products
Lisa Hartman, Jeffrey A. Joines, K.A. Thoney and Russell E. King .............. 410

Modelling the Fabric-Apparel Compatibility using a Specialised Program
Antonela Curteza, Adela Florea and Viorica Scobiola ............................... 416

Optimization of the liquor distribution in yarn package dyeing by CFD
Fabiano Scalora, Carlo Beltramo, Marco Vanni, Giorgio Rovero and
Silvio Sicardi ......................................................................................... 423
SCIENTIFIC PROGRAMME
TUTORIAL
HUMANITARIAN DEMINING: EFFICIENCY BY INTELLIGENT PLANNING AND LOW-COST ROBOTICS
Rinaldo C. Michelini, Roberto P. Razzoli and Udayanga M. Hemapala
PMAR Lab, DIMEC, University of Genova
Via Opera Pia 15/a - 16145 - Genova, Italy
E-mail: michelini@dimec.unige.it; razzoli@dimec.unige.it; hemapala@dimec.unige.it

ABSTRACT
The robot mine clearing, based on cheap and widespread farming contrivances, provides considerably unusual prospects, which aim at enhancing safety, reliability and efficiency, to transform the terrorist anti-personnel mine into almost useless practice. The talk develops along the following lines: • to conceive and implement low-cost robotic equipment, assuring effective, safe and reliable demining: the resources (facilities, functions, fixtures) choice lets devising worthy strategies; • to help out equipment integration, with resort to cheap and widely available (agricultural) fixtures: the on-process adaptive work-cycles lead to better tactical concert; • to provide the modelling and simulation background for the on-duty assessment and training assistance: the alarm (warning, emergency) management brings to efficiency and safety; • to make easier the operators groundwork, involved in self-learning and up-grading duties: the mixed-mode automation grants balanced resources use on the strategic, tactical and execution horizons.

KEYWORDS: Military Applications; Robotic Mine-Sweeping; Remote Steering

INTRODUCTION
The mine reclamation is entangled duty, with demanding socio-economic conditions. Looking at technicalities, we have:
# military mine-sweeping: to allow reliable logistics, specific tracks shall be made safe, by locally aimed technologies and real-time efficient work-organisation;
# humanitarian clearing: to achieve safe re-appropriation of the ground (for farming, etc.), the all land needs to be overseen and recovered, at acceptable costs.
Indeed, the resort to landmine munitions characterises terrorist-driven warfare theatres, where the local political implications have historical reasons; solutions, on merely technical basis, will not work, unless the implications are also neutralised. The humanitarian demining characterises as necessary activity to assure fit-for-use land operation, where the legal frames bear negligible relevance; the task shall be routine highly effective process, having shared community part at mutual decentralised level (Hemapala 2006).
The aspects suggest series of assumptions. First, no abstract developments are considered; on the contrary, properly aimed measures are dealt with, deeply rooted on the local historical frames. Second, the factual evidence of useful behaviours is deferred to results, and these are consequence of achieved benefits. Third, the routine business involvement shall comply with current habits, utterly neutralising exogenous or illegal biasing damages. Fourth, techniques and workflow schedules should not superimpose, rather organise on shared know-how.
The challenge is to make the mine clearing becoming standard routine, basically suppressing the terrorist threat, with full resort to local contrivances and personnel. Indeed, the project successfulness would mean to offer a technical solution out of the existing socio-economical obstacles.
The humanitarian clearing was, chiefly, fulfilled by front-end personnel, with resort to the harrow/brush. If accomplished by trained and careful operators, it assures comparatively safe and reliable results, having costs covered within acceptable ranges, directly allowed to local people, which could receive proper wages, independent from the actual land productivity.
The unmanned mine clearance permits safer duty progression, since the work-flow does not include front-end operators. The today state of the art distinguishes two techniques:
• heavy armoured vehicles capable to withstand the mine blast, thereafter eliminating the danger, because not requiring the mine handling;
• sophisticated robotic demining, with resort to three step cycle: detection and localisation; reliable picking, handling and removal; neutralisation and reclamation.
Both have the big hindrance in the cost of the special-purpose outfits. To join safety, low-cost and effectiveness, the mixed-mode processing abilities, incorporating front-end mechanised effectors and back-drop man intelligence, ought to be carefully assessed, by tangible technologies transfer and socio-economic impact as for the Civil Service appropriation. Then:
- the robotic mine sweeping should be privileged, enhancing the productivity by means of front-end automation, and lowering the costs and avoiding undue sophistication;
- the manned contribution is included, properly exploiting the on-process decision-making to widen the process adaptivity and flexibility, without increasing cost and sophistication.
On these ideas, the assembly of mine clearing robotic outfits includes the assessment of each step duty request, followed by the integration of the all system, by bottom-up procedures:
• to select the on-the-field resources, depending on the desired clearing tasks, and satisfying low-cost, leanness and know-how appropriateness (Baskin et al. 1999);
• to integrate widespread resources with task-bend enhanced by co-operative technological inventiveness (Cepolina 2005), (Schulz 1999);
• to programme the workflow, adapted to land conditions, mine spreading, effectors type, etc., with remote steering for safe accomplishment (Belotti and Michelini 2007);
• to help facing un-expected occurrences (break-down, dead-stops, blasts, etc.) by autonomic alarm managing and assisted decision-making (Ol ling and Kimura 1992).
The equipment choice quite obviously addresses the area of the agricultural machinery and know-how.

THE LOW-COST ROBOTIC EQUIPMENT
The final robot assembly (Belotti et al. 2008a) shall include:
- self-powered carriers, say suited vehicles, bulldozers, tractors, power-tillers, etc., to confer the proper mobility;
- task-adapted demining outfits, to fulfill effective, reliable and safe mine clearing, with robot carrying out.

The carrier choice has to deal with cost, power, thrust, stability and accessibility. The on-field mobility is, in general, provided by means of power-tillers, light tractors, bulldozers, (armoured) cars, etc., the last being typical for military applications. The power tillers are good choice, widely used and commercially available at low cost, also, as second-hand offer. They request clear modifications, (Belotti et al. 2008b), notably, for steering and stability, through the resort to front trailer, having a further axle. The carrier fitting out shall embed: the special actuation for gear selection and choke setting, to adapt power, thrust and speed; the selective wheels brake, for path tracking; the front and lateral cameras, for visual restitution and course detection; etc., up to the feeding and handling of the front effectors for the mine clearing duties (Cepolina and Hemapala 2007).

The integrated robotic outfits might each other greatly differ, with the only unifying fact of coming from widely available (agricultural-like) machinery (Havlík and Licko 1998). The robot reclaiming duty consists of two tasks (Habib 2007):

- to detect the mines, by proper contrivances (racks, effectors, etc.) or by front-end sensors;
- to disable (remove/destroy) the mines, so that the incurred danger is totally eradicated.

The detection is expensive and, by itself, not decisive; the handling and neutralising actuators, complex and scanty. The duty can use whirring flails, to destroy, rather than to remove, the mines. Either, the outfits should inspire to existing fixtures, such as: the potatoes pulling out, or the mine burst once hit by driven strings. Two outfits are, thus, conceived, Fig. 1:

- ground stripe lifting, and mines singling out by gentle sieve descent of the shifted earth;
- land sweeping by (forward displaced) striking flaps and flounders, inducing the mine blast; the former principally applied on recently laboured pastures or soft sandy lands, the latter commonly used on tough meadows or compact dry areas. The mine lifting device requires design upgrading (by respect to common potatoes risers), to widen the front prospect and to add soil breaking devices. The striking flaps and flounders are derived from existing rigs, and applied to the (added) front axle, with suited forward overhang.

Figure 1: Outfits: mines lifting/singling; soil striking/mine blast

The special purpose mine clearing outfits and procedures are intended to be explanatory options; other fixtures/schedules can be devised on different design concepts and engagement rules; this means that sets of comparable rigs are obtained by resort to the sketched design-and-development approach, as the search just addresses the demands of feasibility and appropriateness, but does not implies the singleness of the outcome. Besides, the frame shows that humanitarian demining is open to lots of issues; technology appropriateness and lean feasibility are only constraints, offering valuable novelty under the methodological points of view, when the clearing process follows, exploiting the intelligent work-organisation paradigms, (Alty and Mikulich 1992, Sata and Olling 1989), once the pertinent modelling and simulation features are developed, and provided for robot exploitation (Zaremba and Prasad 1994).

Moreover, the robot duty closing requires software/hardware to place the operator in safe remote location, with data from in-process diagnostics and on-process cameras. The steering task, Fig. 2, presumes mobility control, performed by the operator on the rear. The choice of the hardware fitting privileges the resort to off-the-shelf devices and standard software aids. The search faces two fundamental demands:

- friendliness, to guarantee easy operability to hastily trained operators and leanness to reach simple maintainability and rugged on-the-field equipment;
- low-cost, to make possible fruitful resources assembly from on-the-market offers, and effortlessly integration of cheap and well assessed technologies.

Figure 2: Remote communication hardware/software set-up

The architecture with the duplicated processing, on-site and at the operator location, permits closing control loops, that only need the friendly afterward acknowledgement of the subsidiary recovery/withdrawal manoeuvring. The remote governing of the robotic outfits avoids cables and umbilical, with resort to self-powered carriers and readily transferable overseeing stands. The governing system has two computers: the on board processor; the operator station. The on board processor is in charge of controlling path and ground processing; it provides, also, wireless link to the operator station. The communication channel uses standard Wi-Fi (IEEE 802.11 b).

**EQUIPMENT & COMMANDS INTEGRATION**

The robotic equipment integration addresses simple and cheap equipment, privileging leanness, not multi-task machines, so that there is large chance to find out basic devices, readily to transform into helpful outfits, solely adding proper kits. The project requires series of actions, such as:

- to characterise the picked up farming machine, assigning the looked-for work-cycles and specifying the duty-driven rigs to achieve demining capabilities;
- to implement the specified fixtures, preserving technological consistency, so that the resulting outfits are ran, up-kept and maintained by shared know-how;
- to programme the suited work-cycles (with targets, thresholds and timings), to forecast the productivity, on the strategic, tactical and execution horizons;
- to appraise and check the achieved performance (safety, efficiency, reliability, etc.) for steady running, with account of unexpected anomalies (incongruities, failures, etc.).
The all is standard engineering task, and happens to be fulfilled detailing all features, with behavioural issues, covering:

- duty-steered functioning on the strategic horizons, to prove suitableness (terrorist counter-measure) and appropriateness (shared habits acceptance);
- occurrence-driven efficiency on the tactical horizons, to select optimal off-process plans, to achieve high productivity, on steady running conditions;
- anomaly-coerced progress on the execution horizons, to show adaptivity and flexibility, by on-process decision duties done by the operators.

The equipment and command integration brings to entangled system assessments, (Tzaferas 1997), leading to the (mobility and clearing) fixtures choice, (off-process) strategic schedules, (on-duty) tactical agendas and (real time) executional resetting, and dressing flow-charts and resource provision details. For explanatory purposes, the example connection of a power tiller and a (specially conceived) mine digger and singlet-out, Fig. 3, is addressed. The itemized list in the work-flow incorporates all the diverse branching and recognition agendas incidents; e.g., with the ground stripe lifting, the appearance of a mine-looking body, the stop/go on impasse, the bursting/inert-stuff singling out (and safe removal), the blast and damage evaluation checks, and so on. The task complexity defines the instrumental setting at the detection levels (camera, for the visual recognition, etc.) and at the govern steps (remote control of the forward motion, of the ground stripe depth, of the sieving action, etc.).

The front effectors present with a series of tilting prongs, for gentle soil loosening, Fig. 4. The digging depth is 15-20 cm deep; the lifted ground strip is moved on a shaker, to extract the solid objects, framed by a camera. The power tiller pushes through two couples of bumpers, having a central link fixed by engaged hooks. The bumpers are pneumatically actuated, to assure proper preload, for (damped) negligible carrier/effectors oscillations. The differential biasing of the powered couple of bumpers is, further, exploited to modify the robot trajectory, assuring the requested (rectilinear) path tracking.

The analysis of the duty-cycle requirements allows to specify the key functions that the robot has to fulfil. Being followed the ground stripe lifting strategy, the fixture is similar to the potato lifting and retrieval mechanisms (to be modified for the sorting and storing section). Besides, the presetting of the remote control set-up, the location of the detection devices (cameras, etc.) and the management of the shelter and reliability cautions (up to safe picking and mine neutralisation) ought to be watchfully dealt with, in view to lay down effective operation conditions.

The outcome is shown by the block-schema of the expected functional assets, with account of the parallel information flow. To help explaining the procedure, the functional flow, Fig. 4, gives the basic sample choices. The clearing strategy is sort out balancing operation requirements and resources availability. At the tactical level, path planning and ground processing follow, once known the robotic equipment (Fig. 3).

![Figure 3: Digital mock-up of the example robotic equipment](image)

![Figure 4: Tilting prongs (easily replaced when damaged)](image)

![Figure 5: Functional flow of the robot clearing work-cycles](image)

With clearing productivity and technological appropriateness in mind, Fig. 5, the winning solution shall incorporate:

- the intelligent task-driven paradigms of flexible automation, with mixed-mode steering schemes that make use of on-process operators to adapt the on-going work schedules;
- the available agricultural machines to devise implementing mobile robotic helps, capable to incorporate the (innovative) task-adapted specially developed demining fixtures;
- the ideation and adjustment of simple front-end effectors, heading to the safe and reliable land reclamation, with high productivity and protected man involvement;
- the (purposely conceived) remote-control apparatus, for the direction-steering and navigation of the mobile robotic outfits, governed by the on-line operators;
- the adaptive process-planning agendas, to rule/manage the strategic/tactical/execution flexibility opportunities, allowing to achieve enhanced operation performance.

For practical purpose, the overall integration and planning of the on-the-field mine clearing constructs needs to require fast assembly/disassembly and easy maintenance/restoring abilities, suitably achieved by parts modularity and joint standardisation. This means to explore off-process versatility, by, e.g., changing the front-end fixtures. All in all, the availability of alternative carriers and end-effectors is only necessary step, not sufficient
to assure good productivity.
The exploration needs forcibly to incorporate advanced checks on the really achieved productivity, at different environmental conditions and competing operation plans. The project looks after suitable simulation aids, based on the modelled dynamical behaviour of material resources and planned decision logics supplied to the human operators. The modelling and simulation permit comparing different resource lay-outs, outfit changes, planning agendas and random alarms, so that the collected data provides reliable anticipations on the on-the-field behaviour of the timely experimented facilities and fixtures. Actually, the counter-measures merit is given by the mine clearing success, and the present examination is forced to devise how the goal can be achieved according to factual measures, directly derived from the locally shared abilities and instrumental aids. The simulated testing is useful, allowing checks on:
- the facility set-ups, by means of the purposely developed digital mock-ups, so that the behaviour of the rigs is duplicated, up to the suited detail accuracy;
- the job agendas, by means of the devised control logics, tactical decision aids and recovery routines, so that the different alternatives are compared.
The transfer of the acquired knowledge into process steps up is greatly enhanced by the modularity of outfits and fixtures, and by the sectional implementation of the govern modes. The latter chance is fostered by resort to virtual instrumention concept.

**MODELLING AND SIMULATION FEATURES**

Once detailed demining resources and management policies, the operation planning is obtained with resort to modelling and simulation features, distinguishing three flexibility horizons;

# the strategic horizon deals with the off-process versatility, and, for example purpose, the checks aims at maximising the process effectiveness comparing the series of the mobility providers and front-end effectors, actually, implemented; the programmes take into consideration the remote control accomplished by the on-the-field operator, which has the direct visibility of the governed robot, through the connected instrumental data (course, speed, thrust, etc.);

# the tactical horizon deals with the on-process adaptivity, and, for example purpose, the mine clearing, done by the power-tiller endowed by the ground strike lifting, are detailed, defining competing agendas; the pertinent decision aids are developed, with the issues brought to the attention of the operator, who might switch between the agendas, and re-initialise the all duty-sequence;

# the execution horizon deals with the unexpected occurrences (equipment failure, mine deflagration, course stop dead, etc.), and, for example purpose, the outlined software/hardware setting is studied on multiple-level (warning and emergency) alarms, depending on the relative risk and frequency, each time showing restoring/healing tracks and occurrence-driven recovery stops of the on-going agenda, and requiring the operator consent for the subsequent steps.

Once the modelling features are assessed, Fig. 6, the equipment behaviour simulation permits to recognise, whether off-the-shelf items (directly or slightly modified) or innovative devices have to be used. The enquiry operates with the resort to locally available pieces, after the throughout understanding of how these might be adapted and exploited. For explanatory purpose, the following alternatives are checked:

- two standard carriers, say: given type of power-tiller or tractor for the open-fields clearing;
- two standard mine clearing techniques: the soil strips lifting, or the flailing and tilling strings.

![Figure 6: Block-schema of the robotic clearing work-cycles](image)

The strategic setting, thus, permits four (off-process) resource selections, fully retaining the low-cost and leaness requisites, with best allocation and management of facilities, functions and fixtures. Then, the duty-cycle analysis distinguishes:
- the path planning and course acknowledgement, by speed, thrust, direction, etc. adjustment;
- the mine clearing processing, through the detection, singling, neutralisation, etc., tasks.

The top-down analysis is accomplished the find-out the optimal sweeping strategies, depending on the land conditions and on the fixture availability. The bottom-up analysis allows selecting the appropriate operation tactics, once allocated the resources and acknowledged the in-progress operation conditions, with account of the up-surging warnings and emergencies.

![Figure 7: Example steady agenda and course planning](image)

Defined the expected engagement, e.g., with soil lifting rigs, the mine-clearing agendas distinguishes:
- the steady running conditions, to be accomplished in the autonomic mode under operator overseeing, unless special
warnings appear on the monitor;
- the emergency running conditions, when the operator selects
  the single-step mode, or when the front-end sensors switch-
  off of the on-going agenda, at unexpected occurrences.
The steady running conditions avail by suited «macros», shown
on the monitor by proper keys, The analysis is fulfilled at both
ranges. The preliminary attempts run with the remote control. In
the hardware selection, the instrumental architecture of the
distributed sensing and actuating devices widely profits by off-
the-shelf components, to preserve the low-cost final integration.
The up-dating, therefore, is easily obtained changing or adding
new modules, warranting enhanced visibility on the process or
supplying wider versatility and/or dexterity. The next measures
address the rethinking of the operation planning.
The information flow during the execution of the standard duty-
cycles exploits macros, with coded operation sequencing, to be
accessed as single (upper level) command. The idea is to permit
the operator to collect data from the surroundings to forecast
coming scenarios, leaving the robot in autonomic mode, Fig. 7,
dealt with by «macro». The careful and extended checks make
possible the detection of local soil or vegetation warning signs,
to be associated with the recent laying down of mines, so that
the steady agendas are devised and the course planning
progresses in the autonomic mode. The path steering, Fig. 8, is
forecast off-process, having account of resources and ground
features, with the course later adapted by the operators,
according to the on-duty rising warning/emergency signals.

Figure 8: Back-and-forth path steering for full coverage

The careful understanding of the currently enabled functional
settings shows that the main constraints, for the project success,
distinguish a series of objectives:
- the set-up of the effective management infrastructure, with the
  bottom-up (cultural) spur of the local society, to establish the
  promoting, steering and checking Civil Service;
- the setting of the standard work-flow charts, to reach the
  industrial organisation productivity in the mine clearing, with
  pre-assigned work-in-progress targets and due times;
- the conception of consistent task-oriented outfits, derived by
  commonly shared technologies and apparatuses, to warrant the
  high on-process helpfulness;
- the truthful assessment of the process-agendas, with resort to
  strategic, tactical and execution flexibility, for the bottom-up
  mine-clearing programming.
The emergency flow-chart permits the bottom up assessment of
the chosen technologies, with discussion of cost, performance
and reliability, to acknowledge the productivity decrease by
respect to the strategic agendas and, possibly, to revise the
resources procurement at the strategic range. The unexpected
occurrences represent the typical interrupts in the steady runs,
depending on internal or external reasons (hardware/software
failures, duty-cycle stops dead, landmine explosion, etc.), and
need to be acknowledged and detailed in the flow-chart
strategies, so that the alarm management and duty recovery
become integral part of the task requirement schedules, at the
execution flexibility level. The robot path planner, at the in-
progress position, Fig. 7, has to chose direction and speed or to
stop, as it is pre-selected by the embedded ruler or established
by the remote operator, according to the visual display (e.g.,
soil aspect) and the sensors feedback (e.g., thrust on the
vehicle). In any case, the mixed-mode automation permits the
in-progress change and adaptation of the pre-planned path track
by the simple switching to alternative «macros».

In fact, the finally rigged robotic equipment is commanded and
overseen, with the operator standing into safe location (around
100 m from the robot). The full perception of duty progression
is provided by the duplication of the on-board command setting
and independent cameras sights. The agendas sequencing can
be switched to the autonomic mode: the progression is deferred
to «macro», transparently coded in the client-server restitution.
Alternatively, can be executed by single-step mode, wanting
consent each time; or might turn to the alarm mode, in front at
emergencies, enabling bottom-up safety measures. Besides, the
three levels hierarchy permits to steadily enhance the operator’s
attention to the dangerousness of the on-going sequencing, and
the switching from one to the other mode is currently shown on
the monitor by the pertinent markers.

PERSONNEL GROUNDWORK AND ENRICHMENT

The landmines are weapons developed to be disseminated on
the ground, or close beneath the soil, and to explode because of
contact or proximity of a person or vehicle. One distinguishes:
AP anti-personnel and AT anti-tank landmines, depending on
the disruptive potential (and cost). Both have military purpose
to create forbidden zones, where the enemy could not enter.
Actually, the AP landmines low-cost transforms them into
munitions, spread in the land, making highly risking the further
resort to civil (agrarian, etc.) exploitation. On these facts, the
(Ottawa Convention 1997) bans their manufacture, trade, use
and stockpile and requests their demolition, because the armed
forces objectives cannot justify the falls-off with the human
disruption. The Convention promoted by the ONU is ratified by
several Nations, but left aside by other ones and ignored by the
population where guerrilla-like warfare is going on. The mine
risk, then, is typical problem of the under-developed countries.
The humanitarian demining does not possess viable answers,
even if long concerns and lots of efforts have been undertaken
to solve it. The idea here is to tackle it by bottom up approach,
trying to operate with smooth continuity by respect the local
technological know-how, so that the overseeing and recovery
processes could develop with appropriateness. This will help
expanding the people consciousness to become the master of
their own safety, by qualifying commonly accepted devices and
techniques and turning them towards demining aids.
The study addresses the Sri Lanka situation, assuming facts of
the in-progress guerrilla and special details on the on-the-field
conditioning peculiarities. The Govern of Sri Lanka, GoSL, has
the capability to capture the territory, as shown by the recent
east region operations, in spite of mined areas by the Liberation
Tigers of Tamil Eelam, LTTE, to stop the Government troops.
But the holding proves hard, pushing to new mine spreading, as
doable hindrance against the LTTE mobility. The humanitarian
demining is done over again, as the land is temporarily seized,
but very slowly, because of lack of facilities and of new buried
mines at troops withdrawal. The charge transforms into endless
business, not even fulfilled, before it needs to be started once
more. The business funding, Fig. 9, shows that, in the 2002-
2005 period, the mine clearing efficiency decreases and the cost
quickly rises. Finally, the cost per mine is higher than the mean
household income in Sri Lanka. This is 12 802 Rs (125 $) in
2002. It consumes huge (international and local) money,
without efficient return, out of the temporary wages for the
(ceaseless) demining. The business cannot reside in the clearing
operations, rather in restoring the community conditions for the
economical growth. If the political frame cannot give any
chance, the problem needs to be tackled by different ways.

<table>
<thead>
<tr>
<th></th>
<th>Funds</th>
<th>Cleared</th>
<th>Cleared</th>
<th>Cost for mine</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>($ million)</td>
<td>mines (N)</td>
<td>Area (m²)</td>
<td>removal ($)</td>
</tr>
<tr>
<td>2002-2003</td>
<td>6.0</td>
<td>45 000</td>
<td>16 356 445</td>
<td>133</td>
</tr>
<tr>
<td>2003-2004</td>
<td>15.8</td>
<td>44 000</td>
<td>2 155 364</td>
<td>359</td>
</tr>
<tr>
<td>2004-2005</td>
<td>23.6</td>
<td>34 000</td>
<td>3 831 553</td>
<td>694</td>
</tr>
<tr>
<td>Total</td>
<td>45.4</td>
<td>123 000</td>
<td>22 543 202</td>
<td></td>
</tr>
</tbody>
</table>

Figure 9: Funds received for humanitarian demining

The poverty and lack of future prospects are major motivation
behind landmine terrorism. The technology-driven solutions are
unsuited in 3rd world countries. The clearing cannot be solved
by the advanced robotics. Most devices work well for clean
and truthful tasks. If the environment downgrades, sophistication
and cost rise, while effectiveness and appropriateness decrease.
When the price to performance ratio is too high, the robots are
academic toys. The demining projects need to involve the local
peoples. The operators shall be familiar with the equipment,
they are required to use. Then, it is easy to teach them how to
use the rigs, and education avoids the necessity to hire foreign
specialists. The resort to foreign people is high costly: as a rule,
their salary is over 1 000 $, while the local operators work for
ten times less. People who live in the north east Sri Lanka are
very poor, and any job helps for earn some money, relieving the
most urgent necessities.
The local population’s technical knowledge is limited, and the
access to high technology fixtures is almost nonexistent. Local
materials, local manufacturing and local manpower should be
used to fulfill the demining operations and to maintain the
involved equipment. On one side, low complexity and safe-use
machines are highly desirable: an example is the simple tractor
or other agricultural machinery. The other side, the operators,
instead of endless demining, need to focus on the reclamation
effectiveness, to the quickest coming back at their farming
duties. God helps those, who help themselves.
The search of solutions, for the landmine problem in the Sri
Lanka situation, permits to identify a series of aspects to be
taken as characterising lines. Both GoSL and LTTE believe to
obtain benefits from landmine giving out, and are confident
that, under the current situation, the damages produced to the
antagonistic part result into many internally displaced people,
with severe effects on the social and economic growth of the
spoil populations, thus, with huge damage due to terrorism.
The idea behind the prospected approach is to change the
landmine spreading into basically useless intervention, so that,
out of the existing socio-political implications, no real profit is
won by terrorist actions. On the contrary, the economical falls-
off would be poor, without helpful outcomes: nearly no indirect
benefit, leaving clear the responsibility helplessness.
On the said facts, the engineering approach to the mine clearing
business reduces to devising the instrumental process, and the
work organisation, to be enabled as routine demining duty, is,
each time, crucial. The operators are enrolled on place. The
means exploit the local know-how, with resort to the standard
agricultural equipment. The efficiency assures high reaching.
The process safety and reliability are fit for the required duties.
Should these figures be reached, the terrorist effects disappear,
and the mine spreading ceases to be winning fact, during the
tactical occupation of the enemy lands, as the routine duty
minimises threats and fully avoids injuring upshots. The idea is
to bring industrial efficiency on-process, assuring high clearing
productivity, by the intelligent work-flow programming and
high reliability by special-purpose low-cost robotic equipment.
The analysis is standard issue, with the following statements:
- the engaged technologies need to use special purpose outfits,
having duty-driven consistency, and to exploit operators
adapted uniformity;
- the work-flow pre-setting ought to detail the work-cycles and
standard achievements, and to specify the on-going failure
protection rules;
- the operators’ instruction and training aim at off-process
optimised work-flows, notably, to circumvent the emergency
of risky engagements;
- the effectiveness comes from organised routine jobs, fulfilled
by the work-force diligent activity, in entire conformity to the
allotted tasks;
- the local Civil Service is entitled of the authority to promote
the mine clearing operations, and the involved community is
solidly concerned.

Figure 10: Example man-machine visual interface

The hazardous work conditions absolutely forbid the presence
of human personnel in the front effectors neighbourhoods. The
remote steering is simple, with today hardware/software tools. The large availability of personal computers allows preliminary tests on the govern logic (the real-time job allocation is fulfilled by process-computers), having full consistency with the setting. The remote monitoring, steering and governing goals are here chosen, because of:

- the capability of working in hazardous zones, with the operators in safe location;
- the ability for the remote setting and the duty autonomy (at least, to face emergencies);
- the measurements restitution, to provide human perception of the remote operations;
- the experimentation on the faults autonomic management and recovery potentials;
- the exploration of the highest level of reliability, in on-the-field operation contexts.

Explanatory hints add. The duty autonomy allows starting recovery cycles with no time delay. The operator perception is enhanced by client-server request, and remote processor updating the on-process parameters. The visual restitution (from the front cameras), Fig. 10, uses graphical interface, where the damage impact is fully understood. The reliability is achieved by the build-up of a state machine, in which the tasks are executed under the real-time processor supervision, and by the command logics, at three levels: macro (decentralised); single step (sequential); and warning (remote steering consent). The three levels, moreover, permits to steadily enhance the attention of operator to the operations dangerousness. The display keys, Fig. 10, exemplify the inter relationship between the defined control modes. The initialize mode applies to the equipment for the manually start. The control mode is in standby; when in start, the equipment is ready to perform the allotted task; when in working mode, the machine is performing the all duty-cycle. The emergency mode is helpful to detect and recover from any urgent situation, occurred while the duty-cycle progresses. In shutdown mode, the machine is in total stop. The all control modes can be by-passed by operator. However, the working and emergency modes have limited autonomous performing abilities, under the operator’s overseeing.

The resource integration, as already mentioned, is off-process operation, mainly, depending on the carrier availability. The effectors choice is more related to the ground processing state: flail outfits are successful at lower speeds to carefully sweep every region, and reaches effectiveness, if the mine blast does not affect the on-process equipment; the lifting effectors require more complex agendas, with stop-and-go manoeuvres, resort to soil loosenning and watchful picking of the dangerous items, achieving, however, high clearing reliability. The direct involvement of the local front-end operators means, also, their active decision-keeping as for the resource combination and for the work-agendas planning.

CONCLUSION

The personnel landmines are scourge of third world peoples: humanitarian demining quickly turns into ceaseless palliative, without actual issues, if the cost of the engaged techniques are too large, either the efficiency of the current operations are too low, as compared with the mine spreading potentials. The novel approach aims at joining lean and cheap contrivances and workers, with safe, reliable and effective robotic mine clearing, providing unusual productivity prospects, with widespread and friendly means and methods. The devised solution develops along four lines:

- to conceive and implement low-cost robotic equipment, assuring effective, safe and reliable demining: the resources (facilities, functions, fixtures) choice permits devising worthy strategies;
- to help out equipment integration, with resort to cheap and widely available (agricultural) fixtures: the on-process adaptive work-cycles lead to better tactical concert;
- to provide the modelling and simulation background for the on-duty assessment and training assistance: the alarm (warning, emergency) management allows joining efficiency recovery and safety;
- to make easier the operators involvement in self-learning and up-grading duties: the resources exploitation at the strategic, tactical and execution horizons exploits the mixed-mode automation.

The background is in reinterpreting some flexible-automation concepts, in keeping with the effectiveness of the dramatic productivity of the industrial organisation. The said lines are, thereafter, moved back to four steps:

# to implement effective duty-driven architectures, with the modular build-up of the robotic resources: - the mobility providers, for explanatory purpose identified by light tractors or power tiller; - the front end effectors, assuring safe and reliable mine clearing on properly wide soil extensions; the two, providing effective processing functions, according to related task-oriented proficiency;

# to integrate equipment and processes, to assure appropriate operation work-cycle and command logics, say: - ground stripe lifting, and mines singling out by gentle sieve descent of the shifted earth; - land sweeping by forward displaced flail and tiller apparatuses, inducing mines bursts; joining the task orientation with the steady/alarm processing sequences documented evidence;

# to manage the resources by suited modelling and simulation features, aiming at intelligent automation: - with the proper versatility, to face all the basic requirements on the strategic horizons; - with high effectiveness to deal with the steady agendas, on the tactical horizons; - with safe reliability, to face warning and emergencies, on the execution horizons;

# to motivate the front-end workers, by individual commitment and collaborative rewarding, enhanced by: - technological appropriateness, all the resources being directly provisioned from local equipment; - operation friendliness, all procedures being constantly adapted by the operator loyalty; together promoting the information loops incessant closure, with non-stop engagement of local people.

The paper summarises the main aspects of this challenging bet, starting by the implementation of the robot architectures, the related modularity, permitting the inter-changeability of the effectors, so that, e.g., the same carrier can be differentely fitted, depending on the soil conditions. This makes significant the integration step, to be standard practice for the off-process mine clearing strategic setting. The modelling and simulation step is even more important, being the fundamental means to manage the flexibility, by effective switching to the most appropriate engagement conditions. Additional details and comments might usefully better explain the whole project, and will be outlined during the presentation with explanatory example at the four steps: resources choice, robot integration, modelling and simulation features and operator motivation and training.
The focus on the industrial robotics capabilities is deemed the winning weapon, making possible to involve the local people in the intelligent work organisation, directly going beyond the old scientific work setting, typical of deeply structured societies. The objective is consistent with the old meaning of “industry”, “diligence, assiduous activity at any work”, spurring the really damaged farmers to tireless involvement, and with its classic denotation, namely, the «structured organisation, or systematic work or labour». The blend leads to modify, in the “trend to innovation”, the role of ideas behind the technologies (say, chiefly, the ability to prearrange assiduous labour, the scientific work organisation that assures the economic growth), because the diligence of the front-end operators (not the mastery of the individual craftsman or scientist), is winning weapon.

Today, we recognise several “industry” levels, replacing the men, with robots, managed by intelligent work organisation of self-governing facilities, each time the business awards yield from fixed assets. Nonetheless, the industrial revolution lies in the aptitude of ruling the work-force totality, by pre-assigned job allocation, once the processing outfits are properly chosen, and the pertinent know-how is duly widespread and accepted. Exclusively in such imperative situation, the technologies reach appropriateness, and lead to effective return to the involved populations. Anyway, it is always true that God helps those, who help themselves.

REFERENCES


BIOGRAPHY

Rinaldo C. Micheli, professor at the school of Engineers of the University of Genova, Italy. ASME Fellow, Member of IFIP WG 5.1 (CIME) and WG 5.5 (COVE), SCS (Senior), Russian Academy of Engineering (Correspondent), etc. Author/co-author of over than 50 technical papers/chapters/books, in the areas of integrated design, robotics, expert automation, industrial diagnostics.

Roberto P. Razzoli, researcher at the PMAR Lab of the University of Genova, Italy. His scientific activity is about eco-sustainable design, mobile and modular robots, industrial automation. He is author/co-author of over ninety publications.

Udayanga M. Hemapala obtained the degree in Electrical Engineering in Sri Lanka, and is presently at the PMAR Lab, DIMEC of the University Genova, to attend the PhD school, mechanical engineering section.
SIMULATION METHODOLOGY
SIMULATION AND AI
A Genetic Algorithm method for generating efficiency intervals in DEA with Imprecise data

Said Gattoufi, Ph.D.
College of Commerce and Economics, Sultan Qaboos University, Sultanate of Oman
Gattoufi@squ.edu.om or Algattoufi@yahoo.com

Abstract
This paper proposes an extension to the existing literature in DEA, the authors call Confident-DEA approach. The proposed new approach involves a bi-level convex optimization model, and hence NP-hard, to which a solution method is suggested. Confident-DEA constitutes a generalization of DEA for dealing with imprecise data and hence a potential method for forecasting efficiency. Imprecision in data is defined as two forms, one is bounded data and the second is cardinal data. Complementing the methodology proposed by Cooper et al (1999) which provides single valued efficiency measures, Confident-DEA provides a range of values for the efficiency measures, e.g. an efficiency confidence interval, reflecting the imprecision in data. For the general case of imprecise data, that is a mixture of ordinal and cardinal data, a Genetic-Algorithm-based meta-heuristic is used to determine the upper and lower bounds defining the efficiency confidence interval. To the best knowledge of the authors, this is the first work combining Genetic algorithms with DEA. In both cases of imprecision, a Monte-Carlo type simulation is used to determine the distribution of the efficiency measures, taking into account the distribution of the bounded imprecise data over their corresponding intervals. Most of previous DEA works dealing with imprecise data implicitly assumed a uniform distribution. Confident-DEA, on the other hand, allows for any type of distribution and hence expands the scope of the analysis. The bounded data used in the illustrative examples are assumed to have truncated normal distributions. However, the methodology suggested here allows for any other distribution for the data.

1. Introduction
In the standard Data Envelopment Analysis (DEA) approach originally proposed in Charnes, Cooper and Rhodes (1978) and further detailed in Charnes and Cooper (1985), Decision Making Units (DMUs) are considered to be economic units using inputs to produce outputs. Although the inputs and outputs are usually assumed to be observable and measurable, in many real life situations these factors are not precisely known except (i) to the extent that the true values lie within prescribed bounds, and/or (ii) to satisfy certain ordinal relations. Data falling in the first category are referred as bounded data while the later are known as ordinal data. Cooper et al (1999) refers to any mixture of such kind of data with standard single valued data as imprecise data. We henceforth use the same terminology.

When data are imprecise, the application of the standard DEA leads to a non-linear program and the piecewise linear efficient frontier defined by that approach is not guaranteed. Moreover, ordinal data cannot be considered in the standard model.

Imprecise Data Envelopment Analysis (IDEA\(^1\)), Cooper et al (1999), treats mixtures involving bounded data in addition to ordinal data and ordinal relations among the weights. However, a major conceptual criticism can be addressed to the IDEA approach. The problem lies with the derivation of single-valued measures from imprecise multivalued data. The efficiency measures should, in my opinion, reflect the imprecision in data. In this setting, a range of values for the efficiency measure is more appropriate than a single-value. Furthermore, this range can be considered as a confidence interval for the efficiency measure and hence the name Confident-DEA. The spread of the efficiency interval can be interpreted as an indicator of the degree of volatility for the efficiency of the corresponding DMU. The wider is the efficiency interval, the higher is the volatility of the efficiency and hence the higher is the uncertainty about the relative performance of the corresponding DMU.

Confident-DEA extends the standard DEA to the case of imprecise data while overcoming the above mentioned weakness by producing efficiency confidence intervals. Furthermore, it compliments the IDEA in the sense that it allows considering stochasticity in data. In the case of bounded data, Confident-DEA takes into account the distribution of values of the factors within their corresponding ranges whereas IDEA implicitly assumes the uniform distribution.

2. The Mathematical Modeling
Given the inputs and outputs to be considered, the following model is the standard Data Envelopment Analysis (DEA) model:

\(^1\) The same abbreviation e.g., IDEA for “Interactive Data Envelopment Analysis” is used in Post and Spronk (1999). The abbreviation proposed by Cooper et al. (1999) is becoming more common in the literature.
\[
\begin{align*}
\text{Max} & \quad \delta_i - \sum_{i=1}^{n} \lambda_i x_{iy} \\
\text{Subject to} & : \\
& \quad \sum_{i=1}^{n} \lambda_i x_{iy} - \sum_{i=1}^{n} \omega_i x_{iy} \leq 0 ; j = 1, \ldots, s \\
& \quad \sum_{i=1}^{n} \omega_i x_{iy} = 1 \\
& \quad y_j = (y_{ij}) \in D \\
& \quad x_j = (x_{ij}) \in D \\
& \quad \mu = (\mu_i) \in A^+ \quad \mu_j \geq 0 \\
& \quad \omega = (\omega_i) \in A^- \quad \omega_i \geq 0
\end{align*}
\]

where \( x \) represents the matrix of input values for each DMU. It specifies the values of inputs used in the production process. \( y \) on the other hand represents the output matrix. It specifies, for each DMU, the values of the different outputs that result from the production process. \( u \) and \( \omega \) are the coefficient vectors to be determined by solving the model. \( RD_j^-, RD_j^c, RA_i^c \) and \( RA_i^- \) respectively represent domains for the outputs, inputs, output multipliers and input multipliers.

Real world situations often dictate data, the values of which lie within some prescribed bounds. Moreover, the data may be ordinal rather than cardinal in form, and hence known only to be satisfying certain ordinal relations. In Cooper et al. (1999) these are labeled “imprecise data”. Lastly, the “data” may represent the decision-maker’s judgmental restrictions on the relative weights allowed to each or to some of the factors and/or multipliers. This is known in the DEA literature as the Assurance-Region. A specific domain for the solution search can be imposed and this is known as the Cone-Ratio. The general form presented above allows for all forms of data as well as all forms of restrictions on multipliers. In this paper we deal only with imprecision in Data. Restrictions are out of the scope of the work work.

In the case of imprecise data, the model presented above is not linear any longer. The standard DEA approach cannot be applied, and hence the piecewise linear efficient frontier defined that approach is not guaranteed. Formulating the basic DEA model using imprecise data leads in fact to a non-linear optimization problem.

The early literature dealing with imprecise data was simply devoted to extend the standard DEA for coping with ordinal data. Golany (1988) presented a model incorporating ordinal relations among the weights of the DEA model. Cook et al (1993) presented a framework for incorporating a single input within the standard DEA framework. In a follow-up work, Cook et al (1996) extended their framework to the case where more than one factor is ordinal. Kim et al (1999) developed a procedure for handling both ordinal data and weights preferences. Lee et al (2002) transformed the nonlinear program obtained by considering imprecise data to a linear program, using a series of modification of variables. Cooper et al (1999) developed a unified approach to treating mixtures involving bounded data in addition to ordinal data and ordinal relations among the weights. Their approach, the Imprecise Data Envelopment Analysis (IDEA), extends the standard DEA to cope with imprecise data. In a following-up work, Cooper et al (2001a) presented an illustrative application of their unified approach. Formulating the basic DEA model using imprecise data leads to a non-linear optimization problem. For the linearization, IDEA proceeds in two steps, scale transformations followed by variable alterations. The transformed model has the form of a standard DEA model. The solution for the original model is obtained from that of the transformed model using the reverse variable alterations and scale transformations.

A common criticism in this respect is that these approaches do not reflect explicitly the imprecision of the data within the assessment efficiency represented by the efficiency coefficients provided. That is, the efficiency measure obtained for each DMU is single-valued regardless the data are single-valued or imprecise.

Three major critics can be addressed regarding the IDEA approach. First, the authors, in order to linearize the model obtained from the application of standard DEA to imprecise data, transformed the status of data to variables. That is, the authors consider the factors of data not precisely known as values. This leads to an optimization problem where they decide about data as well as about variables. The basic Operations Research methodology requires a clear identification and separation between the decision variables, object of decision for the optimal level they should have, and the parameters represented by the coefficient defined by the data of the problem.

Second, for a variable defined as having bounded data, the IDEA approach requires that for the DMU used as anchor for the scale transformation and variable alteration, that is the DMU with the highest range for the corresponding bounded variable, the range is transformed into a single-valued. If this “approximation” is not made, the reverse transformations to retrieve the solution for the original problem can not be performed. This reduces some of the generality of the IDEA approach. However, this was corrected in Cooper et al (2001b) by introducing dummy DMUs in the analysis.

Finally, the major criticism is conceptual in nature. The problem with the existing literature dealing with imprecise data is the derivation of single-valued measures from imprecise multi-valued
data. The efficiency measure should reflect the imprecision in data and a range of values for the efficiency measure is more appropriate than a single-value. This range can be considered as a confidence interval for the efficiency measure. Later in this study, a new methodology, called Confident-DEA, is provided. It extends the standard DEA to the case of imprecise data and produce efficiency confidence intervals.

This paper develops a new approach, Confident-DEA that extends and generalizes the IDEA approach in the case of single valued and bounded cardinal data in the sense it allows imprecision in data to be reflected in the resulting efficiency measures. This is achieved by providing a range for the efficiency measures, an efficiency confidence interval and hence the name Confident-DEA, for each DMU instead of the single valued measure provided by the IDEA approach. A generalization to the case of imprecise data, an approach using a Genetic Algorithm based meta-heuristic for determining the bounds of the efficiency confidence interval is proposed in this work.

The upper bound confident efficiency interval for each DMU in the case of bounded data is obtained by solving the following model:

\[
\max_{x_{iy}} \max_{y_{i}} \ h_{i} - \sum_{r=1}^{r} \mu_{r} y_{r0} \tag{2.1}
\]
subject to:

\[
\sum_{r=1}^{r} \mu_{r} y_{iq} - \sum_{j=1}^{m} \omega_{j} x_{ij} \leq 0; j = 1,2, \ldots, m \tag{2.2}
\]

\[
\sum_{i=1}^{n} \omega_{i} x_{io} = 1 \tag{2.3}
\]

\[
y_{ir} \in (y_{r}) \in D_{i}^{+} \tag{2.4}
\]

\[
x_{ij} \in (x_{j}) \in D_{j}^{+} \tag{2.5}
\]

\[
\mu_{r} \in (\mu_{r}) \in A^{+}, \mu_{r} \geq 0 \tag{2.6}
\]

\[
\omega_{i} \in (\omega_{i}) \in A^{+}, \omega_{i} \geq 0 \tag{2.7}
\]

The lower bound is determined by considering the minimization model.

These two models represent a non-linear convex problems and can be written in the general form of a bilevel convex model, discussed in greater detail by Bard (1998). In the two levels of optimization, multipliers are subjects at the lower level while the factors are subjects at the upper level. The model proceeds by determining the optimal multipliers for a given level of the factors. The general mathematical form of a bilevel convex problem, where \( F, G, f \) and \( g \) are convex functions is:

\[
\max_{x,y} \ F(x,y) \tag{3.1}
\]

Subject to \( G(x,y) \leq 0 \tag{3.2} \)

\( \omega(x) = \max f(x,y) \tag{3.3} \)

Subject to \( g(x,y) \leq 0 \tag{3.4} \)

The well known Max-Min problem is a special case of the general bilevel convex problem. While Jeroslows (1985) proved that the Max-Min problem is NP-hard, Hansen et al (1992) proved that the lineal bilevel programing problem is strongly NP-hard. The significant difficulty in solving the general form, convex bilevel optimization problems, justifies the use of heuristics.

3. Confident-DEA: A Genetic-Algorithm-Based solving method for a mixture of data

3.1. The context for the genetic algorithm:

As mentioned previously, when imprecision in data is considered, the standard DEA model is not a linear program anymore. Furthermore, it can be seen as bi-level convex model, an NP-hard problem. This justifies finding heuristics solving methods. Our choice goes to genetic algorithm because, the high predisposition of the model to this meta heuristic.

Holland (1992) and his associates suggested initially in the sixties and seventies the basic principles of Genetic Algorithms. They are inspired by the mechanism of natural selection where stronger individuals are likely to be the winners in a competing environment. Through the genetic evolution method, an optimal, or a satisfactory, solution can be found and represented by the final winner of the genetic game. The name Genetic Algorithm originates from the analogy between the representation of a complex structure by means of a vector of components, and the idea of the genetic structure of chromosomes familiar to biologists. A vector, generally a sequence of 0-1 components, represents a chromosome and each component represents a gene that reflects a specific elementary characteristic. Manipulations made on chromosomes are called genetic operators and the most common are crossover and mutation.

The idea of Genetic Algorithm in optimization can be understood as an intelligent neighboring random search method. While several methods using random sampling have been used, the Genetic Algorithm approach is more flexible and provides a new framework for a variety of problems.

The original version, Holland’s version, of the Genetic Algorithm works by maintaining a population of \( M \) chromosomes considered as potential parents. Each chromosome is evaluated using a given function, and assigned a fitness value. Each chromosome encodes a solution to the problem and its fitness value is related to the
objective function value for that solution. One parent, a chromosome, is selected on a fitness basis (the better the fitness value, the higher the chance of being chosen), while the other parent is chosen randomly. They are then mated by choosing a crossover point X at random, the offspring consists of the pre-X section from one parent followed by the post-X section of the other.

The Genetic Algorithm in general allows a population composed of many individuals to evolve under specified selection rules to a state that maximizes the fitness, a measure of goodness of individuals. It emulates the survival-of-the-fittest mechanism in nature. A mating pool is extracted from the original population of individuals or chromosomes. The Genetic Algorithm presumes that each chromosome, a potential candidate, can be represented by a set of parameters called genes and can be structured by a string of values in binary form. These selected chromosomes constitute the original set of parents.

3.2. Description of the metaheuristic proposed:

The more general case of Confident-DEA proposed in this section uses Genetic-Algorithm to handle a mixture of data involving ordinal, single-valued and bounded. The steps of the meta-heuristic are described in Figure 1, Figure 2 and Figure 3.

As any meta-heuristic, the first step is the encoding process that enables representing DMUs in the standard form for Genetic Algorithm use. For each DMU, a string of numbers is defined (continuous or discrete) representing the values of factors. For the factors presumed to be known exactly (single valued), there will be a single-value substring for each. For the bounded factors, each will be represented by a substring containing all possible values obtained from the discretization of the corresponding range. That is, the final string of numbers representing the DMU will be composed of substrings each one representing the possible value(s) for one factor. The key idea in the Confident-DEA approach is to represent each DMU by a set of chromosomes, binary strings, in which each gene, 1 or 0, refers to whether or not the corresponding value is assigned to the corresponding factor.

Each DMU is split into a set of chromosomes, each one representing a virtual single-valued alternative for the real imprecise DMU.

For the illustration of the splitting-up process and generation of virtual DMUs, let a DMU using two inputs, $X_1$ and $X_2$, to produce two outputs $Y_1$ and $Y_2$. Suppose that $X_1$ and $Y_1$ are assumed to be described by exact data while $X_2$ and $Y_2$ are described by bounded data.

Let $X_1 = 20$ and $Y_1 = 30$ while $1 < X_2 < 5$ and $11 < Y_2 < 15$.

The factor’s order is arbitrarily chosen as $\{X_1 ; Y_1 ; X_2 ; Y_2\}$. The semi-columns are used only for the purpose of explanation. The string of numbers representing this DMU will then be: $\{20 ; 30 ; 2 ; 3 ; 4 ; 12 ; 13 ; 14\}$. The set of chromosomes representing this DMU will then be:


The factors’ value of the virtual DMU represented by $K_i$ are:

$$X_1 = 20; Y_1 = 30; X_2 = 2; Y_2 = 12.$$
meta-heuristic. The cutting of the matrix-individual to define the crossover points is both vertical and horizontal. The size of the initial population as well as the number of iterations is set up arbitrarily at the beginning.

![Diagram](image)

**Figure 2: Genetic modifications: Crossover and Mutation**

An initial population using the *Roulette Wheel* selection mechanisms is generated, and it constitutes the mating pool. An individual is a set of chromosomes, each one representing a DMU. All DMUs are represented in each individual and there is a single representative, a chromosome, of each DMU in each individual. The fitness function is the efficiency coefficient of the base-DMU.

Once the initial population is determined, the next phase is the creation of the next generation. This phase proceeds in three steps illustrated in Figure 2: (i) the mating of two selected individuals, considered as future parents (ii) make crossover with high probability and (iii) make mutation with low probability. All genetic modifications are decided based on the fitness of the individual determined by running a standard DEA model. The fitness measure is the efficiency coefficient of the base-DMU and it is computed for the selected individual at each step. Considering the binary matrix representing the individual, the corresponding virtual DMUs are identified. By solving the corresponding DEA model, the fitness, that is the efficiency coefficient of the base-DMU, is determined.

The process continues until a new generation is obtained. This new generation replaces the former generation and the process in initiated again. Iteration stops when the number of generations reaches the predetermined number.

The meta-heuristic proceeds in depth first, which means that all iterations are run for the first base-DMU to determine the lowest level of efficiency, then the iterations are run to determine the highest level. Once done with the first DMU, the process is iterated for the second base-DMU and so on.

Using this Genetic Algorithm based approach, summarized in Figure 3, an upper bound and a lower bound for the efficiency coefficient of each DMU are defined. Like any heuristic or meta-heuristic, obtaining an optimal solution is not guaranteed.

![Diagram](image)

**Figure 3: Methodological Contribution: Marriage of DEA with Genetic Algorithm Procedure**

3.4. A Simulation-Based Component of Confident-DEA

The third component of Confident-DEA is a simulation based heuristic. It proceeds in three phases:

(i) define the individuals in the same way described for the Genetic Algorithm based heuristic,

(ii) run a standard DEA for each individual in order to determine its efficiency coefficient and

(iii) determine the confidence interval and the distribution of efficiency coefficient for each DMU by using a Monte Carlo type simulation.

Once an individual is chosen, the efficiency coefficient of each one of its virtual DMUs is computed by solving the corresponding standard DEA model. These values are stored for future comparison. In the next iteration, the coefficients obtained are compared with previous results in order to determine the minimum and the maximum efficiency level for each DMU. Once the predetermined number of iterations is reached, the output of the heuristic has three components. First, a confidence efficiency interval for each DMU is
determined. Second, benchmarks for different levels of efficiency are identified. Finally, the distribution for the efficiency coefficient is defined based on the frequency histogram number of hits for each predefined sub-interval of [0-1]. The interval [0-1] is in fact pre-divided in a set of sub-intervals with the equal length. This predetermined length reflects the degree of precision in efficiency measure fixed by the modeler. A counter is placed in each sub-interval to record the frequency of efficiency coefficient corresponding to this sub-interval. A histogram is obtained for each DMU and the corresponding efficiency distribution is determined by smoothing the histogram.

3.5. An Illustrative Example

To illustrate the methodology and for comparative purposes, consider the data contained in Cooper et al. (2001a) summarized in Table 1. Details about the description of the real-world case, an example involving efficiency evaluations of the branch offices of a mobile telecommunications corporation in Korea, can be found in the cited reference.

<table>
<thead>
<tr>
<th>DMU</th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>Y1</th>
<th>Y2</th>
<th>Y3</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMU 1</td>
<td>124</td>
<td>18.22</td>
<td>4</td>
<td>75.53</td>
<td>89.8</td>
<td>180.85</td>
</tr>
<tr>
<td>DMU 2</td>
<td>95</td>
<td>9.73</td>
<td>2</td>
<td>18.43</td>
<td>99.6</td>
<td>185.90</td>
</tr>
<tr>
<td>DMU 3</td>
<td>92</td>
<td>8.07</td>
<td>6</td>
<td>10.29</td>
<td>87</td>
<td>175.80</td>
</tr>
<tr>
<td>DMU 4</td>
<td>61</td>
<td>5.62</td>
<td>8</td>
<td>8.32</td>
<td>99.4</td>
<td>100</td>
</tr>
<tr>
<td>DMU 5</td>
<td>63</td>
<td>5.33</td>
<td>7</td>
<td>7.04</td>
<td>96.4</td>
<td>170.75</td>
</tr>
<tr>
<td>DMU 6</td>
<td>50</td>
<td>3.53</td>
<td>3</td>
<td>6.42</td>
<td>86</td>
<td>190.65</td>
</tr>
<tr>
<td>DMU 7</td>
<td>40</td>
<td>3.5</td>
<td>5</td>
<td>2.2</td>
<td>71</td>
<td>181.85</td>
</tr>
<tr>
<td>DMU 8</td>
<td>16</td>
<td>1.17</td>
<td>1</td>
<td>9.87</td>
<td>98</td>
<td>105.40</td>
</tr>
</tbody>
</table>

The GA based heuristics is used to determine the bounds for the efficiency confidence interval. The results are presented in Table 9. This table also contains the efficiency measures obtained by Cooper et al. (2001a).

<table>
<thead>
<tr>
<th>IDE 1</th>
<th>IDE 2</th>
<th>IDE 3</th>
<th>IDE 4</th>
<th>IDE 5</th>
<th>IDE 6</th>
<th>IDE 7</th>
<th>IDE 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMU 1</td>
<td>1</td>
<td>1</td>
<td>0.994</td>
<td>1</td>
<td>0.976</td>
<td>1</td>
<td>0.995</td>
</tr>
<tr>
<td>DMU 2</td>
<td>1</td>
<td>1</td>
<td>0.863</td>
<td>0.873</td>
<td>0.846</td>
<td>0.845</td>
<td>0.811</td>
</tr>
<tr>
<td>DMU 3</td>
<td>1</td>
<td>1</td>
<td>0.584</td>
<td>0.759</td>
<td>0.265</td>
<td>0.741</td>
<td>0.256</td>
</tr>
</tbody>
</table>

One can notice a small deviation from the exact optimal solution as determined by IDE 1. This is due to the assumption of normality, which gives small weights to extreme values where the highest and lowest values of efficiency coefficients are most likely to be reached.

We define here a new concept: the efficiency dominance. This dominance can have three types:

(i) **First order efficiency dominance**: corresponds to a situation where the lowest value in the efficiency confidence interval of DMUj is greater than the highest value in the efficiency confidence interval of DMUi. In this situation, DMUj is strongly outperforming DMUi. We call this situation *First order efficiency dominance*. This is illustrated by the DMU2 and DMU4.

(ii) **Second order efficiency dominance**: corresponds to a situation where the two efficiency confidence interval of DMUi and DMUj are overlapping. In this situation, DMUi is weakly outperforming DMUj. We call this situation *second order efficiency dominance*. This is illustrated by the couple DMU5 and DMU6. Studying the distribution over each confidence interval will provide additional information about comparative performances. An interesting situation here is when an interval is totally included in the second. The DMU with narrower efficiency confidence interval range is likely to be considered outperforming since it has less volatile efficiency level. This is illustrated by the couple DMU5 and DMU7.

An interesting way of presenting the results is what we call the Efficiency Cartogram, presented in Figure 4.

![Figure 5: Efficiency Cartogram](image)

To determine the distribution of the efficiency measures over their corresponding efficiency confidence interval, the simulation component of Confident-DEA was run 100,000 times. The results are provided in the appendix. The efficiency histogram for DMU6 is in Figure 5.

![Figure 5: Efficiency Histogram for DMU 6](image)

As in the case of cardinal bounded data, the Monte Carlo simulation component for Confident-DEA permits in the case of imprecise data the approximation of the distribution of efficiency values over the efficiency confidence interval.
4. Conclusions

Over its lifespan DEA had had its roots solidly planted in real world soils. I has enjoyed practitioner acceptance. As reported in Gattoufi et al. (2004a and 2004b), its vitality is demonstrated by the high rate of literature growth and by the large number of journals included. The diversity of journals having DEA content speaks to its diffusion and to its relevance. Real world situations often dictate data, the values of which lie within some prescribed bounds. Moreover, the data may be ordinal rather than cardinal in form, and hence known only to be satisfying certain ordinal relations. A new approach for dealing with the imprecision of data in DEA is presented in this paper, to further take into account the real world facts in the DEA analysis. Called Confident-DEA, this approach makes it possible to reflect the imprecision in data in the final efficiency coefficients by providing an "efficiency confidence interval", hence the name confident-DEA, for those coefficients. This generalizes and improves the more traditional IDEA approach suggested first by Cooper et al. (1999). The spread of the efficiency confidence interval in any application may be considered as a measure of the "risk" attached to the corresponding DMU: the larger is the spread of the interval, the higher is the uncertainty in the level of the corresponding DMU’s efficiency and therefore the higher is the risk attached to the corresponding DMU. Also, the spread can be an indicator of volatility for the efficiency. The wider is the spread, the more volatile is the efficiency and hence the less is the stability of the corresponding DMU in terms of efficiency.

This paper presents an original formulation of the general case of DEA with imprecise data, namely single valued and bounded data as well as ordinal data, in the form of a bi-level convex model. This NP-hard problem has no exact solving method in the literature. A genetic algorithm based solving method is suggested and represents an additional original contribution of the paper. The solving method proceeds in two steps. First each DMU is split up into a set of chromosomes, each one representing a virtual single-valued alternative for the real imprecise DMU. An individual is defined by a set of chromosomes determined by choosing, taking into account the distribution of imprecise values to make the approach stochastic, a representative from each set of virtual DMUs with exact data representing a real DMU. It is important to remark that, unlike the standard Genetic Algorithm procedure, an individual here is represented by a binary matrix rather than a binary string. The second step is to do the genetic modifications on the individuals. An originality in the genetic algorithm approach proposed here is to consider horizontal and vertical cutting for the crossover, unlike the traditional multiple cutting on the chromosome string commonly used in genetic algorithm. This is believed to improve the efficiency of the algorithm, although it needs to be proven.

Once the range for the efficiency was determined, a Monte-Carlo simulation based method was suggested to determine the distribution of the efficiency coefficients over the confidence interval. Significantly, IDEA always results in a single valued efficiency measure and implicitly assumes a uniform distribution for the bounded data. Confident-DEA on the other hand allows the use of any distribution for the bounded data. Additionally, the simulation component proposes benchmarks, in terms of inputs and outputs, for any DMU considered and for any desired level of efficiency included in its confidence interval. The use of simulation was dictated by the inexistence of analytical results about the relation between the data distribution(s) over their intervals and the distribution of the efficiency coefficients over the efficiency intervals. This remains an open research topic.

It can be affirmed that Confident-DEA generalizes IDEA in the sense that the efficiency levels identified by IDEA for each DMU coincide with the optimistic point of view in the Confident-DEA approach, in the case of bounded cardinal data, as introduced in Gattoufi (2002 and 2004).

Finally, like any heuristic method, the genetic algorithm solving method provides, one should remind, a satisfactory solution rather than an exact solution. Hence, other solving method can be suggested and compared with the results provided by the solving method described and illustrated in this work.

A potential application for this is analysing the performance of shares in a stock market and providing advice on their future performances. Such work, in the best of our knowledge will represent an originality in the applications of DEA in general and Confident-DEA in particular. Another potential application of Confident-DEA is in predicting efficiency. Given the relative nature of the efficiency measures, forecasting efficiency measures cannot be done directly by using a timeseries methods or econometric modelling. By predicting the production factors, one can generate prediction confidence intervals. These intervals are then considered as imprecise bounded data in a Confident-DEA approach to provide the efficiency confidence interval for each DMU. The results obtained from the simulation component can be used to define a parametric approximation for the distribution of efficiency measures over their corresponding confidence intervals.
References:


OF THE STABILITY OF CLOSED SELF ORGANIZING MAPS (gSOMs)
FOR PREDICTIVE CONTROL

Matthias Reuter
CUTEC Institut GmbH
Leibniz Str. 21+23
D-38678 Clausthal-Zellerfeld, Germany
Matthias.reuter@cutec.de

KEYWORDS
Computing with Activities, Neural Impulse, Neural Impact, Prediction of Waste Incinerator Operating Conditions

ABSTRACT
In the last years the method of the Computing of Activities (CWA) has found its application in several industrial and medical implementations. So we can show that especially predictive control can be handled by this special coding of information on closed self organizing maps (gSOMs). Nevertheless until today a mathematical and application-oriented proof why CWA can guarantee predictive control and especially can guarantee stability was outstanding. Therefore the following paper presents experimental proofs of these premises and show that if a potential oriented formulation of the neural nets is used, smooth acting SOMs for the predictive control can be created. Furthermore, we show that the parameters impulse and impact of the nets leads to a better interpretation of SOMs.

INTRODUCTION
As researches in the last years show for predictive control the principle of Computing of Activities (CWA) found several applications in robotics and production control (Harnett 2005), (Birkenfeld 2008). Based on self-organizing maps CWA means not to account the winner neuron as representative of a classification concept but the global net-activity of a map. Using this principle instead of maximal 16^n concepts it is possible to store 16^n classification concepts in a map of n neurons and a 16-bit colour coding. This large storing rate makes it possible to build predictive control SOMs as now history trajectories are available. Especially if a SOM is placed on a torus, one can expect that the resulting activity pattern of the map shows symmetrical forms and no break down effects on its edges will occur.

2. MATHEMATICAL FRAMEWORK
In the context of the potential oriented formulation of the neural nets (Reuter 1999), (Reuter 2004) it is necessary to define an impulse of the neurons in succession.

Definition:
The impulse $p_i$ of a neuron i is the quantity that shows how much an input-vector $\vec{I}_i$ is changed by a neuron into its activity. Following this definition we get in formulas in the one-dimension state:

$$p_i = I_i - f_{trans}(I_i, w_i)$$

for RBF neurons, (when the Euclidian distance function is used during condition) and

$$p_i = I_i - f_{trans}(I_i, w_i)$$

for MLP neurons (when the vector scalar product is used during condition).

The problem of the given definition is, that when more than one dimensions are given, the resulting vector $\vec{p}_i$ reduces to a scalar. Therefore - from the mathematical point of view - one has to deal with the one-dimensional impulses of the different weights only. To circumvent this problem, it is necessary to introduce so-called classification potentials which are defined as

$$E_{p_i} = \frac{p_i^2}{4m}$$

With this definition we can surround the dimension problem by looking upon the activity of a neuron as a superposition of the potentials $E_{w_i}$ of the weights. Therefore the following relation holds

$$E_{w_i}(\vec{x}) = \frac{\vec{p}^2}{4m^2} = \left(x_i - w_{ij}\right)^2 + ... + \left(x_n - w_{ni}\right)^2.$$
The corresponding activities \( a_{kt} \) the resulting in

\[
\frac{\sum I_w}{\sqrt{\sum I_w - \sqrt{\sum I_w - \sum I_w \left( f_{\text{max}}(I_w - w_{in}) \right)^2}}} = \frac{\sum I_w}{\sqrt{\sum I_w - \sqrt{\sum I_w - \sum I_w \left( f_{\text{max}}(I_w - w_{in}) \right)^2}}}
\]

for RGB neurons and

\[
\frac{\sum I_w}{\sqrt{\sum I_w - \sqrt{\sum I_w - \sum I_w \left( f_{\text{max}}(I_w - w_{in}) \right)^2}}} = \frac{\sum I_w}{\sqrt{\sum I_w - \sqrt{\sum I_w - \sum I_w \left( f_{\text{max}}(I_w - w_{in}) \right)^2}}}
\]

for MLP neurons. Furthermore in the context of the potential oriented formulation of the neural nets, let us assume that the weight vector \( \tilde{w} \) of a neuron \( i \) corresponds to the representative vector \( \tilde{V}(\text{Rep}_i) \)

\[
\tilde{w} = \tilde{V}(\text{Rep}_i)
\]

whereby this weight-constant represents a common winner neuron. For the local coordinate of \( \tilde{V}(\text{Rep}_i) \) then holds

\[
\tilde{x} = \tilde{0}
\]

With this coordination system we can define the impact \( a_i \) of the neuron \( i \) by

\[
a_i = \tilde{p}_i \tilde{x}_i
\]

and the frequency function \( \Psi(x, t) \) of the activities (without the index of neurons) by

\[
\Psi(x, t) = A_0 \cdot e^{-h^{-1} \left( \tilde{p} \tilde{x} - 2m(E_p(x) + U_p(x)) \right)}
\]

whereby \( U_p(x) \) is the condition potential, \( m \) the mass of a neuron (which occurs in theory as a normalisation factor of value \( \frac{1}{2} \)) and \( h \) a semi-definite constant, which bears information about the number-representation of the calculating computer.

Supervising the real term of the last formula we get

\[
\Psi(x, t) = A_0 \cdot \cos(h^{-1} \left( \tilde{p} \tilde{x} - 2m(E_p(x) + U_p(x)) \right)) = A_0 \cdot \cos(h^{-1} \left( \tilde{p} \tilde{x} - p \tilde{x} \right))
\]

which can be observed experimentally as described below. If we deal with time independent activity patterns, \( \Psi(x) \) will be a function of \( a = \tilde{p} \tilde{x} \) only

\[
\Psi(x, t) = -h^{-1} A_0 \cdot \cos(\tilde{p} \tilde{x})
\]

If an input-vector \( \tilde{I}_i \) meets \( \tilde{V}(\text{Rep}_i) \) for \( a \) and \( p \) we get

\[
a = p = 0
\]

and therefore

\[
\Psi(x, t) = -h^{-1} A_0 \psi \eta > 1
\]

It has to be the task of experimental proof to show for which kinds of nets these formulas really hold as the structure of \( \Psi(x) \) demands that the activity pattern involves the stability of the classification concepts if smooth transfer-functions are used. Therefore it demands also stability of classification for all kinds of input-vectors - contrary to the common interpretation of SOMs classification by a winner neuron. If such structures, really exist, one of the basic conditions for predictive control by SOMs would be given (Birkenfeld, 2008).

**EXPERIMENTAL RESULTS**

For experimental proof we code an own software tool written in C++ which can simulate common SOMs with Euclidian distance calculation and vector scalar-product, topologically closed SOMs (gSOMs) with Euclidian distance calculation and vector scalar-product and neural gas with Euclidian distance calculation and vector scalar-product.

**Experimental proof 1: Dependency of the activity-, impact and impulse patterns from the input-vector**

If the input-vector \( \tilde{I}_i \) alters, theory demands that the activity, impact and impulse alter accordingly over the whole structure of the competitive layer. In Figure 2 the experimental layout of the verification software tool is shown. On the left hand side the 2-dimensional space of weights is shown. The circles mark the clusters of the momentary input vector. In the middle from top to bottom the development of the parameters courses of the activities, impulses and impacts of all neurons are shown. On the right hand side one can see the visualisation of these parameters on a common 2-dimensional map.
Analysing Figure 1 one can see, that for gSOMs the values at the beginning and the end of the curves of are equal (arrows left hand side). By that fact we proofed that the net-structure and the resulting parameters are really acting on a closed topological map and the activity patterns can be considered as “endless” without structural instability or jumps. Furthermore it can be pointed out, that the global activity structure in all representations is similar to each other.

The phase shift which occurs between activity- and impulses/impact curves (arrow from down to top) depends on which of the three input-clusters are activated by an input-vector $\vec{I}$, (see Figure 2).

Figure 3 shows the results for a normal SOM (left hand side) or a neural gas net (right hand side). Please note that the x-axis for the neural gas points out the differences of weights of the neurons, if a fixed reference neuron is given. Therefore the quasi harmonic structure of the diagrams shows evidence of the clusters of the gas.

Figure 3: Global parameters of a SOM and a neural gas with Euclidian distance calculation. It is symptomatic for open SOMs that the values differ. This fact can be used to discern SOMs from gSOMs.
Even more results can be observed, if the Euclidian distance is used, as shown in Figure 4.

**Figure 4: Global parameters of a SOM and a neural gas with Euclidian distance**

We got the same result, if the scalar product instead of the Euclidian distance is used.

**Figure 5: Global parameters of a gSOM with scalar vector product**

**Experimental proof 2: Dependence of the impact and the function \( \Psi(x) \)**

Figure 6 shows the impact function \( a \) and the function \( \Psi(x) \). Clearly is smoothing \( a \) especially around zero. Otherwise, if we compare these both functions, we see that \( a \) is more sensitive, and that’s the reason why the impact function bears more information then the activity function. It is this enlarged information which predestines the impact for a more sensitive control of the condition state and for predictive control. Otherwise \( \Psi(x) \) is predestined to detect clusters of neurons acting similar.

**Figure 6: Impact function \( a \) and \( \Psi(x) \) and normal of a gSOM**

**Remark:** In all Figures shown above a harmonic structure can be detected. This results from how we define the diagrams and easily can be pointed out by Figure 7. On the other hand it’s resulting from the fact that CWA is a stable form of classification.


**EXPERIMENTAL EXAMPLE**

Model predictive controlling of complex systems like waste incinerators in many cases is complicated due to the absence of an explicit mathematical model. Using neural networks may produce relief because of their ability to create an implicit model of the subjacent processes in their internal structure. So we used a hybrid neural network, composed of both a special kind of self-organizing map (shown next) and a downstream feed-forward net (not shown here). We show, that this architecture is capable of classifying the process history as well as predicting operating conditions while using comparatively few neurons.

The SOM used for classifying the input data defines a hexagonal neighbourhood on the neurons, which are arranged on a toroidal grid in order to avoid fringe effects. However, the crucial point is that the traditional consideration of the winner neuron [4] is irrelevant concerning the classification result. Rather, the current course of operating conditions described by the input vector is coded in the entire activity pattern of the net. The relation between the current process data and the associated activity pattern shall be demonstrated in Figure 8. To improve clearness, this example uses only four input values (oxygen and carbon monoxide contents, exhaust gas temperature and the produced steam mass). In addition the SOM visualisation uses the traditional rectangular grid instead of a toroidal one. The activity pattern refers to the rightmost twelve time steps in the diagram. The relevant curves concerning the parameter changes are the black one showing the oxygen content and the green one representing the produced steam mass.

![Figure 8: ‘Normality’, stable process behaviour](image)

The operating conditions of the plant and their changing are represented by the structure of the activity patterns and their transitions, respectively. Since different states change continuously during the
process, the SOM patterns show smooth transitions as well due to the topology preserving character of this kind of neural network. This behaviour is shown in Figure 8. We can see, that the „normal jumping“ of common SOMs will not occur any more and predictive controllers can be implemented via neural network structures.

Figure 8: Activity coded course of the process

CONCLUSIONS

As experimental results show, the theory of the CWA and the potential oriented formulation of the neural nets guaranty the stability of classification instead of the common application of winner neurons. Furthermore it leads to the new parameters impulse $p$ and impact $a$ of the neurons which enables to supervise the condition of SOMs, gSOMs and neural gas nets in a more sensitive way. Especially the impact $a$ announces in a very sensitive way, if an input vector $\vec{I}_i$ fits a representative $\vec{V}_{(Rep)}$ and so if a net is trained best. This information of the weight structure and the resulting classification concepts is one of the basic conditions to construct neural net-based predictive controllers, which demand time trajectories without gaps.

REFERENCES

Reuter, M. “About the Quantisation of the Neural Nets”, Lecture Notes in Computer Science, Springer, Heidelberg, 1999, 530-542
Reuter, M. Bostelmann, S. “Computig with Activities IV: Chunking and Aspect Integration of Complex Situations by a New Kind of Kohonen Map with integrated Long-Term Memory Structure”, ISC’2004, Malaga Spain, 2004

BIOGRAPHY

MATTHIAS REUTER was born in Dresden, Germany and went to the Clausthal University of Technology, where he studied physics and geophysics. He worked as scientist at the University of the German Army in Hamburg, the University of Dortmund and the Clausthal University of Technology. He developed a couple of new neural nets structures and learning strategies. His major investigations areas are neuronal net dynamics and robotics. In 2007 he awards the technology price of the Germany Industry of Defence for his developed system for de-mining.
INTEGRATING IMMUNITY FEATURES IN A MULTI AGENT MODEL

Saber DARMOUL
saber.darmoul@ifma.fr
URAII
INSAT, Institut National des Sciences Appliquées
et de Technologie
B.P. 676, centre urbain nord, 1080 Tunis
Tunisia

Sonia HAJRI-GABOUJ
sonia.gabouj@insat.rmu.tn

Henri PIERREVAL
henri.pierreval@ifma.fr
LIMOS UMR CNRS 6158
IFMA, Institut Français de Mécanique
Avancée, Campus des Cézeaux
B.P. 265, F-63175, Aubière
France

KEYWORDS
Modeling, Multi Agent System, Artificial Immune System, Production Monitoring.

ABSTRACT

Multi agent systems have been widely proposed to introduce intelligence in production control systems. Unfortunately, due to their generic conceptual framework, multi agent systems lack specific paradigms for disturbance handling in manufacturing environments. Biological immunity has proven its efficiency in protecting its host against the occurrence of diseases. It has already inspired successful applications in various domains such as robotics and scheduling. Therefore, we are interested here in embedding immunity features in a multi agent system devoted to the monitoring of manufacturing systems.

INTRODUCTION

In today’s manufacturing systems, the occurrence of disturbances may disrupt production flows resulting in “unplanned or undesirable states or functions of the system” (Kuivanen, 1996). Consequently, disturbance handling is a major concern in manufacturing systems (Harlin et al., 1999). Many researchers and practitioners are increasingly directing their efforts towards designing more intelligent, flexible and automated control systems. Multi agent systems have been widely proposed to introduce intelligence in these control systems (Shen et al., 2006). Unfortunately, due to their generic conceptual framework, multi agent systems do not offer specific paradigms suited to handle the disturbances occurring in manufacturing environments.

Recently, the biological immune system (BIS) inspired the design of intelligent problem solving techniques, named artificial immune systems (AIS), which have already been successfully applied in various domains such as robotics or scheduling (Luh et al., 2006, Darmoul et al., 2006). In a production monitoring context, similarly to a BIS protecting its host against the occurrence of diseases, an artificial immune system could be designed to better deal with production disruptions. We are therefore interested here in embedding immunity features in a multi agent system devoted to the monitoring of manufacturing systems.

More precisely, emphasis will be put on how agents could be used in modelling the immune cells of such a system.

BIS: A NATURAL DISTURBANCE HANDLING SYSTEM

BIS activities are directed towards detecting, identifying and eliminating malicious behaviour due to troublmaking pathogens. The BIS is organized around several cell types, such as APC, B and T cells, which are involved in coordinating immune responses to eliminate pathogens. APC cells can capture pathogens, evaluate their virulence and influence the downstream T cell behaviour. T cells are in charge of selecting appropriate response strategies through the activation of downstream effector cells, such as B cells. B cells can neutralize pathogens through the secretion of chemicals specifically tailored to pathogen morphology.

The BIS mechanisms are the basis of several successful existing applications in various domains. For example, the biological self/non self discrimination principle (Kubi, 2002) inspired the design of on line diagnosis approaches to insure anomaly and failure detection in production lines (Dasgupta et al., 1999). The immune network theory (Kubi, 2002) served the development of behaviour arbitration and action selection mechanisms in robotics (Luh et al., 2006). The biological clonal selection and memory principles were useful to several combinatorial optimization techniques for scheduling (Darmoul et al., 2006). In production monitoring and control, Mori et al., (1997) applied immune cooperation mechanisms to insure coordination between the resources of a production line.

Dasgupta, (1999), suggested an intelligent decision support system based on immunological principles. The author particularly emphasized that his multi agent system can provide a framework for general purpose decision support systems where components can be customized according to the application requirements.

We already discussed (Darmoul et al., 2007) the BIS potential to find applications in production disturbance monitoring. We particularly showed that it is worth assimilating a production system to a biological organism, and production disruptions to pathogens. We focus here on how artificial immune cells could be designed as agents so as to better deal with production disturbances.
AGENT DESIGN OF ARTIFICIAL IMMUNE CELLS

Multi agent systems can model and solve large complex problems while introducing flexibility, modularity, distributed intelligence and decentralized decision making (Shen et al., 2006). A multi-agent system is a distributed computing system which consists of interacting agents that coordinate their actions in order to fulfil tasks, either jointly or in a competitive manner (Tambe, 1998). Agents are usually endowed with knowledge, motivation and unique capabilities. They can autonomously act on their surrounding environment according to a partial representation of this environment. Information arrives to an agent, who uses its knowledge base in order to select and apply one of its methods in order to interact with the situation he perceives. Once a decision is made, the agent communicates with other agents in order to let them behave accordingly.

In a production system monitoring and control context, the biological immunity can inspire the development of several groups of artificial immune cells. These cells could be designed as interacting agents to better handle production disturbances. Each agent within this artificial immune system plays a specified role and interacts with other agents in a manner which can be described using the AGR model (Ferber et al., 1998).

As shown in Figure 1, the AGR model is based on three core concepts, namely agent, group and role. An agent is only specified as an active communicating entity which plays roles within groups. The group is considered as atomic sets of agent aggregation. The role is an abstract representation of an agent function, service or identification within a group.

![Figure 1: The Agent Group Role (AGR) Model](image)

To illustrate how this model can describe an artificial immune cell, let us consider for example a machine failure event (cf. Figure 2). Such a disruption event gives rise to a machine failure agent, which belongs to the group of disruption agents. The role of these agents is to disable resource agents. Consequently, the machine failure agent diffuses an aggression message directed towards resource agents. This message particularly stimulates the disrupted machine agent, thus interrupting the normal progress of activities already planned on this resource.

Building henceforth, other artificial immune agents, such as evaluation, decision and action agents, can be further designed. These agents directly map to the main biological immunity cells, namely APC, T and B cells. The capabilities of each agent stem from the corresponding immune cell behaviour.

![Figure 2: Disruption Agents AGR Model](image)

For example, a disruption agent integrates the aggression and stimulation features of a pathogen. Therefore, a pathogen agent aggresses some cell agents, which relate to already planned production system activities, such as material handling, manufacturing or delivery activities.

Activity agents integrate some biological features, such as biological cell lifecycle. Consequently, an activity is consecutively in an active state, an execution state and finally a completed state. Aggression messages from pathogens disrupt this normal cell lifecycle making it deviate to an agressed state. In this state, activity agents send out alarm signals directed towards evaluation and action agents.

Both evaluation and action agents integrate immune detection features, such as detection of aggression signals from pathogen agents and alarm signals from agressed cell agents.

The role of evaluation agents is to determine the consequences of the failure event on the system performance. In the case of a machine failure, evaluation agents detect this disruption event and relate its consequences to manufacturing delays and penalty costs, for example.

The role of action agents is to propose alternative solutions to problems raised by the disruption event. In the case of a machine failure for example, an action agent may be available to examine outsourcing possibilities and look for subcontractors that can provide emergency supply. Another action agent may be stimulated to resequence and/or reassign some activities to alternative resources.

Immune concepts embedded within decision agents are used to select a subset of action agents, which are best appropriate to respond to the disruption among the set of all available action agents. Besides, immune concepts are used to coordinate between selected action agents in order to determine a response well tailored to the disruption.

IMPLEMENTATION ISSUES

A prototype model is currently under implementation using the JAVA language and JADE agent development platform.

The prototype multi agent system aims to interface enterprise databases, such as Enterprise Resource Planning (ERP)
system or Manufacturing Execution System (MES), in order to determine production activities to be monitored and to update the system status. Particularly, pathogen and cell agents are created according to newly introduced activities and events.

In addition, the prototype multi agent system serves as a decision support system to human users. A decision maker supervises the system through, for example, the validation of response strategies, performing requests and analyses on induced disruption consequences, etc.

CONCLUSION

Immunity based systems have much to offer in real world problem solving. This research suggests a way of modelling an artificial immune monitoring system directed towards better handling production disturbances. This system has much to gain from integrating the features that make the success of a biological immune system in protecting its host against diseases.

These features include immune system distributed organisation and mechanisms such as detection, action selection and coordination mechanisms. The research has led us to notice that immune features integrate well within a multi agent system, wherein agent groups directly map to biological immunity cell types. These agent groups are able to interactively evolve towards a response to a disruption. A prototype multi agent immune system is currently under development to show the effectiveness of our proposals. As further research, we are directing our efforts to incorporate immune system danger perception to handle risks in production systems. Besides, we are investigating the ability to generate new situations, either in case of new disruptions, or due to a second encounter with the same disruption but in a different context.

REFERENCES


PROJECT PLAN FORMALIZATION AND MODELING VIA TIMED COLOURED PETRI NETS

Šárka Květoňová and Dušan Kolár
Department of Information Systems,
Faculty of Information Technology, Technical University of Brno
Božetěchova 2, Brno 612 66, Czech Republic
E-mail: {kvetona, kolar}@fit.vutbr.cz

KEYWORDS
Project Plan, Network Diagram, Timed Coloured Petri Nets, Formalization

ABSTRACT
The paper presents a way, how the network diagram representation of a project plan is transformed to a corresponding Timed Coloured Petri Net (T-CPN) working over discrete time. Such a T-CPN model is further equipped with resource modeling and mutually exclusive activity performance modeling. The paper demonstrates an algorithm, how the key main network diagram representations of a project plan are translated to a T-CPN and how the resource modeling is (possibly algorithmically) added next.

INTRODUCTION
The need of efficient project control is required together with increasing complexity of project resource consumption (by project, we mean a collection of mutually bound activities that build up a process). The most used and popular technique used for project planning by these days is based on a network diagram; we can say that those are methods of net analysis used as a tool for project analysis and control. Model of a project, in such a case, is the network diagram (simple directed acyclic graph with a single start node and a single end node). There are several methods of network analysis already: CPM (Critical Path Methods), PERT (Program Evaluation and Review Technique), GERT (Graphical Evaluation and Review Technique), MPM (Meta Potential Method), and others. These techniques are successful in an “offline” model of planning and monitoring—it is quite difficult to monitor and control project (plan) performance and, moreover, to model and detect resource limitations.

Thus, it is necessary to find new possibilities, way, and methods for support of “online” planning and monitoring of project performance. We use Timed Coloured Petri Nets (T-CPN) over discrete time for the purpose. The paper introduces an interactive approach to project control (monitoring of project performance according to its specification and detection of possible collisions, inefficiencies, etc.) based on T-CPN. The main motivation is to get over or to eliminate, respectively, limitations of project modeling based on network diagrams.

PRELIMINARIES
First of all, some terms from management and controlling are presented. Nevertheless, it is expected that a reader is familiar with the concept. Deeper explanation of the terms can be found in (Anbari 2005, Gray and Larson 2005, Kerzner 2005, Morris and Pinto 2007, Spinner 1992).

Definition 1 A project is a temporary effort undertaken to create a unique product or service, or result conforming to certain specifications and applicable standards. See (P.M.I. 2004, Anbari 2005).

There are many definitions of project management. Each of them respects different aspects of it. We mean two basic definitions that express the main foundation and characteristics of this term.

Definition 2 Project management is a procedure of managing and directing time, material, personnel and costs to complete a particular project in an orderly and economical manner; and to meet established objectives in time, costs, and technical results. See (Spinner 1992). Project management is the application of knowledge, skills, tools, and techniques to project activities to meet project requirements. Project management is accomplished through the use of the processes such as: initiating, planning, executing, controlling, and closing. See (P.M.I. [online]; 2004, Schwalbe 2005).

Definition 3 A process is a series of actions bringing about a result. It is a complex of mutually connected resources and activities, which changes inputs to outputs. At present, activities and resources under the project are managed almost entirely like processes. See (Anbari 2005).

A graphical view on a project is usually provided through network diagram a definition of which follows.
Definition 4 A project network diagram, shortly network diagram, is a graphical representation of actions sequences that may compose a project performance. A network diagram shows flow of actions in time during process development. The diagram enables ordering of actions in a sequence or in parallel branches, it shows their mutual collisions, consequences, inefficiencies, time gaps, overall process length, etc. See (Veber 2006).

A formal view on the network diagram is provided via graph theory (see (Šeda 2003)). A standard definition of an oriented graph is used as a basis for the following definition used next in the paper:

Definition 5 A network, $S$, is a quadruple $S = (G, z, s, c)$, where $G$ is an oriented graph, $z, s \in V(G)$ are two special vertices of graph $G$—there is no input edge of the vertex $z$ and no output edge out of the vertex $s$, $c$ is a function, $c : E(G) \rightarrow N$, that assigns to every oriented edge $e$, $c \in E(G)$, a natural number $c(e)$. The vertex $z$ is called source (or start), the vertex $s$ is called sink (or end), and the number $c(e)$ is called capacity (or throughput) of the edge $e$.

The following definitions introduce necessary background to Petri Net theory. Broader and deeper explanation can be found in (Češka 1994, Girault 2003, Rozenberg 1991).

Definition 6 A net, $N$, is a triple $N = (P, T, F)$ such that

- $P$ and $T$ are mutually disjoint sets such that:
  - $P = \{p_1, p_2, \ldots, p_n\}$ is a finite set of places
  - $T = \{t_1, t_2, \ldots, t_m\}$ is a finite set of transitions
- $F$ is a binary flow relation: $F \subseteq (P \times T) \cup (T \times P)$.

Definition 7 A P/T Petri Net, $N$, is a six-tuple $N = (P, T, F, W, K, M_0)$, where:

- $(P, T, F)$ is a finite net.
- $W : P \rightarrow (N \setminus \{0\})$ is a non-negative weight of every net edge.
- $K : P \rightarrow (N \cup \{\omega\})$ is a map denoting capacity of a place.
- $M_0 : P \rightarrow (N \cup \{\omega\})$ is initial marking of the Petri Net such that $\forall p \in P : M_0(p) \leq K(p)$.
- $N$ is a node function defined as $N : A \rightarrow (P \times T) \cup (T \times P)$.
- $C$ is a colour function defined as $C : P \rightarrow \Sigma$.
- $G$ is a transition guard function defined as $G : T \rightarrow EXPR$ such that $\forall t \in T : Type(G(t)) = B \land Type(Var(G(t))) = \Sigma$.
- $E$ is an arc expression function defined as $E : A \rightarrow EXPR$ such that $\forall a \in A : Type(E(a)) = C(p(a))_{MS} \land Type(Var(E(a))) \subseteq \Sigma$, where $p(a)$ is a place in $N(a)$, $C(x)_{MS}$ denotes that function $C$ returns for $x$ a multiset as a result.
- $I$ is an initialization function $I : P \rightarrow CEXPR$ such that $\forall p \in P : Type(I(p)) = C(p)_{MS}$.

Next, if $(p, t) \in F \cap (P \times T)$ then we say that $p$ is the entry place and $(p, t)$ is the entry edge of the transition $t$. Similarly, if $(t, p) \in F \cap (P \times T)$ then we say that $p$ is the exit place and $(t, p)$ is the exit edge of the transition $t$.

Definition 8 Let $N = (P, T, F)$ be a net. For all $x \in P \cup T$ we define:

- $\bullet_x = \{y \mid yFx\}$ to be a preset of the element $x$,
- $\bullet^x = \{y \mid xFy\}$ to be a postset of the element $x$.

Definition 9 Let $N = (P, T, F, W, K, M_0)$ be a Petri Net then

- Mapping $M : P \rightarrow (N \cup \{\omega\})$ is called marking of the Petri Net $N$ if $\forall p \in P : M(p) \leq K(p)$
- Let $M$ be a marking of a Petri Net $N$. Transition $t$, $t \in T$, can fire in marking $M$ if
  - $\forall p \in \bullet^t : M(p) \geq W(p, t)$
  - $\forall p \in \bullet_t : M(p) \leq K(p) - W(t, p)$

A status of a system modeled by a Petri Net is denoted by a marking of such a Net. A dynamic behavior of a modeled system is denoted by the transition firing. Influence of a Net marking on the possibility of transition firing and, vice versa, influence of transition firing on the Net marking evolution is obvious.

Definition 10 A Non Hierarchical Coloured Petri Net (CPN), $N_C$ is a nine-tuple $N_C = (\Sigma, P, T, A, N, C, G, E, I)$ such that:

- $\Sigma$ is a finite set of finite non-empty types called set of colours;
- $P$ is a finite set of places;
- $T$ is a finite set of transitions such that $P \cap T = \emptyset$;
- $A$ is a finite set of edges such that $A \cap P = A \cap T = \emptyset$;
- $N$ is a node function defined as $N : A \rightarrow (P \times T) \cup (T \times P)$;
- $C$ is a colour function defined as $C : P \rightarrow \Sigma$;
- $G$ is a transition guard function defined as $G : T \rightarrow EXPR$ such that $\forall t \in T : Type(G(t)) = B \land Type(Var(G(t))) \subseteq \Sigma$;
- $E$ is an arc expression function defined as $E : A \rightarrow EXPR$ such that $\forall a \in A : Type(E(a)) = C(p(a))_{MS} \land Type(Var(E(a))) \subseteq \Sigma$, where $p(a)$ is a place in $N(a)$, $C(x)_{MS}$ denotes that function $C$ returns for $x$ a multiset as a result;
- $I$ is an initialization function $I : P \rightarrow CEXPR$ such that $\forall p \in P : Type(I(p)) = C(p)_{MS}$.
To complete the definition, we express that type of a variable, \( v \), is denoted as \( \text{Type}(v) \). If \( V \) is a set of variables then \( \text{Type}(V) = \{ \text{Type}(v) \mid v \in V \} \). Let \( \text{expr} \) be an expression, then type of the expression \( \text{expr} \) is denoted as \( \text{Type}(\text{expr}) \). Set of all variables inside an expression \( \text{expr} \) is denoted as \( \text{Var}(\text{expr}) \). An expression \( \text{expr} \) is closed if \( \text{Var}(\text{expr}) = \emptyset \). Let \( \text{EXPR} \) be a set of all well defined expressions in a used inscription language, then \( \text{CEXP} \) is a set of all closed expressions in a used inscription language such that \( \text{CEXP} \subseteq \text{EXPR} \). Finally, \( \mathcal{B} \) stands for set of boolean values, \( \mathcal{B} = \{ \text{true}, \text{false} \} \).

**MAIN RESULTS**

The transformation of network diagram into a T-CPN is performed over edge-oriented network diagram. Key features of such a network diagram are the following:

- **Pros:**
  1. To evaluate overall lasting of a project or to identify the soonest possible starts of particular actions, it is possible to use algorithm of cheapest/most expensive path in a graph.
  2. It saves graph components in comparison with node-oriented network diagram.

- **Cons:**
  1. Quite a difficult network construction—the usual input is a table describing actions and their sequencing.
  2. It is necessary to use fictive vertices to ensure synchronization, quite often.

![Figure 1: Example of Edge-Oriented Network Diagram](image)

The Timed CPN, which is a target of the transformation, is defined as follows:

**Definition 11** A Timed CPN (T-CPN), \( N_t \), is a ten-tuple \( N_t = (\Sigma, P, T, A, N, C, G, E, I, \tau) \) such, so that:

\[ N_C = (\Sigma, P, T, A, N, C, G, E, I, \tau) \text{ is a CPN; } \tau \text{ is a mapping } \tau : T \to \mathbb{N_0} \text{ that assigns to every transition non-negative discrete time period denoting amount of time that must elapse during transition firing.} \]

Transitions can fire at the same conditions as in CPN, a difference is in marking modification. Starting with initial marking:

1. Set up the timer counter to zero.
2. Detect a set, \( F \), of all transitions that can (possibly) fire.
3. According to previous step, denote a set, \( E \), of all transitions that can fire simultaneously—in a fact, all transitions from previous detection can fire, nevertheless, if there is a conflict (for simultaneous firing) in resource sharing then conflicting transitions are resolved in a given way (according to a specified strategy, i.e. random one).
4. Split the set \( E \) into two sets, \( E_0, E_\Delta \) such a way, so that \( E = E_0 \cup E_\Delta, E_0 \cap E_\Delta = \emptyset \) and, moreover, let the set \( E_0 \) contains only transitions with 0 ticks time elapse during transition firing and let the set \( E_\Delta \) contains only transitions with more than 0 ticks time elapse during transition firing.
5. Start firing of all transitions in the set \( E \)—start of a firing represents just removal of tokens from input places of firing transitions and, moreover, reduction of a capacity of output places (of firing transitions) according to the number of tokens that will be assigned to a given place, when the transition firing finishes.
6. Finish firing of transitions in the set \( E_0 \)—increase capacities of transition output places, transfer new tokens to appropriate output places (of firing transitions).
7. If there is any transition that can fire then go to the step 2.
8. Increment the timer counter by 1.
9. Detect a set, \( C \), of all firing transitions a firing of which is finished in a given time period.
10. Finish firing of transitions in the set \( C \)—increase capacities of transition output places, transfer new tokens to appropriate output places (of firing transitions).
11. Go to the step 2.

**Result**

The edge-oriented network is mapped to T-CPN such a way:

1. Set \( P = \{ v \mid v \in V \} \)—every vertex of a diagram is mapped to a specific place.
2. Set \( P = P \cup \{ RS \} \)—place modeling resources is added.
3. Set $P = P \cup \{ <pa> \mid p,q \in V, a \text{ is an action on the edge outgoing from the vertex } p \text{ to the vertex } q \}$—see steps 9 and 10.

4. Set $P = P \cup \{ <aq> \mid p,q \in V, a \text{ is an action on the edge outgoing from the vertex } p \text{ to the vertex } q \}$—see steps 11 and 12.

5. Set $T = \{ r_{a/t} \mid r \text{ is an edge in the network diagram with assigned action } a \text{ lasting } t \text{ time units} \}$—see steps 10 and 11.

6. Set $T = T \cup \{ z^p \mid z^p \text{ is a transition with no assigned action and with zero delay during firing such that it represents all input edges of a vertex } p \text{ from the network diagram} \}$—see steps 12 and 13.

7. Set $T = T \cup \{ z_p \mid z_p \text{ is a transition with no assigned action and with zero delay during firing such that it represents all output edges of a vertex } p \text{ from the network diagram} \}$—see steps 8 and 9.

8. Set $A = A \cup \{ e_{z,v} \mid v \in P, z_v \in T \}$—all outgoing edges of the vertex $v$ are mapped to a single outgoing edge from particular place in the T-CPN to a 0 time transition $z_v$ (step 7).

9. Set $A = A \cup \{ e_{z,v} \mid z_v \in T, <va> \in P \}$—0 time transition $z_v$ represents outgoing edges from a vertex $v$; for details about place $<va>$ see step 3; from such places, appropriate number of edges (the same number as in the network diagram is outgoing from the vertex) is outgoing to fresh (newly introduced) transitions in the step 7.

10. Set $A = A \cup \{ e_{<va>,r_{a/t}} \mid <va> \in P, r_{a/t} \in T \}$—places, $<va>$, representing input to actions on edges of the network diagram (step 3) are connected with actions on them represented by transitions $r_{a/t}$ (step 5).

11. Set $A = A \cup \{ e_{r_{a/t},<av>} \mid <av> \in P, r_{a/t} \in T \}$—transitions, $r_{a/t}$, representing actions with appropriate time consumption (step 5) are connected with places, $<av>$ (step 4), representing output of actions on edges of the network diagram.

12. Set $A = A \cup \{ e_{<av>,z^v} \mid <av> \in P, z^v \in T \}$—places, $<av>$ (step 4), representing output of actions on edges of the network diagram are connected with transitions, $z^v$, representing all input edges of a vertex $v$ (see step 6).

13. Set $A = A \cup \{ e_{z,v} \mid v \in P, z^v \in T \}$—transitions representing all input edges of a vertex $v$ ($z^v$, see step 6) are connected with appropriate places, $v$, representing such vertices.

14. Set $A = A \cup \{ e_{RSr_{a/t}},e_{r_{a/t}RS} \mid RS \in P, r_{a/t} \in T \}$—
Resources are modeled by a single place, $RS$. It is connected by outgoing edges to every non-zero delayed transition (those modeling actions) and by incoming edges from the same transitions.

15. Set $E$, $G$, and $\Sigma$ to appropriate values to establish resource consumption/modeling/sharing/usage/creation.

16. Set $I$ such a way, so that a place $p$, $p \in P$, representing start place of network diagram carries appropriate tokens, moreover, add to $I$ settings for place $RS$ to model resources, for remaining places, set $I$ to carry no tokens at all.

Such a model can be extended with variant paths to model choice on the level of actions, which is not possible in the network diagram unless we use some extensions. Moreover, we can set-up any situation and, using traditional techniques for T-CPN, we can retrieve information about possible further scenarios and time development and overall resource consumption. In such a case, traditional techniques for analysis of P/T Petri Nets are used in the beginning. When it is verified that a Net is correctly designed T-CPN behavior simulation techniques are used to provide required information about timing, resource blocking, etc.

**Example** A network diagram shown on Figure 1 is transformed to the T-CPN demonstrated on Figure 2 using the presented algorithm. Moreover, introduction of resource modeling is sketched on Figure 2 too.

On Figure 2, for the sake of readability, the resource modeling is demonstrated just for two transitions (labeled $A/10$ and $C/12$)—state modeling resources are in the upper left corner of the figure with cross hatching on its background. Moreover, dark black transitions are modeling features of operation flow split/join and they do not have a direct binding to some part of the network diagram (they represent a network diagram substructure-steps 6 and 7 of the mapping algorithm). The light grey transitions stay for edges of the network diagram—they represent actions with the time amount necessary for their completion (step 5 of the mapping algorithm).

The place marked $RS$ models resources available and created during the process development (see beginning of the previous paragraph). Places with slanted hatching on their background represent direct mapping of vertices from the network diagram. Finally, the white places (empty background) do not have a direct representation in the network diagram and they are introduced for synchronization purposes and they represent a substructure of a network diagram (steps 3 and 4 of the mapping algorithm).

Marking of edges is omitted in the example, too, to make it more readable. In a fact, it represents resource usage, consumption, and release/creation during action performance. We expect that it is inserted either by an analyst or, automatically, from a textual description in an appropriate format.

**CONCLUSIONS**

The paper presented a conversion algorithm of a network diagram to a T-CPN in order to achieve all features provided by the network diagram and, moreover, to gain new features via Petri nets. It was demonstrated that some features hard to achieve in a network diagram can be easily obtained in T-CPN. Thus, a new possibilities of analysis are introduced to improve project planning both for newly established and running process. Besides better static analysis, it can be used a dynamic analysis and, especially, resource sharing modeling, which is impossible in the network diagram. Such a way, we study conflicts that can appear during process development and, thus, we are ready to provide a solution that disables such a conflicts and/or does not lead to a critical situation or deadlock.

**OPEN QUESTIONS**

Even if the presented concept of T-CPN is based on a discrete time, it provides good features for both static and dynamic analysis. Nevertheless, there are some questions still open:

- Could we algorithmically simplify/reduce the T-CPN without any distortion of their key properties as timing, resource sharing modeling, etc.?
- Can we unroll processes to sub-processes even if they are cyclic ones and still keep the availability of good and easy-to-do static analysis of T-CPN features?

We expect resolution of these questions in a following work over this topic.

**ACKNOWLEDGEMENTS**

This research was supported by the Research Plan No. MSM 0021630528, Security-Oriented Research in Information Technology.

**REFERENCES**


**AUTHOR BIOGRAPHY**

ŠÁRKA KVĚTOŇOVÁ was born in 1981, Brno, Czech Republic. She has studied Economics and Management at the Brno University of Technology, Faculty of Business and Management, Brno, Czech Republic. Presently, she is Ph.D. student at Brno University of Technology, Faculty of Information Technology, Brno, Czech Republic. Her main research interests are software engineering, processes and project management with focus on Petri Net techniques and their application for software projects management.

DUŠAN KOLÁŘ was born in Hradec Králové, Czech Republic, and went to Brno University of Technology, Czech Republic, where he studied computer science and cybernetics and obtained his degrees in 1994 and 1998. Since then, he has been working at the university, presently at the Faculty of Information Technology. His main research interests are formal languages and automata and formal models with focus on Petri Nets and reactive/triggered models.
IDENTIFICATION OF MULTIVARIABLE NARMA MODELS USING ARTIFICIAL NEURAL NETWORKS

Brahim Tili, Fauzi Bouani and Mekki Ksouri
Ecole Nationale d'Ingénieurs de Tunis
BP 37, Tunis Belvédère, 1002, Tunis
Tunisie
E-mail: Brahim.Tili@enit.rnu.tn

KEYWORDS
Identification, Nonlinear Systems, NARMA Model, Genetic Algorithm, Neural Networks.

ABSTRACT
This paper deals with the identification of multivariable nonlinear systems using the Nonlinear Auto Regressive Moving Average (NARMA) model. This type of model can characterize several classes of nonlinear systems. Besides, it has the advantage of suitably characterizing the system behaviour by exploiting a limited number of terms in the output expression. However, the complicated part of NARMA modeling is the model structure. In this paper, we shall suggest an approach to model complex systems using a procedure based on a single layer neural network with polynomial activation functions which are optimised with a real-coded genetic algorithm. This approach allows the determination of the significant terms in the expression of the NARMA model as well as the estimate of the corresponding coefficients. Two examples were given to validate the performances of the suggested method. The first one is a simple simulated system, whereas the second represents a physical process; a laboratory interconnected three tanks system.

INTRODUCTION
This paper deals with the structural and parametric identification of the NARMA type models for the multivariable systems. NARMA model requires a reduced number of parameters since it utilizes a limited number of terms in the output expression (Liu et al. 2003). Several approaches have been proposed to determine the NARMA type models (Billings and Leontaritis 1982). In (Kortmann and Unbehauen 1988) the authors proposed the modified stepwise regression for nonlinear dynamic system algorithms, which is based on statistical tests to identify terms and parameters which form the NARMA model. The binary genetic algorithm was also used for the identification of NARMA model (Tili et al. 2006). In (Sheng et al. 2001) and (Sheng and Chon 2003), the authors used affine geometry and "hyper surface" distance minimization for the estimation of the NARMA model parameters. This approach is based on the combination of the Optimal Parameter Search (OPS) algorithm and the method of Total Least Squares (TLS). The approach makes it possible to estimate the parameters of the NARMA model even in the presence of a significant noise. However, it is computationally intensive and the analytical solution cannot be calculated for systems of order higher than three (Sheng and Chon 2003).

The NARMA models which are nonlinear with respect to the old output values but linear with respect to the current input values were also developed (Adetona et al. 2004). These models, obtained through a development in Taylor series of neural models are useful for the synthesis of the regulators. The work of Ki and Chon (1997) presents a method which exploits the Artificial Neural Networks with polynomial activation function to estimate ARMA and NARMA models for Single-Input Single-Output (SISO) systems. However, no approach has been suggested for the determination of the activation function coefficients. Note that the quality of the obtained model depends on the adequate choice of the polynomial activation function coefficients. The majority of the methods presented are applicable for SISO or Multi-Input Single-Output (MISO) systems type. Little work was done in the multivariable NARMA models.

In this paper, we propose an approach to model complex systems using a single layer neural network with polynomial activation functions. The coefficients of the functions are optimised using a real-coded genetic algorithm. In this case, the population individuals of the genetic algorithm characterize the coefficients of the polynomial of the activation function. The parameters of the NARMA model are then estimated from connection weights of the neural network after the end of the training phase. The rational behind this approach is that there is no a priori knowledge of the system structure.

The paper is organized as follows. In the second section, we present the NARMA model structure. The third section is reserved for the development of the approach used to estimate the structure and the parameters of the model. The fourth section presents the simulation as well as experimental results. Conclusions are drawn in the last section.

2. THE NARMA MODEL
Consider the class of the nonlinear systems which can be modeled by the following relation:

\[ Y(k) = G(\phi(k)) + E(k) \]  \hspace{1cm} (1)

with \( \phi(k) = [Y(k-1),...Y(k-n)U(k-1),...U(k-m)] \).
\( Y(k) \) and \( U(k) \) are, respectively, the output and the input vectors and \( G \) is a presumed unknown nonlinear function which can contain all the possible combinations (linear and nonlinear) between \( Y \) and \( U \). Parameters \( m \) and \( n \) represent, respectively, the input regression order and the output regression order. \( E(k) \) is a white noise vector.

Consider a multivariable system having \( n_1 \) outputs and \( m_1 \) inputs. Each output can be expressed according to the observation vector by the following relation:

\[
y_i(k) = g_i(\phi(k)), \quad i = 1, \ldots, n_1
\]

(2)

with \( g_i \) is a nonlinear function and \( \phi(k) \) is formed by old measurements of all the outputs and all the inputs:

\[
\phi(k) = [u_1(k-1), \ldots, u_1(k-m), \ldots, u_m(k-1), \ldots, u_m(k-m) \ldots y_1(k-1), \ldots, y_1(k-n), \ldots, y_n(k-1), \ldots, y_n(k-n)]
\]

(3)

The number of possible terms which can intervene in the expression of each output of the model \( \hat{y}_i(k) \) is given by the following relation:

\[
r = \frac{(n+n+m+m+q)!}{q!(n+n+m)!} - 1
\]

(4)

where the parameter \( q \) represents the nonlinearity model order. Consequently, the number of possible models which one can have for a system with \( n_1 \) outputs is:

\[
w = (2^r - 1)^{n_1}
\]

(5)

For a given \( m, n, q, n_1 \), and \( m_1 \), there is a very high number of possible combinations. For example, if the system has two inputs and two outputs (\( m_1=2 \) and \( n_1=2 \)), for \( m=n=q=2 \), one will have 3.0949 \( 10^{26} \) different combinations. It is thus necessary to use a procedure allowing the selection of a model among all the possible models which represents the system behaviour.

The dynamic structure of the general NARMA model output, in a deterministic environment, is given by the discrete mathematical equation:

\[
\hat{Y}(k) = \begin{bmatrix} \hat{y}_1(k) & \hat{y}_2(k) & \ldots & \hat{y}_{n_1}(k) \end{bmatrix}
\]

(6)

Every output \( \hat{y}_i(k) \) \( (i = 1, \ldots, n_1) \) of the model can be written in the following form:

\[
\hat{y}_i(k) = \hat{\Theta}_i^T \Phi(k)
\]

(7)

with \( \hat{\Theta}_i^T \) and \( \Phi(k) \) indicate, respectively, the parameter vector and the observation vector:

\[
\hat{\Theta}_i = \begin{bmatrix} \hat{\theta}_1 & \theta_1 & \theta_2 & \ldots & \theta_r \end{bmatrix}
\]

\[
\Phi(k) = \begin{bmatrix} 1 & v_1 & v_2 & \ldots & v_r \end{bmatrix}
\]

(8)

(9)

where

\[
\begin{align*}
\phi_1(k) &= \phi_{11}(k-1), \ldots, \phi_{1n}(k) = \phi_{11}(k-m), \\
\phi_{n_1,1}(k) &= \phi_{n_11}(k-1), \ldots, \phi_{n_1n}(k) = \phi_{n_11}(k-m), \\
\phi_{n_1n_1}(k) &= y_1(k-1), \ldots, \phi_{n_1n_1n}(k) = y_1(k-n), \ldots, \\
& \vdots \\
\phi_{n_n_n}(k) &= u_1(k-l)^r, \ldots, \phi_{n_n_n}(k) = u_1(n-r)
\end{align*}
\]

(10)

The method used for the identification of the NARMA type models is presented in the next section.

3. IDENTIFICATION OF THE NARMA MODEL

Our method uses the input-output system measurement file: \{\(Y(k),U(k), k=1, \ldots, N\}\} offline. Indeed, the parameters of the NARMA model can be estimated using neural networks with polynomial activation function (Ki and Cohen 1997). The structure of the neural networks is of type “feed-forward” with only one hidden layer. The topology of this network is represented in Fig. 1. The neural network receives as input the elements of the vector \( \phi(k) \) (ref. Equation 3) and delivers at its output an estimate of the current output.

![Neural Network structure with one hidden layer](image)

Figure 1: Neural Network structure with one hidden layer

The outputs of the first layer are given by:

\[
S(k) = f(W^1\Phi^T(k)) = [s_1(k) \ldots s_M(k)]^T
\]

(11)

\( W^1 \) is the matrix of weightings between the input layer and the hidden layer. \( f \) is a polynomial activation function and \( M \) indicates the number of neurons of the hidden layer. The relation (ref. Equation 12) presents a polynomial activation function of order \( q \):

\[
S(k) = f(W^1\Phi^T(k)) = [s_1(k) \ldots s_M(k)]^T
\]
\[ f(x_i) = f_0 + f_1 x_i + f_2 x_i^2 + \ldots + f_q x_i^q \]  \hspace{1cm} (12)

Taking into account the weights of connections between the input layer and the hidden layer, the quantity \( x_i \) will be expressed by:

\[ x_i = \sum_{j=1}^{n_i} w_{ij} (k-j) + \sum_{j=0}^{m_i} w_{ij} x_j (k-j+m) + \ldots + \sum_{j=0}^{m_i} w_{ij} x_j (k-j+m(m-1)+1) + \]
\[ \sum_{j=n_i+1}^{m_i+n_i} w_{ij} x_j (k-j+m(m+n)) + \ldots + \sum_{j=n_i+m_i+1}^{m_i+n_i+n_i} w_{ij} x_j (k-j+m+n(n-1)) \hspace{1cm} i \in [1, M] \]  \hspace{1cm} (13)

The choice of \( M \) depends on the complexity of the system. Indeed, a small number of hidden neurons can lead to a poor model, whereas the choice of a high number of hidden neurons can make the model very complex. In general, the number of hidden neurons is selected in order to carry out a compromise between the precision and the complexity of the estimated model.

The network output equation will then be:

\[ \hat{Y}(k) = \left[ \hat{y}_1(k), \ldots, \hat{y}_{n_i}(k) \right]^T = W^2 S(k) \]  \hspace{1cm} (14)

where \( \hat{y}_j(k) = \sum_{i=1}^{M} w_{ij}^2 f(x_i), \hspace{1cm} j \in [1, n_i] \)

\( W^2 \) is the matrix of weightings between the hidden layer and the output layer.

Replacing the quantity \( f(x_i) \) of the Equation 12 by its value in Equation 14, one obtains:

\[ \hat{y}_j(k) = w_{1j}^2 \left( f_0 + f_1 x_1 + \ldots + f_q x_q \right) \]
\[ + w_{2j}^2 \left( f_0 + f_1 x_1 + \ldots + f_q x_q \right) \]
\[ + \ldots + w_{Mj}^2 \left( f_0 + f_1 x_M + \ldots + f_q x_M \right) \]  \hspace{1cm} (15)

Replacing \( (x) \) by its expression given by (Equation 13), the expression of every output of the network for a nonlinearity degree \( q \) will be as follows:

\[ \hat{y}_j(k) = \Theta_f \hat{D}(k) \]  \hspace{1cm} (16)

where \( \Theta_f = h(f_0, f_1, \ldots, f_q, W^1, W^2) \), \( h \): nonlinear function, \( \dim \Theta_f = [r+1, 1] \).

The weights (\( W^1 \) and \( W^2 \)) of the neural network connections are obtained using the gradient algorithm. The deduction of the parameters of NARMA model is done from weightings (\( W^1 \) and \( W^2 \)) and the parameters of the considered polynomial \( (f_0, f_1, \ldots, f_q) \). Indeed, using Equation 16 and Equation 6, one finds the parameters of the NARMA model. The quality of the obtained model using this method depends on the adequate choice of the polynomial coefficients \( (f_0, f_1, \ldots, f_q) \). For the sake of reducing the number of trials, we suggest to use a real coded GA with the ANN, (Figure 2).

Figure 2: Structure of supervisor based on genetic algorithm

Each individual of the genetic algorithm population characterizes the neural polynomial activation function coefficients. The initial population of the genetic algorithm is chosen in a random way from a fixed interval. Every generation of the genetic algorithm consists of taking the coefficients given by the individual of the population and training the neural network. For every individual, we calculate an evaluation function (i.e. the sum of the quadratic errors between the system output and those of the neural networks). Then, the genetic operators (selection, crossover and mutation) are used to form the next generation individuals (Goldberg 1989). In this work the selection based on the lottery wheel, the arithmetic crossover and the uniform mutation is considered.

The criterion \( J \) which intervenes in the evaluation function (fitness) of the real coded genetic algorithm is given by:

\[ J = \sum_{i=1}^{N} \left( (\hat{y}_i(k) - \hat{y}_i(k))^2 + \ldots + (\hat{y}_{n_i}(k) - \hat{y}_{n_i}(k))^2 \right) \]  \hspace{1cm} (17)

With this method, the calculation of the NARMA model parameters is done after the neural network training phase. The role of the genetic algorithm is to find the adequate coefficients of the polynomial while the ANN takes charge of the adjustment of the connections weights.

4. PERFORMANCE EVALUATION

During the identification phase, the input signal used is a white noise with zero mean and a variance one. The total correlation coefficient \( R^2_{tot} \), is calculated in order to decide if the obtained model is acceptable or not. The model is acceptable if \( R^2_{tot} \) is close to one (Haber and Unbehauen 1990), where:

\[ R^2_{tot} = 1 - \frac{\sum_{i=1}^{N} (y_i(k) - \hat{y}_i(k))^2}{\sum_{i=1}^{N} (y_i(k))^2} \hspace{1cm} (18) \]

The adjustment parameters used in the simulations are summarized in Table 1.
Table 1: Algorithms parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>GA</th>
<th>ANN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of generations</td>
<td>40</td>
<td>-</td>
</tr>
<tr>
<td>Number of individuals</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td>Probability of crossover</td>
<td>0.8</td>
<td>-</td>
</tr>
<tr>
<td>Probability of mutation</td>
<td>0.1</td>
<td>-</td>
</tr>
<tr>
<td>Number of epochs</td>
<td>-</td>
<td>3000</td>
</tr>
<tr>
<td>Coef. of training</td>
<td>-</td>
<td>0.005</td>
</tr>
</tbody>
</table>

In this section two nonlinear (but linear in the parameters) multivariable systems, having two inputs and two outputs, are considered.

System 1: the first system is described by a NARMA type equation given as follows:

\[
y_1(k) = 0.8 y_1(k-1) + u_1(k - 2) + 0.4(u_1(k - 2))^2 \]

\[-1.2u_1(k-1)u_2(k-2) - 0.1y_2(k-1). \quad (19)\]

\[
y_2(k) = 0.5 y_2(k-1) + u_2(k - 2) + (u_2(k - 1))^2 \]

\[+ 0.5y_2(k-2)u_2(k-1). \quad (20)\]

The choice \(n=m=q=2\) makes it possible to have 44 terms in the NARMA model expression. Equation 20 presents the identified NARMA model. The coefficients of the activation function found by the real genetic algorithm are: \(f_0=15; f_1=-4.16\) and \(f_2=13.5\).

It is noticed that the good choice of the activation function coefficients made it possible to achieve a good precision.

\[
y_1(k) = 0.7999 y_1(k-1) + u_1(k - 2) + 0.3998(u_1(k - 2))^2 \]

\[-1.2u_1(k-1)u_2(k-2) - 0.1y_2(k-1). \quad (20)\]

System 2: the second multivariable system, schematised by Figure 3, presents a laboratory process of water regulation level. It consists of three cylindrical tanks that are connected by valves and equipped with outlet tabs. The three thank system is at the Department of mathematical engineering at the Catholic University of Louvain, Belgium.

Figure 3: Schematic diagram of the three tank system

The water level in each tank depends on the quantity of water which enters and which leaves the tank. This quantity depends on the flow given by the valves \(V_i\) and \(V_j\). \(h_i\), \(i = 1,2,3\) indicate the level of water in tank \(i\). \(q_j\) and \(q_2\) are the flows in tanks 1 and 2. \(S_j\) represents the section of the column \(j\), \(q_0\) represents the flow of the water which passes from tank \(i\) to the tank \(j\) and \(q_0\) represent the outgoing flow of the tank \(i\).

In our case, the outputs of the NARMA model are the levels in tanks 1 and 2 \((y_1=h_1; y_2=h_2)\) and the inputs are the flows of \(u_1=q_0\), \(u_2=q_2\). The NARMA model is characterized by 44 terms \((n=m=q=2)\). A file of 1600 measurements (inputs/outputs) is used to model the system, the first 1100 measurements are used for the identification and the last 500 measurements are used to validate the model.

Table 2 presents the terms selected as well as the estimated parameters of the NARMA model. The activation function coefficients found by the real genetic algorithm are: \(f_0=5\); \(f_1=26.3\) and \(f_2=35\).

Table 2: Terms and parameters estimated for multivariable system 2

<table>
<thead>
<tr>
<th>Possible terms</th>
<th>(y_1(k))</th>
<th>(y_2(k))</th>
<th>(u_1(k-2))</th>
<th>(u_2(k-2))</th>
<th>(y_1(k))</th>
<th>(y_2(k))</th>
<th>(u_1(k-1))</th>
<th>(u_2(k-1))</th>
<th>(y_1(k))</th>
<th>(y_2(k))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(v)</td>
<td>0.0123</td>
<td>0.0010</td>
<td>0</td>
<td>-0.0003</td>
<td>-0.005</td>
<td>-0.0034</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(u_1(k-1))</td>
<td>0.0038</td>
<td>-0.003</td>
<td>(u_2(k-1))</td>
<td>-0.0012</td>
<td>-0.004</td>
<td>-0.0049</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(u_1(k-2))</td>
<td>0.0028</td>
<td>0.0036</td>
<td>(u_2(k-2))</td>
<td>-0.0006</td>
<td>0.0036</td>
<td>0.0047</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(u_1(k-2))</td>
<td>-0.0029</td>
<td>0.0019</td>
<td>(u_2(k-2))</td>
<td>-0.0004</td>
<td>0.0087</td>
<td>0.0013</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(u_1(k-2))</td>
<td>0.0045</td>
<td>-0.0063</td>
<td>(u_2(k-2))</td>
<td>-0.0024</td>
<td>-0.0015</td>
<td>-0.0057</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(y_1(k))</td>
<td>0.4668</td>
<td>0.1762</td>
<td>(u_2(k-2))</td>
<td>-0.0013</td>
<td>-0.004</td>
<td>-0.0298</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(y_1(k))</td>
<td>0.3238</td>
<td>-0.0403</td>
<td>(u_2(k-2))</td>
<td>-0.0005</td>
<td>-0.0021</td>
<td>0.0492</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(y_1(k))</td>
<td>0.0203</td>
<td>0.4004</td>
<td>(u_2(k-2))</td>
<td>-0.0007</td>
<td>-0.0035</td>
<td>0.0478</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(y_1(k))</td>
<td>0.0618</td>
<td>0.4916</td>
<td>(u_2(k-2))</td>
<td>0.0031</td>
<td>0.0007</td>
<td>0.0213</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(u_1(k-1))</td>
<td>0</td>
<td>0</td>
<td>(u_2(k-2))</td>
<td>0.0021</td>
<td>-0.0002</td>
<td>0.0263</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(u_1(k-1))</td>
<td>0</td>
<td>0</td>
<td>(u_2(k-1))</td>
<td>0.0005</td>
<td>0.0020</td>
<td>0.0517</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(u_1(k-1))</td>
<td>-0.0004</td>
<td>0.0007</td>
<td>(u_2(k-2))</td>
<td>0.0004</td>
<td>0.0025</td>
<td>0.0367</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(u_1(k-1))</td>
<td>0</td>
<td>0.0010</td>
<td>(u_2(k-2))</td>
<td>0.0029</td>
<td>-0.0006</td>
<td>-0.0391</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(u_1(k-2))</td>
<td>0</td>
<td>0</td>
<td>(u_2(k-1))</td>
<td>0.0016</td>
<td>0</td>
<td>-0.0446</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(u_1(k-2))</td>
<td>0</td>
<td>-0.0002</td>
<td>(u_2(k-1))</td>
<td>0.0015</td>
<td>-0.0024</td>
<td>-0.0540</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(\)
Table 3 presents the values of $R^2_{tot}$ and the computation time of the identification, using a PC Pentium 4, 3.2 GHz. It is noted, according to this table, that the values of $R^2_{tot}$ are very close to 1. This indicates that the models found are precise; however, the computation time is very high.

<table>
<thead>
<tr>
<th>System 1</th>
<th>System 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_1(k)$</td>
<td>$y_2(k)$</td>
</tr>
<tr>
<td>0.9953</td>
<td>0.9998</td>
</tr>
<tr>
<td>Time (sec)</td>
<td>546.08</td>
</tr>
</tbody>
</table>

In order to validate this model the evolutions of the system outputs with those of the identified model are compared. Figure 4 respectively presents the inputs ($u_1$, $u_2$), the outputs system/model, the error of the first output $y_1$ and the outputs system/model, the error of the second output $y_2$. It is noticed that the modeling errors are very small. This allows one to conclude that the model obtained characterizes suitably the dynamic of the physical system.

The use of such method is very required when we have no a priori knowledge of the structure of the system. This advantage allows this approach to treat the complex systems. As prospects with this work we suggest the simplification of the NARMA model complexity, by eliminating the terms having a weak contribution in the model output, and the optimization of the parameters and the structure of the neural network.

**Acknowledgment**

We make a point of warmly thanking Mr V. Wertz Professor at the Catholic University of Louvain, Belgium, who enabled us to carry out practical tests within his laboratory at the Department of mathematical engineering. Our sincere thanks also go to Mr B. Bouzouita, Researcher at the Laboratory “ACS” at the National Engineering School of Tunis, who carried out these practical tests and enabled us to validate our results with a practical application.

**REFERENCES**


DISTRIBUTED SIMULATION
PIELOINED PARALLELISM SCHEDULING IN HOMOGENEOUS AND HETEROGENEOUS ENVIRONMENTS

Carmen Odubășteanu
Department of Computer Science
“POLITEHNICA” University of Bucharest
Spl. Independentei 313, Bucharest, Romania
e-mail: carmen_od@yahoo.com

Călin Munteanu
Department of Automatics and Industrial Information
“POLITEHNICA” University of Bucharest
Spl. Independentei 313, Bucharest, Romania
e-mail: mc_aurel@yahoo.com

KEYWORDS
Query optimization, parallel databases, pipeline parallelism scheduling, computer-aided analysis.

ABSTRACT
Pipelined parallelism scheduling problem is very important in the area of parallel query optimization. The scheduling problem is modeled by using a Pipelined Operator Tree (POT), which is a tree whose nodes represent query operators that can be run in parallel and edges represent communication between adjacent operators; we must find a schedule for the POT that minimizes the total response time, a problem which has been shown to be NP-hard. This paper presents algorithms (some of them proposed by authors) for pipelined parallelism scheduling for homogeneous and heterogeneous environments and compares their by computer-aided analysis, simulating their behaviors.

INTRODUCTION
Today we are challenged with sophisticated applications on parallel database systems, such as decision support systems and data mining. Therefore, the minimization of the query response time is more than ever necessary. The complexity of this problem is reduced if we used a two-phase approach (Chekuri et al. 1995; Florescu et al. 1996): join ordering and query rewriting followed by parallelization and scheduling. In the second phase, atomic units of the query (operators) are extracted and then scheduled to provide the minimum response time. One of the most important issues that must be considered is the parallelism-communication trade-off (Dewitt and Gray 1992; Hasan 1996).

A query will be represented as a weighted operator tree in which each node represents an operator and each edge represents the timing constraints between operators (Hong 1992; Hasan and Motwani 1994). A timing constraint is either a precedence or parallel constraint. The parallel constraint introduces a pipelined parallelism and requires that the two adjacent nodes start and terminate their works approximately at the same time, behaving as a producer-consumer system.

Algorithms for managing pipelined parallelism are an essential component of an optimizer because pipelining is sometimes the only way of speeding up a query not just a useful supplement to partitioned parallelism (Dewitt and Gray 1992). For example, when each reduced relation of a query that join a large number (say 10) of relations and apply external functions, grouping and aggregation is small, partitioned parallelism ceases to be a viable option and pipelined parallelism is the only source of speedup.

Scheduling of a Pipelined Operator Tree (POT - weighted operator tree in which all edges represent parallel constraints) (Hasan and Motwani 1994) is different from the classical scheduling problems because of the communication. Brute force algorithms are impractical for scheduling pipelines due to the extremely large search space.

This optimization problem can also be viewed as to find a schedule that minimizes the maximum load of the processors where load of a processor is the sum of the weights of the operators assigned to it plus the weight of the edges that connect nodes on this processor to the nodes on other processors.

POT scheduling problem was first introduced by Hasan and Motwani for identical processor systems and was shown to be NP-hard (Chekuri et al. 1995; Hasan and Motwani 1994). They proposed several approximation algorithms. Four of them are presented and compared from different points of view: Modified LPT, BalancedCuts, Hybrid and LocalCuts. Also, in this paper are described three recent algorithms OptimBalancedCuts, OptimHybrid and FiLocalCuts designed by authors.

Efficient solution for parallel query optimization in heterogeneous processor systems has been considered to be a challenging task (Florescu et al. 1996; Hasan 1996). Systems based on network of workstations or PC clusters are good examples of such systems. The workstations may differ in processor speed, amount of memory and the speed and the number of attached disks. Moreover, the commodity interconnections make pipelining as a practical way and sometimes the only way to speed up query execution.

Heterogeneity of resources has been considered before in the context of the problem that we are dealing with only in (Termehchy and Ghodsi 2001, 2003).

In this paper, we consider processors with different fixed speeds, called uniform processor system (Gonzalez et al. 1977) and we introduce two algorithms for POT scheduling.
on uniform processor system, UniformModLPT and UniformHybrid.

The paper is organized as follows: first, an overview of the model and problem definition are discussed. Then, nine algorithms for scheduling pipelines parallelism for homogeneous and heterogeneous environments are shortly described. Finally, the experimental results are presented and analyzed.

A MODEL FOR THE PROBLEM

The following definitions are based on earlier models presented in (Chekuri et al. 1995; (Hasan and Motwani 1994). A POT is represented as a weighted operator tree $P= (V, E)$ with $n$ nodes. The weight $t_i$ of the node $i$ is the time to run the operator in isolation assuming all communications are local. The weight $c_{ij}$ of the edge from node $i$ to node $j$ is the additional CPU overhead that both $i$ and $j$ will incur for inter-operation communication if they are scheduled on different processors. A schedule of $P$ on $p$ processors is a partition of $V$, the set of $n$ nodes, into $p$ sets $F_1, F_2, ..., F_p$ such that set $F_k$ is assigned to processor $k$. The load of processor $k$, or $L_k$, is the cost of executing all nodes in $F_k$ plus the overhead for communicating with nodes on other processors. That is, $L_k= \sum_{i \in F_k} t_i + \sum_{(i,j) \in E} c_{ij}$. $L$, the total load, is $\max_{1 \leq k \leq p} L_k$.

Two operations are used to modify the POT: collapse $(i,j)$ replaces adjacent nodes $i$ and $j$ by a single node $i'$ having weight of $t_i = t_i + t_j$. Operation cut $(i,j)$ deletes edge $(i,j)$ and add its weight to those of node $i$ and $j$. Collapse and cut operations should be interpreted as decisions to allocate nodes to the same or distinct processors respectively.

As shown in (Hasan and Motwani 1994), we can convert each POT into a POT with no worthless edges, called monotone tree, by collapsing all its worthless edges using the GreedyChase algorithm that “chases down” and removes parallelism that is “worthless” irrespective of the number of processors. A GreedyChase algorithm is used as a pre-processing step in all described algorithms. Then we schedule the monotone tree. In a monotone tree we use also the following notations: net weight of a node $i$ is $R_i = t_i + \sum_{j \in V} c_{ij}$ and $R = \max_{1 \leq i \leq p} R_i$; total weight, $W = \sum_{i \in V} t_i$.

PIPELINE SCHEDULING ALGORITHMS FOR HOMOGENEOUS ENVIRONMENTS

Scheduling pipelined operator tree is an intractable problem (Graham et al. 1979) and the space of schedules is super exponentially large. Thus any algorithm that finds the optimal is likely to be too expensive to be usable. The following algorithms are fast heuristics that produce near-optimal schedules.

Modified LPT Algorithm

Modified LPT algorithm (Hasan 1996) simply preprocesses away worthless parallelism by running GreedyChase before running LPT (Graham et al. 1979). LPT assigns the job with the largest running time to the least loaded processor, repeating this step until all jobs are assigned. The algorithm is still oblivious to the tradeoff between parallelism and communication. Edges in a monotone path can have high weights and the algorithm is unaware of the savings that can occur when two nodes connected by an edge with a large weight are assigned the same processor.

BalancedCuts algorithm

BalancedCuts algorithm (Hasan 1996), is finding the optimal connected schedule. A connected schedule requires the nodes assigned to any processor to be a connected set.

The algorithm for finding the optimal connected schedule for trees in which all edge weights has two steps which are repeated until the resulting number of fragments is no more than $p$:

Step 1: Given a bound $B$ and a number of processors $p$, the BpSchedule algorithm (Hasan 1996) will be run to find a connected schedule with a response time of at most $B$, if such a schedule exists. Algorithm simply picks a mother node (a node is a mother node if all adjacent nodes with at most one exception are leaves) and traverses the children in the order of non-increasing $t_i - c_{im}$. Then, children are collapsed into the mother node as long as the weight of the mother stays below $B$ and then cut off the rest. The process is repeated until no more mother nodes are left. If the resulting number of fragments is no more than $p$, a $(B,p)$-bounded schedule was found, otherwise no such schedule is possible.

Step 2: For an unsuccessful run of BpSchedule we will revise $B$ as being the minimum of $B_i$ (for each fragment $F_i$ produced by BpSchedule, let $B_i$ be the cost of the fragment plus the weight of the next node that was not included in the fragment). Initially, $B$ will be set to a lower bound on the response time.

OptimBalancedCuts algorithm

OptimBalancedCuts algorithm is an optimization of the BalancedCuts algorithm. A more careful analysis (and implementation) of the following idea gives us a bound of $O(np)$. Whenever the $B$ value is updated, the total work done in finding a new candidate solution can be charged to the nodes which migrate from a component to a previous one. It is easy to verify that the implementation cost works out to be $O(1)$ for each such node migration. Since any one node can migrate at most $p$ times, the total work can be bounded by $O(np)$.

The algorithm picks a mother node, traverses the children in the order of non-increasing $t_i - c_{im}$ collapses them into the mother node as long as the weight of the mother stays below $B$ and then cut off the rest saving for each node the context before the corresponding cutting step (a context is defined by the current state for: tree, mother node, son and
the number of cuts). The process is repeated until no more mother nodes are left or the number of cuts is not p-1. If the cost of the last fragment is no more than B, a (B,p)-bounded schedule was found, otherwise no such schedule is possible.

B is revised by minimum of B, and we repeat the process of collapsing nodes to their mothers beginning from the context corresponding to the node C (the one choosed for the new B value, which has the minimum value from the last iteration cutting nodes). So, the algorithm is run directly from the iteration corresponding to the mother node of C, skipping that way the steps already made before the cutting of the node C.

**Hybrid Algorithm**

BalancedCuts performs poorly on stars since the constraint of connected schedules is at odds with load balancing (Hasan 1996). While the algorithm is cognizant of communication costs, it is poor at achieving balanced loads. On the other hand, LPT is very good at balancing loads but unaware of communication costs. The Hybrid algorithm (Hasan 1996) resulted by combining these two algorithms: BalancedCuts to cut the tree into many fragments and then schedule the fragments using LPT. LPT can be expected to “cleanup” cases such as stars on which connected schedules are a bad approximation.

**OptimHybrid Algorithm**

Hybrid algorithm has the best performance ratio in our experiments. Also, it has the worst execution time. So, we developed a new algorithm, OptimHybrid, based on Hybrid, which has a better complexity. OptimHybrid uses OptimBalancedCuts algorithm instead of BalancedCuts algorithm to cut the tree into fragments which are then scheduled using LPT.

Algorithm OptimHybrid:
1. T’ = GreedyChase
2. for i = p to n do
3. \( F_1, ..., F_i \) = OptimBalancedCuts(T’, i)
4. schedule = LPT(\{F_1, ..., F_i\}, P)
5. end for
6. return best of schedules found in steps 2 to 5

**Approximation algorithms: LocalCuts and FiLocalCuts algorithms**

For this algorithm we use a two-stage approach: fragmentation, and the actual scheduling. For scheduling, it was used LPT algorithm.

In order to obtain better performances we used the golden number, (known also as Fibonacci number, Divine section, or Phi) which has an approximate value of 1.618. This number is often met all around the world, from the ancient and modern art and architecture to the organization of nature, including human beings too. It is said that Phi represents a measure of harmony, a divine proportion known and used from the antiquity.

LocalCuts algorithm (Chekuri et al. 1995) repeatedly picks up a leaf and determines whether to cut or collapse the edge from the leaf to its parent. It selects proper operation based on the ratio of the leaf weight to the weight of the edge to its parent. If the ratio is greater than an input parameter \( \alpha \), it will cut the edge, since this operation does not considerably increase the weight of the resulting fragments. If the ratio is less than \( \alpha \), the leaf is collapsed to the parent node. This is because the weight of the parent node will not increase substantially.

In FiLocalCuts algorithm we used for \( \alpha \) golden number based values.

Algorithm FiLocalCuts:
1. T’ = GreedyChase
2. while there exists a mother node m with child j do
3. if \( t_j > \alpha \cdot t_m \) then cut(j,m)
4. else collapse(j,m)
5. end-while
6. return schedule

**PIPELINED SCHEDULING ALGORITHMS FOR HETEROGENEOUS ENVIRONMENTS**

In this paper, we consider processors with \( p \) processors having different fixed speeds \( s_1 \geq s_2 \geq ... \geq s_p \), called uniform processor system (Gonzalez et al. 1977).

**The UniformModLPT algorithm**

The UniformModLPT algorithm is designed for the uniform processor system (UPS) and assigns tasks to processors in the decreasing order of their processing times. A task is taken from this list and is assigned to a processor whose finishing time is the earliest. Also, the UniformModLPT algorithm preprocesses away worthless parallelism by running GreedyChase. This results in a significant improvement.

**The UniformHybrid Algorithm**

The OptimHybrid algorithm has the best performance ratio in our experiments. So, it was developed a new algorithm, UniformHybrid, based on OptimHybrid for uniform processor systems. The UniformHybrid uses OptimBalancedCuts algorithm cut the tree into fragments which are then scheduled using a variant of LPT.

The UniformHybrid Algorithm:
1. T’ = GreedyChase
2. for i = p to n do
3. \( F_1, ..., F_i \) = OptimBalancedCuts(T’, i)
4. schedule = LPT(UPS, \{F_1, ..., F_i\}, P)
5. end for
6. return best of schedules found in steps 2 to 5
EXPERIMENTAL RESULTS

Based on experimental data, the presented algorithms will be compared in this section from different points of view. Over 5000 trees were generated for experiments, from which, after applying the algorithm for conversion into monotone trees, only those who have 10 nodes were kept, meaning 1000 trees. The generated trees have a value domain ranging from 1 to 20 for node weight and for edge weight. All presented algorithms were simulated for every monotone tree and for a number of processors p ranging from 1 to maximum 10.

For performance analysis of each algorithm, performance ratio was defined as the ratio between experimental value and optimal value. For homogeneous environments the optimal value was considered the maximum values from $R_{\text{max}}$, $\Sigma s_i$, and $(W+C_E)/p$ where $C_E$ is the sum of the weights of the cheapest p-1 edges (Hasan 1996). For uniform processor systems we used the maximum values of the two lower bounds of the optimal solutions for the optimal values (Termehchy and Ghodsi 2003). These two bounds are $W/\Sigma s_i$ and $R/\max(s_i)$.

Performance ratio

The results for average performance ratio are presented in a graphical (comparison) form in figure 1 and figure 2.

For homogeneous environments, the most efficient algorithms are Hybrid and OptimHybrid (same performance ratio). Also, FiLocalCuts presents better performance ratio than LocalCuts, the algorithm which is derived from.

For all the 1000 tested trees, the minimum performance ratio value was 1 for all the presented algorithms. The maximum performance ratio values (and the number of processors for which are obtained) are in Table 1. Also, the domain length for performance ratio is detailed.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>P</th>
<th>Max</th>
<th>Max-Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modified LPT</td>
<td>p = 3</td>
<td>1.74</td>
<td>0.74</td>
</tr>
<tr>
<td>BalancedCuts,</td>
<td>p = 3</td>
<td>1.84</td>
<td>0.84</td>
</tr>
<tr>
<td>OptimBalanced</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hybrid,</td>
<td>p = 3</td>
<td>1.53</td>
<td>0.53</td>
</tr>
<tr>
<td>OptimHybrid</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LocalCuts</td>
<td>p = 4, 5, 6, 9, 10</td>
<td>2.15</td>
<td>1.15</td>
</tr>
<tr>
<td>FiLocalCuts</td>
<td>p = 7, 8</td>
<td>1.59</td>
<td>0.59</td>
</tr>
<tr>
<td>UniformModLPT</td>
<td>p = 8</td>
<td>2.64</td>
<td>1.64</td>
</tr>
<tr>
<td>UniformHybrid</td>
<td>p = 9</td>
<td>1.82</td>
<td>0.82</td>
</tr>
</tbody>
</table>

For homogeneous environments Hybrid and OptimHybrid present the minimum for both maximum performance ratio (1.53) and domain length (0.53). A closed value (0.59) for domain length is also obtained by FiLocalCuts which presents a significant improvement from this point of view relative to LocalCuts (1.15).

The algorithms for heterogeneous environments have the maximum performance ratio for 8, respectively 9 processors; this result denotes that the approximation of the optimal schedule is not the most suitable for a number of processors approximately equal to the number of nodes for this case.

Average time for generating the solution

The best time (figure 3) is obtained by LocalCuts and FiLocalCuts (homogeneous environments) respectively UniformModLPT for heterogeneous environments.
time and memory need related to OptimHybrid respectively UniformHybrid, the algorithms with the best performance ratio.

Using the golden number, Phi, for FiLocalCuts we improved LocalCuts performances for both maximum performance ratio and its domain length. Thus, we can say that this first use of golden number in pipelined parallelism scheduling problem was a successful one and it proves that golden number implications in pipelined parallelism are alike to the implications which appears in other domains (especially a “harmonious distribution”).

FURTHER RESEARCH

A possible further research is the use of golden number for other approximation algorithms. Also, it will be useful and interesting too to extend these algorithms for heterogeneous environments and to find a more appropriate approximation of the optimal schedule for a number of processors approximately equal to the number of nodes.

REFERENCES


An efficient method for Distributed Constraint Satisfaction Problems resolution

KAIS BEN SALAH
KHALED GHEDIRA

SOIE, 2010 Campus Universitaire
la manouba, Tunisia
E-Mail : bensalah.kais@planet.tn
khaled.ghedira@isg.rnu.tn

ABSTRACT

The expansion of the multinational corporations, the multi-sites production and the globalization of the economy increased the importance of the Distributed Constraint Satisfaction Problems (DCSP), which explains the necessity to find a new method that handles these problems without loosing the properties of the geographical (natural) distribution of the initial problems. In this paper, we present our model DOC for new Distributed meta-model for Optimization under Constraints, that solves distributed problems and prove its efficiency theoretically and experimentally.

INTRODUCTION

The evolution of the researches in Distributed Artificial Intelligence (DAI), allowed a great expansion of the Multi-Agent Systems (MAS). Many applications in MAS are concerned with determining a consistent arrangement of agent behaviours like the Distributed Constraint Satisfaction Problems (DCSP). In fact, the highly combinatorial character of these problems makes them a privileged domain of DAI. One of the most effective approaches in these cases is the Multi-Agent systems. The agents will take advantage of the distributed character of the DCSP that matches very well with their architecture to solve them.

An important family of the DCSP, consists of the geographically distributed problems, that are present everywhere, especially in the domain of production management, such as extended enterprises, network companies, “multi-sites” production, etc.

Despite the fact that there is a very important cooperation between the subsidiaries of the same company in order to achieve a global goal, they behave independently and sometimes concurrently to accomplish their local objectives. In this kind of environment, the privacy and security are a main issue. So, the centralisation of the distributed problem is absolutely unacceptable. The communication between the sites, decision-centers or subsidiaries, must only concern the variables involved in the inter-sites (decision-centers, subsidiaries) constraints.

So, the most natural distribution that insures the privacy is the one that will associate each sub-problem (production site, decision-center, subsidiary, etc) to an agent that will be responsible of its resolution (see figure1). Each site or node is considered as a decision center and formalized as a Constraint Satisfaction Problem (CSP); hence, it has to find an instantiation of its local decision variables that satisfies both internal and external constraints (see figure2).

![Figure1. An example of distributed problems](image)

The problem in distributed environment is how to ensure that the chosen algorithm is complete and terminates. One
of the most effective approaches in these cases is the Multi-Agent systems. The agents will take advantage of the distributed character of the DCSP that matches very well with their architecture. In fact, the DCSP are defined as a number of Constraint Satisfaction Problems (CSP) connected by external constraints or a CSP in which variables and constraints are distributed among multiple automated agents [9]. Knowing that agents have knowledge, behavior and they interact with each other to reach local and global goals, they constitute the main research issue of the DSCP. Hence, the most efficient algorithms in the literature are based on the Multi-Agent Systems, like the asynchronous Backtracking, the asynchronous weak-commitment search [5] and the asynchronous aggregation search [11].

But, the major drawback of these algorithms is their need to centralize first the problem, before choosing their own distribution. For instance, in the asynchronous Backtracking or the weak-commitment search; each agent is responsible of one variable and in the asynchronous aggregation search; an agent is responsible of one constraint (or more).

In our meta-model DOC, we associate each sub-problem (site, decision-center, subsidiary) to an agent called “Decision-Maker”. Another class of agents called “Advisor”, in charge of the external constraints (i.e. intercenters) is added (see figure2). Therefore, DOC proposes a new Multi-Agent architecture and a new Multi-Agent dynamic which doesn’t rely on any centralized algorithm. Thus, DOC is different from all the algorithms, also based on DCSP, such as ABT-Like algorithms [5], asynchronous aggregation Search algorithm [11].

A Distributed meta-model for Optimization under Constraints

DOC for new Distributed meta-model for Optimization under Constraints presents a new Multi-Agent architecture and dynamic.

Our model consists of three classes of agents. The first one, comprise a single agent called the Interface agent. The second class corresponds to the “Decision-Maker” agents and the third one to the “Advisor” agents. Each Decision-Maker is in charge of a sub-problem (see figure 2) involving its sub-set of variables and constraints that will be called respectively internal variables and internal constraints. As regards to the external constraints, i.e. inter-Decision-Maker constraints, they are under the responsibility of the “Advisor” agents. The allocation of external constraints to the “Advisor” agents will be explained in the next section.

![Figure 2. An example of DCSP](image)

It’s necessary to have an Interface between this society of agents and the user in order to:
- Create and initialize the “Decision-Maker” and the “Advisor” agents,
- Determine the order between the agents,
- Recognize that the problem was solved by this society and
- Inform the user about the final result.

Each agent is endowed with a simple structure: its acquaintances, its knowledge and its behavior.

The approach that we will use to solve distributed problems consists in establishing first the global consistency via the external constraints and then locally solving the sub-problems at the level of Decision-Maker agents. The global consistency is maintained in a totally distributed manner and this by the intermediary of the Advisor agents.

Indeed, each Decision-Maker receives from its potential Advisors a consistent search partial direction (i.e. satisfying its external constraints). Each partial direction corresponds to a partial instantiation that must satisfy the external constraints involving the variables concerned by this instantiation.

If the Decision-Maker is satisfied, then it provides to the Interface agent the obtained results. Else, according to the priority of the agents, it asks one of its Advisors to propose another partial direction.
Let us note that the Interface agent is satisfied when the system is in a stable state, i.e. when there is no more communication between Decision-Makers and Advisors.

Before exploring the details of our meta-model, let us explain how we determine efficiently our agents:

Since we consider that the problem is already distributed (as the case of geographically distributed problems), i.e., portioned into sub-problems, each one will be associated to a Decision-Maker agent (see figures 1 and 2). Hence, there will be as many Decision-Maker agents as sub-problems.

Each sub-problem is formalized as a Constraint Satisfaction Problem (CSP).

The choice of the Advisor agents is made in a way to distribute the guidance of the search, as much as possible, to minimize the communication between agents and to keep the same strategy of guidance for all the system. Then, associating each external constraint to a different Advisor, increases the communication, needs an advanced negotiation and increases considerably the complexity. But, choosing to associate all the external constraints to a unique Advisor will centralize a great part of the problem and we will not benefit of the distributed aspect of the Multi-Agent systems. Another drawback can be found, if we affect two or more constraints having common variables. In fact, there will be a conflict in the task of allocating values to these variables. So, we choose three basic elementary configurations:

- One Advisor per external connex (related) component, i.e. component formed by nodes (variables) belonging to external constraints. A component can correspond to an association of binary external constraints. For instance, the set of constraints \{\{X_{13}, X_{42}\}, \{X_{42}, X_{31}\}\} of figure 2, is a component, that will be affected to a unique Advisor.
- One Advisor per pair of Decision-Makers: if we have one or more external constraints between two Decision-Makers, the subset that doesn’t belong to any connex component, will be affected to a unique Advisor. For instance, (X_{44}, X_{33}) will be under the supervision of a unique Advisor.
- If we have the exception of an isolated Decision-Maker, i.e. not connected to other Decision-Maker, this one will become its own Advisor. It chooses randomly one of its internal constraints, which becomes a “virtual” external constraint.

According to its external constraints, a Decision-Maker can have one or several Advisors. And, it’s important to note that an external constraint can’t belong to more than one Advisor. But one Advisor can be responsible for several constraints at the same time, even if they involve more than two Decision-Makers.

The version DOC-SATISFACTION

DOC solves the Distributed Constraint Satisfaction and Optimization Problems (DCSOP) and DCSP.

DOC-SAT is the version of DOC limited to constraint satisfaction. We established its foundations in a previous work [8].

In fact, with the advent of distributed reasoning and the progress of parallel computing, some papers based on these considerations have been published. Three families are identified: parallelization, Multi-Agent systems and hybrid techniques.

One of the parallelization techniques consists in using the Depth First Search Tree (DFST) strategy in order to exploit its natural parallel aspect for the backtracking and for the self-stabilizing property relative to the communication protocol [14]. Another technique consists in dividing the search space between the available processors [2]. It is efficient, but it needs both centralized information and control.

With regards to Multi-Agent systems, all the proposed algorithms are based on DCSP formalism: distributed consistency achieving [1], and ABT-Like algorithms, namely, asynchronous backtracking algorithm, Weak-Commitment search algorithm [9, 5], and the Asynchronous aggregation search (AAS) [11].

Finally, hybrid techniques have been used on the base of CSP, DCSP and parallelization strategies [12].

Let’s remark that the basic idea of all these algorithms is to distribute and/or parallelize the well known centralized algorithms such as backtracking, backjumping, etc, while DOC doesn’t rely on any centralized algorithm. Moreover, DOC can benefit of the asynchronous aspect of the Multi-Agent dynamic and it doesn’t impose the use of a special algorithm. In other words, each agent can use any CSP technique which allows us the use of various algorithms at the level of the same system.

In DOC-SAT, Decision-Maker and Advisors cooperate and interact by exchanging requests in order to satisfy intra-Decision-Makers constraints and inter-Decision-Makers constraints (see figure 4). Indeed, each Decision-Maker agent receives from each one of its Advisors a search partial direction and then it tries to initiate its internal variables (see figure 2) using one of the possible CSP techniques such as MAC [7], Forward Checking [7], etc. If it can’t, it asks one of its Advisors to propose another partial direction. As for the behavior, we have defined a standard agent structure: acquaintances, knowledge, mailbox and a standard behavior.
But, before giving more details about the structure and the behavior of the Decision-Maker and Advisor agents, we have to know that it is very difficult, if not impossible, to insure that an algorithm is complete and terminates in a distributed environment without establishing an order between the agents [7].

In our model, the Interface determines the order between the agents according to five rules:

- The Decision-Maker that has the minimum number of Advisors has the higher priority,
- If two Decision-Makers have the same number of Advisors, the one who has the minimum number of variables involved in external constraints, has the higher priority,
- If two or more Decision-Makers have the same level of priority after applying the two first rules, it’s the one who has the smallest domain’s size, has the higher priority,
- If after this, we have two or more Decision-Makers have the same level of priority, the alphanumeric order will be considered,
- Finally, the Advisor having the highest level of priority is the one who has the Decision-Maker having the highest priority in his list of acquaintances.

The current instantiation: this instantiation concerns its variables (interior) as well as its external variables (frontier).

- The state of satisfaction.

- The set of the non requested Advisors SetNonReqAD(DMi), this set is sorted by the order-priority sent by the Interface agent. This set is updated each time DMi processes an “exploreDirection” message or when DMi sends a “changeDirection” message. When this set becomes empty, it is reinitialized to ADL(DMi).

- The set of local solutions.

Behavior:

The communication of the Decision-Maker with its Advisors is based on two sent messages and three received ones (see figure 5).

At the initial phase, the Decision-Maker receives from each one of its Advisors a partial direction in a message called “ExploreDirection”. After the reception of all the partial directions, i.e. all its frontier is instantiated, it tries to instantiate its interior, using any one of the centralized CSP algorithms. According to the result of this attempt, a Decision-Maker can have, one of two states: satisfied or unsatisfied. It is satisfied in two cases:

a- When its frontier and its interior are instantiated in a consistent manner. This means that there is a local solution. To propagate this solution, the Decision-Maker will ask its Advisors, via a message called “InfoDecision” (involving the partial direction correspondent to each Advisor), about the consistency of the actual values, i.e. if they satisfy the external constraints or they become obsolete.

b- When it finishes to browse all the possible search directions

In this state, its behavior consists in informing the Interface agent about its both current instantiation and its state of satisfaction. In the opposite case, the Decision-Maker chooses the first Advisor (having the highest level of priority) in the list of not yet requested Advisors, sending a message called “ChangeDirection”, this message include, the last partial direction sent by this Advisor (that gathered with the other partial directions, doesn’t lead to any local solution) and asks it to provide a new partial direction. Finally, the Decision-Maker updates it’s the SetNonReqAD(DMi). Moreover, if the Decision-Maker (DMi) has a single Advisor, it informs this Advisor to put this partial direction in the Nogood direction set (see figure 3).
As we said, there are three kinds of messages that can receive a Decision-Maker (DMi). The first one is a request from one of its Advisors; comprising a search partial direction (partDir) to browse. This message is in the form: ("ExploreDirection", partDir), Adj, DMi). The second message comes also, from its Advisors: ("NogoodPartialDirection", partDir), Adj, DMi) and induces an update of the Explored Search Directions List: ExploredSet(DMi), i.e. the deletion of the search directions that involve a tuple or a combination of tuples that becomes “Nogood” (the term nogood means that this partial direction doesn’t lead to a consistent instantiation). The third message is an answer to the "InfoDecision" message. It has the following form: ("AnswerInfoDecision", partDir, Decision (OK/Not OK)), Adj, DMi).

When the Decision-Maker has several Advisors, the list of non-requested Advisors insures that the Decision-Maker (DMi) will visit all the possible search directions (which is necessary to find all the solutions). Hence, if the DMi can’t find a possible consistent instantiation (i.e. satisfying all the internal constraints), it sends to the first Advisors existing in the SetNonReqAD(DMi) (having the highest level of priority), a message "ChangeDirection" and updates as usual, in this case, the SetNonReqAD(DMi).

If the Decision-Maker (DMi) finds a local solution, it request all its Advisors about the values affected to variables of its frontier (variables involved in external constraints), if they are valid (consistent) at the global level and this via a message: ("InfoDecision", partDir), DMi, Adj). In the opposite cases (i.e., the Decision-Maker can’t find any local solution), after visiting all the existing directions, when the size of ExploredSet(DMi) becomes equal to 0 (ND(DMi)), the Decision-Maker informs the Interface agent, about its state of satisfaction, and the absence of solution.

This behavior will be repeated until the size of the ExploredSet(DMi) reaches the ND(DMi).

While size (ExploredSet) \(\leq\) ND(DMi) DO
If All the PartDir are received Then
Update(CurrentInst(DMi))
Apply Backtrack(CurrentInst(DMi), ICL(DMi), Domains(Intr(DMi)))
If an internal Solution is found Then
DMi.State = “Satisfied”
For each Adj of ADL(DMi) DO
Send("InfoDecision", PartDir, Adj)
Do Wait ( )

While All “AnswerInfoDecision” are not received
If each Decision in “AnswerInfoDecision” message is equal to “OK” Then
Send (DMi.State, Solution, Interface)
Else
DMi.State = “unsatisfied”
Else
DMi.State = “unsatisfied”
If DMi.State = “unsatisfied” Then
If size (ADL(DMi)) = 1 Then
Send("ChangeDirection", PartDir, “Nogood”, ADj)
Else
Send("ChangeDirection", PartDir
SetNonReqAD(DMi).First)
Update (SetNonReqAD(DMi))

Figure 3. General behavior of a Decision-Maker Agent

The Advisor Agent

Acquaintances: its acquaintance list Acq(Adj), i.e. a list formed by its own Decision-Maker agents and the Interface agent.

Knowledge:
- The mail box.
- The constraint list, i.e. the constraints that are external to its Decision-Maker agents.
- The list of the tuple of values LTV(Adj) authorized by the constraints that are external to its Decision-Maker agents. Each tuple or combination of tuples is called “search direction”.
- The list of “nogood” directions NogoodD(Adj), i.e. this list comprises the search partial directions that make impossible the instantiation of the interior of one or several of its Decision-Maker agents.
- The list of “tabu” directions TabuDL(Adj): inspired by the tabu method[4]; this list comprises the temporarily forbidden search partial directions. In fact, in order to not propose successively, the same partial direction many times and necessarily to diversify the search, so to cover the search space at the most, each Advisor holds a tabu list, in which it puts the proposed search partial direction.
- The state of satisfaction.

Behavior
After receiving the message from the Interface agent, containing the priority order of their Decision-Maker
agents, each Advisor sends to each one of its Decision-Maker, the first element of its partial direction list and waits for eventual messages (see Figure 4). An Advisor can receive two kinds of message from its Decision-Makers. The first one is a request to change the partial direction called “ChangeDirection” message. This message can also involve a partial direction qualified as nogood (can lead, in any case, to a local solution), so, it will be ranged in the nogood partial direction list. This nogood will be propagated, via a message called “NogoodPartialDir”, involving this partial direction.

The second message received by the Advisor (sent by one of its Decision-Makers), called “infoDecision” message. It is a request of information about the consistency of a partial direction. As a Decision-Maker, an Advisor can have one of two possible states: satisfied or unsatisfied. It is satisfied when it hasn’t received a request to change partial direction, or when all the partial directions received in the “InfoDecision” messages are consistent (satisfying all the external constraints supervised by the Advisor ADj). In this case, it sends a message to all its Decision-Makers, informing them about the consistency of their partial direction, via a message called “AnswerInfoDecision” with “OK” as parameter, the message will have the following form: (“AnswerInfoDecision”, “OK”, PartDir, DMi).

In the contrary case, its dissatisfaction behavior depends on the received message. When the Advisor receives a “ChangeDirection” message, it chooses a partial direction from its list of the tuple of values LTV(ADj), that neither belongs to the tabu list TabuDL(ADj), nor to the nogood list NogoodDL(ADj) and sends it to its Decision-Makers. When the Advisor receives an “infoDecision” message, its behavior consists in sending an “AnswerInfoDecision” message with “NotOK” as parameter, the message will have the following form: (“AnswerInfoDecision”, “NotOK”, PartDir, DMi), accompanied by an “ExploreDirection” message involving a new partial direction (see figure 5).

```
Do
  Wait()
While received message
  If message.item = "ChangeDirection" Then
    If message.involve "Nogood" Then
      Add PartDir to NogoodD (ADj)
    For each DMi in Acq(ADj) DO
      Send("NogoodPropagation", PartDir, DMi)
    PartDir1 = Choose (First (LTV (ADj)), TabuDL (ADj))
    /*PartDir1 is the first partial direction that belongs to LTV*/
  Else
    Update (LTV (ADj), PartDir)
    Add PartDir1 to TabuDL (ADj)
    Send("ExploreDirection", PartDir1, DMr)
  EndIf
Else
  PartDir1 = Choose(First(LTV(ADj)), TabuDL (ADj))
  Add PartDir1 to TabuDL (ADj)
Else
  If PartDir received in "InfoDecision" message is consistent Then
    Send("AnswerInfDecision", PartDir, "OK", DMr)
  Else
    Send ("AnswerInfDecision", PartDir, "NotOK", DMr)
    PartDir1 = Choose (First (LTV (ADj)), TabuDL (ADj))
    Add PartDir1 to TabuDL (ADj)
  EndIf
EndIf
```

Figure 4. General Behavior of an Advisor Agent

We present two sequence diagrams (of AUML [13]), to clarify the communication between Decision-Makers and Advisors on one hand (see Figure 5) and the Interface agent and the other agents on the other hand (see figure 6).
Experiments

One of the most important advantages of our model is the fact that it doesn’t rely on any centralized algorithm.

We programmed a DCSP generator that has four parameters: number of variables, size of domain, tightness and density. The number of agents depends on the method.

We chose to generate four Decision_Maker agents. This choice was made after several experimentations. In fact, we generated problems with 20 variables (see figure 7) and problems with 28 variables (see figure 8), varying the number of variables affected to each Decision-Maker agent. And the best results were perceived in the case of 4 Decision-Makers, i.e. affecting 5 variables per agent in the case of problems with 20 variables (see figure 7) or 7 variables per agent in the case of problems with 28 variables (see figure 8).

This can be easily explained. In fact, in the case of several Decision-Makers (few variables per agent), the cost of communication will be a handicap. In the other case (several variables per agent), solving the local problems becomes very complex.

We chose to apply the Backtrack algorithm as the solver of the Decision-Maker’s local CSP.

One of the reasons to choose this algorithm is because it is complete.

We generate problems with a static level of density, equals to 0,7. At each level of tightness, we generate ten problems. The average of the resolution (CPU) Times will be considered.

We chose to compare our model to the Asynchronous Backtracking algorithm. Because this algorithm [5] is one of the most known and efficient complete methods in the DCSP domain. Many papers proved its efficiency, and many methods were based on this algorithm, like the Asynchronous weak commitment search [5], etc (ABT-Like algorithms).

In these algorithms, each agent will be responsible of one and only one variable. So, there will be as much agents as variables.

The priority order between agents is determined by the alphabetical order of the variables identifiers. Each agent communicates its tentative value assignment to neighboring agents. Each agent maintains the current assignment of other agents [5].

Figure 7. Testing DOC-SAT (20 variables problems) with several distributions
As it shown in the figures 9, 10 and 11, the Asynchronous Backtracking is more efficient in the lower levels of tightness and when the size of the problems is smaller (contrarily to the cases of large problems and high levels of tightness). That can be explained by the fact that the Asynchronous Backtracking is simpler, each agent is responsible of one and only one variable and it uses only two messages to communicate (“OK?” message to communicate its value and “Nogood” message to communicate a new constraint) [5, 9]. So we tried to improve our model, by reducing its complexity.

So we tried to improve our model, reducing its complexity.
after the reception of each new “ExploreDirection” message and this when the size of the ExplDSet(DMi) exceeds two. This program uses the ExplDSet and the “ExploreDirection” message (particularly the PartDir), to create new instantiations (tuples) and tests if they satisfy all the local constraints.

In the positive case, that means the presence of new solutions. In the other case, that means that the program has explored new directions. Hence, the acceleration of research is reached.

The sub-program Change, will create new directions (using the history of the research, i.e. the explored directions set and the new partial directions), replacing the newly received partial direction (PartDir), in its position (p), in the first (size (ExplDSet (DMi)) – 2) directions belonging to the explored Directions set.

CTNT (ExplDSet(DMi), PartDir, p)
BEGIN
For u = 1 ToExplDSet(DMi).length – 2 DO
NewTuple u = Change (ExplDSet(DMi), PartDir, p)
If NewTuple u Leads to a Consistent instantiation Then
Update(Solutions(DMi))
Update(ExplDSet(DMi))
FND

Figure 12. CTNT program

In order to prove the efficiency of our model, we chose to generate problems with 20 variables, 24 variables and 28 variables. For each level of tightness, we generated ten (10) problems. The average of the resolution (CPU) Times will be considered.

![Comparison (20 variables)](image)

Figure 13. General comparison (problems with 20 variables)

![Comparison (24 Variables)](image)

Figure 14. General comparison (problems with 24 variables)

The “PotentialSet” is a new parameter added by the Decision-Maker to the “ExploreDirection” message and sent to one of its Advisors. This “PotentialSet” can be a singleton or a set and that depends on the nature of the “frontier” of each Decision-Maker. This set is formed in a way that if the partial direction (PartDir) sent by the Advisor, belongs to this set, the “interior” of the Decision-Maker (DMi) can be instantiated in a consistent way (i.e. at least one local solution will be found). Thus, the “PotentialSet” presents two advantages:

- The Advisors will send to their Decision-Maker agents, partial directions easily extensible to local solutions.
- The absence of solutions is quickly detected (when the Advisor doesn’t have partial directions that belong to this set).
The three figures; 13, 14 and 15 prove the contribution of
“Create and Test New Tuples” program in the cases of high
size problems and the highest level of tightness. We can
note the excellent improvements brought by the
“PotentialSet” at all the levels of tightness.

**Conclusion**

We presented in this paper a new efficient method that solves
Distributed Constraint Satisfaction Problems. It has many
advantages: in fact, it doesn’t need to centralize the distributed
problem and like the real cases (geographically distributed
problems), each sub-problem is solved by a decision center. This
advantage insures the confidentiality and the privacy at the level
of each decision center. The decision centers exchange only
the values of variables involved in external constraints.

The completeness of the DOC-SAT (i.e., finding all the existing
solutions and terminates if there is no solutions) is guaranteed and
was proved [8].

In our model, as it was explained in the agent’s behavior sections,
the communication between agents doesn’t only rely on values
exchanging. And we believe that improve significantly the quality
of our model. In fact, many papers studied the impact of agent’s
communications on the performance of distributed algorithms.
Fernandez et al. investigated the effect of communication delays
on the performance of DCSP algorithms [6]. Another very
interesting paper [3] has proved that DCSP strategies with
additional information-exchange can lead to big speedups.

Moreover, the CTNT program and the “PotentialSet”, as shown
in the figures 13, 14 and 15, have improved the efficiency of our
model. In fact, the more that the size of the problem and the
tightness increase, the more the CPU time is improved.

But, the most important advantage of our model is that it can use
any CSP technique as a solver for the Decision-Maker’s local
CSP. So we can improve more our model, by using for example,
the Forward Checking to solve the local problems of the
Decision-Maker agents. Moreover, DOC-SAT can easily deal
with large problems contrarily to the ABT-Like algorithms.

In their paper, Yokoo and his team explained that the ABT-Like
algorithms are neither efficient nor scalable to large problems [5].
Moreover, in another paper [10], experimentation showed that
the unification of the Asynchronous Backtracking algorithm and
the asynchronous distributed optimization (ADOPT), to solve
Distributed Constraint Optimization Problems, was efficient in the
cases of non large problems. But, the major drawback, in large
problems, was the colossal number of exchanged messages
between agents.

**REFERENCES**

[1]: Ahlem Ben Hassine and Khaled Ghédira 2002, “How to
establish Arc-Consistency by reactive Agents”, ECAI 2002,
Lyon.

[2]: B.Burg, “Parallel Forward Checking”, TR-594, Institute

[3]: Fernandez, C., Bejar, R., Krishnamachari, B., Gomes, C.,
Selman, B.: Communication and computation in distributed
csp algorithms. In: Lesser, V., Ortiz, C., Tambe, M., eds.: 
(2003)

[4]: Glover F. “Tabu search” 1990. ORSA Journal on computing
2, 4-32.

to Over-Constrained Distributed Constraint Satisfaction
Problems: Distributed Hierarchical Constraint Satisfaction»,
ICMAS 2000.

Communication in Solving Distributed Constraint
Satisfaction Problems”, CEEMAS 05, 15-17 September

[7]: Ionel Muscalagiu, Jose M.Vidal, Vladimir Cretu, Popa Horia
Emil and Manuela Panoiu; “The effects of Agent
Synchronization in Asynchronous Search Algorithms”,
KES-AMSTA 2007, Wroclaw, Poland, May 31 – June 1,
2007.

of DOC-SAT ». AAMAS, Autonomous Agents &
MultiAgent Systems 2002. Workshop 12: Distributed
Constraint reasoning.

[9]: Makoto yokoo and Katsutoshi hirayama, “Distributed
Breakout Algorithm for Solving Distributed Constraint
Satisfaction Problem” Second International Conference on
Multi-agent Systems 1996.

[10]: Marius C. Silaghi, Makoto Yokoo, “Nogood based
asynchronous distributed optimization (ADOPT ng)
Proceedings of the fifth international joint conference on
Autonomous agents and multiagent systems (AAMAS), May 08-12, 2006, Hakodate, Japan.


PROCESS SIMULATION
IMPLEMENTATION OF THE PROCESS INTERACTION APPROACH IN A GENERAL-PURPOSE LANGUAGE
Hans P.M. Veeke, Jaap A. Ottjes, Gabriëll Lodewijks
Delft University of Technology, The Netherlands
e-mail: H.P.M.Veeke@tudelft.nl; J.A.Ottjes@tudelft.nl; G.Lodewijks@tudelft.nl

Keywords: Process Interaction

ABSTRACT

Two types of approaches are distinguished in the process interaction world view. Depending on the type of elements that own a process, the transaction view and the real process interaction view are defined. In the transaction view all activities are performed by the flowing elements. In the real process interaction view the resources are active. The process interaction view is widely used in the simulation community, but the majority of applications use the transaction view. Contrary to this, the development of business models is nowadays completely based upon process models, which closely connect to the real process interaction view. This paper shows a single-threaded implementation of the real process interaction in Delphi by means of stack contents swapping. By this it becomes possible to translate business models almost directly into simulation models. Above that, the implementation in a general purpose language like Delphi is fast, cheap and extendable. The major problems of implementing process interaction will be explained and tackled. The implementation as a whole is available as free source.

INTRODUCTION

In this paper, we discuss a research project to develop an implementation of discrete event simulation in order to combine structural and behavioural business modelling. This combination is often missing, although simulation literature always starts with “structural modelling”; which means in this case structural simulation modelling. Simulation models consist of a structural (or static) part and a behavioural (dynamic or rather time-dependent) part. In the structural part it is nowadays common use to apply object-orientation that should closely match structural business modelling and, with respect to the dynamic part, many languages are said to be process oriented. However, simulation is still considered to be a specialty for experts. We are convinced that simulation should be an integral part of a design project, where the construction of a simulation model is a shared responsibility of a multidisciplinary project group [Veeke, 2003].

In business science the systems approach evolved and as a sign of its growing application it is nowadays common use to construct a business or logistic design in terms of processes. Although the design is still often split in a technologic, organisation and information design, the partial designs have this process approach in common. If simulation is involved in the design project, the simulation expert should not construct a model from scratch, because the structural model is already available. Checkland [2000], as founder of the Soft Systems Approach, argues that the conception of a ‘system to be designed’ is subjective, depending on the background and goal of the researcher. It is therefore to be expected that a simulation model is a subjective view of the simulation expert and the specific contents of the model cannot be communicated with other members of the project team. Probably it is for this reason that simulation packages focus on advanced animation techniques to ‘show’ the results of the model, to distract the attention from the contents (on the pretext of ‘it has to be seen to be believed’) and offer many predefined components so the user will not be bothered with basic (but therefore complex) programming problems originating from the pure event, activity or transaction approach. By this the (forced) translation of a system model into such packages has become an art on its own. The real process approach however is straightforward, but apparently difficult to implement in a programming environment. This is confirmed by Abu-Taieh and El Sheikh[2007] when they state “Which indicates that the implementation of Process-Interaction was very difficult.”.

We looked for a simulation package that supports this way of describing a system in Business Science. Moreover, it should be possible to implement innovative control systems, which could be implemented –without further translation- in the real control system. And finally, as members of an academic community, the package should be usable for education. This last requirement also includes that the package should be not too expensive.

Many packages have been evaluated, but none fulfilled the requirements of direct translation, costs and real process interaction view. Some packages work with their own language (e.g. PROSIS [2005], Simsplot [Rice, Marjinski, 2005]). Their (connection to standard) reporting and graphical facilities are rather inadequate. The requirement of implementing innovative control systems is a major drawback for advanced graphical packages like emPlant[Boer, 2005]. First of all, they are always very expensive. Above that, they do offer the possibility to “program” parts of the control, but a user you will never know the structure and behaviour of the environment of the control, which consists of predefined (and therefore inaccessible) entities.

So we decided to use a general-purpose language and we selected Delphi (i.e. visual Pascal) as the supporting language for educational purposes. Implementations in JAVA are all multithread implementations (e.g. Silk [Healy, Kilgore, 1997]), and these implementations tend to get very slow even with a moderate number of components.

It appeared to be necessary to construct the basic simulation mechanism from scratch. Many simulation tools have been developed in the past, but they all lack the mechanism for the real process interaction view (except Simula, but this requires a
specific compiler again). Micro Passim (Barnett, 1986) uses the –then already-outdated ‘goto’ mechanism to keep track of activation points. SIMCAL (Malloy and Soffa, 1986) creates an extra step in compilation, while other implementations are based on the transaction world view or lack the capability of ‘conditional events’.

We will first explain the requirements for implementing a real process interaction approach. Then the implementation in Delphi is shown, which is called TOMAS [Veeke, Ottjes, 2000], the resulting simulation tool.

IMPLEMENTING THE PROCESS INTERACTION APPROACH

In the real process interaction view the resources are the active components. In order to describe their activities, a resource owns one (and only one) “process”. A process can be considered as a structured description of events, all being initiated by the resource itself. In order to express periods between events, terms like ‘work’ and ‘wait’ are being used. These terms describe a type of “activity” during a period. At the end of such a period, some kind of state change (event) occurs, caused by the component that owns the process description. Between events the model’s clock time advances. The model’s clock time is the time all processes are synchronized to.

A state change is represented by an event and a (finite) description of the property values changing at that moment in time. For example, a product is delivered to stock or a customer has been served.

A period requires an event to start. From that point on, it is sufficient to specify the moment in time of the first future event to fully define the coming period for this specific component. Knowing this moment in time, the period can be described by “advance model’s clock time until next event”.

The fact that during the period no event occurs, does not mean the process is “doing nothing” or “makes no progress”. It only expresses that the current state remains unchanged with respect to the significant properties, but the current state can be ‘working’, or be just ‘waiting’. So “work until next event” or “wait until next event” all express a period. We will use the general verb “advance” to express a period.

To finish a period, three situations are distinguished:

1. The event time of the next state change is exactly known.
2. The event time of the next state change is unknown, but the condition for it to occur is known.
3. Both the event time and the condition are unknown.

Ad 1. If the event time is exactly known the duration of the period is exactly known, so it is fully defined by “advance x time units”. In this case, the process is called to be in the state “scheduled”.

Ad 2. Here the event time is unknown, but the condition is known. In this case an unambiguous expression of this condition is sufficient. A condition can be expressed in two ways:

- the period continues as long as a condition holds; the period is defined by “advance while condition”
- the period ends if a condition is satisfied; the period is now defined by “advance until condition”

The process is again called to be in the state “scheduled”. Ad 3. Both the event time and the condition are unknown: the length of the period is indefinite. For these cases the word “advance” without further specification is used. No event caused by this component, will occur, unless another process (in the surrounding system or the environment) creates it. It can only be awakened by external circumstances. In these cases the process is called to be in the state “suspended”.

To conclude this part, there is one state of a process left that needs to be defined. At the very moment in time of an event there is no period defined at all. Therefore at the moment a process handles an event the process is called “current”. A process automatically becomes current at the end of a period in its process description.

If a process is suspended, it will not proceed without intervention from the process’s environment. Therefore, facilities are required to intervene with the progress of the process. Interventions can be divided in two categories:

- regular interventions
- disturbance interventions.

Regular interventions occur at regular events of other processes. For example, a customer generator has generated a new customer and a suspended process of a clerk should be reactivated to serve this customer, if it is suspended. For this purpose the expression “Resume” is introduced.

By using Resume, the process resumes progress with the line directly succeeding “Wait” (or Suspend). Resume introduces a new event from outside the process. The causing process must assure itself, that the receiving process is able to respond to this event (i.e. is “suspended”). If the process would be in a scheduled state, the period would have been defined already (in other words the process is watching the clock time or a condition and nothing else). It would not recognize the event introduced.

There are however two regular interventions that necessarily are caused by the environment of the process:

- the process must once be started
- there may be circumstances where the process must be interrupted.

To start the process at some moment in time, the expression: “Start process at T” is used (the process now changes from a suspended into a scheduled state). As soon as the clock time T is reached, the process will start with its first line and follow its description.

To stop a process the expression “Stop process at T” is used and the process description is abandoned at clock time T.
Before starting a process and after stopping a process, the state of the process is considered suspended.

Disturbance interventions interrupt a process, whatever state it is in, and the process must react immediately to it. The only exception is a process being current. Such a process cannot be disturbed, because the model’s clock does not advance. Mostly the reaction of a process will be to do nothing until the disturbance disappears. This is a typical reaction in case of bad weather conditions or technical disturbances. Repairing disturbances can be a complex task of course, but they are usually performed by other processes (maintenance or service). A disturbance will be expressed with “Cancel process”. The state of the process immediately changes to “suspended”. Resuming the process after resolving the disturbance can occur in three ways:

- the process should finish its period that was proceeding before the disturbance. The process returns to the original state. If it was scheduled, the duration of the period will be the remaining period at the moment it was cancelled. In terms of the process description the process “proceeds with the line it was executing”. In between, the causing process can perform its actions to deal with the disturbance.

- The process should resume, but skip the period it was in at the moment the disturbance occurred. For example a product that was being handled, is removed during the disturbance and transferred to another process. The process may now proceed as if the period finishes and deal with the next state change. Again in terms of the description, the process resumes with the line immediately following the period before the disturbance. Now the sentences will read:

- Finally it may be required to restart the process completely.

The process interventions cause transitions of the process state. They are shown in figure 4. It shows that cancel and proceed can be requested for any state of the process that was initiated with an advance statement. A stop request changes the state of the process into suspended. ‘Start’ forces the process to execute its description from the first line, ‘resume’ forces it to continue its description with the line immediately following the point where it was cancelled, and ‘proceed’ continues the description by completing the ‘advance’ statement, which was being executed at the moment of cancel.

**PROCESS INTERACTION IN DELPHI**

In order to implement the structure of Fig. 4 in Delphi, there are several conditions to be met.

First of all we need a mechanism that is able to sort and control ‘events’, which are introduced primarily by ‘advance’ statements on a single time-axis. This “sequence mechanism” should gain control at some point in a general Delphi-application where the user wants to start the simulation and it should return control to the application when the simulation is finished.

Secondly we need to be able to jump across the processes in a way as determined by the sequence of events that are added to the time-axis by means of the ‘advance’ calls inside the process descriptions and external interventions (this is a different sequence than processor execution would have followed without advance).

For meeting these conditions in a single threaded way, two approaches are known. One may apply the approach of method splitting, which actually would require some kind of preprocessor in order to analyze the process descriptions and split them into several event descriptions at the point of ‘advance’ statements. Our goal however is to provide a simulation mechanism within a general application, so we choose the other approach of ‘stack swapping’. We have interpreted this approach as ‘stack contents swapping’. The sequence mechanism is the only specific mechanism in simulation having control of the time axis. All other useful mechanisms like queues, distributions, graphics and animation are not involved in time control.

We therefore decided to split the simulation tool into two levels: the kernel on one side, containing the sequence mechanism, and a user-friendly shell, which enables users to use the sequence mechanism in a convenient way, where error detection takes place along with other facilities (like queues and distribution sampling). The kernel was called TomasProcess, the shell was called TOMAS, which stands for Tool for Object oriented Modeling And Simulation. In this paper TomasProcess is explained.

Every call to ‘advance’ “schedules” the calling process at some moment in future. The term “schedules” means that the execution of the statement sequence is stopped at the advance-
statement and should resume at the determined event time with
the next statement. In the mean time (as depicted by
the simulation clock) other processes may take control. In its basic
form we should thus be able to interrupt the current processor
execution and continue at a point, which is determined by the
sequence mechanism. We should be able to return to this point
in this process when the event time will occur. In order to
continue the process, we must save the return-address of the
calling process i.e. the address of the next statement.

Whenever Delphi source is compiled (and this is analogous for
c++), a computer code in assembler is generated. Delphi also
offers the facilities to the programmer to add lines in assembler
code by means of the inline assembler. We decided to use these
facilities to save the return address after a call to ‘advance’ and
to jump to an address which is determined by the sequence
mechanism.

Calling ‘advance’ will enter the advance-procedure. Before the
execution of the first statement in a procedure, Delphi takes
‘invisibly’ care of local variables and parameters; after the
execution of the last statement in a procedure, Delphi returns to
the calling procedure. Delphi accomplishes this by pushing
local parameters and the return address on the “stack”. In
general, after a call to a procedure, the stack will look like
figure 5.

Fig. 5 shows the stack for a procedure with 3 parameters.
Furthermore Delphi maintains a register EBP to reflect the
stack-address for finding parameters and local variables inside
the procedure. The old value of EBP is pushed on the stack
first and EBP is initialized to become the base pointer for
accessing information on the stack frame.

Finally local variables (like “SomeTemp” in fig. 5) are also
being pushed onto the stack.

We decided not to interfere with these conventions, so the only
thing we do is putting a label after the last statement of the
advance procedure and marking the label address for the
sequence mechanism as the return address to the calling
process. After this label Delphi performs its normal assembly
execution to return to the calling process. The only thing left
to do now is to guarantee that the stack at the moment of
returning is exactly the same as on entry to the advance
procedure.

The sequence mechanism will control the application during
the simulation. It gains control at the start of the simulation by
means of a call to “StartSimulation”. Everything that happens
on the stack from this point is under control of the sequence
mechanism. Inside the StartSimulation procedure the stack
pointer ESP and all registers are saved. ESP is saved in
InitialStackPosition. On return from StartSimulation (which
signals the finish of the simulation) all registers and ESP are
restored and the application regains control. During the
simulation, values can be pushed on the stack from the point
InitialStackPosition. Processes call the ‘advance’ procedure
and then three things happen:

- According the descriptions in the Delphi manual some
  registers should be saved; this is implemented by
  pushing them on the stack.
- All values on the stack between the InitialStack
  Position and the current stack position are saved
  locally with the calling element (the process is
  implemented as a method of an element class).
- The return address within the ‘advance’-procedure is
  also stored locally.

Then the sequence mechanism is called for the selection of the
next event. After selection ESP receives the value of Initial-
StackPosition and the locally saved data are moved to the stack
again. Finally the mechanism jumps to the local return address
as defined earlier.

The big advantage of this approach is that the original Delphi
calling conventions are kept intact and thus a call to ‘advance’
may appear in procedures or functions at any depth within the
process.

IMPLEMENTATION IN TOMAS

The implementation in TOMAS uses slightly different terms:

- Advance(T) is replaced by Hold(T), but implemented
  as described earlier. T is an interval starting now;
- Advance(while condition) asks for a different
  implementation, because passing a condition as a
  parameter requires professional skills in Delphi and
  that is impossible with starting programmers (and we
  want to educate the process interaction view to our
  students). However, Delphi offers a ‘While condition
  Do’ - construction, which can also be used. Replacing
  ‘Advance’ by ‘StandBy’ (expressing the state of the
  element more clearly) a statement like ‘While
  condition Do Standby’ is able to implement
  conditional events. In TOMAS the conditional events
  are checked after every event. This is simply
  accomplished by returning to the standby-calling
  process after each event. Then the condition is
  checked again in the user environment and the process
  calls StandBy again if the condition is still TRUE.
- Advance without any parameters is called ‘Suspend’ in TOMAS and is implemented as described before. Only another process is responsible for resuming the process now.

To complete the picture, we have to deal with different return addresses, after the process has been cancelled. As explained, we save the address of the statement following ‘advance’. However, the element can be cancelled by another element and can proceed in three ways: Start, Proceed and Resume. In the case of Resume the return address is the saved address. In the case of Proceed the element is again inserted into the event chain by the sequence mechanism and waits for the moment of its event time, to proceed at the saved address. In case of Start the element uses the address of the Process itself, which is saved at the time of creation of the element. This is the only reason why the method’s name containing the process description should always be ‘Process’.

Application programming with TOMAS only uses the TOMAS shell, which incorporates the kernel TomasProcess. So the Standby and Hold procedures call procedures in the shell. There the kernel procedures StandByProcess and HoldProcess are called, but before some error checking is added. In this way user friendly messages can be provided.

The structure of TOMAS and TomasProcess is object oriented. We have introduced the standard class TomasElement to introduce standard simulation attributes (like ArrivalTime, Status and Process). The default method Process does nothing more than producing a warning that the element doesn’t have a process, and it can be easily overwritten by the user-defined process description (preceded by reintroduce). Each TomasElement can have only one process. Above that a TomasElement is provided with extremely fast queuing facilities, referring to the implemented TomasQueue element class. This class offers all methods, which are required for simulation such as Length, FirstElement and Remove. For statistics, data like MeanLength and MeanWaitingTime are automatically gathered.

Finally TOMAS provides a well-known random generator and implements several statistical distributions for sampling purposes. All basic requirements for simulation are available in TOMAS, but in a normal Delphi application. One can use therefore for example the extended reporting facilities of Delphi and its basic interface with Microsoft Word and Excel. Numerous components have been developed for Delphi and they can be used also. We have added a 3D animation tool to TOMAS, which was based on freeware animation procedures in the internet community and combined it with the TomasElement class definition.

CONCLUSIONS AND FURTHER RESEARCH

This paper shows the implementation of the real process interaction view for simulation in a general purpose language. By this several advantages are accomplished:

- The modeling process for simulation can be integrated in general business modeling, which applies the systems approach, where it is common use to think in terms of processes. Integration also means that simulation can be used from the very start of a design process.
- Every participant of a design process can be involved and unexpected situations can be solved together.
- The simulation model can be implemented (almost without change) into the designed system and function operationally.
- Although programming skills are required, novice programmers are able to master the simulation principles and to develop quite complicated models very quickly.
- All facilities and developed extras of the general purpose language are available. The simulation expert will be not limited by the tool being used.

We actually applied this in a project called ‘Maasvlakte II’ [Kim and Gunther, 2007], and all team members participated in the discussion on the design of the simulation model. The results of the model were considered very reliable and simulation became an integral part of the design project.

The described implementation of TOMAS in Delphi is as free source available at www.tomasweb.com. It is not a commercial tool, but primarily used for education and consultancy. TOMAS has been used now for several years and no errors have been reported for the last two years.

TOMAS has already been adapted to support distributed simulation, which was not a real problem, because we had full control on the sequence mechanism. In fact a two-level time-axis was the only thing to be provided. One time-axis combines the first event times of all models participating and each model keeps its own time-axis but asks for permission to proceed. Currently, we are extending the sequence mechanism to support mixed discrete continuous simulation.

TOMAS is now extensively being used in our university. The main reason for this is its direct connection to process modelling that is applied by different disciplines.

REFERENCES


MULTIPLE-MODEL DESIGN AND SWITCHING SOLUTION FOR NONLINEAR PROCESSES CONTROL

Ciprian Lupu
Dumitru Popescu
Catalin Petrescu
Alexandru Ticlea
Bogdan Irimia
Catalin Dimon
Andreea Udrea

Automatic Control and Computers Science Department
University “Politehnica” of Bucharest
313, Splaiul Independenți, Bucharest,
Romania
E-mail: cip@indinf.pub.ro

KEYWORDS
multiple models, switching algorithm, manual-automatic bumpless transfer, real-time control systems

ABSTRACT

Using multiple model structures is one of the successful solutions for the real-time control of nonlinear or multi-regime processes. Additionally to designing the multi-controller corresponding algorithm, using this structure, imposes solving some specific problems, like best algorithm selection or control algorithm switching. The main goal of this paper is to provide a method for switching the algorithms of the multiple-models structure, based on the principles of manual to automatic bumpless transfer. The applicability of the method is proved using a real-time structure with an RST control algorithm. The results are tested on a special designed hardware and software experimental platform.

Related to classical control loops, multiple-model based systems need addressing some supplementary specific aspects:

- Dimension of multiple-model configuration;
- Selection of the best algorithm;
- Control law switching.

From multiple-models control systems point of view, two application-oriented problems can be highlighted:

- Class of systems with nonlinear characteristic, which can not be controlled by a single algorithm;
- Class of systems with different operating regimes, where different functioning regimes don’t permit the utilization of a unique algorithm or this would be a very complex one with special problems on implementation.

Depending on the process particularities, several multiple-models structures are proposed (Balakrishnan, 1996). One of the most general architectures is presented on Figure 1.

![Figure 1: General scheme for multimodel structure](image-url)

INTRODUCTION

The essential condition for the real-time functioning of a control system is preserving the closed-loop performances in case of non-linearity, structural disturbances or process uncertainties. A valuable way to solve these problems is the multiple-models or multicontroller structure. The first papers mentioning the “multiple-models” structure/system have been reported in the 90s. Balakrishnan and Narendra are among the first authors addressing problems of stability, robustness, switching and design for this type of structures in their papers (Narendra and Balakrishnan, 1997).

Research refinement in this field have brought extensions to multiple-model control concept. Parametric adaptation procedures – Closed-Loop Output Error (Landau and Karimi, 1997), use of Kalman filter representation (Lainiotis and Magill, 1967), the use of neural networks (Balakrishnan, 1996) or of the fuzzy systems (Dubois, Dieullant and Borne) are some of the important developments.
On Figure 1, the blocks and variables are as follows:
- Process – physical system to be controlled;
- Command calculus – unit that computes the process control law;
- System’s state or position identification – component that provides information about the model-control algorithm “best” matching for the actual system’s state;
- Mod. 1, Mod. 2, ..., Mod. N – previously identified models of different regimes or operating points;
- Alg. 1, Alg. 2, ..., Alg. N – control algorithms designed for the N models mentioned above;
- SWITCH – mixing or switching between the control laws;
- SELECTOR – based on adequate criteria evaluations, provides information about the most appropriate model for the system’s current state;
- y and y1, y2, ..., yN – output of the process and outputs of the models, respectively;
- output generated by Command calculus block;
- u and u1, u2, uN – output of the Command calculus block and outputs of the N control algorithms, respectively;
- r – system’s set point or reference trajectory;
- p – disturbances of physical process.

As noted above, function of the process particularities and the approach used to solve the “control algorithms switching” and/or “the best model choice” problems, the scheme can be adapted on the situation by adding/eliminating some specific blocks. This paper focuses on the “switching” problem.

CONTROL ALGORITHMS SWITCHING

Corresponding to multi-model structure’s functioning logic, after finding the best algorithm for the current process’s functioning point, the next step consists on switching the control algorithm. Two essential conditions must be verified with respect to this operation:
- To be designed so that no bumps will appear at the application of the control law;
- To be (very) fast.

Shocks determined by the switching operation cause non-efficient and/or dangerous behaviors. Moreover, slow switching determines boiling down the control algorithm’s action zone, which involves system’s performances alteration.

These are the main problems to be solved when designing the algorithms’ switching block. Firstly, from structurally point of view, this block may contain all algorithms’ implementation, or, secondly, at least the algorithms’ coefficients.

Classical solutions

Present solutions (Landau, et al., 1997), (Dumitache, 2005) solve more or less this problem and they are based on maintaining in active state all the control algorithms, also called “warm state”. This supposes that every algorithm receives information about the process output y(k) and set the point value (eventually filtered) r(k), but only the control law u(k) is applied on the real process, the one chosen by the switching block. This solution does not impose supplementary function logic for the system’s architecture and, for these reasons, it gives the possibility of switching very fast the algorithms. The drawback of this approach is that when designing the multi-model structure several supplementary steps are necessary.

These supplementary conditions demand the matching of the control algorithm outputs in the neighborhood switching zones. The superposition of models identification zones accomplishes this aspect, that can be seen on Figure 2. As a result of this superposition, the multi-model structure will have an increased number of models.

Figure 2: Superposition of identification zones for two neighbor-models and their corresponding control actions

Other approaches (Dussud, et al., 2000), (Pages, et al., 2000) propose mixing two or more algorithms’ outputs. The “weighting” of each control law depends on the distance from the current process’ operating point and the action zone of each algorithm. Based on this, the switching from an algorithm to another one is done using weighting functions with a continuous evolution in [0–1] intervals. This technique can be easily implemented by using the fuzzy approach. An example is presented on Figure 3.

Figure 3: Algorithms weighting functions for a specified operating position

This solution involves solving control gain problems, generated by the mixing of the algorithms’ outputs.

Proposed solution

This paper presents a solution that provides very good results for fast processes with nonlinear characteristics. The main
idea is that, during the current functioning of multiple-models control systems with N model-algorithm pairs, it is supposed that just a single algorithm is to be maintained active, the good one, and all the other N-1 algorithms rest inactive. The active and inactive states represent automatic, respectively manual, regimes of a control law. The output value of the active algorithm corresponds to the manual control for all the other N-1 inactive algorithms. In the switching situation, when a “better” Aj algorithm is found, the actual Ai active algorithm is commuted in inactive state, and Aj in active state. For a bumpless commutation, manual to automatic transfer problems must be solved; a solution is proposed in section 3.

The system can be implemented in two forms – first - with all inactive algorithms holding on manual regime, or – second - just a single operating algorithm (the active one) and activation of the “new” one after the computation of the currently corresponding manual regime and switching on automatic regime. Both variants have advantages and disadvantages. Choosing one of them necessitates knowledge about the hardware performances of the structure. After a general view, the first variant seems to be more reasonable.

In all situations, it is considered that the active algorithm’s output values represent manual commands for the “new” selected one.

**Practical considerations about the real-time algorithm implementation**

Consider the process’s discrete model:

\[ A(q^{-1}) y(k) = B(q^{-1}) u(k) \]  

where \( A(q^{-1}) \) and \( B(q^{-1}) \) polynomials are:

\[ A(q^{-1}) = 1 + a_1 q^{-1} + \ldots + a_n q^{-n_A} \]  

\[ B(q^{-1}) = b_0 + b_1 q^{-1} + \ldots + b_n q^{-n_B} \]  

with \( n_A \leq n_B \). For this model, an RST control algorithm is used (Fig. 4):

\[ S(q^{-1}) u(k) + R(q^{-1}) y(k) = T(q^{-1}) y^*(k) \]  

where: \( u(k) \) - algorithm output, \( y(k) \) - process output, \( y^*(k) \) - trajectory or filtered set point at moment k. The corresponding polynomials are:

\[ S(q^{-1}) = s_0 + s_1 q^{-1} + \ldots + s_n q^{-n_S} \]  

\[ R(q^{-1}) = r_0 + r_1 q^{-1} + \ldots + r_n q^{-n_R} \]  

\[ T(q^{-1}) = t_0 + t_1 q^{-1} + \ldots + t_n q^{-n_T} \]  

The closed-loop control representation is given on Figure 5.

**MANUAL – AUTOMATIC BUMPLESS TRANSFER**

The implementation practice highlights important problems like manual-to-automatic/automatic-to-manual regime commutations, respectively turning in/from the control saturation states. Of course, these problems exist in analogical systems and have specific countering procedures, which are not applicable on numerical systems.

The process operation begins on “manual” regime, this procedure being used as long as the process did not reach the nominal functioning zone. When commutation is done, it is recommended having a very good matching between the set point and process output values. This procedure releases the system of the shocks sent to the actuators. In the following, these facts will be illustrated using a RST control algorithm (Foulooy, et. al., 2004).

![Proposed solution](image)

**Figure 4: Proposed solution**

![RST algorithm, two freedom-degrees closed-loop canonical form](image)

**Figure 5: RST algorithm, two freedom-degrees closed-loop canonical form**

The control algorithm in (3) can be rewritten as follows:

\[ u(k) = \frac{1}{s_0} \left[ \sum_{i=1}^{n_S} s_i \mu(k-i) - \sum_{i=0}^{n_T} t_i y^*(k-i) \right] \]  

where \( n_S, n_R, n_T \) express the corresponding polynomials degrees and also the memory dimension for the software implementation of the algorithm. For example, if \( n_S=2 \), then it should be reserved three memory locations for the
process’s output: y(k), y(k-1), y(k-2). Respectively, the same rule applies for u(k) and y*(k).

When necessary, an imposed trajectory can be generated using a trajectory model generator:

\[ y^*(k + 1) = \frac{B_m(q^{-1})}{A_m(q^{-1})} r(k) \tag{8} \]

with \( A_m \) and \( B_m \) like:

\[ A_m(q^{-1}) = 1 + a_m q^{-1} + \ldots + a_{mn} q^{-n_A} \tag{9} \]
\[ B_m(q^{-1}) = b_m q^{-1} + \ldots + b_{mn} q^{-n_B} \]

In order to implement this solution, one must be interested in the control algorithm and, eventually, the trajectory’s model generator. One single iteration of the continuous monitoring program implies the following steps:
- Process’s data acquisition;
- Trajectory computation (if necessary);
- Control law computation;
- Sending the controls to the actuators;
- Process evolution graphical display;
- Actualization of the algorithm’s memory for the next iteration.

For example, the control law computation, when \( n_z = n_s = n_f = 2 \) and without trajectory generator \((y^*(k) = r(k))\), is like in the following:

\[ u(k) = \frac{1}{s_0} \left( -s_y u(k-1) - r_y y(k) - \right. \]
\[ \left. -c_y y(k-1) + t_y y^*(k-1) \right) \tag{10} \]

and (11) gives the algorithm’s memory actualization for the next iteration:

\[ u(k-1) = u(k); y(k-1) = y(k); y^*(k-1) = y^*(k). \tag{11} \]

**Manual/automate transfer**

In real functioning, manual to automate transfer is preceded by “driving” the process in the nominal action zone. To avoid command’s “bumping”, one must respect the following two conditions:
- Process’s output must be perfectly matched with the set point value;
- Accordingly with the algorithm’s complexity (function of the degrees of controller polynomials), the complete algorithm’s memory actualization must be waited of.

Neglecting these conditions lead to “bumps” in the transfer because the control algorithm’s output value is computed using the actual, but also the past, values of the command, process and set point, respectively.

At the same time, there are situations when the perfect “matching” between process’s output and set point value is very difficult to be obtained and/or needs very long time. Hence, the application of this procedure becomes impossible in the presence of important disturbances.

In this context, since the algorithm’s output is the manual command set by operator and the process’s output depend on command, the set point remains the only “free” variable in the control algorithm’s computation. Therefore, the proposed solution consists in the modification of the set point value, accordingly with the existent control algorithm, manual command and process’s output.

Algorithm’s memory actualization is done similarly as in the automatic regime. For practically implementation it is necessary a supplementary memory location for the set point value. From (8), it results the expression for the set point’s value:

\[ y^*(k) = \frac{1}{s_0} \left( \sum_{i=0}^{9} s_i u(k-i) + \right. \]
\[ \left. r \sum_{i=0}^{9} s_i y^*(k-i) \right] \tag{12} \]

When the set point (trajectory) generator (9) exists, keeping all the data in correct chronology must be with respect to the following relation:

\[ r(k) = \frac{B_m(q^{-1})}{A_m(q^{-1})} y^*(k) \tag{13} \]

System’s functioning scheme is presented on Fig. 6.

**Figure 6: Computation of the set point value for imposed manual command**

Concluding, this solution proposes the computation of that set point value that determines, accordingly to the algorithm’s history and process’s output, a control equal to manual command applied by the operator. At the time instant of the M→A switching, there are no gaps in the control algorithm’s memory that could determine bumps. An eventually mismatching between the set point and process’s output is considered as a simple change of set point’s value. Moreover, this solution can be successfully used in cases of command limitation.
EXPERIMENTAL RESULTS

We have evaluated the achieved performances of the multi-model control structure using a hardware and software experimental platform. The software was developed using National Instruments’s LabWindows/CVI. In figure 7, one can see a positioning control system. The main goal is the vertical control of the ball’s position.

Figure 7: Process simulator software application

The nonlinear relation between the position Y (%) and actuator command U (%) is presented on Fig. 8.

![Nonlinear diagram of the process](image)

Figure 8: Nonlinear diagram of the process

One considers three operating points P1, P2, and P3 on the plant’s nonlinear diagram (Fig. 8). Three models can be identified: M1 (0-30%), M2 (30-70%) and M3 (70-100%); these models are the regions for corresponding algorithms function.

Accordingly to the models-algorithms matching zones (Fig. 2), we have identified the models M1, M2 and M3, as being appropriated to the following intervals (0-55%), (45-90%) (80-100%), respectively. For a sampling period T_s=0.2 sec, the least-squares identification method from Adaptech/WinPIM platform identifies the next models:

\[
M_1 = \frac{0.0964 - 0.19647q^{-1}}{1 - 1.06891q^{-1} + 0.22991q^{-2}}
\]

\[
M_2 = \frac{0.01297 + 0.05397q^{-1} + 0.03674q^{-2}}{1 - 0.76251q^{-1}}
\]

\[
M_3 = \frac{0.02187 + 0.05668q^{-1} + 0.06048q^{-2}}{1 - 0.93161q^{-1} + 0.02741q^{-2} + 0.09863q^{-3}}
\]

In this case, we have computed three corresponding RST algorithms using a pole placement procedure from Adaptech/WinREG platform. The same nominal performances are imposed to all systems, through a second order system, defined by the dynamics \( \omega_0 = 1.25, \xi = 1.2 \) (tracking performances) and \( \omega_0 = 2, \xi = 0.8 \) (disturbance rejection performances) respectively, keeping the same sampling period as for identification.

All of these algorithms control the process in only their corresponding zones.

\[
R_s(q^{-1}) = 1.863259 - 2.027113q^{-1} + 0.520743q^{-2}
\]

\[
S_s(q^{-1}) = 1 - 1.000000 - 0.554998q^{-1} + 0.445002q^{-2}
\]

\[
T_s(q^{-1}) = 3.414848 - 4.931505q^{-1} + 1.873910q^{-2}
\]

\[
R_t(q^{-1}) = 2.309206 - 1.624937q^{-1}
\]

\[
S_t(q^{-1}) = 1 - 0.815278q^{-1} - 0.106427q^{-2} - 0.078295q^{-3}
\]

\[
T_t(q^{-1}) = 9.645062 - 14.928993q^{-1} + 5.968200q^{-2}
\]

\[
R_t(q^{-1}) = 1.72482-1.611292q^{-1} - 0.03784q^{-2} + 0.292903q^{-3}
\]

\[
S_t(q^{-1}) = 1 - 0.725187q^{-1} - 0.095205q^{-2} - 0.179608q^{-3}
\]

\[
T_t(q^{-1}) = 7.192692 - 11.645508q^{-1} + 4.821405q^{-2}
\]

To verify the proposed switching algorithm, it was designed and implemented a multi-model controller real-time software application, which can connect with the process simulator. The user interface is presented on Figure 9.

![Multi-model controller real-time software application](image)

Figure 9: Multi-model controller real-time software application

On the top of Fig. 9, there are respectively the set point, the output and control values, manual-automatic general switch, general manual command and graphical system evolution display. On the bottom of Fig. 9, one can see three graphical evolution displays corresponding to the three controllers (R_s, S_s, T_s, i=1...3). The colors are as follows: yellow – set point value, red – command value, blue – process output value and green – filtered set point value.

Using this application, there have been made several tests in order to verify the comportment of the system while switching between two algorithms. The switching procedure is determined by the change of the set point value. These tests are:

a) from 40% (where algorithm 1 is active) to 60% (where algorithm 2 is active). The effective switching operation is done when the filtered set point (and process output) becomes greater than 50%. Fig. 10(a) presents the evolutions.
b) from 80% (where algorithm 2 is active) to 100% (where algorithm 3 is active). The effective switching operation is done when the filtered set point (and process output) becomes greater than 85%. Fig. 10(b) presents the evolutions.

Figure 10: a) switching test; b) switching test.

In the tests showed above, it can be seen that there are no shocks or at most very small oscillations in the control evolution by applying this approach. Increasing the number of models-algorithms to 4 or 5 could eliminate even these small oscillations.

CONCLUSIONS

The method was successfully tested on a hardware and software experimental application with nonlinear characteristic, using a 3 multi-model/controller real-time software application. Moreover, the first variant (with all algorithms active) of the approach was implemented, ensuring the fast switching (one step) between algorithms.

With regards to the results obtained in the paper, the switching method can be successfully recommended in multi-model real-time control structures for fast processes.

ACKNOWLEDGEMENT

This work was supported by IDEI Research Program of Romanian Research, Development and Integration National Plan II, Grant no. 1044/2007.

REFERENCES


Pages, O., P. Mouille and B. Caron. 2000. Multi-model control by applying a symbolic fuzzy switches, IFAC.

BIOGRAPHY

CIPRIAN LUPU studied automatics at the Politehnica University of Bucharest. Since 2007 he is reader at Automatic Control and Computer Science Faculty from Politehnica University of Bucharest. His main research interests are real time control systems, nonlinear control and process modeling and simulation.

DUMITRU POPESCU is dean of Automatic Control and Computer Science Faculty and head of Research Center ACPC. His main research interests are real time continuous control systems, large scale systems and nonlinear control.

Catalin Petrescu studied automatics at the Politehnica University of Bucharest. Since 2007 he is reader at Automatic Control and Computer Science Faculty from Politehnica University of Bucharest. His main research interests are embedded systems, assembler programming and system identification.

ALEXANDRU TICLEA studied automatics at the Politehnica University of Bucharest. Since 2008 he is lector at Automatic Control and Computer Science Faculty from Politehnica University of Bucharest. His main research interests are system theory and numerical methods.

BOGDAN IRIMIA graduated in 2006 Automatics and Computer Science Faculty at the Politehnica University of Bucharest. Since 2007, he is a research assistant and PHD student at the same faculty. His research interests are: nonlinear process control and real-time programming.

Catalin Dimon graduated in 2006 Automatics and Computer Science Faculty at the Politehnica University of Bucharest. Since 2006, he is a teaching assistant and PHD student at the same faculty. His research interests are: continuous control systems and system identification.

ANDREA UDREA studied automatics Automatics and Computer Science Faculty, University Politehnica of Bucharest and graduated in 2006. She received the master of science degree in Advanced Automatics in February 2008. Since 2007, she is a teaching assistant and PHD student and her research interests include process control and fractal modeling techniques.
SIMULATION IN MANUFACTURING
LARGE SCALE PRODUCTION SYSTEMS
THE OPERATIONAL STRATEGIES IN FOCUSED FLEXIBLE MANUFACTURING SYSTEMS

F. Nucci, A. Grieco
Department of Innovation Engineering
University of Lecce
73100 Lecce
Italy
email: francesco.nucci@unile.it

KEYWORDS
Discrete event simulation, focused flexibility system, manufacturing system, performance evaluation

ABSTRACT

In this paper a simulation model able to provide a complete framework to the performance evaluation of Focused Flexible Manufacturing Systems is illustrated. The goal of the work is to identify all the operational open question in the management of FFMS and to provide different strategies to manage part transport systems in FFMS. Two different alternatives are proposed and evaluate in order to assess the disadvantages of FFMS versus FMS in terms of system performance (resource saturation and throughput). The proposed approach gives good results in terms of system performance and the presence of two alternative part transport systems seems to adequately cover the increase of part transport system in FFMS. An experimental campaign supports this conclusion.

INTRODUCTION

Manufacturing Flexibility is considered as the main aspect to face markets in which frequent volume changes and product requirements are present even if the investment in FMS solution is capital-intensive. Recently, a new idea related with the design of manufacturing system architectures endowed with the correct level of flexibility in order to face the specific production problem has been introduced. The proposed approach, referable under the term of Focused Flexibility Manufacturing Systems (FFMS) (Tolio 2008), tunes system flexibility on the production problem, in order to address the uncertainty related to product evolution.

Compared with the classical approach in the design of Flexible Manufacturing Systems (FMS), the new one considers Flexible CNC machines suitable only for a limited portion of the entire part program whereas dedicated machines are exploited for certain classes of processing. The main idea under the proposed approach is to tune the level of flexibility on each specific case on the basis of the demand scenarios, technological information of the parts under consideration and economical aspects. If on one side the FFMS solutions may lead to economical savings on the other the adoption of dedicated machines reduces the parallelism level and can lead to bottleneck in the production flows. For example, focusing the part transport systems, the number of pallet transfers increases; this occurs because the pallet is processed on multiple machines. Indeed, considering FFMS, the number of transport units is higher than FMS case.

A relevant management question consists in configuring the transport system in order to face the higher rate of pallet transfers. Since a classical transport solution adopted in FMS consists in a railway system, the passage from one transport unit to two units leads to increasing of the traffic because of interference effect. Respect to this particular example and in order to maximize the system performance (resource saturation and throughput), the goal may be to avoid from an operational point of view to have the transport system and the dedicated machines as a bottleneck of the system.

In general term, the FFMS solution may be of interest if the FFMS performance may be comparable to the FMS ones. Indeed, since machines in an FMS have a higher flexibility index (e.g. degrees of freedom) than FFMS machines, the machine cost for FFMS is lower than FMS case. In this work, different management policies for the transport system have been designed in order to maximize performance for the FFMS and to adapt such a production system to several production scenarios. In this way, the best management policy allows correctly exploiting the limited flexibility level that is available. Since the complexity of the system to be evaluated (setups, distributed processing of pallets, lot demands, tool and pallet management rules), a simulation model has been developed in order to evaluate the performance in different scenarios.

The paper is organized as follows. First a description of FFMS is reported. Second a literature review is presented. Then, the application of the simulation theory is analyzed and the simulation tool is reported. Finally, a case study is inferred from real test cases considering
different production scenarios and goals.

FOCUSED FLEXIBILITY MANUFACTURING SYSTEM

In the last decade, the production of mechanical components to be assembled in final products produced in high volumes has undergone deep modifications due to the overall changes in the way companies compete. In this situation it is possible to notice the directions adopted by the firms. First, strategic components tend to be manufactured by the companies that produce the final products. For these strategic components firms define long term plans. As a consequence, the technological characteristics, even if in continuous evolution, can be predicted with high reliability. Second, less critical parts tend to be externalized. In a context of continuous cost reduction, the producers of components try to obtain economies of scale by enlarging their size while specializing on specific types of components.

For this reason, it is important to note that single companies work on quite stable product categories produced in high volumes but, at the same time, they must cope with frequent product modifications and short product lifecycles. These frequent changes, even if restricted to a defined domain, have an impact on production processes. In this situation the acquisition of production capacity is particularly difficult. On the one hand, dedicated manufacturing systems, even if very competitive from the point of view of costs, are not adequate to accommodate continuous product changes. On the other hand, flexible manufacturing systems have excessive flexibility which remains almost unused and has a negative impact on costs.

Therefore, there is the need to address this trade-off creating new manufacturing systems with the minimum level of flexibility required by the production problem. Within this project we will call them systems with focused flexibility (Tohio 2008, Terkaj and Tohio 2007, Tolio and A.Valente 2006, Tolio and Valente 2008, Tolio et al. 2007). To solve this problem it is necessary to develop a methodology to assess the requirements in terms of flexibility by analyzing the productive problem of the firm, and to develop a system design methodology which starts from the flexibility levels required.

From a scientific point of view, the problem is particularly complex because focused flexibility, even if it provides economical advantages, it reduces some of the safety margins which allow to decouple the phases of manufacturing system design. Moreover, focused flexibility asks for a very careful risk appraisal. As a consequence, all the activities ranging from the detailed definition of the manufacturing strategy to the configuration and reconstructions of production systems must be strictly integrated. As a consequence, there is the need of combining and harmonizing different types of knowledge which are all essential to obtain a competitive solution.

LITERATURE BACKGROUND

The definitions adopted to define the term flexibility, both in the case they describe a reactive or a proactive attitude of the firm toward the competitive environment, do not remove the misconception that flexibility is always a desirable characteristic of a system. However, it is rather frequent to find in the literature the description of industrial situations where flexible systems have unsatisfactory performance (Koren et al. 1999, Landers 2000), cases where available flexibility remains unused (Matta et al. 2000, Sethi and Sethi 1990), or cases where the management perceives flexibility more as an undesirable complication than a potential advantage for the firm (Steece 1985). Reference (Kulatilaka and Marks 1988) shows that, at strategic level, flexibility can even be detrimental under certain circumstances particularly when uncertainty can be limited by means of proper agreements and contracts.

First of all, it is generally recognized in the literature that manufacturing flexibility is multidimensional in its nature (Sethi and Sethi 1990, Browne et al. 1984, Gerwin 1993, Hyun and Ahn 1992, Koste and Malhotra 1999, Upton 1994, Beach et al. 2000) and that a specific form of flexibility can be tied to each dimension. Therefore, a dominant theme in the literature regarding the manufacturing flexibility in the last twenty years is the creation of taxonomies to define and classify different forms of flexibility (Browne et al. 1984, Sethi and Sethi 1990).

A way to limit the drawbacks of the described situation is to define more precisely and in quantitative terms the required forms of flexibility. To this aim, many contributions in the literature try to identify precisely the required flexibility profiles (Koste and Malhotra 1999, Upton 1994, Sharma 2002). The idea followed in the works of this area, is to vary the level of a specific form of flexibility and to analyze the resulting performance of the production system (Bengtsson 2001, Hutchinson and Pfuhleof 1994, Matta et al. 2000). However, the strong dependence of the obtained results on the types of flexibility implemented and on the type of manufacturing system considered, does not allow the definition of general quantitative relations between implemented flexibility and performance of the system.

Summarizing, each type of flexibility has an impact on performance. However, this improvement is normally not proportional to the change in flexibility levels. Therefore, during the system design phase, it is necessary to evaluate if the increase in performance justifies the extra-costs required by flexibility. At the moment, this problem remains largely unsolved, in particular respect to instances representing actual applications. Consequently, it is necessary to find ways to express explicitly the flexibility needs of the firms starting from
the analysis of the productive problem.

SIMULATION FOR EVALUATING PERFORMANCE OF FFMS AND FMS

Since the paradigm to perform all the operation on a flexible machine can be discarded, multiple machine sets perform processing a single pallet in the new FFMS architectures. For this reason, several transportation requests are issued by the pallet during a complete part program execution: in addition to transportation missions between Load/Unload (LU) station and machines, pallet transportation between machines are present. The comparison between FFMS and FMS can be performed from different points of views. From one hand, an economic evaluation can be made. From the other, the performance can be evaluated in order to take into consideration parameters such as throughput and resource utilization. Since the goal of the proposed parameter is to provide a tool to obtain a performance analysis of FFMS architectures, qualitative consideration about FMS versus FFMS are provided in the following. Because, in FFMS, pallets require an additional transport from a machine to another one to complete the part program, the transporter saturation is higher than the FMS. Moreover, on each machine the part program length is shorter and this leads to lower machine saturation. In terms of part production rate, the FFMS has lower performance than FMS. Indeed, the additional time, necessary to move the pallets between machines, leads to an extra time that directly increases flowtime and, consequently, affects work in progress and throughput. The main characteristic of the FFMS paradigm consists in dividing the pallet part program on different machine sets. On one hand, tool sharing is not a critical aspect because of the lower number of tool copies. Indeed, a static allocation of tool can be made on the basis of the machine set for each operation. FFMS tool policy is simpler than the FMS in which tool can be requested on any machine. On the other, FFMS pallet management rule is fundamental because of the higher time to complete the part program. Indeed, transport system utilization and pallet entering rate in machine are higher than FMS case. The importance of the pallet management module implies the necessity to carefully consider the pallet path in the system in order to improve the final performance. In order to evaluate such a performance it is necessary to exploit the discrete event simulation tool. In this way, different pallet management rules can be included in the simulation model logic and the related performance indexes compared. Since the complexity, in terms of rules and detail level, of the system to be analyzed, it is not possible to formulate and solve an optimization model to assess and maximize performance. The transport system is a key aspect in FFMS plants, because the higher saturation of pallet carrier significantly affects performance. This led to the necessity to improve the performance of pallet transport system by adding another pallet carrier unit. Carrier units share the same railway and can not outrace. For such a reason, one carrier is on the left side and the other on the right one. Transporters may be in conflict when they have to perform missions in which a carrier is on the path that the other has to cover. Deadlock may occur if carriers move along the same path in different ways. A specific logic is necessary to avoid a collision that is (i) blocking a carrier until the other mission is completed or (ii) moving the carrier through a different (longer) path than the original one. The issue instant of a transport mission is when the part manager requires the transport mission. The transport mission is in critical state when the delay of the mission execution leads the machine (or LU station) to be stopped due to lack of pallets. The order the transport missions are performed with influences the efficiency of the transport system. Missions should be assigned to a specific carrier unit not to a generic one because this can change the efficiency of the transport system. For example, the position of the two carriers should be taken into account in order to avoid collisions between the transport units. In general, transport mission attributes are:

- Starting station
- Arrival station
- Issue instant: when the resource asks for the carrier.
- Critical time: period of time until the transport mission becomes critical.

The collision avoidance policy has to manage the priority of the carrier in order to avoid interference. In particular two approaches are suitable:

- Alternating priority policy. Each time a carrier begins a mission, it has to avoid collision with the carrier which is already moving.
- General priority policy. A carrier has the priority on the other to avoid deadlock; priority can be changed over time or depending the system state.

The order the pallet transport missions are executed significantly affects the resource saturations. A specific policy has to be selected. Such a policy allows:

- balancing the lot production volumes,
- balancing the machine/LU station resource utilization,
- increasing the number of missions performed by the part transport system.
The first aspect considers the respect of demand lots. Indeed, the lot production rate depends on the priority the corresponding pallets are moved in the system. Usually, in order balancing lot production volumes, the “NopartsProduced/PartDemand” ratio is used to set the priority of each pallet. The second takes into account the resource state in order to select the next pallet to be move. For example, a policy can select the next mission depending on the issue instant or the critical time. In particular, the machine/LU station resource states are considered. Finally, the optimization of transport system has to be analyzed in order to optimize carrier path; this consists in ordering of missions depending of departure and arrival stations.

A hybrid approach has been adopted in our work; this policy considers the N most critical missions in terms of lot and resource balancing and selects that one optimizing the transport system. In this way, the optimization of the transport system is limited only to a very limited number of missions (the most critical ones) and reduces the length of each transportation mission.

CASE STUDY

The validation activity of the simulation model is a long task in which various functionalities have to be checked in both qualitative and quantitative terms. The considered case study represents a typical production problem in which FFMS approach can be an effective alternative to the classical FMS. The objective of this study consists in comparing the performance of FMS and FFMS approaches taking into account both configuration and management policy of the pallet transport system.

Case study description

The system is made of 6 CNC machines and a Load/Unload station. In the classical FMS approach, all 6 machines are capable to perform any operation of the part programs. Instead, for the FFMS, machines are divided in two sets (3 machines per set) and are capable to perform only a limited part of the work plan. Each operation is performed only on a machine set (set 1 or set 2). LU station is supposed to be centrally located within the layout in-line; the distance among machines is 4 m. Machine characteristics are different as it is possible to note in Table 1.

The pallet transport system can use up to 2 carriers sharing the same in-line path; carrier speed is 45 m/min and carrier pallet load time is 15 sec. A suitable logic is adopted by the simulation model in order to avoid collision between two carriers. Tool carrier characteristics are irrelevant because no tool sharing is required. Five different lots have to be produced considering the same demand value. Each lot is machined on a specific setup sequence. Each setup is associated to a pallet. In Table 2, the number of parts loaded on each pallet is reported together with the pallet machining sequence. For example: lot #1 is processed with 5 setups and pallets #1, #2, #3, #4 and #5 are adopted for such setups; each pallet contains 30 parts of lot 1.

At the LU station, 6 different working units are used to load/unload parts from the pallet fixtures. The load/unload time depends on the number of parts loaded on the pallet: each part is loaded in 24 sec; e.g. pallet #1 is loaded with 30 parts in 12 min. In FFMS, for each pallet operation, the machine set, on which the operation is performed, is specified in Table 3; e.g. pallet #1 requires 4 machining operations: operation #1 and #2 have to be performed on machine set 2, whereas operation #3 and #4 are executed on machine set 1. Instead, if FMS is adopted, any machines can process the entire part program of any pallet. For each operation, the experiment cutting times are inferred from a uniform distribution with min/max values equal to 3 and 6 min; once inferred, such numerical values are used for all the simulation runs. For this reason, the simulation is deterministic. For each tool type, up to six copies are available. The tool life is equal to 5000 min and the regeneration time is 20 min.

Experimental campaign

The parameters of the experimental campaign are reported in Table 4. As it is possible to note the two paradigms FMS and FFMS (parameter D) have been investigated under different conditions concerning: carrier number (parameter C), carrier policy (parameter P) and tool configuration (parameter T). In order to assess system performance (resource saturation and throughput), four different indicators have been calculated. The first one is the average value of machine saturation. The second gives the average of the transport system saturation, that is calculated considering the overall utilization of the total number of carrier available; e.g. if two carriers are present in the system, 100% transport system saturation value is obtained if both carriers are always seized. The last one is the throughput of the system considering all the completed parts.

<table>
<thead>
<tr>
<th>Element</th>
<th>FMS set 1</th>
<th>FFMS set 1</th>
<th>FFMS set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shuttle rotation time [sec]</td>
<td>10</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>Tool change time [sec]</td>
<td>2.5</td>
<td>2.5</td>
<td>3</td>
</tr>
<tr>
<td>Tool Magazine</td>
<td>200</td>
<td>200</td>
<td>100</td>
</tr>
</tbody>
</table>
Table 2: Lot - pallet association

<table>
<thead>
<tr>
<th>Lot</th>
<th>Pallet setup #1</th>
<th>No. parts on setup #1</th>
<th>Pallet setup #2</th>
<th>No. parts on setup #2</th>
<th>Pallet setup #3</th>
<th>No. parts on setup #3</th>
<th>Pallet setup #4</th>
<th>No. parts on setup #4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>30</td>
<td>2</td>
<td>30</td>
<td>3</td>
<td>30</td>
<td>4</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>28</td>
<td>6</td>
<td>24</td>
<td>7</td>
<td>24</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>24</td>
<td>9</td>
<td>24</td>
<td>7</td>
<td>24</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>36</td>
<td>11</td>
<td>30</td>
<td>12</td>
<td>30</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>13</td>
<td>24</td>
<td>14</td>
<td>18</td>
<td>15</td>
<td>18</td>
<td>16</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 3: FFMS machine set for each pallet operation

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 4: Parameters of the experimental campaign

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>indicates whether the system adopted the FMS or FFMS approach</td>
<td>• 0: FMS approach (only one generic machine set), &lt;br&gt;• 1: FFMS approach (two machine set)</td>
</tr>
<tr>
<td>C</td>
<td>number of pallet carriers</td>
<td>• 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• 2</td>
</tr>
<tr>
<td>P</td>
<td>carrier management policy; the rule selecting the next pallet to be moved</td>
<td>• 1: the most critical transport mission is performed; &lt;br&gt;• 2: the mission optimizing the transport system is chosen, &lt;br&gt;• 3: the mission optimizing the transport system, considering only the top-5 critical missions, is selected.</td>
</tr>
<tr>
<td>T</td>
<td>number of copies per tool type</td>
<td>• 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• 6</td>
</tr>
</tbody>
</table>
Results

By varying the reproduced system, which depends on D/C/P/T parameters, the performance values are calculated from simulation runs in terms of machine and transport system saturation rate and throughput. The model has been implemented in SIMAN language (Pegg et al. 1995) with the package Arena (Rockwell) ver.7 under OS Windows XP SP2. The simulation model exchanges input and output data with a database in MS Access XP, consequently a generic experiment is described by the corresponding database. Moreover, a graphical interface allows checking the simulation progress (see Figure 1). The logic has been implemented using the VBA programming language that is embedded in Arena. The level of detail for such a simulation model considers the pallet as the main entity and simulates all the operations performed on various step (Loading/Unloading and Machining) with different tools.

Since the system is deterministic, one replication is necessary. The warm-up period length is 480 min (8h), whereas the replication length is 7200 min (120h). Execution time of a simulation run is about 5 min on PC Intel core 2 CPU, 2.33 GHz, 2 GB RAM. The most significant 8 experiments \((id = 1, 2,\ldots, 8)\) are reported in Table 5.

Numerical results show that the FMS paradigm \((D=0)\) allows reaching the maximum performance level (see Table 5, \(id = 1)\). Transport management policy is not relevant because the transport carrier is not critical \((id = 2)\). Indeed, an additional carrier unit is not necessary \((id = 3)\). If the number of tool copies is lower than the machine number, the lack of tool can significantly reduce system performance \((id=4)\). Note that the influence of the tool management policy has not been investigated.

On the other hand, with the FFMS approach \((D=1)\) the number of tool copies can be theoretically limited to 3. If only one carrier is exploited, transport system is a bottleneck for the manufacturing plant \((id=5)\). Instead, if a 2-carrier transport system is adopted, then the machine saturation and throughput significantly increase \((id > 5)\). In this case, it is important to correctly manage the transport system in order to avoid interference between the two carriers. Indeed, carrier moves along the single rail and, sometime, has to perform longer path in order to avoid the collision with the other carrier. The policy \(P=3\) allows to balance the machine mission requests and carrier movements \((id = 8)\).

CONCLUSION

The objective of the work is to measure the performance of an FFMS with a double part transport system versus FMS with a single part transport system. The goal is to measure the system performance in order to provide the necessary performance data to an economical evaluation model. Results show the necessity to adopt a double carrier transport system when FFMS paradigm is adopted, instead of a classical FMS; this allows recovering the performance gap between FMS and FFMS. Such a gap can be further reduced if a suitable pallet management policy is adopted.

Further investigation will be performed in order to develop a scheduling model to maximize FFMS performance taking into account both machine, part transport system and lot demand.

ACKNOWLEDGEMENTS

This research has been funded by Italian Ministry of Education and Research (MIUR) within the project PRIN 2005 titled "Methodologies and tools to design production systems with focused flexibility" and supported by AITeM (Associazione Italiana di Tecnologia Meccanica, www.aiitem.org).

REFERENCES


Table 5: Numerical case study results

<table>
<thead>
<tr>
<th>id</th>
<th>D</th>
<th>C</th>
<th>P</th>
<th>T</th>
<th>Machine saturation [%]</th>
<th>Transport system saturation [%]</th>
<th>Throughput [parts/min]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>100%</td>
<td>61%</td>
<td>2.47</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>100%</td>
<td>62%</td>
<td>2.46</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>100%</td>
<td>32%</td>
<td>2.43</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>100%</td>
<td>38%</td>
<td>1.13</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>67%</td>
<td>95%</td>
<td>1.34</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>83%</td>
<td>79%</td>
<td>2.12</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>80%</td>
<td>75%</td>
<td>2.09</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>85%</td>
<td>79%</td>
<td>2.16</td>
</tr>
</tbody>
</table>


BIographies

**Francesco Nucci** studied Computer Science at the University of Lecce in 1998 and got his Ph.D degree at Polytechnic of Milan. Since 2001 he is research assistant at the group of Production System in the Department of Innovation Engineering of the University of Lecce. He is mainly interested in simulation under uncertain conditions and manufacturing system scheduling.

**Antonio Grieo** received his degree in civil engineering from the Polytechnic of Bari in 1992. From 1992 to 1995 he worked in the Institute of Industrial Technology and Automation of the National Research Council of Italy. Since 1996, he has been a researcher in technology and manufacturing systems at the Department of Innovation Engineering of the University of Lecce. His interests are in the areas of management and scheduling in manufacturing systems (in particular FMS plants) and Fuzzy Sets theory application in technological and management questions.
Figure 1: Partial display of the simulation model
VALIDATION OF THE DESIGN OF A PROPOSED STEELMAKING FACTORY

Nicolas Pipard
Corus RD&T
PO Box 10000
1970 CA Ijmuiden
The Netherlands
E-mail: nicolas.pipard@corusgroup.com

Steve Thornton
Corus RD&T
Eston Rd., Grangetown
Middlesbrough, TS6 6UB
United Kingdom
E-mail: steve.thornton@corusgroup.com

KEYWORDS
Manufacturing, Decision-making, Layout optimization, Validation, Simulation interfaces.

ABSTRACT
Corus was approached to assist in designing a new steelmaking installation. A major need was to use simulation to evaluate and optimise proposals for several phases of the development, including the verification and validation of the design, the sizing of the different treatment operations and the defining of the layout.

The circumstances of the project presented a great number of challenges for the simulation engineer; such as the limited information due to the high confidentiality of the project, the complexity of the project including the multi-phase approach and the need for a clear, accurate animation to facilitate communication and rapid decision making in the project team. The biggest challenge of all, however, was the need to accomplish this under a very tight deadline, requiring a pragmatic and disciplined approach for knowledge capture and modelling.

This paper presents an overview of the project and the modelling results, highlighting the challenges posed and methodologies and techniques applied to address these.

INTRODUCTION
The manufacture of steel is based either on the melting of recycled steel scrap in electric arc furnaces, or more often via the blast furnace and oxygen steelmaking process (integrated steelmaking). For the integrated case, iron ore, coke and lime are fed continuously into the blast furnace and this iron ore is reduced to a molten iron containing about 4% carbon and other elements such as manganese, silicon, phosphorus and sulphur. In the subsequent linked steelmaking process, the carbon content is reduced through the exothermic oxidation process and phosphorus removed to produce iron with about 0.1% carbon, some manganese and other elements. Still in the liquid state and transferred in up to 350 tonne ladles, the steel is then refined further in secondary steelmaking processes to fine tune the chemical composition and add alloys to meet the product specification. Processes in this stage include vacuum degassing to remove hydrogen and further reduce carbon content, ladle furnaces for reheating (temperature control is a very important factor), and various types of material injection. Several ladles of steel will be in process at the same time in the secondary steelmaking phase. The liquid steel is then solidified in the continuous casting process into a semi-finished product which has a cross section appropriate to the rolling operations which are decoupled from the steelmaking (although efforts are often made to speed transfer to retain heat and thus reduce energy needs in reheating for the rolling processes). The design of the continuous casting process is such that successive steel ladles can be poured through a water cooled mould to produce a continuous strand of steel section which is cut off into lengths of several metres to be processed downstream (sequence casting). Thus the continuous blast furnace process feeds the batch steelmaking process, which is close coupled to the continuous casting process, and hence to the decoupled batch orientated rolling process.

Figure 1: Sequence of Stages in the Steelmaking Process (Bureau of Energy Efficiency n.d.)
OBJECTIVES
This project was concerned with the redevelopment of an existing low production outdated steelmaking facility with brand new Oxygen Steelmaking, secondary steelmaking processes and multi-strand continuous casters. The implied change for the existing installation was enormous and therefore it was essential that the proposed solution was clearly validated and communicated very effectively. Discrete event simulation modelling presents an excellent medium to facilitate a multi-disciplined project team to achieve shared understanding about a relatively complex scheme, and guiding them towards good decisions very quickly. The Arena software from Rockwell Software has been used for this project. Specific objectives were as follows:

- Determine overall feasibility and sensitivity with respect to constraints and bottlenecks.
- Specify size and numbers of process installations to achieve the required throughput.
- Optimise ladle sizes and numbers to support sequence casting through the continuous caster.
- Validate proposals for a phased development of the scheme.
- Test various scenarios for proposed operational strategy and process timings.
- Recommend crane numbers and positions and other transfer options for moving ladles through the steel plant.
- Model impact of unplanned events, e.g. returned steel due to temperature loss.

The time scales for the exploration of these aspects was extremely short and undertaken against a background of high confidentiality placing some special challenges on the simulation modeller.

CHALLENGES AND APPROACHES TAKEN

Project Planning And Management
Good planning was essential to make sure that there was sufficient time spent in the three main stages of the simulation project, knowledge capture, model development and experimentation.

Launching The Project
The initial project meeting made clear that only a very experienced modeller could achieve the complex project tasks within the tight deadline. As the most experienced modeller was however fully occupied with other projects, a solution needed to be found to free his time. After careful consideration of all options, it was decided that the less critical of his running simulation projects would be slightly delayed and that one less-complex project be handed over to another simulation expert. By keeping only parallel projects that were too important to delay or hand over, the modeller was able to free sufficient time to take on the challenge of this new project.

In the next step objectives were specified and deadlines agreed upon. To avoid promising the impossible and to ensure the customer was aware of the length and importance of the different development phases, this step was completed with particular care.

Understanding The System
Once the project was launched, the modeller was challenged with the task of getting to know a production system he was not familiar with. In order to do so, the modeller gathered basic information about the system from various sources, incl. the contact person and internet sites such as steeluniversity.org. In this project activity charts helped the knowledge capture process and the initial visualisation of activity cycles around each installation. This also helped the modeller to further evaluate the likely duration and complexity of the project.

Not being an expert in the real system can be beneficial for the modelling process, as the knowledge capture tends to be more complete by asking many questions, consulting multiple sources and thus minimizing the risk to miss essential detail. Further the system is initially modelled at a much higher level, with complex rules only added later when the objectives of the study become more specific. This way the model does not become overly complicated too early: “the complexity needs not exceed that required to accomplish the purposes for which the model is intended” (Banks et al. 2005)

High Confidentiality / Communication
A great challenge throughout the project was posed by its high confidentiality and the consequent limitations on knowledge resources and communication. In fact there was access to only one contact person, the future user of the model. Therefore the quality and frequency of the communication with that contact became extremely important and keeping the contact involved in all the steps of the project was crucial. Like in any simulation project, regular meetings were held with the customer to make sure the project stayed on the right path. That meant also strategic planning of all meetings to ensure efficiency with regard to capturing missing information and minimising travelling. Updates regarding the project progress were spread carefully in order to keep the customer interested.

Model Development
Because of the time constraints on the project, the model development phase needed to be limited in time and effort. At times, model builders fall so much in love with the model that they keep on refining and refining. Then time is up and the presentation to management needs to be done (Centeno and Carrillo 2001). In order to prevent exceeding the time allocation to this activity, the right level of complexity for the different installations needed to be planned.
Modelling Of The Overhead Cranes
We can take for example the modelling of the overhead cranes, which cannot cross, because they are moving on the same overhead rail system. After putting together in a table all the possible positions of the cranes and the distances between them, the project duration did not allow enough time to model all the crane interactions complexity. In order to still make the right decision about the cranes, the interactions were recorded and outputted together with the crane utilisation figures.

Incorporating Animations
A clear, moderately accurate animation facilitates communication and rapid decision making in the project team. It offers a practical way of validating the model by allowing for a walk through of the model. Also animations help to achieve customer buy-in that the model is valid (Centeno and Carrillo 2001). Making animations too realistic, however, can easily take up a great proportion of modelling time. It was therefore crucial that the modeller chose the right level of animation. He achieved this with only a smaller amount of effort by using elements from the CAD drawing as a base for the animation features.

![Figure 2: Print Screen of Model Animation](image)

Scenario Evaluation And Optimisation
Due to the multi-phase development approach of the facility, it was clear from early on that many scenarios would need to be run to answer all study questions. As there was not enough time to maintain more than one model version, all configurations and possible scenarios had to be incorporated within a single model. In order to facilitate entering inputs and changing parameters both for the modeller and the customer, an active X control interface was developed. Active X control has the advantage to be flexible and interactive, which allows experimental capabilities directly based on the simulation workspace.

Further a simple output interface was created in Excel displaying performance parameters and main results from the simulation runs. Excel was chosen because of its pre-formatting facilities and availability on most computers making it easy to share and compare results from different scenarios.

With these interfaces/displays in place the project contact could easily be involved in the evaluation of the various scenarios, which was greatly beneficial to the evaluation due to his higher level of expertise. Also this active participation had the advantage of keeping the contact person involved, up-to-date and motivated.

OUTCOMES AND DEVELOPMENT OPPORTUNITIES

Outcomes
The results from the simulation study facilitated the communication of the people involved in the design of the installation by providing a platform to test the different design scenarios. The simulation project helped greatly to forecast the number of cranes, and the number of ladles. It also helped determining the number of ladle furnaces needed in order to satisfy the number of sequences on the different casters. The results from the simulation work helped sizing the process installations, influenced layout changes such as moving the location of casters, ladle furnaces, and ladle preparation area, and confirmed the production capability of the proposed industrial unit.

Determination of the number of cranes
As mentioned above the number of needed overhead cranes was a major factor to investigate in particular because of their cost and the difficulty of their implementation. The results from the simulation study confirmed the number of cranes needed on the first development phase of the facility, but showed a problem on the second phase. The crane utilisation was very high with a lot of hindering between the cranes. This led to a drastic change in the layout with the intention to reduce their utilisations and interactions. The model updated with the new layout confirmed the intended improvement.

Maximization Of Caster Productivity
The practice of sequence casting (casting a series of heats in succession without interrupting the casting process) is used to improve caster productivity as it avoids the frequent setting-up of the casting machine. A plant wide strategy is therefore required to achieve optimum results. (Craig et al 2001). If for any reason a ladle is too late, the casting has to be interrupted and the caster has to be prepared again, which results in loss of valuable production time (Hamoen and Moens 2002).

The ladle furnace installation, upstream operation of the caster, is used for treatment and reheating of ladles but it can also be used to hold ladles at a certain temperature, crucial for the casting operation. Enough ladles need to be held at the ladle furnace station before starting a casting sequence in order to ensure that the last ladles of the sequence will arrive in time at the caster. Since one ladle furnace will hold only one ladle, the number of ladle furnaces is crucial. The optimised number of ladle furnaces was found to be two per caster in order to achieve the required sequence.
Development Opportunities Identified

- Development of templates/simulators to reduce the time spent on model building and to have the opportunity of the customer developing simple models helping initial knowledge capture.
- Web deployment of tools and simulation models.
- Usage of data mining tools to pre and post process data captured and extract some complex rules from models.
- Integration of costing aspects in simulation models.

REFERENCES


AUTHOR BIOGRAPHIES

NICOLAS PIPARD, born in France, completed his BEng in Chemical Engineering (1999) and his MSc in Process Manufacturing Management (2000) at the University of Teesside, UK. When he graduated in 2000, he became researcher in Corus group, where he has been working in the field of simulation ever since. The projects he has worked on range from decision support on strategic projects through layout optimisation of production processes, to improvement of bottlenecks in supply chains.

STEVE THORNTON is an Engineering Fellow in Corus Research, Development and Technology, specialising in the area of Business Integration. Steve has a first degree in Mathematics with Engineering from Nottingham University (1984), and an MBA from Warwick University (1996). After time spent in the fields of mathematical modelling and process measurement and control, his current role involves the development and application of technologies including data mining, simulation and knowledge engineering for integration and optimisation of manufacturing and business systems.
EVALUATING THE IMPACT OF THE LEAN APPROACH ON A FOOD PROCESSING LINE

Flavio Tonelli, Roberto Revetria, Ernesto Revello
Department of Production Engineering, Thermoenergetics and Mathematical Models
University of Genoa
I 16145, Genoa, Italy
E-mail: flavio.tonelli@dipitem.unige.it

KEYWORDS
Food Processing, Lean Production, Simulation, Action Based Research.

ABSTRACT

Lean is now widely recognized as one of the most effective means for increasing competitiveness and improving operational efficiency. Anyway, what could be the benefit in adopting a Lean approach in the food processing industry, is not still clear. For these reasons, the Authors decided to develop an action research, inside a large food company sited in Italy, collaborating with the Plant Technical Support (PTS) unit to verify applicability and benefits of Lean in food processing. According to the action research methodology Authors provided specific expertise to the PTS people developing a simple but practical conceptual model and several research hypotheses. In order to investigate these hypotheses Value Stream Mapping and Discrete Event Simulation have been extensively used. Results confirm some of the generated hypotheses showing interesting improvement areas. Furthermore, PTS people realized that Lean results could be obtained quicker and to a greater extent through the use of simulation technology. As this paper shows, through predictive simulation modelling, the time to implement Lean is greatly reduced and hidden forms of waste become much more apparent.

INTRODUCTION

The analyzed food company produces a wide range of different snacks and biscuits. Depending of the transformation process different production lines can be adopted. A food processing line can be considered as a production flow involving transformation processes over a set of interconnected phases; from this point of view, a food processing line is implicitly Lean for many different aspects. In order to focalize the analysis on a specific case, Authors selected one single product realized on a single food processing line. The product is a unique fried snack sold in different packages. By analyzing the production process, from raw material to the package and shipment, several areas of improvement have been pointed out by the Plant Technical Support People:
- High working scraps,
- High quality variability
- High production lead-time.
- Many not value added activities (internal movements, set-up, packaging).
- Issues in managing stocks of semi finished product.
- Low processing line saturation.
- Low productivity.

Looking at the similarities between the aforementioned problems and the Lean approach, appropriateness of Lean in reducing these problems arose spontaneously from Technical People. Major doubts were related to the difficulty of connecting Lean techniques to the improvement areas in terms of quantitative justification. The literature review research was not satisfactory since very few contributions address the Lean approach and limitations with respect to the food industry (Cox and Chicksand 2006), probably because a food processing line is implicitly considered lean. Nevertheless, concepts like waste reduction, production levelling, work standardization, 5S, kaizen, and more, are quite simple and intuitive to understand but difficult to apply in a real context without having an indication about estimated costs and benefits. This is particularly true when the required interventions have to be made on a semi-automatic complex and expensive processing plant. Through predictive simulation modelling, the impact of a Lean approach can be estimated and the time to implement is greatly reduced since hidden forms of waste (poor operational planning, suboptimal use of resources, etc.) become much more evident. With this in mind Authors decided of proposing simulation to test and verify proposed technical solutions to create value and eliminate waste, all in a risk-free environment. Authors strongly believe that simulation achieves Lean better and quicker, but in order to understand how simulation can be beneficially applied to Lean initiatives, it is essential to understand the relationships between the Lean approach and the food industry. What tools are more appropriate to achieve food industry objectives? How are processes affected from Lean initiatives and what are the impacts? These are questions that should be answered in order to get an adequate picture of what Lean is all about the food industry. Following the next section will introduce the reader how simulation is being used to help realize the full potential of Lean even in the food industry. Then a section will illustrate research hypotheses and conceptual model. A further section proposes a Value Stream Map approach and outlines its limitation in this case. The last section provides a description of a discrete event simulation approach able to catch the desired insights and verifying the developed hypotheses. In the remaining part of
the paper results are reported and conclusions drawn putting in evidence the limitations of using Lean in food processing industry.

RELATIONSHIPS BETWEEN THE ANALYZED FOOD INDUSTRY AND THE LEAN APPROACH

The modern food industry is the result of technological and cultural changes that have occurred over the last 150 years. The modern food industry relies far more on technology, particularly on mechanization and biochemistry. In this way, food is raised, manipulated, preserved and moved around, resulting in a food industry that is to a great degree global in nature, with food and related resources travelling great distances. This modern food system relies heavily on technology, transportation, management and logistics for physical fulfilment, and on marketing and government regulation for maintaining an efficient consumer market. The food industry is used in a broadly inclusive way to cover all aspects of food production and sale. Food processing is the methods and techniques used to transform raw ingredients into food for human consumption. The food processing industry utilizes these processes. High quality and hygiene standards must be maintained to ensure consumer safety and failures to maintain adequate standards can have serious health consequences. Processing food is a very costly process, thus increasing the prices of foods products. When designing processes for the food industry the following performance parameters may be taken into account:

- Hygiene, e.g. measured by number of micro organisms per ml of finished product.
- Energy consumption, measured e.g. by “ton of steam per ton of sugar produced”.
- Minimization of waste, measured e.g. by “percentage of peeling loss during the peeling of potatoes”.
- Labour used, measured e.g. by “number of working hours per ton of finished product”.
- Minimization of cleaning stops measured e.g. by “number of hours between cleaning stops”.

Concerning this paper the following points have been investigated:

- Minimization of waste: particularly work-in-progress and finished products scraps.
- Labour used: number of shifts and productivity scenario analysis.

In order to understand the improvements achievable for these two points with respect to a Lean approach a brief description of finished product and production process is provided. The base product is a fried chip. The production process begins with raw material silos; the days of inventory can vary between seven and ten. The raw materials are spilled and blended according to the recipe (3x340 kg/batches treated for 60 minutes). Before the extrusion process the mixer processes the blended material for 20 minutes by adding 1% of oil and 30% of water. At this point the material is cooked for 5 minutes in the extrusion machine (800-900 kg/hr production rate). The extruded material is then cut according to the pellet size and pneumatically sent to three flake-machines. After the flaking process, always by using a pneumatic transport, pellets are sent to two dryer machines for the drying phase (20 minutes from cutting to drying). The dryer machine treats the pellets for 105 minutes and then a conveyor move the pellets to a Big Bag containing 740 kg. The Big Bags are stored in a specific warehouse featuring a stock coverage of three production days (from 86 to 400 Big Bags). From the Big Bag warehouse, by using forklifts, the bags are moved and cleared out through a vibrator and a conveyor feeding the fryer silos. The fryer machine can work a variable quantity between 500 and 850 kg/hr and the frying process lasts 19 seconds. The lead-time between Big Bag positioning and finished product exiting is about 15 minutes. The product is then aromatized and moved to a horizontal buffer through a conveyor belt. Finally, the product is packaged spending a variable time between 20 and 45 minutes according to the package type. The product can be packaged in three different ways:

1. Metallic bag inside a printed box.
2. Metallic small printed bags.

Single packages are stocked in big boxes and put on pallets to be shipped to distribution centres and dealers. Human operators for the following areas supervise the production process: extrusion, flaking, pre-drying, drying, and frying.

From a quick literature survey (Womak and Jones, 1998) it is evident that Lean provides numerous benefits including:

- Reduced lead/cycle time
- Decreased work-in-process (WIP)
- Reduced cost
- Increased resource (equipment, operator) utilization
- Easier scheduling
- More streamlined flow
- Reduced floor space, and
- Improved quality

All these benefits sound interesting to the above described production process so that Plant Technical Support People decided to find an answer to the following research question: “Lean methodology and tools, if adopted for a food processing line, can improve daily productivity and scrap reduction?”. Since the considered production context is partly and implicitly lean, at least in terms of continuous flow, Technical People was, mainly, interested in preserving that flow. From very practical considerations, continuous flow can only be achieved when disruptions in the system are eliminated. Common disruptions included late deliveries, equipment failures, operator delays, defective work, machine setups and operator error. In a process that supports continuous flow, operations are tightly coupled so that any disruption ripples through the system affecting both upstream and downstream operations. It therefore becomes crucial, wherever disruptions occur, to minimize the impact of these disruptions. The most common way to buffer a process from disruptions is to set up inventories in the process. This is usually done by establishing a storage whose level is regulated by some sort of signal (kanban) to ensure that the right amount of inventory is maintained to buffer the process from disruptions. Taking into account this strong orientation, the general Lean approach has been focussed on the following Lean operational techniques:

- work standardization,
• load balancing,
• kaizen,
• 5S,
• point of usage storage,
and tools:
  • Value Stream Mapping,
  • Kanban.
To evaluate improvement results six measure of performance have been identified (Parsec Automation Corp. 2006):
  • Scraps.
  • Quality.
  • Total cycle time.
  • WIP.
  • Line saturation.
  • Productivity.

The connection of independent variables (production factors) to expected performance (dependent variables), has been done through the intervention variables (the Lean techniques) and moderation variable (operator training), leading to the following conceptual model. The operator training has been considered a moderation variable able to amplify or inhibit the expected performances, but not explicitly considered in the modelling phase.

**Figures 1: Conceptual Model**

From the depicted conceptual model, and considering the described context, Authors developed several assumptions and hypotheses to validate through the quantitative modelling phase. In the following there is an explanation of the Lean operational techniques with respect to the proposed research purposes.

**Work standardization:** because of high automation level of this type of plants the manual component influences the total productivity of the production line. The work standardization, involving a complete documentation and application of best practices, assures that operations follow a standard sequence for different sub-processes benefits in scrap reduction, general process quality improvement (waiting times reduction), operational efficiency, leading to a possible work shifts increment thanks to the fact that repeatable operations are done in the same manner by different user at different times. The related improvements can be evaluated looking at productivity and scraps performance indexes.

**Load balancing:** flow bottlenecks can reveal in poor throughput performance and work-in-progress degeneration. The slowest phase/machine has to be tuned to assure the required tuck-time (at maximum production speed/maximum sales rate), but the remaining phases/machines have to set up close to that production rate, unless different batch consideration have to be done. Even in this last case a good load balancing results in a better flow and improved plant/resources saturation. In the analyzed scenarios this technique has been applied to extrusion, frying, and packaging phases since batches are decoupled with the remainder. The load balancing activity allows waiting times, production lead-time and work-in-progress reduction.

**Kaizen:** the continuous improvement achievable through the kaizen approach doesn’t require huge investments and production interruptions; people working on kaizen can fix small but significant problems increasing their productivity and maximizing the overall operational efficiency reducing times and scraps. Hence, in the analyzed scenarios, small percentage improvements of productivity and scrap indicators have been considered.

**5S:** the workplace organization according to the 5S principles improves efficiency acting on tools, material, and equipment optimal positioning as well as on cleaning the location. In the analyzed scenarios Authors considered a small increment in location keeping time versus a benefit in terms of time variability reduction. Optimized workplace allows a reduction of not-value added activities influencing positively the total flow value.

**Point of usage storage:** positioning the material, where the process will use it, reduces the internal transportation times and facilitates the inventory level visual management activity. In the analyzed scenarios Authors applied this technique to pellet stocking area in order to move the Big Bag unloading point close to the fryer machine reducing the previously required transportation time. The expected benefits are related to material work-in-progress and to loading time corresponding to a not-value added operations reduction.

**Value Stream Mapping:** VSM is used to statically visualize, analyze and improve process and information flows. The adoption of VSM tool facilitates the elimination of not-value added activities. The VSM allows underlining, in visual and intuitive way, every phase of the productive process focused on the family of products. In particular, the implementation of a VSM, is articulated in four distinct steps:

- selection of the product family;
- creation of a current-state map;
- creation of a future-state map, using lean techniques;
- implementation of the future-state map through structured continuous improvement activities.
Improvements typically include inventory reduction, cell creation, batch size reduction, and a more pull-oriented production. A VSM is created for a single product or product family and therefore presupposes that resources are dedicated to this process (like in this particular context). Often, however, resources (transport vehicles, maintenance crews, etc.) are used across multiple value streams. This certainly has an impact on process performance that should be taken into account. Practical adoption of VSM, to create the current-state map, outlined the continuous flow characterization of this particular context providing a flow value close to 1 (1,0227). In figure 2 the future-state situation has been mapped by using a VSM tool (flow value = 1,0112).

**Figures 2: VSM of AS-IS production line**

As it could be expected, the processing line steps are already organized in order to reduce the not value added activities. So a deeper investigation has to be done in order to understand how the selected Lean techniques could operate to reduce the waste and increase the productivity. Unfortunately, a VSM is effective for making tactical decisions such as whether continuous flow can be implemented and where “pull” needs to be used, but it is ineffectual in working out the operational details of the process such as allocating workers, prioritizing tasks and assessing the impact of variation on flow. A VSM presents only a static view of a process based on sketchy process information (e.g., operation times and setup times) and only basic performance data (e.g., cycle times and inventory levels). What it fails to capture is the statistical variability and detailed interactions in the process that often create almost unpredictable behaviour. While this may be okay for a current-state or future state-map, consideration of these factors becomes essential if process changes are to be evaluated.

**Kanban:** for example the introduction of kanban to regulate production and movement between process steps where continuous flow isn’t possible, is not a risk free trial and error activity, since its implementation requires a not negligible cost, and the possible failure could produce significant damages from the production point of view. Finding where kanban needs to be used and what kanban sizes should be it is not a trivial enabling different kanban sizes to be evaluated to see exactly what the tradeoffs will be in terms of cycle time, WIP, throughput and response time.

**SIMULATION AND LEAN**

With an understanding of the basic phases and targets of the considered process, it is possible to explore the application of simulation in the implementation of Lean. Simulation provides an effective means of achieving the goals of Lean on many different levels, helping organizations to reach the full potential of Lean by facilitating a better decision, faster. There are different ways in which simulation helps achieve the goals of Lean:

- training,
- project prioritizing,
- value stream mapping,
- kanban sizing,
- ongoing improvement.

The action research target, reported in this paper, was to clarify effective impact of Lean methodology on the considered food processing line. So, skipping the training the project priority goals, and value stream mapping (the flow value already results close to 1), avoiding generic kanban considerations (on the considered food processing line the kanban could be used only in the final phases), Authors and Plant Technical People decided to investigate deeper the problem developing a discrete event detailed simulation model of the food processing line in order discover operational waste and effective ways to eliminate it. In the first phases of Lean transformation, much of the waste is so glaringly obvious that the solution is almost self-evident. By reducing setup times and regulating flow through pull triggers, and implementing single-piece flow, work-in-process and cycle times can be greatly reduced (that is the analyzed case). To achieve continual improvement, however, requires being somewhat of a sleuth or detective. This is where simulation modelling comes into play. Simulation captures the actual dynamics of a process and allows one to conduct experiments, to test the impact of alternative changes, determining which improvements have the biggest impact on process performance, and it quantifies the amount of improvement one can expect. The simulation framework to develop the model was Simul8; in figure 3 a snapshot of the model is shown.
ANALYSIS AND RESULTS

The simulation model, once is built, can be used on an ongoing basis to find problems and test possible solutions corresponding to different hypotheses. The hypotheses have been defined trying an association between expected improvement and Lean technique adoption, according to suggestions received from Plant Technical People. From this process six hypotheses have been generated:

**Hypothesis 1 (Scraps decrease):** application of standardized work, kaizen, and 5S.

**Hypothesis 2 (Quality variability decrease):** standardized work, kaizen.

**Hypothesis 3 (Total Cycle Time decrease):** standardize work, work balancing, 5S, and point of use storage, kanban.

**Hypothesis 4 (WIP decrease):** work balancing, point of use storage, kanban.

**Hypothesis 5 (Line saturation increase):** standardized work, kanban.

**Hypothesis 6 (Productivity increase):** work balancing, kaizen.

For each hypothesis three scenarios have been evaluated:
1. current situation with average deterministic processing values,
2. current situation with stochastic processing values,
3. future situation with stochastic processing values.

For each scenario a set of possible real or estimated values (minimum, average, maximum), for each production factor (i.e. extrusion speed), have been explored functionally to the Lean technique introduced (i.e. work balancing).

In the figure 4 is reported a graphical summary of obtained results.

![Figure 4: VSM of AS-IS production line](image)

A specific evaluation of each scenario has been performed comparing scenario 2 versus 3 with respect to 8 Key Performance Indicators (KPI) defined by the Plant Technical People:

1. WIP (big bags number),
2. WIP (pellets number),
3. Line production rate (boxes/hr),
4. Total cycle time (min),
5. Fryer blocked time (%),
6. Line production rate (big bags/day),
7. Fryer rate (big bags/day),
8. Shipping rate (shipment/day).

From the numerical comparison the following considerations can be derived:

- A productivity improvement and a scraps reduction can be achieved by implementing standardized work, kaizen, and 5S.
- The production line saturation occurs when standardized work, kanban, and work balancing are implemented.
- A WIP reduction is related to the work balancing, and kaizen.

The influence of kaizen is present but limited by scenario estimations in improving operational steps. The simple reduction of processing time for only one production step reflects in a worse flow, hence WIP increase and fryer blocks (buffer with automatic blocking system). The kanban control adoption improves buffer discontinuities in material flow, i.e., where you can’t achieve continuous flow because of an unforeseen disturbance.

CONCLUSIONS

The action research reported in this paper presents some interesting findings for Lean practitioners and Plant Technical People. Firstly, it provides a simple but practical conceptualization of Lean in the food processing industry, outlining techniques and tools more suitable to face typical food industry targets. Secondly, it shows that VSM approach cannot provide useful insight at a processing line level (this is a quite obvious observation for Lean expert but less evident to technical practitioners). Lastly, it confirms, once
again, the power of simulation approach to gain a deeper insight of a process, especially when the process performance can be afflicted heavily by small, yet uncertain, disturbances. This situation is particularly critical with continuous flow processes where intermediate buffers tend to be reduced to the minimum level, often without having a complete understanding of the trade-off between uncertainty and buffers. A different research approach could be used (i.e. model base research), to define a more general theory about Lean adoption in food industry, leading to a more detailed conceptual model and hypotheses, but it would require a more extended sample of food manufacturers.

REFERENCES

MAINTENANCE PERFORMANCE SIMULATION
AVAILABILITY MODELLING FOR OPTIMIZING THE COMPLEX SYSTEMS MAINTENANCE

Radhouane DJERIDI, Mathieu GLADE, Samir GHELAM, Frédéric MATHERON
Eurocopter ETAVM / ETAST – Maintenance Research Department
Aéroport Marseille Provence
13725 Marignane - France
Email: [Radhouane.djeridi; Mathieu.glade; Samir.ghelam; Frederic.matheron] @ eurocopter.com

Aline CAUVIN
Laboratoire des Sciences de l’Information et des Systèmes (LSIS)
UMR 6168
2 cours des Arts et Métiers
13 617 Aix en Provence - France
Email: Aline.cauvin@lsis.org

KEYWORDS
Maintenance Planning Documents, Complex system, Availability, Maintenance, Scheduling.

ABSTRACT

Several researches have been realized in the production field. It’s only recently that the safety studies have been developed to optimize the maintenance management. The safety studies have to maintain the quality requirement of a system during his life cycle. This sciences field is broken down into 4 concepts: Reliability, Maintainability, Safety and the Availability. Our problematic concerns the modelling and the management of the availability in order to improve the complex systems performance. This paper presents a modelling method based on scheduling maintenance to optimize the availability.

CONTEXT AND PROBLEMATIC

In a high competitive market, the companies tend to provide sophisticated systems with reduced maintenance efforts while following safety requirements from the industry branch (rail transportation, shipbuilding, aerospace, power station...). These complex systems (aircraft, train, boat, rocket...) have the particularity of generating important operating costs in comparison with life cycle cost (Glade 2005).

Maintenance costs represent an important part of operating costs. They take into account the labour cost, depreciation, spares, consumable which can be reduced by optimizing the maintenance program (Busssel and schontag 2002). The maintenance program of these systems is synthesized in a Maintenance Planning Document (MPD) defined during the design phase. The MPD role is mainly to plan maintenance operation. It is based on reliability safety and maintainability studies. More precisely it provides details on maintenance operations needed for system maintainability such as maintenance intervals (Service Life Limit (SSL)), Time Between Overhaul (TBO), extreme climatic and derived conditions of use and maintenance inspections (Maintenance tasks and workcards) (Maisononneuve et al. 2007).

In this context, the helicopters are complex machines which have one of the most important maintenance man hours per utilization cycle (MMH ratio) comparing to other industrial applications. These unavailability time due to maintenance can lead to several losses (penalties). These losses can be financial in transportation service (offshore) case or human in search and rescue mission case. In all cases, the unavailability time is translated into a loss of earnings (Figure 1).

![Figure 1: Aircraft availability problem in regard to the maintenance effort](image_url)

Faced with this constraint, the helicopter industry is developing availability models to improve the fleet management and the maintenance policy alignment with the customer activities.

In this way, the availability modelling must lead to implement a customized maintenance planning document with the following actions:

- Study the maintenance inspections impact considering the actual use of the aircraft
- Identify maintenance tasks which can increase time between inspections and thus impact on the aircraft availability.
- Reduce maintenance tasks and adapt inspections to customer activities (Saving labour, time to repair)
PROPOSAL OF AN AVAILABILITY MODELLING METHOD

In this paragraph, we will present firstly a definition of the availability concept and the brief literature review of the availability modelling. Secondly, we present the global approach to model the availability and to optimize the maintenance scheduling.

Availability modelling : Definition

In this article, we exploit two availability concepts: operational availability and service availability.

"The operational availability is defined as the probability that, when used under stated conditions, a system will operate satisfactorily at any time. The operational availability also includes standby time, administrative and logistic delay time" (US D.O.D 1991). The service availability represents the probability of a successful service delivery in specific condition, on the user and / or the operator demand (Djeridi and Cauvin 2007). An other availability concept is used in our method to model the fleet availability; instantaneous availability which is used for follow up process.

Many methods are currently used to assess the operational availability (Petri Network, Bayesian Network (Smati and al. 2003), Neural network (Rajpal and al. 2005), Genetic algorithms (Celso and al. 1999), Monte Carlo (Labeau and Zio 2002)...). These method presents advantages for computing but need important means materials and license to treat complex system.

According to the method, the availability assessment can follow two methods: predictive method and a method based on feedback data processing. The predictive method is based on a theoretical analysis of the operating product and on design data of the system. The method based on feedback data consists in modelling the realistic availability and analyzing the unavailability time due to maintenance.

For our method, we used the MonteCarlo approach. This approach enables to simulate easily with any programming language and enables a flexibility for the optimisation step.

Availability modelling method

We proposed an analysis method which is structured around a predictive method using the random processes theory to define a theoretical maintenance planning documents and a consolidation method that enables to validate the Customized Maintenance Planning Documents created (Figure 2).

The using of the random processes is necessary to analyse the product failure occurred and the necessary time to restore the aircraft (Bussett and Zauijer 2004). To simulate the failure occurred, we used a failure distribution defined by segment of activities (Search and Rescue, Military, Oil & Gas,...) (Figure 3).

This analysis is realized with the Monte-Carlo simulation for the availability assessment and heuristics for the optimisation of the maintenance planning.

The Monte Carlo simulation is used to assess the operational availability and to generate a maintenance planning. These planning are generated with stochastic data which take account the safety margin of the TBO, flight hour rate, inspections milestones. These simulation aims to generate several scenarios in order to obtain the best solution.

The optimization is based on the comparison of the maintenance plan generated and criteria defined by the designer and the customer. These criteria are proposed in regard to the experience feedback data. To validate the feedback data we will use an availability tree which is a breakdown of required times in operation. The availability tree is used to focus on the main drivers and to set the parameters will be used by the customer.

The results of the modelling enable two improvement for the aircraft fleet. Firstly, a customized planning document can lead to reduce and to prevent the unavailability time. With this planning, it is possible to forecast the spares demand and to adapt the store policy of spares (Djeridi and Cauvin 2008).

Secondly, the maintenance planning enables to optimize the use of the fleet with regard to the mission (Figure 4). It's
possible in this way to optimize the utilization period of the aircraft in order to avoid an important maintenance inspection.

CONCLUSION AND PROSPECTS

The purpose of this article is to propose a scheduling method in order to adapt the maintenance inspection milestones defined by the MPD to maximize the operational availability.

The scheduling method is structured on several modules. The first module consists in the operational availability assessment with the standard state of Maintenance Planning Documents. The second module enables to generate maintenance planning. These planning are generated with random processes which take into account the safety margin of TBO, flight hour rate, inspections milestones. These simulation aims to generate the most possible cases in order to obtain promptly a sufficient solution. The last module is based on the comparison of the maintenance plan generated and criterion defined by the designer and the customer.

Others work concern the proposal of service availability and the modelling of logistics availability spares. These works tend to propose a methodology of supply chain design with regard to the service availability.

REFERENCES


Maisonneuve P.L., S. Ghelam, M. Glade., JP. Derain,, P. Lyonnet. 2007. “Maintenance credits : From the monitoring of helicopter mechanical parts to a dynamic maintenance planning”. In CEAS’ 07.


US DOD.1991. “DOD requirements for a logistic support analysis record”.

BIOGRAPHY

RADHOUANE DJERIDI was born in Cannes, France and went to the Ecole Polytechnique Universitaire de Marseille, where he studied Engineering Sciences and obtained his engineering thesis in the industrial organization field in 2006 and his Master of Sciences thesis in supply chain modeling. Since 2006, he prepares a PhD in Industrial Engineering and works for Eurocopter Company where he studies to model the maintenance management of the helicopter fleet. He carries his research activity in the Laboratory of Sciences of Information and Systems where he tries to formalize a design for availability methodology.
EVALUATION OF MAINTENANCE SCHEDULE OF A LIGHT METAL INDUSTRY PLANT USING PETRI NETS

Vrassidas Leopoulos
Charalampos Tsogkas
Evaggelos Champiromatis
Leonidas Tsikinis
National Technical University of Athens
Iroon Polytechniou 9
15780 Zografou

KEYWORDS
Maintenance, Light metal industry, Petri Nets, Simulation.

ABSTRACT

This paper presents the experience of a project aiming the evaluation of a risk based maintenance tool in a traditional production line of light metal industry (collapsible tubes). Maintenance for the industry under consideration is based on a well-defined process, for both corrective and preventive interventions, and failure rates for the components of equipment are based on existing historical data. The model calculates, through simulation, the expected output of the production line, for a specific production plan, taking into consideration the interruptions due to the maintenance work. The results are used in order to evaluate different maintenance plans. For the model development Generalised Stochastic Petri Nets (GSPN) have been used and the software package ARTIFEX. The project is co-financed by GSRT.

INTRODUCTION

The markets’ requirements for sophisticated high quality products in small amounts and quick response to the demand necessitate production systems with high flexibility. A major concern of the capital intensive industries is the availability of the plants as long production interruptions may result in customer dissatisfaction. Apart from spare capacity the answer to the problem is the proper maintenance (Katchmar, 1997).

Maintenance of the plant under consideration is divided into two major types: corrective and preventive, the former absorbing an important percentage of the resources, leaving restricted resources to the latter. The efforts to optimise preventive maintenance processes, necessitate a system supporting the prioritization of the work orders. Risk based maintenance methods can be used to provide the necessary ranking.

Evaluation of such methods necessitates simulation techniques (Adamyan and He, 2004) as the criticality of each equipment component depends often on the current production plan. In addition the nature of equipment failures is stochastic and the duration of maintenance tasks uncertain because the necessary activities are known only after the stoppage and inspection of the components (Ogaji, 2002, Tatsiopoulos et al, 1999). Several studies propose models based on Petri Nets for validation of maintenance plans (Schneeweiss, 2001, Yang and Liu, 1998). The resulting models for real maintenance processes are often too complicate to allow comprehensive graphs (Antoniadis and Leopoulos, 2002).

In the present study, the production line is modeled using Generalised Stochastic Petri Nets (GSPN). Generalised Stochastic Petri Nets (GSPN) contain both immediate and stochastic transitions. They can model activities that occur almost instantly, along with processes that take a longer and random amount of time. The latter is useful for modeling processing, failure and repair times of individual machines thus the physical operation of the production system. Two interconnected views have been developed:

1. The physical view which models the process transforming the raw material (aluminum disk) to the final product (collapsible tube).
2. The informational view which models the production orders and the feedback from the execution of a production plan.

The model is used in combination with a risk maintenance tool, developed to support risk based maintenance. The results of the simulation inform the planning manager about the throughput time of a production plan taking into consideration the availability of technicians for corrective maintenance.

MAINTENANCE PROCESS

The company under consideration has been in the aluminium business since 1965 with the first production line for collapsible tubes. In the years that followed the company developed the manufacturing capabilities, by successively installing production lines for aerosol cans and monoblock bottles. The plant operates a slugs operation and lubrication unit, a collapsible tubes production section, a monoblock aerosol cans production section and a monoblock bottle production section.

The Greek market of collapsible tubes remains traditionally under the direct control of the big multi-national companies of the cosmetics and pharmaceutics sectors. The market being very competitive, a major concern of the company’s management is the availability of the plant as long
production interruptions may result in serious problems. Therefore proper maintenance is critical for the company’s operation.

Maintenance is defined as a set of activities intended to keep a functional unit in, or to restore it, to a state in which it can perform a required function. It includes activities such as monitoring, tests, measurements, replacements, adjustments, repairs, and in some cases administrative actions (ISO, 1997).

The production line of the collapsible tubes section, the operations and the equipment used, is presented in the following Table 1.

<table>
<thead>
<tr>
<th>Operation</th>
<th>Equipment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extrusion</td>
<td>Press</td>
</tr>
<tr>
<td>Cut</td>
<td>Burr remover</td>
</tr>
<tr>
<td>Annealing</td>
<td>Annealing furnace</td>
</tr>
<tr>
<td>Internal laquering</td>
<td>Sprimag</td>
</tr>
<tr>
<td>Polymerisation</td>
<td>Polymerisation furnace</td>
</tr>
<tr>
<td>Buffer</td>
<td>Moving shafts</td>
</tr>
<tr>
<td>Printing</td>
<td>Printing drum</td>
</tr>
<tr>
<td>Color heating</td>
<td>Printing furnace</td>
</tr>
<tr>
<td>Buffer</td>
<td>Moving shafts</td>
</tr>
<tr>
<td>Stoppering</td>
<td>Cap screw machine</td>
</tr>
<tr>
<td>Rubber jointing</td>
<td>Rubber jointing machine</td>
</tr>
<tr>
<td>Handling</td>
<td>Conveyor</td>
</tr>
<tr>
<td>Quality control</td>
<td>Optic</td>
</tr>
<tr>
<td>Packaging</td>
<td>Packaging machine</td>
</tr>
</tbody>
</table>

**Table 1. Production line of the collapsible tubes**

Maintenance of the light metal industry under consideration is divided into two major types: corrective and preventive. The former comprises the majority of the maintenance activities; repairs are made to the equipment only when actual failure occurs, aiming to get the equipment into production as quickly as possible. The latter is intended to prevent breakdowns and minimise shutdowns through systematic inspection of equipment, adjustments, and scheduled repairs before failures occur. The resources for preventive maintenance are limited thus the parts that require intervention of a maintenance team are selected using a composite indicator. This indicator is based on the estimation of the risks associated with the part, including the probability of failure. The duration of the intervention is estimated using historical data.

The determination of this composite multi-criteria indicator helps the maintenance group to rank the maintenance actions in accordance with their significance, estimate the duration of each action and form a schedule of the maintenance actions for the “critical” parts of equipment.

In the present study a simulation tool is proposed, in order to evaluate the maintenance schedule thus produced. The influence of the schedule on the current production plan is taken into consideration for the final decision. The minimum necessary availability of the technicians for corrective maintenance is calculated, so the rest of the maintenance workforce can be assigned to the preventive activities.

**MODELLING, SIMULATION AND RESULTS**

Because of the locality principle on states and actions (transitions), Petri net models can be locally modified, refined or made coarse, without altering the rest of the model (Proth, 1993). This means that nets can be synthesised using top-down and bottom-up approaches. Both approaches have been used, in order to produce a model comprehensive and appropriate for simulation results.

**Figure 1. Production line – first level model**

A top down decomposition method has been used for the development of the first level model and each piece of equipment is modeled as an object. A part of the first level model is presented in Figure 1. A bottom-up (modular) composition method has been used in order to model each object using three generic modules, a mono-operational machine, a buffer and a transportation system. The model of the mono-operational unreliable machine is presented in Figure 2.

**Figure 2. Mono-operational unreliable machine**

A further top-down refinement, models the appearance of a failure or the interruption of the machine for maintenance purposes. These situations are modeled using the modules. Each transition ENTRANCE can be fired either sending a token to the place PHASE_A, beginning thus the process of piece on the machine, or sending a token to the sub-module
FAILURE INTERRUPTION interrupting the operation of the machine.
The firing of the transition “FAILURE” has two results: Firstly, removes the tokens from places “OK_1” and “OK_2”, and second it sends signal via the place of “TECHNICIAN REQUIREMENTS”, asking for a technician to repair the machine. The other output place “MACHINE UNDER REPAIRING” models the situation of a machine under repair.
If there is no technician available and consequently no token present in the place “TECHNICIAN AVAILABLE”, then repair work is not beginning. The duration of repair, that is to the firing of the transition “REPAIR” is a random variable.
When the transition “REPAIR” fires the tokens appear in places “IN”, “OK_1” and “OK_2”, and the machine returns in the regular function. The model of the machine failure and repair is presented in Figure 3.

Figure 3. Machine failure and repair

The interruption of the operation of the machine for maintenance purposes is modeled using the module presented in Figure 4.

Figure 4. Machine interruption
The number of the available technicians is given as the initial marking in the place TECHNICIAN AVAILABLE. The sub-module that models the technicians’ work is presented in the following Figure 5.

Figure 5. Technicians’ activities
The priority of the transition “FAILURE” to fire, over the transition “ENTRANCE” is ruled by a random variable. The reliability function of the variable is calculated using historical data. The following figure 6 presents the reliability function for the piece of equipment “press”.

Figure 6. Reliability function for the piece of equipment “press”
For each production plan simulation is run taking into consideration the machine failures. Each simulation set is run for a different number of technicians and the duration of the plan execution stored. The minimum necessary availability of the technicians for corrective maintenance is thus determined, leaving the rest of the maintenance
workforce available to be assigned to the preventive activities.

CONCLUSIONS – FURTHER RESEARCH

This paper presents a case study where Generalised Stochastic Petri Nets (GSPN) have been used for the evaluation of the maintenance activities in a light metal industry plant. Both top-down analysis and bottom-up synthesis have been used in order to produce a comprehensive model. Simulation is proposed in order to take into consideration the stochastic nature of both failures and duration of maintenance tasks. The model based on the failure rates of the equipment, the duration of maintenance tasks and the availability of the technicians calculates, through simulation, the duration of the execution of each production schedule, enabling the production managers to decide about the workforce size of the maintenance personnel, and to promise more accurate due dates to customers. Since the modeling of the physical system (components and sub-components of the plant) has been created a further step should be the combination of a production plan with a maintenance plan. The interaction between the production and maintenance plan will give the possibility to evaluate different maintenance schedules.

REFERENCES


AUTOMOTIVE ENGINEERING
SHEET HYDROFORMING OF AUTOMOTIVE B-PILLAR AND
COEFFICIENT OF FRICTION EFFECTS ANALYSIS

Javad Marzbanrad and Mir Javad Peyman
Automotive Engineering Department,
Iran University of Science and Technology
Tehran, Iran
E-mail: marzban@iust.ac.ir

KEYWORDS
Sheet hydroforming, Automotive B-Pillar, Coefficient of friction, Finite element method, Strain hardening coefficient.

ABSTRACT

Hydroforming is the fabrication process that uses a fluid medium to form a piece by using high internal pressure. In sheet hydroforming, a sheet is placed between two dies, sealed and pressurized water or oil up to 100 MPa is injected, deforming the sheet in the cavity form of the dies. The advantages of the hydroforming over the traditional process are: (a) weight reduction; (b) project flexibility increase, making possible to create new piece shapes; (c) stiffness and rigidity increase by obtaining tube pieces without weld; (d) dimensional repeatability. The weight reduction obtaining by the hydroforming process can be allied to the weight reduction by the substitution of materials of high density by material of low density. In this study, the hydrostatic force is used to forming a raw sheet to automotive B-pillar using the finite element method (FFM). The modeling and simulation were developed with the FE software. The results including thickness distribution of the formed product and friction coefficient effects are presented and compared to each other in the same conditions. In addition, the results show that the smaller the friction force at contact surfaces, the more uniform for the thickness of the final part.

INTRODUCTION

Due to a shortage of energy and increasing demands for environment protection, the manufacturing industry is obliged to develop advanced production technologies that permit savings of energy and raw materials. The automobile industry in particular needs new manufacturing methods for cars with reduced pollution without a significant cost increase (Schmoeckel and Geiger 1999). Sheet hydroforming, the basic principle of which is to supply a liquid medium between usually one or two metal sheets and then to form the sheets into the inner cavity of a die, is one of these new technologies to satisfy the industrial requirements. Compared with the conventional stamping process which involves the process sequence of deep drawing of two shell halves and subsequent joining, sheet hydroforming has both technological and economical benefits. Conventional stamping requires at least two additional stamping punches and a blank-holder while sheet hydroforming requires just two forming dies comprising an upper and a lower half (Geiger et al. 1999). A technological advantage of sheet hydroforming over stamping is a more uniform thickness distribution over the component due to the introduction of hydrostatic pressure. The hydroformed component also satisfies very high accuracy and surface-finish requirements so that it can be used without any further finishing operations (Hein and Vollertsen 1999). This paper surveys the processes of sheet hydroforming for production of an automobile B-pillar using a commercial explicit FEM code. The results are illustrated in figures and tables for comparison.

FINITE ELEMENT MODELING

In this research, CATIA software is used for modeling and ABAQUS code is applied for analysis the B-pillar. A flat sheet is used as raw material and pressurized hydraulically by simulation to convert to an automotive B-pillar component under different conditions.

The modeled rigid die is shown in figure 1. After die modeling, export it to ABAQUS software for finite element analysis. Similarly, deformable plate and rigid upper die are modeled and exported to analyzer software. Clamping forces applied on the upper die plates for prevention from standing up raw sheet during forming process. These parts and assembled collection are shown in figures 2-4.

![Figure 1: Modeled Rigid B-Pillar Die](image)

The material employed on the simulation was the steel alloy (IF). Mechanical Properties of this alloy are presented in table 1. Also true stress-strain diagram for this alloy shown in figure 5. According to material properties as shown in table 1, it seems selected steel alloy is adaptable for sheet hydroforming process. Increasing elongation of material and strain hardening coefficient, will be cause the thickness distribution more uniform.
The flow stress of the material was obtained by a tensile test and is expressed by a power law of its equivalent strain, i.e., \( \overline{\sigma} = KE^n \), where \( K = 580 \text{ MPa} \) and \( n = 0.32 \).

Loading path for the best forming and minimum trial and error, approximate by analytical formula simultaneously with consideration of material characteristics. Loading with fluid pressure in finite time period is shown in figure 6.

### Table 1: Mechanical Properties of Steel Alloy (IF)

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density (( \rho ))</td>
<td>7800 (Kg/m(^3))</td>
</tr>
<tr>
<td>Young modulus (E)</td>
<td>205 (GPa)</td>
</tr>
<tr>
<td>Poisson coefficient (( \nu ))</td>
<td>0.31</td>
</tr>
<tr>
<td>Strength coefficient (( K ))</td>
<td>580 (MPa)</td>
</tr>
<tr>
<td>Strain hardening coefficient (( n ))</td>
<td>0.32</td>
</tr>
<tr>
<td>Yield stress (( \sigma ))</td>
<td>147 (MPa)</td>
</tr>
<tr>
<td>Elongation</td>
<td>60 (%)</td>
</tr>
<tr>
<td>Initial plate thickness</td>
<td>1.5 (mm)</td>
</tr>
</tbody>
</table>

Simulation and analyzing repeated for three different coefficient of friction conditions and results compared in these situations. Coefficient of friction assumed \( \mu = 0.01, 0.1 \) and 0.2 in each case and Coulomb friction law is adopted.

**RESULTS OF FEM MODEL FOR \( \mu = 0.01 \)**

During performing, the lower die (figure 1) keeps static and the upper die (figure 3) moves downward to prevent standing up raw sheet and then the sheet is preformed into...
the shape as shown in Figure 8. After hydroforming, the thickness distribution and stress variation are shown in Figures 9-10.

**RESULTS OF FEM MODEL FOR $\mu=0.1$**

All of mentioned steps of analysis for $\mu=0.01$, repeated for $\mu=0.1$ and the results shown in figures 11-12.

![Figure 11: Stress Distribution During Process with $\mu=0.1$](image1)

![Figure 12: Thickness Variation During Process with $\mu=0.1$](image2)

In figure 11, element A is critical element in comparison with other elements. Maximum and minimum stresses occur in element A and B respectively. Also according to figure 12, minimum thickness of plate at the end of process is approximately 1.2 mm. In other words, maximum thinning percent is 20 %.

**RESULTS OF FEM MODEL FOR $\mu=0.2$**

Base analysis repeated for $\mu=0.2$ and the results shown in figures 13-14.

![Figure 13: Stress Distribution During Process with $\mu=0.2$](image3)

In figure 9, element A is critical element in comparison with other elements. Maximum and minimum stresses occur in element A and B respectively. Also according to figure 10, minimum thickness of plate at the end of process is approximately 1.43 mm. In other words, maximum thinning percent is 4.6%.
in element A and B respectively. Also according to figure 14, minimum thickness of plate at the end of process is approximately 1 mm. In other words, maximum thinning percent is 33%.

Figures 15-18 are shown stresses distribution and thickness variation in shaped B-Pillar at $\mu=0.1$ and 0.2.

Total comparisons between three cases presented at table 2. Increasing coefficient of friction will be cause increase stress and plate thinning percent.
Figure 16: Final Thickness Variation of Process with $\mu=0.1$

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Min. STH (fraction = 0.1) (Avg: 75%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+5.087e+08</td>
</tr>
<tr>
<td>2</td>
<td>+4.680e+08</td>
</tr>
<tr>
<td>3</td>
<td>+4.272e+08</td>
</tr>
<tr>
<td>4</td>
<td>+3.864e+08</td>
</tr>
<tr>
<td>5</td>
<td>+3.457e+08</td>
</tr>
<tr>
<td>6</td>
<td>+3.049e+08</td>
</tr>
<tr>
<td>7</td>
<td>+2.641e+08</td>
</tr>
<tr>
<td>8</td>
<td>+2.233e+08</td>
</tr>
<tr>
<td>9</td>
<td>+1.826e+08</td>
</tr>
<tr>
<td>10</td>
<td>+1.419e+09</td>
</tr>
<tr>
<td>11</td>
<td>+1.010e+09</td>
</tr>
<tr>
<td>12</td>
<td>+6.026e+07</td>
</tr>
<tr>
<td>13</td>
<td>+1.943e+07</td>
</tr>
</tbody>
</table>

Figure 17: Final Stress Distribution of Process with $\mu=0.2$

Figure 18: Final Thickness Variation of Process with $\mu=0.2$

Table 2: Total Results in Different Friction Condition

<table>
<thead>
<tr>
<th>Coefficient of friction</th>
<th>Max. Stress (MPa)</th>
<th>Min. Thickness (mm)</th>
<th>Thinning percent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>320</td>
<td>1.43</td>
<td>4.66</td>
</tr>
<tr>
<td>0.1</td>
<td>440</td>
<td>1.2</td>
<td>20</td>
</tr>
<tr>
<td>0.2</td>
<td>470</td>
<td>1</td>
<td>33</td>
</tr>
</tbody>
</table>

CONCLUSIONS

Because it has several advantages over conventional methods, sheet hydroforming is a new field of investigation in the manufacturing industry. In this paper, a numerical simulation using elasto-plastic FEM is carried out and an automobile B-pillar is produced using sheet hydroforming instead of the stamped-welded method used previously. Use of the finite element method to analyses a complex sheet metal forming and hydroforming problem has also proven to be effective as a design analysis tool.

The friction force between the die surface and raw sheet is more important parameter for the best forming. Increasing coefficient of friction (C.O.F) will be cause increase stress and plate thickness percent. High C.O.F causes great tangential forces due to very high internal pressure. The
tangential forces caused by friction hinders the flow of the material and decreases of thickness occurred. More thinning part of sheet that placed in bottom of die and finally rupture occurs. If there is no friction between sheet and die wall (theoretical case, $\mu=0$), the sheet deform more uniformly and thickness is constant along B-pillar. In reality, friction force at the interface always occurs and consequently non homogeneous deformations take place. The biggest wall thinning takes place at die bottom. Improving lubrication is main solution to obtain more uniform thickness distribution and decreasing stresses.

REFERENCES


BIOGRAPHY

JAVAD MARZBANRAD is an assistant professor in Automotive Engineering Department at Iran University of Science and Technology. He was born in 1964 in Tehran, Iran. He got his Ph.D’s Degree in 2001 from Tarbiat Modares University. His research interests are Automotive, Control and Vibration and Solid Mechanics.

MIR JAVAD PEYMAN was born in 1982 in Tehran, Iran. He got his Master’s Degree in 2007 from Iran University of Science and Technology, Automotive Eng. Dept. He works in brand and localization department of Iran Khodro Company (IKCO), the biggest sedan vehicle manufacturers in Iran. His research interests are Body and Structure, Engine, Power train and suspension systems.
A NEW APPROACH FOR MODELING OF SPOT WELDS FOR NUMERICAL ANALYSIS

Abolfazl Masoumi  Amir Najibi
Department of Mechanical Engineering  Automotive Engineering Department
Faculty of Engineering  Iran University of Science & Technology
University of Tehran  Tehran
Tehran 11155, Iran  Tehran, Iran
E-mail: amasomi@ut.ac.ir

KEYWORDS
Spot Weld, Finite Element, Damage, Material Property

ABSTRACT
The use of spot welds for the bonding of metal sheets is an extremely common practice in the automobile industry. The number of such bonds in a typical vehicle is on the order of several thousand. The use of connector elements to model spot welds in full-vehicle analyses leads to efficient finite element models that are able to capture the structural response of these local features with optimal computational effort. However, load-displacement curves required for the modeling of spot welds may not be readily available. Furthermore, the number of experimental tests required for the proper calibration of a complete set of spot weld pairs in a vehicle can be prohibitive since the mechanical response of these local mechanisms is dependent on both the geometric data, such as the thickness of the metal plates and the radius of the spot weld, and the material properties of the plates being welded. Virtual testing can generate the necessary modeling parameters when experimental data are not available. In this paper, a range of failure mechanisms typically observed in spot welds have been modeled numerically. The results from these virtual experiments are in acceptable agreement with laboratory-obtained data.

INTRODUCTION
The use of three-dimensional continuum elements and intricate material models (elastic-plastic and damage constitutive behavior) to reproduce experimentally observed load-displacement curves of tested spot weld specimens; and demonstrating how virtual experiments allow for the generation of load-displacement data of structural components from readily available geometric and material data have been studied. The load-displacement curves can be used subsequently in calibrating connector behavior for efficient use in large-scale models, such as full-vehicle analyses.

The material data used in the input files were obtained from the actual material data by subjecting it to a number of transformations that preserve the trends in the overall shape of the curves without revealing the exact material behavior. However, the results published in this paper use the actual data for comparison with the physical tests.

The correct modeling of the elastic-plastic deformation of spot welds is a prerequisite for realistic predictions of subsequent failure mechanisms (Werner et al. 2004). The authors showed that by taking into consideration material property changes in the weld nugget, one could obtain different failure modes involving peeling or shearing failure of the spot weld. Furthermore, they suggested the use of hardness measurements as a possible indicator for the change in properties of the welded material. The change in hardness between the center of a spot weld and at a distance far away from the center depends on the material grades joined, as well as their thicknesses.

The failure of aluminum-alloy sheets which occurs in thin-walled extrusions, results from one or more of the following mechanisms (Hooputra et al. 2004): nucleation, growth, and coalescence of voids, shear bands, and necking. Damage due to initiation, growth, and coalescence of voids leads to ductile failure in metals; the formation of cracks within shear bands leads to shear failure.

EXPERIMENTAL WORK

Geometry

The geometry of the patented test specimens used (Hahn et al. 1996 and Hahn and Rohde 2004) are shown in Figures 1 and 2. A single spot weld of radius 2.65 mm connects two steel plates that are 1.4 mm thick. These steel plates are 50 mm long and are bent over radii of 4.0 mm.

Materials

All specimens used in this study are made of galvanized high-strength steel H340LAD+Z100. This material behaves in an elastic-plastic manner during the initial loading stages. As the material is further loaded, it can either display a ductile damage response (caused by growth and coalescence of voids) or it can display a shear failure mechanism (caused by the formation and growth of cracks within shear bands). The material data used as the input files associated with this paper are approximative. The data were obtained by transforming the actual material test data (Werner et al. 2004) to preserve the overall trends in the behavior without revealing the true material behavior.
Boundary Conditions and Loading

In the physical tests the vertical sides of the specimens (Werner et al. 2004) were longer than those shown (Figure 1), and they were rigidly clamped in the clamping device (not shown) of the tensile testing machine. In the analysis these clamping conditions were modeled by fixing the bottom edges of the specimens and imposing a constant velocity of 0.15 m/s along the top edges in the vertical global 3-direction. For the physical tests and analysis, a hinge had been located 400 mm from the top edges of the specimens along the vertical global 3-direction allowing the fixed edges to rotate about the horizontal axis. During the pull and peeling analysis, the specimens were aligned so the top edges were initially orthogonal to the vertical direction, whereas during the shear test the edges were initially parallel to the vertical direction.

FINITE ELEMENT MODELING APPROACH AND SIMULATION TECHNIQUE

A total of 18 different simulations were performed corresponding to the three test cases (pull test, shear test, and peeling test). Each of the simulations was solved with a coarse mesh and a fine mesh using three different scaling factors for the thermal influence modeling, as shown in Table 1.

<table>
<thead>
<tr>
<th>Case 1</th>
<th>Pull test.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 2</td>
<td>Shear test.</td>
</tr>
<tr>
<td>Case 3</td>
<td>Peeling test.</td>
</tr>
</tbody>
</table>

Mesh Design

All simulations were performed with 8-node, linear brick, reduced integration elements (C3D8R). The density of the meshes increases toward the center of the plate where most of the deformation occurs. Each test was performed with a coarse mesh using four elements through the thickness of each plate and a fine mesh using six elements through the thickness of each plate. The coarse mesh and fine mesh used for the pull and shear tests are shown in Figures 3 and 4, respectively. The initial undeformed configuration of the coarse mesh used in the peeling test simulation is shown in Figure 5, while the initial configuration of the fine mesh is shown in Figure 6. The same meshes are used for the pull and shear simulations, since they are based on the same model geometry.
Elastic-Plastic Behavior

In this paper, both the elastic and the plastic behavior were assumed to be isotropic with the yield surface described by a Mises yield function. Different hardening curves were considered to encapsulate thermal effects near the spot weld. For simplicity, the specimen was partitioned into three zones corresponding to the different thermal exposures observed in the vicinity of the spot weld during the welding process. Different scaling factors for the stress-strain curves were used in the three zones as derived from hardness measurements. The geometry of each zone can be prescribed according to the welding process parameters. The scaling of the yield curve is accomplished with the use of a field variable defined as constant throughout each region, and three scaling magnitudes were considered for each simulation. The scaling factors given below are approximative, but they reflect the trends of the elastic-plastic behavior near the spot weld:

1. A baseline configuration, where the original material properties (no scaling) are assigned to the specimens in all three zones. While this choice is not realistic, it provides an extreme solution for comparison purposes.

2. A second configuration uses a scaling of 1.2 of the yield curve in Zone 1 and a scaling of 1.1 in Zone 2.

3. A third configuration uses a scaling of 1.4 of the yield curve in Zone 1 and a scaling of 1.2 in Zone 2.

These three scaling factor configurations demonstrates the effects of the thermal influence zone in capturing the correct behavior, as discussed below.

Damage Initiation and Evolution

Abaqus offers phenomenological damage initiation criteria for both of these mechanisms. The ductile criterion is specified by providing the equivalent plastic strain at the onset of ductile damage as a function of stress triaxiality and
strain rate. Similarly, the shear criterion is specified by providing the equivalent plastic strain at the onset of shear damage as a function of shear stress ratio and strain rate.

Damage evolution occurs once the damage initiation criteria are satisfied and further loading is applied. A plastic displacement based on linear damage evolution law was used for each damage initiation criterion. The value of the plastic displacement at which the damage variable reached 1.0 (complete degradation) was taken as 0.1, based on data from independent base material testing. The default maximum degradation rule was used, and the elements were removed from the mesh once the maximum degradation occurred. Damage initiation and evolution were assumed to be the same in all three thermal influence zones described above, a simplifying modeling assumption.

**Initial Conditions**

As discussed in the material model, a field variable was used to scale the yield surface and capture the thermal effects of the welding process on the yield strength.

**Boundary Conditions**

The specimens were loaded by fixing their bottom edges and imposing a constant velocity of 0.15 m/s along the top edges in the vertical global 3-direction.

**Constraints**

The top and bottom edges were constrained by kinematic couplings to model the rigid clamping in the testing devices. A TRANSLATOR connection was used to model the stiffness of the loading apparatus.

**Interactions**

General frictionless contact was defined between all surfaces, an appropriate approximation since friction forces were small when compared with the forces in the weld nugget.

**Analysis Steps**

All simulations consist of one explicit dynamic step. All analyses consider geometric nonlinearity and utilize mass scaling to model quasi-static loading conditions.

**Discussion of Results and Comparison of Cases**

The results derived from simulation work in this paper were compared with the actual material data which presented in (Werner et al. 2004).

The Mises stress contour and final deformed shapes of the pull, shear, and peeling simulation with the coarse mesh and base material (no scaling of the yield curve) are shown in Figures 7, 8, and 9. The final deformed shapes in pull and peeling derived from simulation and experimental tests are shown in Figures 10 and 11 which indicate good qualitative agreement between experimental and simulation results. Shear test results predict failure of the spot weld rather than

on the surrounding plate, regardless of the amount of scaling that was applied to the yield curve (Figure 12). This behavior was seen in some but not all experimental results (Werner et al. 2004), and it does not seem to influence the overall bearing capacity of the structure in shear.

![Image 7: Pull Test, Coarse Mesh, Baseline Material (no Yield Curve Scaling): Final Deformed Shape of Specimen and Mises Stress Contours.](image)

![Image 8: Peeling Test, Coarse Mesh, Baseline Material (no Yield Curve Scaling): Final Deformed Shape of Specimen and Mises Stress Contours.](image)

![Image 9: Shear Test, Fine Mesh, Baseline Material (no Yield Curve Scaling): Final Deformed Shape of Specimen and Mises Stress Contours.](image)

The load-displacement history obtained from the simulations were compared with the experimental results (Figures 10 to
12). A good match is observed for the pull and shear tests with the yield curve scaled by 1.2 in Zone 1. Results obtained with the baseline material underpredict the peak load capacity, and results obtained with the 1.4 scaling overpredict the peak load of the structures (Figures 10 and 12). The thermal zone scaling does not have a significant impact on the qualitative character of the loading curves. Good mesh convergence is also observed, indicating that acceptable results can be obtained even with the coarse mesh discretization.

![Figure 10: Pull Test with Fine Mesh: Comparison of Reaction Force Versus Imposed Displacement for Different Yield Surface Scaling.](image)

Results obtained for the peeling test do not show such good agreement with the experimental results even the baseline material simulation results overpredict the peak load capacity. This overstiff behavior is shown even in the purely elastic levels of deformation (in very early deformation stages), which indicates that modeling of the loading apparatus for this loading configuration may be inexact.

![Figure 11: Peeling Test with Fine Mesh: Comparison of Reaction Force versus Imposed Displacement for Different Yield Surface Scaling.](image)

![Figure 12: Shear Test with Fine Mesh: Comparison of Reaction Force versus Imposed Displacement for Different Yield Surface Scaling.](image)

**CONCLUSION**

In conclusion, the results from both the quasi-static spot weld connector simulations match the experimental pull and shear data very well. Changes in the material properties induced by the welding process are not essential to the capture of the peak loads during the spot weld failure.

**REFERENCES**


CUTTING SIMULATION SOFTWARE FOR MOLD INDUSTRY

Su-Jin Kim  
School of Mechanical and Aerospace Engineering  
Gyeongsang National University  
Gajwa-dong 900, Jinju, Gyeongnam, 660-701, South Korea  
E-mail: sujinkim@gnu.ac.kr, http://ma.gnu.ac.kr  
Tel: +82-55-751-6075, Fax: +82-55-762-0227

KEYWORDS  
Mold Production, NC Milling, Machining Simulation, Tool Path Optimization

ABSTRACT

The mold machining simulation software that will be introduced in this paper is developed to make NC machining more effective. This software includes many digital technologies related to plastic injection mold or press mold machining. It changes the cutting conditions automatically for every different working status. Many mold and press die companies are using the simulation software to get about 27% productivity increase. Not only machining time but also easier management of machining process is the effect of the simulation software.

1. INTRODUCTION

Mold industry market has been demanding higher quality and shortening manufacturing time. Those demands will be intensifying in the future. In these market circumstances, NC milling processes should be improved to innovational ways for satisfactory of the mold industry market and improvement of competitiveness. NC cutting of the mold manufacturing takes the main portion of time and money. But there are many difficulties about the NC machining. Machining time and finishing quality come out differently according to CAM data quality and machine operator's skill. Operating NC machine can not progress fluently by lacking of the cutting information. Therefore, it is hard to expect to reducing error rate, and it can't get to maximize productivity through ATC (Automatic Tool Change) unmanned operation. A cutting simulation software was developed by the help of NC machining engineers in mold & die companies to solve those difficulties. During simulation, it regenerates CAM data into optimized NC data in practical way. It makes a NC machine to perform higher productivity, and workshop floor can be change to be better environment by stress less machine operation.

2. CUTTING SIMULATION

The software read tool path generated by conventional CAM software and simulate NC machining. It controls and modifies cutting conditions of tool paths automatically to increase productivity during simulation. In this chapter the implemented algorithms of it is introduced.

2.1 Control Feed Rate Automatically

Cutting load is continuously changed along the work shape. Original tool path use normal feed rate would be F1000 mm/min. The simulation software controls the feed rates, from F800 to F4000 automatically. Faster and safer machining is possible.

2.2 Control Spindle Speed

Spindle speed should be low when cuts massive volume with outer blade and high when cuts slight volume with inner blade. But practically, the real cutting spindle should be set to the left side case. It changes spindle speed for the high efficient manufacturing with corresponding feed rates value.

2.3 Arc Fitting for Old Controller

There are huge amount of small line segments at corner of parts and these are problems for old-type controller. As all NC engineers know, there are many problems of cutting quality because of sudden feed decrease and hesitant milling motion at corners. It translates massive line segments into
simple arc paths for smooth ac/deceleration and fine finishing. The processing time reduced 30% and the quality increased in case of old-type controller.

![Old Controller (Decreased Feed, Machine Chattering)](image1)

![New Data (Smooth Movement without Chattering)](image2)

Fig. 3: Arc Fitting for Old Controller

2.4 Add Tool Path at Over Loaded Regions

If the machine makes sounds loud in cutting, feed rate should be reduced or stop the machine to change tool path pattern at CAM software. Some cases can not be solved with feed rate control and spindle speed control. Develop simulation system solve that with self-creating tool path in extremely loaded area. As it shows on figure, it adds tool paths at overloaded area. NC milling can be safe enough to run at night without an operator.

![Path added](image3)

Fig. 4: Add Tool Path at Over Loaded Regions

2.5 Delete Air Cut Tool Path

During the core machining, air cut paths reach 5%, up to 30%. In the figure, the air cut tool paths should be deleted. Only necessary tool path should be remained. Developed simulation system deletes unnecessary air cut paths. There are many other algorithms, such as output tool and holder setting length information, modifying tool paths for preventing the bottom collision of the cutter.

![Sawed stock, Original data, Aircut delete](image4)

Fig. 5: Delete Air Cut Tool Path

3. APPLICATION FOR A MOLD

The difference between general way and proposed way of the NC machining is explained in this chapter. A technical role is also explained.

3.1 A Tested Mold

A mold in figure is difficult to cut because of various size slots and bottom corners between slop faces and curved surfaces. An engineer using CAM software should deeply consider of tool patterns and cutting load when creating tool paths.

![A Tested Mold](image5)

Fig. 6: A Tested Mold

3.2 Shorten Tool Path Length

General mold cores have been cut off by pre-roughing process before NC milling, and it is good to make short machining time and prevents deformation of the stock model. In those cases, tool path of CAM software includes air cut paths a lot. Developed system simulates these tool paths and deletes itself to reduce milling time. Also programming task of CAM could be less stressful.

![Sawed stock, Original data, Aircut delete](image6)

Fig. 7: Delete Air Cut Tool Path
3.3 Protect Tool from Large Cutting Load

Tool breaking when the path met large volume happens frequently because previous step of NC data did not cut off. CAM software couldn’t know that kind of big volume area. NC operator should be careful all the time to prevent a tool breakage. This tool breaking problem will never happen with it, because it recognizes remained stock volume and creates additional tool paths.

![Fig. 8: Add Tool Path at Over Loaded Regions](image)

3.4 Protect Flat Tool from Bottom Collision

Pencil process or re-machining process are usually used to cut off the corner volume repeatedly after the roughing or semi-finishing. But this tool pattern is pencil with Ø16 0.4 tool tip-radius cutter. It is only possible to use for the practical cutting since developed system check up any possibilities to have problems and modifies every tool paths that it may happen any interruption. As all engineers know, flat cutter has no bottom blades. It is very dangerous to use the cutter at –Z direction tool paths. Generally the cutter is not using for the pencil tool pattern. At the picture, tool paths after the simulation have lamp lead-in paths which the system creates by recognizing of –Z direction paths. It prevents tool damages or over-cut. The pencil process is the fastest pattern for removing corner volume and developed system makes it possible, even practically safer.

![Fig. 9: Add Slope Flange Tool Path](image)

3.5 Adjust Rapid Height

Adjusting rapid height makes lower rapid path height to the shape. It makes G00 rapid motion time minimize for higher efficiency. Every rapid path will always have 5mm gap from the surface. Developed system adjusts the G00 rapid motion height and also it rectifies the G1 approach length for the safe machining.

![Fig. 11: Adjust Rapid Height](image)

3.6 Control Feed Rate

Cutting load is continuously changed along the work shape. The developed system commands the optimum feed rates value at every NC data blocks for the safe and the fast machining. The tool path color shows feed rate differences. Operator doesn’t need to control the feed override dial.

![Fig. 12: Control Feed Rate](image)

3.7 Compare to CAD Model

It detects differences between the CAD model and the simulated model. It outputs comparison result by color. The blue shows the under cut, and the red is the over cut area.
The blue area needs the additional electro discharge machining process.

Fig. 13: Compare to CAD Model

3.8 Comparative Results

Comprehensive view about this case, CAM programming time and task stress could be cut short and tool consumption and milling time could be also saved. All functions, such as adding tool paths at overloaded regions, air cut path delete, feed control, are working at once during that developed software simulates every NC data with a stock model. NC data from it could be operated safe and fast. It will be the best way to achieve the productivity with the higher efficiency.

Table 1: Comparative Results

<table>
<thead>
<tr>
<th></th>
<th>Conventional style</th>
<th>Developed system</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAM</td>
<td>Lots of NC data files are needed</td>
<td>Simple tool path style</td>
</tr>
<tr>
<td></td>
<td>Very complex tool path style with tool selection</td>
<td>Less than a half time for CAM tasks</td>
</tr>
<tr>
<td></td>
<td>Stressful tasks of CAM</td>
<td></td>
</tr>
<tr>
<td>Tool</td>
<td>Use more than 10 kinds of tool used</td>
<td>Less than 6 kinds of tool used</td>
</tr>
<tr>
<td>Milling time</td>
<td>About 20 hours</td>
<td>About 10 hours (50%)</td>
</tr>
<tr>
<td></td>
<td>Frequent tool break</td>
<td>Full ATC operation possible</td>
</tr>
<tr>
<td></td>
<td>ATC operation disable</td>
<td></td>
</tr>
</tbody>
</table>

4. CONCLUSIONS

Developed system upgrades NC machining efficiency practically. Almost all of technologies and techniques about NC milling operation for the mold and press die making are converted into the digital database of it. And it will change the cutting conditions for every different working status automatically. NC machining process can be finished faster and safer by it. It will be helpful to NC operating engineers as a secretary.

Machining time, tool consumption, easier management of manufacturing process, and uncountable benefits including reducing the task stress will make mold makers get positive effect. The result simulation software will give mold & die manufacturers the alternative competitiveness. It will be helpful for mold & die industry.

REFERENCES


S.J. Kim, 2000, “A study on the feed optimization and NURBS tool path for CNC high speed machining”, *KAIST*, M.S. Thesis


BIOGRAPHY

SU-JIN KIM is an assistant professor of mechanical engineering at Gyeongsang National University in South Korea. He obtained a BS degree at Seoul National University and MS and PhD degrees at KAIST in 1998, 2000 and 2005 respectively. He was a postdoctoral candidate at MIT in 2005. His research interests include Computer-Aided Manufacturing and NC Machining Simulation.

ACKNOWLEDGEMENT

The software tool introduced in this paper is developed through the industrial research process between the author and NCB Co. Ltd (http://www.ncbrain.com). Thanks to president Du-Jin Kim who supported the research and allowed the figures and texts of the company to be used for ISC’2008.
SIMULATION AND PERFORMANCE IMPROVEMENT PROCESS AT FORD POWERTRAIN OPERATIONS

HABTOM MEBRAHTU
Anglia Ruskin University
Chelmsford Campus, Chelmsford
Essex CM1 7QG, UK
Email: h.mebrahtu@anglia.ac.uk

JOHN LADBROOK
Ford Motor Company,
Dunton Engineering Centre, Basildon,
Essex SS15 6EE, UK.
Email: jladbroo@ford.com

ABSTRACT

This paper describes the use of simulation in supporting system design and in improving manufacturing performance at the Ford powertrain operations, Manufacturing Engineering division. Simulation modelling of engine assembly lines is carried out using the Ford Assembly Simulation Tool (FAST) which is an Excel based front-end user interface to the Witness simulation system. FAST is a component-based simulation system that enables engineers and non-simulation experts to develop engine assembly models easily and quickly. The model building is also enhanced by the use of specialised sub-models (Witness modules). Assembly line models are automatically built using a Witness shell that imports data from Excel data sheets, produces Witness elements and populates the elements with relevant data. The importance of the quality of input data and their sources has been emphasised. The models are generally verified using expected cyclic throughput values, and validated with real manufacturing data from the shop floor. Validated models are used to assess and enhance performances of engine assembly lines.

Key words: Simulation, Performance Evaluation, Component-based Simulation, Model Verification & Validation.

1 INTRODUCTION

The use of simulation modelling to support the machining and assembly lines in the automotive industry especially in defining strategies, production line design, analysing and improving performance of existing lines, and in experimenting with different operational options (scenarios) has been well documented. The importance of simulation can be underlined by the fact that most of the automotive manufacturers world-wide and, in particular, the three big companies (Ford Motor Company, General Motors, and Chrysler Corporation) required all new and modified manufacturing system designs to be verified by simulation analysis before they are approved for final equipment purchase (Ulgen and Upendram, 1997). The case still holds in most automotive industries today.

The design and operation of manufacturing systems in automotive industry can be classified into conceptual system design phase, detailed design phase, launching phase, and fully operational phase. The use of simulation particularly in the detail system design and the launching phase of engine assembly lines at PowerTrain Operations Manufacturing Engineering (PTO ME) of Ford Motor Company is the centre of this paper.

This paper introduces the use of simulation at the PTO ME within one of the widely known automotive industries, Ford Motor Company. It briefly describes typical problems that call for the use of simulation and then explains how the relevant data are collected. The core of the paper, how the integrated simulation tool Ford Assembly Simulation Tool (FAST) works, its benefits and limitations are then explained. Finally, outputs from engine assembly simulation runs are discussed and conclusions drawn.

1.1 Background

Ford Motor Company’s PowerTrain Operations, Manufacturing Engineering division has benefited from using simulation for more than twenty years. The benefits range from improved productivity to significant cost savings and more efficient asset utilisation (Pocock, 2004).

PTO ME is the Ford’s biggest user of simulation with more than 120 users world wide. PTO ME has a series of generic modelling tools for ongoing process improvement across its manufacturing operations. For its engine assembly modelling, PTO uses a component-based simulation system whereby predefined modules are used to build assembly line models, and simulations are created at run time to meet specific requirements (Winnel and Ladbroke, 2004). The core simulation engine is Witness - a discrete event system developed by Lanner Group (Lanner Group, 2007). The model building process is further enhanced by a spreadsheet based front-end interface that is familiar to engineers and allows quick building of model layouts even by non-specialist Witness users. The integrated simulation system is known as Ford Assembly Simulation Tool (FAST).

1.2 Powetrain Operations ME

The powertrain operations (PTO) deals with the engine and transmission elements of an automobile. The engine and the transmission are the essential subassemblies involved in generating and transmitting power, thus the name powetrain. The operations involved in machining and assembling of components to make the engine and the transmission of a vehicle thus fall into this category.
The powertrain manufacturing plants of Ford Europe include Dagenham and Bridgend in the UK, Cologne in Germany, Valencia in Spain, Chihuahua in Mexico, Taubate in Brazil and Strandale in South Africa. The engineering work of the engines made in these plants is controlled from the Ford Research and Development Centre, Dunton, UK. This paper concentrates on the engine assembly side of PTO ME.

2 THE ENGINE ASSEMBLY LINES AND PERFORMANCE MEASURES

Automobile engines are made up of the major subassemblies: Cylinder blocks, Cylinder heads, Crankshafts, Camshafts and a Connecting rods. These are commonly known as the 5 C’s. Some of these major subassemblies are machined/assembled in their respective production sections and delivered to the main engine assembly line.

The main engine assembly line is made up of a series of stations in a loop interlinked with accumulating conveyors between the loading and unloading stations. The base element is the cylinder block and all other subassemblies and parts are assembled to the block at different stations within the line. The engine block is initially loaded to a carrier (or pallet), known as a platen, at the loading station. Different assembly operations are carried out on the stations, in the loop. Once the engine is fully assembled, it is unloaded from the platen at the unload station and then loaded by a new block at the loading station which is normally adjacent to the unload station. The platens continue circulating and are always kept in the loop.

Most of the stations are on line and carry out operations without moving the platen off the conveyor. Few stations are offline and mainly work on subassemblies of the engine such as putting RTV onto the cylinder head.

The main performance measures in an engine assembly line are the average number of Jobs completed Per Hour, JPH, and the availability of the assembly line known as Uptime percentage. As the name implies, the JPH is the throughput of completed engine assemblies measured per hour. And the uptime is the usefully utilised percentage of resources (mainly stations) in the whole line. The cycle time of each operation in the line is determined based on the volume of engines required and the uptime.

The performance measures are mainly influenced by the number of engines rejected (measured by First Time Through, FTT), equipment breakdowns and human efficiencies (measured by overcycles).

3 DATA COLLECTION

In any simulation process the quality of data obtained is of paramount importance. In the Ford PTO ME simulation process, data for model building are mainly gathered from the assembly line layout plan (layout drawing), the work standard, and the Ford real-time computer monitoring system.

The layout gives a scaled drawing of the assembly plant layout and includes information such as loading/unloading area, operation numbers, relative sizes of stations, repair areas, team sections and, in some cases, line cycle times. The model is to be laid out in a similar proportion on the Witness window screen.

The work standard, generally provided as a spreadsheet, is the main source of most of the data needed at the model construction stage. The data obtained from the work standard include the names and types of engines to be assembled, the cycle times of the line and individual stations, details of operations carried out in each station, whether the station is automatic/semi-automatic or manual, efficiency of each station, the number of operators in each station, and auxiliary elements such as prestops, buffers and section/zone names.

The Ford real time monitoring system is useful especially in providing historical data of a given assembly line. Data from the shop floor are continuously logged into this system and stored. The data stored include historical data on equipment faults, overcycles of operators and the FTT. The data on equipment faults would be extracted from the real-time monitoring system and processed to give machine breakdown values. Similarly, the overcycles and the FTT would be used for generating the operator efficiency and the quality values respectively. Another important use of real-time monitoring system is in generating surrogate (comparative) data for a new line from a similar or close fit line or equipment. The surrogate data would be entered into a design-stage simulation model and the results would be used to estimate performance, justify proposed design or to decide on alternative options.

4 THE FORD ASSEMBLY SIMULATION TOOL (FAST)

FAST, as its name implies, is the simulation tool used to model engine assembly plants by the Ford PowerTrain Operations Manufacturing Engineering (PTO ME). It is an Excel interface to Witness created by Ford PTO ME to facilitate the creation of models of assembly lines. FAST consists of a set of Excel worksheets whose cells are filled by the user with the data needed for the simulation model such as operation data, buffer data, cycle times, machine breakdowns and human overcycles data, quality related data, batch sizes, quantity and size of platens, repair and teardown data, and layout co-ordinate data.

FAST is designed to make model construction of assembly lines easier, with an interface that is readily understood by end users. The FAST Master page and the most important worksheet, the LayoutFile, are shown in Figure 4a. Usually LayoutFile1 has details of the main assembly line and subassembly lines are detailed in LayoutFile2, LayoutFile3, etc. Based on the dimensions and sequencing of elements in the Layout files, the model is constructed in Witness showing the same shape and proportion of layout as the real manufacturing line. Most of the production parameters and reliability data are also put into the Layout files.
FAST also contains construct Witness modules designed for building the assembly line model elements into Witness. Reusability of these modules is one of the major advantages in the ease and speed of model building. The entire line model is automatically created by inputting data from the FAST directory into a Witness model shell at run time. The whole process of model building is controlled by Visual Basic and Witness command language.
FAST is a component-based simulation system where pre-constructed witness modules (sub-models) are reused. These modules have codes that, along with the Witness shell, build elements, name them, perform relative positioning, define input/output rules (including the direction of flow), and populate the elements with variables and data. Most of the variables’ values are called in and assigned from the FAST data sheets using the coding in the Witness shell at run time. A typical Witness model that was built using FAST is shown in Figure 4b.

Figure 4b. Example of a Witness Model Built Using FAST
The whole model building process using FAST can be depicted as shown in Figure 4c.

![Diagram of Model Building Process Using FAST](image)

Figure 4c. Model Building Process Using FAST

5 MODELS VERIFICATION/VALIDATION

The issue of verification, validation and testing throughout the life cycle of a simulation project is an essential element of all credible modelling and simulation processes (Robinson 2004, Law and Kelton 2000, Balci 1994).

Witness models that are built using FAST are generally accurately represented provided the input data from Excel are accurate; the conversion is done automatically with the help of verified modules. Any building error can also be visually identified from the way the elements are constructed in Witness. The main verification process, therefore, is the running of the Witness model for a specified period of time (plus the warming up time) and checking that the simulation throughput is consistent with the cyclic throughput. This is done by first disabling all causes of losses including breakdowns, overcycles and FTT prior to running. The simulation is then run and the output in JPH compared with the cyclic JPH which is calculated as 60 divided by the longest cycle time (in minutes) of the line. If the two values are the same, then it can safely be said that there are no logical errors in the model.

Once the model has been verified, the model is reset to include all the relevant data (including the losses) and rerun. The simulation output is then compared with the real data from the manufacturing floor. The acceptable variation to validate the model is generally 1 JPH or 5% on the average JPH.

6 SIMULATION RESULTS AND ANALYSIS

Simulation models are generally run to an equivalent of 30 day manufacturing which is considered long enough to represent realistic results. Simulation outputs are mostly generated using witness command lines (WCL) in the experiment section of Witness. This command lines allow different scenarios to be included such as the potential outputs of zones, the effects of breakdowns on the output, the effects of both breakdowns & overcycles on the output, etc., and most importantly the overall throughput of the assembly line. Zones are control segments of the assembly line and usually have separate team leaders. Having the overall report and individual zone reports helps in identifying the bottleneck zone, understanding the degree of interference between the zones and their effect on the overall JPH.

The findings are presented in a diagram showing the distribution of daily average JPHs for the zones, the overall simulation and the actual shop floor JPH (see Figure 6a). Once the bottleneck zone has been identified, the individual operations in the zone are then analysed to identify the bottleneck operations. This is usually supported by an analytical method that calculates the potential output of each operation using the cycle time, the breakdown, the overcycle and the FTT values. Two typical charts showing the potential outputs at operation level are shown in Figure 6a.
The simulation outputs are discussed with manufacturing personnel and experimenting decisions set out. The fundamental consideration is how the causes of the bottleneck could be alleviated. As a general example, if operation A is the bottleneck, then what would the JPH be if the breakdown/overcycles/FTT values are reduced by x percent and/or buffers are added and/or cycle times are reduced at the bottleneck? If the bottleneck has shifted into another operation, similar steps would be followed until the required JPH is attained.

7 LIMITATIONS AND POTENTIAL IMPROVEMENTS

One of the main challenges (although rare) with using FAST is the debugging process. The user has to be familiar with the error messages to understand where the causes of the problem could be. The complexity and comprehensiveness of FAST require considerable time to gain familiarity with all the variables, functions and associated codes of the model. Therefore, the personnel with expertise and experience to solve this type of problems are small in number.

Another challenge (external to FAST) is the process of analysing simulation reports, making sense out of them and then acting on potential problems. FAST is commendable in terms of ease of building Witness models but does not have the support at the rear end. On top of the difficulty understandably caused on non-expert users, the time spent by simulation experts in processing the simulation reports is significant. Therefore, an introduction of a mechanism that can support in analysing simulation reports and assessing performance could ease the burden at the rear-end and save significant amount of expert’s time.

Some commercial methods such as Optimiser and OptQuest, as applied to manufacturing, have proven to be good search engines and may produce satisfactory results. But their use mainly depends on systematic search from a pool of possible models that takes a long time especially on big models such as that of Ford assembly lines (Mebrahtu, et al 2004). Neither do they help in identifying problem areas such as the bottleneck. A mechanism that analyses and assesses simulation report in a manner similar to that of an expert is what would be helpful.

8 CONCLUSIONS

Simulation has been a rewarding tool to Ford powertrain operations, manufacturing division, especially in the system design and launching phases. The component-based nature of their specialised simulation system (FAST) has made building of accurate assembly line models very easy even to end users with limited simulation expertise. This has helped engineers to concentrate more on the analysis and performance evaluation of simulation results, and in
examining ways of improving performance. It is hardly surprising therefore that simulation is used within Ford powertrain operations to take the risk out of business change.

This paper discussed the use of simulation at Ford PTO ME and included the description of the sources of quality data, elements of FAST, the main performance measures, the verification and validation processes, the analysis and presentation of results, and the assessment and performance enhancing procedures. As complete as the simulation process is, it also has areas where improvement could be made such as in the debugging process and introducing an automatic supporting mechanism at the rear-end of the simulator.

On the whole, FAST has proven to be an effective assembly line simulation tool. This may have been facilitated by the focussed nature of the system to engine assembly domain.

9 REFERENCES


Lanner Group, 2007, Witness 2007 online, Lanner Group, UK


Robinson S., 2004, Simulation: The Practice of Model Development and use, John Wiley & Sons Ltd.


AUTHOR BIOGRAPHIES

HABTOM MEBRAHTU is a Senior Lecturer and Pathway Leader of Engineering at the Anglia Ruskin University, UK. He received his PhD degree in Manufacturing Simulation and Optimisation from the University of Bath, Mechanical Engineering Department. He is a Chartered Engineer, Fellow member of the HEA and a member of the IET. He has recently worked (as sabbatical secondment) with the simulation team of the Ford powertrain operations, Manufacturing Engineering division at Dunton, UK. His areas of interest include performance optimisation with simulation, and web-based simulation.

JOHN LADBROOK has worked for the Ford Motor Company since 1968 where his current position is European Simulation Technical Specialist. In 1998 after 4 years research into modelling breakdowns he gained an M.Phil (Eng) from the University of Birmingham where he has lectured on a part-time basis since 1996. He has used and promoted the use of discrete event simulation since 1982. In his role he has been responsible for sponsoring many projects with various universities. This resulted in an appointment as Fellow with Cranfield University in 2001. He chaired the WITNESS Automotive Special Interest Group, and is in the Editorial Board of the Journal of Simulation (a publication of the OR society).
Hybrid Powertrain Simulation

Valerian CROTORESCU
Andrei MACIAC
Mircea OPREAN
Alexandru MARIN
Cristian ANDREESCU

University POLITEHNICA of Bucharest
AUTO Integrating Platform, Automotive Engineering Department
313 Splaiul Independentei st.
6th Sector, PC 060042,
Bucharest, ROMANIA

E-mail: valerian.crotorescu@gmail.com, andreii302@gmail.com, mircea_oprean@yahoo.com, marin@hydropub.ro, andreeescu.cristian@gmail.com

KEYWORDS
Transportation, Dynamic modelling, Model design and testing

ABSTRACT

The EURO 5 requirements (in the 1st decade of 2009) and EURO 6 norms (maybe in 2012 or later) in Europe, US and Japan requirements to reduce the harmful gases and CO₂ vehicle emissions increased the interest of vehicle manufacturers in developing new technologies for automotive propulsion systems, amongst which the hybrid propulsion systems is one of the most important alternative.

The hereby paper aims to create models and simulate the operation of several hybrid architectures, trying to find out the strengths and weaknesses of each model.

The following types of hybrid powertrain solutions have been analyzed: Serial hybrid with clutch between two electric motors, Combined hybrid with clutch between two electric motors, Parallel ‘through the road’ hybrid (TTR), Parallel hybrid with motor-generator ‘Between two clutches’ as well as a mixt hybrid powertrain with ‘power splitting device’ (Wolfgang Reik et al 2006). Some standard components and common parameters which remain the same for all hybrid solutions have also been considered.

SIMULATION OF HYBRID POWERTRAIN SOLUTIONS

Serial hybrid with clutch between two electric motors

In the case of the serial hybrid (Figure 1), the powertrain can be completely separated in mechanical terms and the power can be transmitted by purely electric means (Oprean 2003).

![Figure 1: Serial Hybrid with clutch between two electric motors](image)

The electric motors are arranged in series. They act as a power generator giving power and also as a converter. An automatic transmission or a continuously variable transmission (CVT) could equally be used. The continuously variable transmission has the advantage of using only the electrical motors at appropriate dimensions for the reverse gear (Figure 1).

The clutch placed between the two electrical motors was controlled using a function. It was engaged and then disengaged when the car velocity reached a previously defined value; this value was introduced in the expression of the function. In order to avoid the slip of the clutch, it was necessary to slow the command of the function using an order lag, where the time constant and the gain had to be introduced.

![Figure 2: The function and the lag](image)

In the simulated model a continuously variable transmission and a special controller were used as shown in the picture below (Figure 6). The CVT-H controller is using the load of the engine and the velocity of the vehicle for a good transmission control.

![Figure 3: Series Hybrid vehicle with clutch between two electric motors using CVT](image)

The continuously variable transmission used in this model was controlled by a ‘smart controller’ depending on the engine load and on the velocity of the vehicle.
There were taken into account the flywheel and the pulleys moments of inertia. The clutch was included in the CVT.

The engine started 138 seconds after the beginning of the test and stopped after the 157th second, then started again after second 333 and ran for 20 seconds, and so on until the cycle finished. In Figure 5 the fuel consumption curve is constant, while the engine was working, and the increased when the engine was working. To be noted that the total fuel consumption at the end of the cycle was 383 g while the battery state of charge was 84%.

Figure 6 shows the position of the working points during the test in the engine torque characteristic. It can be observed that the engine was working mainly at 1200 rev/min, at different loads from idle to maximum, due to the functional characteristic of the CVT.

**Combined hybrid with clutch between two electric motors**

The following description presents the model of a combined hybrid equipped with a special ECU (Figure 7), including a parallel ECU and a series ECU, working separately depending on the velocity of the car. The command of the new ECU was made by a switch. The method was the same as the one previously used in the case of the series hybrid with clutch between electric motors; however, instead of the CVT, the system was equipped with an automatic gearbox.

At the end of the cycle, the fuel consumption was lower than 380g and the battery state of charge was almost 65% (Figure 8).

**Parallel ’through the road’ hybrid (TTR)**

In this solution, the internal combustion engine and the electric motor are placed on different axles (Wolfgang Reik et al 2006). For the front wheel drive the transverse powertrain arrangement remains unchanged. In addition, the rear axle is driven by an electric motor, as shown in Figure 9.
The model used for simulation in this case was shown in Figure 10. The continuously variable transmission was controlled by a smart controller depending on the engine speed and the velocity of the vehicle. This controller was able to change the CVT transmission ratio between the limits 0.4 \( \ldots \) 2.5. The results of simulation for the NEDC are shown in the following charts.

Figure 10: ‘Through the road’ hybrid solution

As in the models shown before, the CVT was controlled by the same controller especially designed to use this kind of transmission without any architectural or mechanical problems. Using the CVT has helped the internal combustion engine to function close to the economic pole. The velocity of the car is taken directly from the wheel with a velocity sensor.

It can be noticed that during 813 seconds the vehicle’s internal combustion engine did not function; after that moment the battery’s state of charge reached a value lower than 60%. Afterwards, the engine started and the battery was charging until the time reached second 970; afterwards the engine started again and continued to function until the 1126\(^{th}\) second, close to the end of the cycle. Again, to be noted that the engine function points observed the usual order and orientation, being close to the economic pole.

Figure 11: Total fuel consumption [g] / Battery state of charge [%] / Vehicle linear velocity [m/s] during the NEDC (1180s) – TTR

Figure 12: The position of the working points during the test – Parallel ‘through the road’ hybrid

‘Between two clutches’ hybrid architecture

This architecture is based on the idea that the internal combustion engine can be shut down and decoupled from the transmission, resulting less friction losses during its cycle. The clutch (usually a dry clutch) is not used as a starting clutch.

Figure 13: Powertrain with a starter between two clutches

Figure 14: The model using automatic gearbox
Figure 15: The model using continuously variable transmission

This kind of architecture includes an automatic transmission or a continuously variable transmission, as shown in the pictures above. The automatic gearbox is controlled by a special ECU, designed for automatic gearboxes, achieving a simple dynamic modelling of an ‘n-ratio’ (Figure 16). The continuously variable transmission is controlled by a smart controller depending on the load of the engine and on the velocity of the vehicle (Figure 17).

Figure 16: An ‘n-ratio’ automatic gearbox with integrated torque converter and the ‘ECU’

Figure 17: The continuously variable transmission and the special controller

For the automatic gearbox, the torque converter, its moments of inertia (impeller and turbine) and the powered axle were taken into account. Parallel with the torque converter a lockup clutch was added, bypassing the torque converter when the rotary velocities of the turbine and impeller were close to each other. For the continuously variable transmission, the flywheel and the pulleys’ moments of inertia were taken into account. A clutch was included in the CVT.

Figure 18: Variation of engine speed and rotary velocity of the electrical motors during NEDC (1180s) – AT

It can be observed that in both cases the vehicle had only used its electrical motor for 831 seconds, and only afterwards the engine had started. The engine speed was higher in the case of the automatic gearbox (~ 4200 rev/min) compared to the CVT (~2700 rev/min).

Power Splitter hybrid architecture

This device, usually referred to as the ‘PSD’, was the core of the full hybrid system in Prius. Using this device, the gasoline engine and two electric motors were connected. Due to the fact that all of the components are permanently engaged, the power is transferred similarly to a common differential, rather than a traditional automatic transmission providing remarkably smooth operation and rapid responsiveness.

When the ‘PSD’ operates, it splits and/or combines power to achieve the greatest efficiency. The large electric motor sometimes provides and sometimes captures power, following that it sometimes just spins without any electricity flowing in either direction.

Figure 19: Variation of engine speed and rotary velocity of the electrical motors during NEDC (1180s) – CVT

Figure 20: Model simulations of ‘Power Splitter’ Hybrid architecture
It could be noticed that the engine started before second 150, the consumption at the end of the cycle reached 358.7 g and the battery state of charge got under 60% (55.25%).

CONCLUSIONS

Using AMESim it is possible to simulate different hybrid powertrain solutions in order to obtain data about fuel consumption and other parameters, and compare these solutions even if they were not yet physically achieved. The next table shows the fuel consumption in NEDC cycle, for every hybrid architecture simulated. Although the instantaneous velocities were kept inside the tolerances of NEDC, the total displacement differs from one model to another due to functional characteristics specific to each model.

The lowest fuel consumption model was the 'Parallel through the road, using CVT', because, in this case, the battery state of charge was kept above the lowest accepted limit of 60%. Therefore, this model seems to be the most economic in the NEDC conditions, but its dynamic performances outside the NEDC have to be further studied in the future.

Another interesting architecture is 'Between two clutches, using CVT', but the fuel consumption was 7.5% higher than in the previous case; the eventual higher dynamic performances that may result from this higher consumption will be verified.

In the case of the architecture 'Clutch between two electric motors' the battery state of charge at the end of the cycle was the highest, but the fuel consumption was also the highest.

The hybrid architecture 'PSD' did not meet the lowest acceptable threshold of 60% for the battery state of charge that ensures a long-lasting battery life.

Table 1: The most important parameters resulted after the simulation

<table>
<thead>
<tr>
<th>Hybrid architecture</th>
<th>N.E.D.C.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total displacement [km]</td>
</tr>
<tr>
<td></td>
<td>Total fuel consumption [g]</td>
</tr>
<tr>
<td></td>
<td>Battery state of charge [%]</td>
</tr>
<tr>
<td></td>
<td>Mixt fuel consumption [l/100 km]</td>
</tr>
<tr>
<td>Between two clutches, using CVT</td>
<td>11037.2</td>
</tr>
<tr>
<td>Between two clutches, using AT</td>
<td>1106.6</td>
</tr>
<tr>
<td>Combined hybrid</td>
<td>11086.6</td>
</tr>
<tr>
<td>Clutch between two electric motors</td>
<td>11009.9</td>
</tr>
<tr>
<td>Through the road, using CVT</td>
<td>11047.9</td>
</tr>
<tr>
<td>Power splitter device</td>
<td>11061.1</td>
</tr>
</tbody>
</table>

AKNOWLEDGEMENT

The present paper has been conducted within the framework of the 'IDEI' contract, ID_1091/2007, financed by the Romanian Government, two of the authors improving their skills in AMESim Software at LMS in Lyon. The AUTO Integrating Platform of laboratories from the University POLITEHNICA of Bucharest has supported and enabled the simulation, by purchasing the AMESim software.

REFERENCES

SIMULATION IN ROBOTICS
COMPONENTS
SIMULATION-
NAVIGATION
SYSTEMS
AND
MOVEMENT
A SIMULATION-BASED METHOD FOR FAULT DIAGNOSIS IN HYDRAULIC COMPONENTS

C. Angeli
Department of Mathematics and Computer Science
Technological Institute of Piraeus
Konstantinoupoloios 38, N. Smirni
GR-171 21 Athens, Greece
E-mail: c_angeli@otenet.gr

A. Chatzinikolaou
S. Patsi 62
GR-118 55 Athens, Greece
E-mail: achatzi@otenet.gr

KEYWORDS
Modelling, fault diagnosis, hydraulic systems, on line process

ABSTRACT
Fault detection methods for drive and control systems are very important for the modern production technology because they improve the reliability of automated systems. In this work, simulation results and data acquired from an actual system are compared to obtain the information needed for the detection of possible faults in proportional valves. The diagnostic system is based on the inconsistency between the actual process behaviour and its expected behaviour as described by an analytical model. The efficiency and accuracy of the simulation based methodology has been verified by experimental results for a proportional valve in a hydraulic system.

INTRODUCTION
Electro-hydraulic drives and systems are extensively used in applications of the automation technology from robotics and aerospace to heavy industrial systems. The performance of the hydraulic components affects the reliability of the total system. The operation of proportional valves is related to the main function of the hydraulic system and the development of diagnostic procedures for these components is very significant for the operation and maintenance of the hydraulic driven machines.

Fault detection and diagnosis have been considerably interesting in recent years because they are a crucial issue in modern automation processes. Modelling information involved in a diagnostic method is considered as a quite effective diagnostic technique and various kinds of approaches have been published for automated industrial processes over the last years such as (Frank 1996; Gertler 1998; Patton et al. 2000; Kinnaert 2003; Korbicz et al. 2004; Isermann 2005). Model-based fault detection scheme is based on the evaluation of the deviation between measured and estimated values or outputs and the nominal ones leading to residuals or analytical symptoms. The relation between symptoms and faults can be exploited in a diagnosis part, leading to information about fault sources.

In this paper an approach for model based fault detection and diagnosis is presented and applied to an electro-hydraulic valve. Real-time measured data from the actual process are compared with the relevant values of the simulation process. The data acquisition software permits the qualification of the numerical measured data and as consequence makes possible the on-line interaction with the symbolic information stored in the knowledge base of the expert system that permits fault diagnosis based on several sources of information. Previous research work on fault diagnosis methods for these systems has been published, among other researchers by (Angeli and Atherton 2001; Angeli and Chatzinikolaou 1999; Chen and Lu 1995; Feuser 2004; Gao et al. 2003; Pung et al. 2006).

DESCRIPTION OF THE ACTUAL SYSTEM
The actual system that was used as a basis for this work was a hydraulic system where a proportional valve (1.3) controls the actual flow rate to hydraulic motor (2.1) and determines the angular velocity, the acceleration and the deceleration of a rotated mass according to the input voltage to the amplifier of the valve by a constant Δp. Figure 1 presents schematically the hydraulic system.

The proportional valve used for the experimental part of this work, is a directly operated directional proportional valve with electrical feedback. This valve consists basically of a housing, a spool, two return springs, two proportional solenoids A and B and an inductive position transducer. The proportional solenoids convert an electrical signal into a proportional movement of the spool in conjunction with an inductive position transducer. The position of the spool in either direction is determined from this electrical input signal. This input signal is a voltage $U = 0$ to $\pm10$ Volt to an electrical amplifier that is connected to the proportional valve. The amplifier converts this voltage signal to a current signal $I_3$ to the relevant solenoid A or B. The force generated through this current moves the main spool against a spring. The springs centre the spool in a zero flow position when the input signal is zero.
MODELLING PROCESS

The proportional valve is modelled using the operating curves of the technical specifications data sheets given from the manufacturer. A relation that can be used for the mathematical description of the proportional 4-way valve in a steady state can be derived from the operating curves. The operating curves represent the relation between flow rate through the valve and input current to the solenoids for various pressure drop values $\Delta p$. Figure 2 shows these operating curves.

Figure 2: Performance curves of the proportional 4-way valve for various pressure drop values $\Delta p$. 
We have the following relationships for the flow through the proportional valve:

The flow through the valve \( q_{val} \) for a \( \Delta p \) value is equal to the flow \( q_{val} \) for \( \Delta p = 10 \) bar multiplied by \( \sqrt{(\Delta p/10)} \). This relationship is valid for every \( U \) value. For \( U = U_{val} = 10 \) V and \( \Delta p = 10 \) bar is \( q_{val} = Q_{val} = 10.8 \) l/min. For another set of \( U \) and \( \Delta p \) values the corresponding \( q_{val} \) value is equal to \( (Q_{val} / U_{val}) \cdot U \cdot \sqrt{(\Delta p/10)} \).

In consequence, we have the following relationships for the flow through the proportional valve:

\[
q_{val} = \frac{Q_{val}}{U_{val}} \cdot \frac{\Delta p}{10} \quad \Delta p_a = p_a - p_a
\]

\[
q_{val} = \frac{Q_{val}}{U_{val}} \cdot \frac{\Delta p_b}{10} \quad \Delta p_b = p_b - p_{tank} = p_b
\]

where \( Q_{val} \) is the nominal flow of the proportional 4-way valve and \( U_{val} \) is the nominal voltage to the amplifier of the proportional 4-way valve.

For those components which contain large quantities of oil as pipes and hydraulic actuators it may be assumed that the change in pressure in the dynamic state is proportional to the net inflow of oil, that is:

\[
dp / dt = (E / V_o) \cdot \Sigma Q
\]

where

\( E \) is the elasticity module of the oil plus the included air,

\( V_o \) is the Volume of the pipe plus a part of the volume of the attached actuator,

\( \Sigma Q = Q_{in} - Q_{out} \)

\( Q_{in} \) is the incoming flow to the volume \( V_o \) of a connecting pipe and

\( Q_{out} \) is the outgoing flow from \( V_o \).

That means that the pressure increase in a pipe element at a junction is proportional to the algebraic sum of the incoming and outgoing flows and inverse proportional to the included oil volume.

The pump flow to the hydraulic motor is given by:

\[
Q = V \cdot n \cdot \eta \quad \text{where}
\]

\( V \) is the displacement (flow per revolution) of the pump

\( n \) is the number of the revolutions per minute of the pump shaft and

\( \eta \) is the volumetric efficiency of the pump

The derived equations were used for the development of the mathematical model of the proportional valve and together with the equations derived from the technical characteristics of the other hydraulic elements build the mathematical model of the total system. The description of the dynamic behaviour takes into account the non-linear character of hydraulic systems and the incompressibility of the hydraulic fluid in the pipes as well as the special characteristics of the hydraulic elements used. In consequence the model represents the behaviour of the system elements more accurately.

**FAULT DIAGNOSIS PROCESS**

The fault detection process is based on the model of the normal behaviour. Faults are diagnosed by determining changes in variables that have significant contribution to the diagnostic process and by relating these variables to specific faults. Final fault diagnosis is based on variable changes and related experiential knowledge suitably formatted.

The data acquisition system runs in parallel to the actual system and acquires data for the angular velocity and the pressures while the speed is changing. For the measurements process a generator module controls the proportional valve with a signal of a definable frequency and amplitude through a PCI-muti I/O card which is represented by the “PCI-D/A” module of Figure 3.

The control voltage is represented in the digital display “Input”. Three A/D channels of the PCI-card are used for the measurement of the pressures \( p_o, p_a, p_b \). In order to eliminate the high frequency noise of the pump a DASYLab low pass filter module was used so that the “Y/T Chart” module can represent the real form of the pressure curves. Another A/D channel of the card measures the actual position of the inductive transducer of the proportional valve. This position is expressed in V and presented in the digital display module “Feedback”. If the difference between the indications of the DASYLab “Input” and “Feedback” digital display modules exceeds a threshold that is defined in the technical characteristics of the valve an error message appears that can be evaluated by the expert system.
The simulation program is running simultaneously with the real system and calculates the values of the angular velocity and the pressures for the same period of time. The comparison of measured and calculated values is then performed through another DASYLab “experiment” and the differences are written to output files after their transformation in linguistic information. This information can interact with the knowledge base of the system for the final diagnostic process. In addition important values from digital signals are used complimentary for the detection of possible faults.

The result of the data acquisition process for the pressures $p_a$, $p_b$ and $p_c$ at critical points of the hydraulic system is presented in Figure 4.
CONCLUSION

One of the main benefits of using simulation-based methods in a diagnostic system is the higher degree of reliability of the operation of the system by detecting small or unexpected faults. In this paper, a fault diagnosis method for proportional valves is presented where modelling information, actual process data and stored knowledge are used co-operatively.

This method enables quick and reliable detection of faults in hydraulic components and improves the efficiency of a diagnostic system for technical processes by permitting the diagnosis of faults at an early state. The experimental results provide evidence that the system is reliable and capable to detect faults of proportional valves and the method is applicable to the real world systems.

REFERENCES


FORMATION PRESERVING NAVIGATION OF AGENT TEAMS IN 3-D TERRAINS

Ali Galip Bayrak and Faruk Polat
Department of Computer Engineering
Middle East Technical University
06531, Ankara, Turkey
E-mail: {aligalip, polat}@ceng.metu.edu.tr
May 15, 2008

KEYWORDS
Formation Control, Military, Gaming, Continuous Simulation, Coordination

ABSTRACT

Navigation of a group of autonomous agents that are needed to maintain a formation is a challenging task which has not been studied much in especially 3-D terrains. This paper presents a novel approach to collision free path finding of multiple agents preserving a predefined formation in a 3-D terrain. The proposed method could be used in many areas like navigation of semi-automated forces (SAF) at unit level in military simulations and non player characters (NPC) in computer games. The proposed path finding algorithm first computes an optimal path from an initial point to a target point after analyzing the 3-D terrain data from which it constructs a weighted graph. Then it employs a real-time path finding algorithm specifically designed to realize the navigation of the group from one way point to the successive one on the optimal path generated at the previous stage, preserving the formation and avoiding collision both. A software was developed to test the methods discussed here.

INTRODUCTION

Navigation of a group of mobile agents in coordination is a popular problem studied in different areas such as robotics, computer games and military simulations. Several path finding algorithms were proposed for single and multiple robots (Balch and Arkin 1998) (Chia et al. 2005) (Desai et al. 1999) (Lewis and Tan 1997) (Ngo et al. 2005). These algorithms were generally developed for navigation in 2-D environments and hence restricted as they cannot be used or efficiently adopted to 3-D terrains. In computer games, especially in strategy games, algorithms are designed for moving a group of soldiers, or vehicles in the simulated battlefields (Dawson 2002) (Tomlinson 2004) (Verth et al. 2000). In military simulations, there has been growing interest in modelling behaviors at both individual and unit level to simulate decision making and tactics used in real-life for simulation based training, analysis and acquisition (ONE-SAF, SWARMM, Tebat, MODSAF). In that respect, navigation of individual soldiers/vehicles and that of a unit require efficient algorithms to be embedded in Semi-Automated Forces (SAF) (Reece 2003). There are only a few published works for the so-called path finding task in the areas of military simulations and computer games. Motivated by this, we developed algorithms for formation preserving navigation task of multiple autonomous agents in 3-D terrains and tested them on real 3-D maps.

In this paper we developed a novel algorithm for finding an efficient path between two points for a group of agents that are also required to maintain a pre-defined formation and avoid collisions with each other and with obstacles in 3-D terrains. We divided the problem into two main parts; finding the path and maintaining the formation. In path finding step, we first construct a weighted graph where nodes corresponds to way points that will define the route, edges to the accessibility of way points connected, weights to the cost between the way points connected. Then with the help of an off-line planner we determine the optimal path in this graph. Note that this is done before the agents actually start to move. After that, the agents, in parallel, follows the optimal path found by the high-level planner, by using a real-time algorithm that plans and navigates agents between two successive way points on the solution path, avoiding collisions and maintaining the formation.

In the next section, we define the environment and describe how terrain features are extracted from the 3-D terrain data. Following section contains an overview of our formation preserving path finding method. Then, steps of off-line path determination phase is given. After that, the on-line path finding algorithm is given in details. Next, experimental results and sample runs are given. The last section contains conclusions and future research directions.
DESCRIPTION AND REPRESENTATION OF THE ENVIRONMENT

The environment is a 3-D virtual world and represented as a height map. The natural and man-made objects such as trees and buildings are represented with polygons and they are treated as obstacles. Mobile agents are situated in the environment and aim to reach to a specified target point with a predefined speed. In order to maintain formation and avoid collisions they may tune their speeds. Concerning the representation of the environment, the most common approach is the grid representation where the environment is divided into a number of square cells of the same size, where each cell is marked as obstacle, or not. The search algorithms for finding a path from one cell to another, runs on the set of midpoints of these cells. The drawback of this method is that the size of cells should be determined carefully. If the size is small, the number of points in the search space will be huge. If the size is big, the homogeneity of the cells will decrease and even it may result in failure of search. There are some other methods like triangulation, Voronoi diagrams and visibility graphs, but these methods can be costly in 3-D terrains having huge number of contour lines, that result in huge number of points to be considered in the search. So, there is a need to modify the grid representation that minimizes the number of points for search and preserves homogeneity. In (Kambhampati and Davis 1986), Kambhampati and Davis propose a multi-resolution representation with quad trees. In this paper we use a simple but successful method that analyzes the terrain height map to extract useful features from which it identifies waypoints to be used by the path planner.

Concerning the representation, we consider the terrain as a 3-D surface and find the critical and singular points of the smoothed terrain, and construct a search graph using these points as vertices of the graph. Since the terrain may not be smooth enough, some equidistance points like in grid method are taken and the terrain is considered smooth in between these points.

First, we divide the map into small cells and take the center points of these cells, as in the grid method. The size of a cell should be as small such that it could be considered as homogeneous. Let sizeOfCell denote the size of a cell. Then, mark the points that are local maxima and local minima according to their height. The marking process adds the point to the set which will be used in the search. A point can be considered local maxima if its height is greater than any of its neighbors (4 or 8 neighborhood can be assumed for simplicity) and local minima if its height is less than any of its neighbors. Also, mark the points that are partially extrema. A point can be considered as partially extrema if it is either maxima, or minima on every opposing pair and it is not local maxima or local minima. To exemplify, considering 4 neighborhood, if the height of a point is greater than the height of its south and north neighbors and less than the height of its west and east neighbors it is partially extrema. The points that are marked (local maximum, local minimum and partially extremum points) will form the set of extremum points. Figure 1 shows examples for each of maxima (a), minima (b) and partially extrema (c) respectively. In the figure, the numbers in the cells represent the heights. And the points in the center cells are the extremum points.

![Figure 1: Extremum Points](image)

The terrain may be rough and hence there may be some very small hills or cavities. To eliminate such extremum points, the following condition is checked against every marked point:

\[
\text{(heightDifference (thePoint, N) < k) \land (heightDifference (thePoint, S) < k) \land (heightDifference (thePoint, E) < k) \land (heightDifference (thePoint, W) < k)} \tag{1}
\]

where N, S, E, W denotes the nearest marked point in its north, south, east and west, respectively. If the above condition evaluates to true for a point, then we unmark that point. \( k \) can be considered as the height of an obstacle that the agent can pass over easily. In this way, we finished determining local extremum points.

If a point is not accessible from at least one of its neighbors, then we call it singular point and mark it accordingly. A point is inaccessible from another point if the slope between them prevents the agent movement. Figure 2 contains a sample terrain where extremum points are shown with small black circles and singular points are shown with small black squares.

If the agent only considers the points marked so far in determining the path to its destination, the path realized may not be natural and the length of the path may be much longer than that of optimal path. Hence, we need to mark some additional points as follows. For each marked point \( u \), we divide the environment into \( l \) zones by \( \frac{l}{2} \) lines passing through \( u \) in the neighborhood directions where \( l \) is the number of neighborhood and we select the nearest accessible marked point \( v \), if exists, from each zone. Then we say that \( v \) is adjacent to \( u \). Additional points will be marked as follows. For each adjacent point pairs, we draw a line joining them. We mark such points on this line that, two consecutive
marked point on the line has height difference greater than or equal to a predefined value. Two consecutive marked points on the line are considered adjacent from now on. Figure 3 contains an example where height difference threshold is 1. The points shown with ‘X’ are additionally marked.

![Figure 2: Sample Environment](image)

![Figure 3: Marking Additional Points](image)

Determination of the set of the points which will be used in the search will be finished after marking and including the corners of the polygons representing the obstacles and the initial and goal locations of the agent group.

The marked points will form the vertices of the graph on which the search will be performed. The edges of the graph are determined as follows. We consider the corners of each obstacle in turn and introduce an edge between two corners if the projections of them to 2-D plane are in the line of sight of each other. Then, for each marked point (i.e., vertex) \( u \), if any adjacent point \( v \) is accessible from \( u \), an edge incident from \( u \) to \( v \) is created. A point is accessible from another if the slope between them is less than a threshold which can be prescribed by the physical capability of mobile agent and there is no obstacle in between them.

The cost of an edge is calculated as follows. We divide the edge into small pieces each of which has the length \( sizeOfCell \), and then sum the costs of all these small pieces. Cost of a small piece is initially assumed to be the flying distance between the two corners of the piece. Then, we multiply this cost with a function of slope such that the higher is the slope the higher is the cost. The last factor that affects the cost of an edge is whether the desired formation can be maintained along it. Note that, this is just a heuristic for calculating the costs of edges of the search graph and the calculation is done before the team actually start to move. Considering the formation as a rectangular box and checking whether this box can move along the edge can be a solution. However, moving along the whole edge for all edges will be computationally costly. So, we used a computationally cheap method given in Algorithm 1, which checks whether this box can pass through the two vertices incident with the edge. This does not always give the desired result but is a good approximation. Figure 4 shows an example for Algorithm 1 to determine whether the formation can pass through a way point. The algorithm finds two points, \( leftPoint \) and \( rightPoint \), on the line perpendicular to the edge, denoting the left and right boundaries of the line accessible from the way point and checks whether the formation can fit into the particular area between them.

![Figure 4: Checking Whether the Formation Can Pass Through a Point](image)

If this algorithm returns false, multiply the cost of the edge with a function of the distance between \( leftPoint \) and \( rightPoint \), such that, the smaller is the distance the higher is the cost. Repeat the same procedure for the end point of the edge. After this step, construction of the graph for searching is completed.

**THE PROPOSED FORMATION PRESERVING PATH FINDING METHOD**

In this section, we will give an overview of our method given in Figure 5. We first construct the search graph as described above. In this search graph, we then use an informed search technique (i.e., A*) to find an optimal path. If there exists a path, we use a smoothing algorithm in order to make the found path smooth, which may be jagged. After that, for each way point in the path we determine each agents’ position at that way point, assuming that the team is able to arrive at the way point. All these steps are executed off-line before the team actually starts to move. At this point, we have a path (sequence of way points) for each individual agent. Then, in real time, each agent moves between
Algorithm 1 CanFit(formationWidth, edge) : boolean
1: startPoint ← edge.start
2: lineP ← a line perpendicular the edge and passing at startPoint
3: for i = 1 to formationWidth do
4:   currentPoint ← point that is i * sizeOfCell far from startPoint on the line lineP and to the left of the startPoint
5:   if impassable(startPoint, currentPoint) then
6:      break
7:   end if
8: end for
9: leftPoint ← point that is (i-1) * sizeOfCell far from startPoint on the line lineP and to the left of the startPoint
10: for i = 1 to formationWidth do
11:   currentPoint ← point that is i * sizeOfCell far from startPoint on the line lineP and to the right of the startPoint
12:   if impassable(startPoint, currentPoint) then
13:      break
14:   end if
15: end for
16: rightPoint ← point that is (i-1) * sizeOfCell far from startPoint on the line lineP and to the left of the startPoint
17: if distance(leftPoint, rightPoint) ≥ formationWidth then
18:   return true
19: else
20:   return false
21: end if

its own way points in coordination with its teammates maintaining the formation and avoiding collision. During online path finding, the group reorganizes itself in case some agent loses its mobility. In the following sections we will describe each of the steps of off-line and online path planning in details.

OFF-LINE PATH FINDING

Since the number of points to be considered in the search in a realistic application can be huge, an informed search technique is used. We employ A* algorithm on the constructed weighted graph to produce an optimal path. However, the path generated by A* may be jagged. In order for the path to be more realistic, it should be smoothed. For path smoothing, we used the method described in (Ngo et al. 2005) given in Algorithm 2, with some little changes. The basic idea is that we remove a point from the path if its predecessor is visible to its successor.

Since the terrain is 3-D, the cost may sometimes can grow during this smoothing process. In order to prevent this, step 4:1. of the algorithm is revised as follows:

4:1. If the end node is not the goal, check whether \( \text{cost(start node, end node)} \leq \text{cost(start node, immediate precedent of end node)} + \text{cost(immediate precedent of end node, end node)} \) holds. If it holds, then choose the node successive to the end node as the new end node. Otherwise, record the immediate precedent of the end node as a new way point and choose end node as the start node with its successive as the end node. Go to step 2

Formation Representation

The spatial structure of a formation should be represented precisely. In (Lewis and Tan 1997), Lewis and Tan introduced the concept of virtual structure, considering the agent group as a rigid body and all the plan-
Algorithm 2 SmoothPath(wayPoints)
1: Choose the first path node as the start node and its successive node as the end node. Record the first path node as the new way point
2: Check the visibility between the start node and the end node
3: If they are not visible to each other, record the immediate precedent of the end node as a new way point and choose this node as the start node with its successive as the end node. Go to step 2
4: If they are visible to each other, then
   1. If the end node is not the goal, choose the successive to the end node as the new end node. Go to step 2
   2. If the end node is the goal, record it and exit

ning is done for this body. Another popular method is to let each agent decide its own path according to its relative position with respect to other agents. In (Desai et al. 1999), Desai et al. presented a graph theoretical approach where each agent determines its location according to the locations of other agents and there is a leader agent who does not follow any other agent but leads the group. The relation between two agents consists of relative distance and orientation between them. There are several methods that make use of priorities among agents to describe a formation in a team (Ngo et al. 2005).

In our work, we defined a formation by specifying relative positions of agents and their priorities. Agents are given ID’s from 1 to n where n denotes the number of agents. The agent with lowest ID (i.e., 1) has the highest priority and is called the leader. The lower the ID, the higher the priority. The priority is used to determine the order of movement within the next time period. The leader agent is given an ID of 1 first. An agent’s relative position is given with respect to that of the other agents having smaller ID. Relative position of agent a with respect to agent b is defined with two variables, Δdepth(a,b) and Δwidth(a,b). Δdepth(a,b) is the distance between a and b in the movement direction of the group. Δwidth(a,b) is the distance between a and b in the axis perpendicular to the movement direction of the group. In addition, the numbering of agents brings about another constraint that if agent a is closer to the front of the group in depth compared to agent b, a has lower ID than b’s. Figure 6 shows how some commonly used formations (User can design any formation using our representation) are represented. In the figure, circles are the agents and numbers in the circles are ID’s of the agents. A directed edge from agent a to agent b means that position of a is described with respect to position of b in the formation. a is called predecessor of b and b is called successor of a.

Figure 6: Formation Representation

Determining Agent Positions at Every Way Point

At any way point found by off-line planner, we use the following method to determine each agent’s position in accordance with the pre-defined formation. First, we determine whether the formation can fit into the area around a particular way point. By using Algorithm 1, for a way point, we determine the two points leftPoint and rightPoint, considering the edge between the way point and its successor. If the formation can fit between these two points, then we determine the midpoint of leftPoint and rightPoint and superpose the front midpoint of bounding rectangle of the formation on this midpoint. At the way point, goal position of each agent is the position of it in the superposed bounding rectangle. Bounding rectangle of the formation can be defined as follows. Draw a line passing through the leader agent which is perpendicular to facing direction of the formation. Then, draw a parallel line passing through the agent that is farthest from the first line. Then, draw two lines passing through the leftmost and the rightmost agents with respect to the moving direction. These 4 lines will form the bounding rectangle. An example is given in Figure 7.

Figure 7: Determination of Agent Positions at a Way Point
If the formation cannot fit between leftPoint and rightPoint, we use Algorithm 3 to shrink the formation as much as necessary to fit in.

Algorithm 3 FormationFit(wayPoint)
1: Let $d$ be distance between leftPoint and rightPoint
2: Let’s say two agents are on the same row if their distances to leader in depth are equal. Repeat the following steps starting from the first row (the leader’s row) until no row is left:
3: If the distance between the leftmost agent and rightmost agent of the row $\leq d$, place the agents to the current row
4: Else scale all the distances between the neighboring agents on the row until all the agents fit the row. If all agents can not fit, or any two agents became so closed that can affect movement, take out the agent with highest ID from the row and repeat this step
5: If there exists any agent taken out from the row, shift all rows beginning from the next row, introduce a new row, and put all the removed agents to this new row. Take the next row and go on step 2

Using this algorithm, agent positions at any way point are determined. Agents that appear on the same row perpendicular to the moving directions are considered one by one. The algorithm attempts to scale down the coordinates of agents in any row so that they can fit between the leftPoint and the rightPoint. The scaling algorithm also considers the constraint that no two agents should be placed within a predefined distance to guarantee that one’s movement will not be harmed by the others’. The algorithm enforces this constraint and identifies agents that must be removed from the row so that the row can fit between the leftPoint and the rightPoint. The removed agents will be placed on a new row that will be inserted next to the current row.

ONLINE PATH FINDING

Having the position calculated with the help of Algorithm 3 for each agent in the team at any way point, each agent navigates between its own way points in real time, avoiding the collisions. Agents plan their next moves and execute them at discrete time points where the time between two successive time points is very small. Note that at any time step agents plan and execute their moves in the order of their priorities (i.e., IDs).

We introduced an online path finder algorithm, low level planner, that navigates an agent from one way point to another which may take more than one time tick. At each time tick, each agent determines its moving direction by using Algorithm 4 and its speed by using Algorithm 5 concurrently and moves on that direction with the determined speed.

The basic idea behind the Algorithm 4 is to move the agent towards the next way point at each time tick and if this cannot be accomplished, get closer to the line from last visited way point to the next way point since it is guaranteed that the next way point is accessible from the last visited one. The reason why all agents do not move along this line is to move in formation.

Algorithm 4 OnlinePathFinder(nextWayPoint)
1: If the current location of the agent is near to the nextWayPoint, finish this procedure
2: Draw a line from the current location to nextWayPoint and let’s call the direction of this line as upper direction
3: Determine whether the agent can move without colliding with any other agent or obstacle with its current speed in upper direction
4: If it can move, then move the agent and go to step 1 for next time tick
5: If it will collide with any other mobile agent, then wait for it in this time tick and repeat step 2 in next time tick
6: Otherwise, apply these steps (i.e. step 2.1 - 2.3) for upper-left, upper-right, left and right directions instead of upper direction, respectively
7: If step 2 failed (agent cannot move in mentioned directions), check whether the agent is on the right side or the left side of the line drawn from last visited way point to nextWayPoint
8: If the result is right, try the procedure in step 2.1 for upper, upper-left, left, lower-left directions respectively
9: Else, try that procedure for upper, upper-right, right, lower-right directions respectively
10: If move is possible, first move to the direction found above (3.1 or 3.2), then if result is upper, upper-left or upper-right, go to step 1 for next time tick; else repeat this step (i.e., step 3) in next time step
11: Else if it will collide with any other mobile agent, then wait for it in this time tick and repeat step 3 in next time tick
12: Otherwise, wait a predefined amount of time (because the path may have been unavailable for an amount of time), and then go to step 2

In order to move smoother, the agents can change their direction a few steps before facing obstacles. This can be done by changing the step 2.1 as follows:

2.1: Determine whether the agent can move with its current speed in the direction of line, without colliding with any other agent in 1 unit time and without colliding with any obstacles during the next t unit times (where t determines the lookahead value, if it is huge the efficiency is affected negatively, if it is
very small (i.e. 1) it is same as main algorithm)

Agents tune their speeds at each time tick in order to preserve the formation while moving. Algorithm 5 is executed by the agents sequentially with increasing ID. According to the distance between an agent and its predecessor, it tunes its speed and tells the predecessor to do so when the distance exceeds a threshold.

Algorithm 5 TuneSpeeds(positionsOfAgents, directionOfAgents)

1: // p is a pre-defined threshold (a real number between 0 and 1)
2: // that specifies to what degree the relative position of any agent
3: // wrt its predecessor can be altered.
4: for i = 2 to numberOfAgents do
5: dist ← distance in depth between the agent i and agent predecessor(i) according to the direction(predecessor(i))
6: if dist < ∆depth(i, predecessor(i))*(1-p) then
7: decelerate(i)
8: else if dist > ∆depth(i, predecessor(i))*(1+2*p) then
9: decelerate(predecessor(i))
10: else if dist > ∆depth(i, predecessor(i)) * (1 + p) then
11: decelerate(i)
12: else
13: normalizeSpeed(i)
14: normalizeSpeed(predecessor(i))
15: end if
16: end for

If an agent gets more than one contradictory requests, the decelerate request has higher priority over accelerate request. However, an accelerate request has higher priority over a normalize request.

In the case of mobility loss of an agent, the other agents should rearrange their role according to the Algorithm 6. The team re-organizes by making each agent get the role (relative position and ID) of its predecessor. This is only the state change for the team, the formation will be restored in time as the team moves.

EXPERIMENTAL RESULTS AND SAMPLE RUN

The experimental results of the single agent path determination algorithm is given in this section. Algorithm was run on two different terrain data, 1000 m (meters) x 1000 m and 1500 m x 1500 m, on a PC with Intel Centrino 1.7GHz processor. 10 different random, but far from each other, start point and end point pairs are selected for each terrain. Then, for each pair, A* algorithm was tested on grid method(G), proposed environment representation with 1 m resolution, with 2 m resolution, with 5 m resolution and with just critical and singular points(C). For each terrain, average running times (in seconds), average path lengths (in meters) and the number of points in the search graph are given in Table 1 and Table 2.

<table>
<thead>
<tr>
<th>Method</th>
<th>Number Of Points</th>
<th>Time (sec)</th>
<th>Cost (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1m</td>
<td>54455</td>
<td>27</td>
<td>1818</td>
</tr>
<tr>
<td>2m</td>
<td>26585</td>
<td>15</td>
<td>1954</td>
</tr>
<tr>
<td>5m</td>
<td>13250</td>
<td>6</td>
<td>2062</td>
</tr>
<tr>
<td>C</td>
<td>8478</td>
<td>2</td>
<td>2172</td>
</tr>
<tr>
<td>G</td>
<td>1000000</td>
<td>104</td>
<td>1680</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>Number Of Points</th>
<th>Time</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1m</td>
<td>134724</td>
<td>77</td>
<td>2661</td>
</tr>
<tr>
<td>2m</td>
<td>70096</td>
<td>39</td>
<td>2587</td>
</tr>
<tr>
<td>5m</td>
<td>39730</td>
<td>20</td>
<td>3028</td>
</tr>
<tr>
<td>C</td>
<td>29279</td>
<td>13</td>
<td>3262</td>
</tr>
<tr>
<td>G</td>
<td>2250000</td>
<td>889</td>
<td>2581</td>
</tr>
</tbody>
</table>

Results show that the proposed terrain representation scales well with increasing sampling rate. The number of points considered only approximately doubles from no sampling to 5 m, then to 2 m, then to 1 m. From the results we also note the significant improvement gained by the our terrain representation over Grid method in terms of time and path length.

To test all the proposed methods, especially maintenance of formation, we implemented a 3-D software using OGRE as graphics rendering engine. Screen-shots from sample runs of three common formations, line, column and wedge, are given in Figure 8(a), Figure 8(b) and Figure 8(c). In figures, the red lines show the paths travelled by agents.
CONCLUSIONS

In this work we developed an algorithm for planning a path for a group of autonomous agents that need to move in a specified formation in 3-D environments. First, considering the group as a rigid body, a search graph is constructed by analyzing and identifying important terrain features in 3-D height map. Then a high level planner that uses A* algorithm determines an optimal path from an initial location to a target location. Then with the help of a low level on-line planner, each agent in the team navigates between way points on the solution path, avoiding collision with each other and with environmental objects. We defined a representation for group formation on which algorithms were developed to maintain formation while moving from one way point to another. We implemented the proposed algorithms and observed a significant gain in both performance and solution quality over grid based terrain representation. We come up with a software system that can be used in real-life applications. The proposed method has also ability to repair formation when an agent or some agents in the team loses mobility, which is quite possible in real-life applications. The proposed method can be used in robotics, computer games, military simulation and unmanned vehicle teams. As a future work we are planning to extend our work to helicopter team movement in 3-D terrains.

REFERENCES


Fusion of INS with GNSS and its application to Mobile Robot Navigation

L. Domenech, L. Armesto, J. Tornero
Polytechnic University of Valencia
email: luidobal@posgrado.upv.es, leoran@isa.upv.es, jtornero.isa.upv.es

KEYWORDS
GNSS, MEMS, INS, Fusion, Navigation

ABSTRACT
This paper investigates the fusion of Inertial Navigation Systems (INS) and Global Navigation Satellite Systems (GNSS). Both systems present interesting complementary characteristics to increase pose and orientation estimation performance, in comparison to separated approaches. In the paper, two different fusion strategies are discussed and compared, namely the Uncoupled and Loosely Coupled, where clearly the Loosely Coupled gives better estimations. The application of this paper is to estimate the motion of a mobile robot in different contexts such as mobile robot manipulators for facade restoration tasks and Automated Guided Vehicles (AGVs) in industrial outdoor environments.

INTRODUCTION
The accurate and continuous position is fundamental for navigation applications. The accuracy provided by the Global Navigation Satellite System (GNSS) has significantly improved, reducing the error from meters to centimeters. However, the environment in which the system works takes a great importance in the quality and availability of the measurements. Moreover, the speed at which the sensor works is too slow to be considered as a continuous trajectory.

On the contrast, Inertial Navigation Systems (INS) use Inertial Measurement Units (IMU) to sense accelerations and angular rates at a high frequency and their availability doesn’t depend on the environment conditions, but the Gravity field. However, the measurement errors, mainly bias, produce a degradation that increases with the number of measurements.

Furthermore, this precise IMU sensor is highly expensive. However, these days, given that IMU sensors are being manufactured with MEMS technology, the situation is changing dramatically. MEMS IMU provides a more affordable solution for medium and low cost applications. At present, MEMS IMU sensors are still not performing as well as conventional high-grade sensors in terms of precision. Therefore, frequent calibrations using external aiding sources are essential, such as magnetometers for orientation and GNSS for position.

The fusion of GNSS/INS allows avoiding most of the problems that appear when any of the previously discussed systems are used separately. This is due to the complementary characteristics of sensors, such as sampling frequency, data missing and robustness, accuracy of measurements, etc. In this sense, when the GNSS signal isn’t available, the INS can provide enough information for precise navigation, while the GNSS signal can correct the accumulative error that is generated by the IMU sensor. Theoretically, GNSS is capable of providing precise positioning information and is appropriate for being used as an INS aiding source. However, the GNSS doesn’t contribute directly to the orientation estimation, therefore the orientation estimation should be aided by means of data, such as the magnetic field.

This paper investigates and applies the fusion of GNSS and INS based on MEMS technology providing more precise and accurate estimations in comparison with separate approaches.

In this paper the implementation of the proposed fusion (GNSS/INS) has been applied to self-localization mobile robots in two different contexts:

1. Facade restoration with a mobile manipulator (robot arms mounted on a mobile base): In this context it is required to known the position of the mobile robot end-effector to properly apply manipulation strategies for restoration, as shown in Figure 1a.

2. Navigation of a fleet of industrial vehicles (AGVs) which operates in warehouses with bulk-storage systems in outdoor environments. Figure 1b shows the vehicle which has been automated for this purpose.

GNSS NAVIGATION
Let be $F^e$ a inertial frame such that its origin is in the earth center, its $Z$-axis is the earth rotation axis, and the axis $X,Y$ defines the plane of the Equator. And be $F^i$ a not inertial frame such that its origin is in the earth center, the $Z$-axis is the earth rotation axis, the $X$-axis is in the Equator plane and passes through the Greenwich meridian, and the $Y$-axis is perpendicular to $Z$-axis and $X$-axis. Then

$$x^i = C_{ei}x^e$$ (1)
where \( \Omega_{ic}^e \) is the matrix describing the rotation between body frame (mobile robot) and \( F^e \) frame, and \( a^b \) is the acceleration in the body frame.

That equation can be represented in a general way in that form:

\[
\ddot{x}^e = C_b^e a^b + f(\Omega_{ic}^e, \dot{x}^e, x^e)
\]

where in general, the terms grouped in the vector function \( f(\cdot) \) are much less than the values of the first term, mainly due to the slow angular rate of the Earth. That fact allows to use a simple integration algorithm, such as Simpson’s rule with accurate results.

The GNSS Navigation is implemented with a filter in which the \( a^b \) is treated as system noise, and the sensor measurement is the GNSS data.

GNSS Navigation Equations:

\[
\begin{bmatrix}
\dot{x}_e \\
\dot{y}_e \\
\dot{z}_e
\end{bmatrix} = \begin{bmatrix}
0 & 1 & 0 \\
-(\Omega_{ic}^e)^2 & -2\Omega_{ic}^e & 0 \\
0 & 0 & 0
\end{bmatrix} \begin{bmatrix}
x_e \\
y_e \\
z_e
\end{bmatrix} + \begin{bmatrix}
0 \\
C_b^e \\
W
\end{bmatrix}
\]

\[
Y = \begin{bmatrix}
1 & 0 & 0
\end{bmatrix} \begin{bmatrix}
x_e \\
y_e \\
z_e
\end{bmatrix} + R
\]

The measurements supply the position in the \( F^e \) frame in polar coordinates. There are many methods that allow to convert the polar coordinates to Cartesian coordinates. One of the most widely used is WGS84 GRS (2004), which converts \( \{\phi, \lambda, h\} \) to \( \{X, Y, Z\} \) as follows:

\[
\begin{bmatrix}
X \\
Y \\
Z
\end{bmatrix} = \begin{bmatrix}
(N + h) \cos(\phi) \cos(\lambda) \\
(N + h) \cos(\phi) \sin(\lambda) \\
(N(1 - e^2) + h) \sin(\phi)
\end{bmatrix}
\]

where \( a = 6378137 m, f = 0.00335281, e = \sqrt{2f - f^2} \) and \( N = \frac{a}{\sqrt{1 - e^2 \sin^2(\phi)}} \).

Errors in GNSS Navigation

The GNSS Navigation has a set of well known disadvantages that make difficult its application to mobile robot navigation. In this sense, it has a high sensibility to the meteorological conditions, that means that the satellite’s signal can become too weak, avoiding the localization, or delayed, generating an error in the calculation of the position.

Moreover, the elements of the environment can also have influence on the signal reception, since they can block the signal and cause rebounds (Multiroute), failing in the pose estimation or causing errors.

Finally, the main disadvantage that makes GNSS Navigation a poor solution to vehicle self-localization, lies on the slow frequency at which the sensor works (10Hz), which implies that the uncertainty between measurements is too big for accurate localizations.
INS NAVIGATION

INS Navigation is based on the relative localization by means of the integration of the accelerations measured by the accelerometers, and angular rates measured by gyroscopes to obtain the orientation. The IMU (Inertial measurement Unit) is the set of inertial sensors, which joined with the movement equations defines the INS (Inertial Navigation System).

In general, the INS systems can be classified according to their reference frame:

1. **Gimbal**: The IMU frame corresponds to a global reference frame. The main advantage is that it is not necessary to transform coordinates, making calculation easier.

2. **Strapdown**: The IMU frame coincides with the frame of the vehicle, measuring the local accelerations and angular rates. It is necessary to calculate the attitude of the vehicle in order to transform the measured values to the global frame. MEMS technology uses a strapdown configuration, which is the one adopted for this paper.

The position equations that govern the movement of a vehicle are the same as the explained above, see eq (3). However, their implementation in the filter is not the same, since there is no measurement equation, and the \( a^b \) becomes the input term.

\[
\begin{bmatrix}
\dot{x}_e \\
\dot{\alpha}_e
\end{bmatrix} = \begin{bmatrix}
0 & 1 \\
-(\Omega^b_{re})^2 & -2\Omega^b_{re}
\end{bmatrix} \begin{bmatrix}
x_e \\
\alpha_e
\end{bmatrix} + \begin{bmatrix}
0 \\
C^b_r
\end{bmatrix} a^b
\]

Moreover, the INS let us also obtain the orientation of the vehicle, since it also measures the angular rates. The representation most widely used is quaternions, which have a well known dynamics equation.

\[
\dot{q} = \frac{1}{2} \Omega^b_{re} q
\]

where \( \Omega^b_{re} \) is the angular rate of the body in the \( F^e \) frame.

To resolve the equation (10) many algorithms can be used, easy implementations such as Simpson rules or others more complicate such as Runge-Kutta. In Bortz (1971) Ignagni (1990) Jiang (1992) are developed specific algorithms that increases the precision and reduces the computation cost. In Jekely (2000) a comparison of them is done, and the third order Runge-Kutta algorithm is justified to be enough accurate to reduce the estimation error below the sensor error.

Errors in INS Navigation

The main disadvantage of INS Navigation lies in the fact that there isn’t any measurement equation in the filter. That implies that position is not an observable variable, which means that the uncertainty of the position is unbounded and the position consequently diverges. Under the point of view of implementation that means that the accuracy of the INS is only valid over a finite time, depending on the accuracy of the IMU sensor.

Another disadvantage of precise IMU systems is the price, which is not valid for low-cost applications. However from few years, the MEMS technology has evolved, allowing to join microelectrical and micromechanical pieces in a only one integrated low-cost chip.

GNSS/INS FUSION

The navigation techniques that use only one of the previously mentioned sensors have been proved to be too poor for real-time mobile robot navigation. Integration of GNSS and INS is a widely accepted solution in providing position, velocity and attitude in modern navigation and geo-referencing applications. The combination of these two sensors have the advantages of long term positioning accuracy, high update rates, robustness to GNSS signal jitters and interference sources, and continuous calibration of INS errors during operations so that the inertial navigation performance could be improved during GNSS outages.

Strategies of Fusion

1. **Uncoupled**: In that case, both sensors process the measurements independently, and the integration between both systems is implemented by a decision algorithm, which according to the availability selects one sensor or other. That is, the GNSS is used as primary system, since it guarantees the stability, and the INS is reset with each measurement of the GNSS.

2. **Loosely Coupled**: It is possible to get an improvement replacing the decision algorithm by a fusion filter, such as Kalman filter. That allows us to estimate an therefore to correct sensor errors, such as accelerometers biases.

3. **Hardly Coupled**: This strategy consist of a complete couplement between the information from the INS and the GNSS in only one filter which includes all the equations of movement of both sensors, and the bias of each measurement, in Weiss (1996) there is a detailed explanation of that technique. That allows as main advantage that although there aren’t visible the minimum number of satellites to get the position from the GNSS sensor, it is possible to use the available satellites to correct, as much as possible, the INS estimation.

There are different filters that can be used to implement the data fusion between both sensors. Since it is a system with perfectly known equations, but not all are
linear, the most widely filter used is EKF, since in general non-linearities are quadratic terms. In Marins et al. (2001) Beghi (2005) Yunchun Yang; Farrell (2003) Bevly (2006) are show different implementations of EKF. However, there are other filters that are also used, in El-Sheimy (2005) an study of Neuronal Networks is done. In Jr. (2003) a comparative between EKF and UKF in systems with quaternions is done. Finally, in Vanicek (1999) an specific filter for Vehicles Navigation is developed.

**Filter Variables**

The set of variables that are implemented as state variables is a fundamental decision in the filter design. That happens because the filter is composed by some non-linear equations, so that, if all the non-linearities are extracted from the filter (avoiding the error propagation) making it as linear as possible, the estimation obtained has not accuracy. While if all the non-linearities are treated inside the filter, not only the computation cost increases, but also the complexity of the filter. In Marins et al. (2001) that effect is showed perfectly, since 2 proposes are done, the first has a highly non-linear measurement equation, and the second has the identity equation. Our implementation is as follows:

The Position Filter has the next dynamic equation:

$$
\begin{bmatrix}
\dot{p} \\
\dot{v} \\
\dot{B}
\end{bmatrix} = \begin{bmatrix}
0 & I & 0 \\
-(\Omega_{ie})^2 & -2\Omega_{ie} & C_{eb} \\
0 & 0 & 0
\end{bmatrix} 
\begin{bmatrix}
p \\
v \\
B
\end{bmatrix} + 
\begin{bmatrix}
0 \\
C_{eb} \\
0
\end{bmatrix} a^b
$$

(11)

with $I$ the 3x3 identity matrix.

While the measurement equation is:

$$
Y = \begin{bmatrix} I & 0 & 0 \end{bmatrix} 
\begin{bmatrix}
p \\
v \\
B
\end{bmatrix}
$$

(12)

And the Orientation Filter:

$$
\dot{q} = \frac{1}{2} A q
$$

(13)

where $A$ is the quaternion matrix form:

$$
A = \begin{bmatrix}
0 & w_x & w_z & w_y \\
-w_z & 0 & w_y & w_x \\
-w_y & -w_x & 0 & w_z \\
w_y & w_z & -w_x & 0
\end{bmatrix}
$$

(14)

where $w_x, w_y, w_z$ are the angular rates measured by the IMU.

**RESULTS**

An experiment in an outdoor environment has been done, using a IMU strapdown XSens MT9-B and a GPS receiver TGPS Rasant. Data fusion algorithms have been implemented in MATLAB, and computed off-line where both, Loosely Coupled and Uncoupled strategies are compared. MEMS INS initial alignment is obtained by means of a magnetometer integrated into the IMU sensor. During the field test, GPS data obtained from the TGPS Rasant receiver and IMU raw data obtained from the XSens MT9-B were time synchronised and logged using the same hardware.

In Figure 3, it is shown a reconstruction of the trajectory using the Uncoupled algorithm, comparing it with Real Data provided by GPS receiver, while in Figure 4 it is shown the reconstruction of the trajectory based on a Loosely Coupled algorithm. It is obvious that the Coupled solution is more accurate that the Uncoupled. The main reason of this difference is a consequence of the good estimation of the accelerometer bias that takes place in the Coupled solution. Bias estimation can be seen in Figure 5, where clearly converges to not zero values, as expected.

**CONCLUSIONS**

In this paper we have investigated strategies for GNSS/INS fusion in the context of mobile robotics. In particular, two different approaches have been implemented, discussed and compared, namely the Uncoupled and Loosely Coupled. For each case, the fusion has been implemented in a different way, showing that
the Loosely Coupled gives better estimations, because it does not perform a continuous resetting of previous estimations.

In this paper the implementation of the proposed fusion (GNSS/INS) has been applied to self-localization mobile robots in two different contexts: Facade restoration with a mobile manipulator (robot arms mounted on a mobile base) and Navigation of a fleet of industrial vehicles (AGVs) in outdoor environments.

Further work is oriented in fusing magnetometer data as well as data coming from a vision system. With this proposal more precise estimations are expected. In addition, the vision system will provide additional data for 3D modelling and reconstruction.

Acknowledgments

This work has been supported by the Spanish Government (Ministerio de Ciencia y Tecnología), research project BIA2005-09377-C03-02 and Local Government (GVA-IMPIVA), research project IMPCV/2007/8.

REFERENCES


Vanicek P; Omerbasic M., 1999. Does a navigation algorithm have to use a Kalman filter?


APPENDIX A

The Newton’s law is only defined for inertial frames, so it can only be applied for $F^i$ frame,

$$\ddot{x}^i = \dot{C}^i_{\tilde{e}} x^e + C^i_{\tilde{e}} \ddot{x}^e$$

(15)

According to the properties of the rotation matrices

$$\dot{x}^\tilde{e} = C^i_{\tilde{e}} \dot{x}^i x^e + C^i_{\tilde{e}} \ddot{x}^e$$

(16)

Then, deriving to work with accelerations

$$\ddot{x}^i = \dot{C}^i_{\tilde{e}} \Omega^e_{\tilde{e}} \dot{x}^e + C^i_{\tilde{e}} \dot{\Omega}^e_{\tilde{e}} x^e + C^i_{\tilde{e}} \ddot{x}^e + \dot{C}^i_{\tilde{e}} \ddot{x}^e + C^i_{\tilde{e}} \ddot{x}^e$$

(17)

taking account that $\dot{\Omega}^e_{\tilde{e}} = 0$

$$\ddot{x}^e = C^i_{\tilde{e}} \Omega^e_{\tilde{e}} x^e + 2C^i_{\tilde{e}} \Omega^e_{\tilde{e}} \dot{x}^e + C^i_{\tilde{e}} \ddot{x}^e$$

(18)

isolating the term $\ddot{x}^e$

$$\ddot{x}^e = -2\Omega^e_{\tilde{e}} \dot{x}^e - \Omega^e_{\tilde{e}} \Omega^e_{\tilde{e}} x^e + C^i_{\tilde{e}} \ddot{x}^i$$

(19)

By definition, $\ddot{x}^i$ is the acceleration of the vehicle in the reference frame $F^i$, so it can be expressed in the frame $F^e$ by multiplying by $C^i_{\tilde{e}}$

$$\ddot{x}^e = -2\Omega^e_{\tilde{e}} \dot{x}^e - \Omega^e_{\tilde{e}} \Omega^e_{\tilde{e}} x^e + a^e$$

(20)

where $a^b$ is the acceleration measured in the body frame $F^b$. 
TRAJECTORY PLANNING BASED ON ROBOT DYNAMICS WITH ARTIFICIAL POTENTIAL FIELDS

José A. Bernabé
Department of Systems Engineering and Control
Universidad Politécnica de Valencia
46022 Valencia (Spain)
e-mail: joberchu@doctor.upv.es

Marta C. Mora
Department of Mechanical Engineering and Construction
Universitat Jaume I
12071 Castellón (Spain)
e-mail: mmora@emc.ujl.es

Ana Ruescas and Josep Tornero
Department of Systems Engineering and Control
Universidad Politécnica de Valencia
46022 Valencia (Spain)
e-mail: jtornero@isa.upv.es

KEYWORDS

ABSTRACT
In this paper, two different strategies for implementing robot trajectory planning using artificial potential fields are described. The standard recursive Newton-Euler algorithm for robot dynamics computation has been modified in order to include the attractive and repulsive forces produced by the Artificial Potential Fields approach. It results in an iterative process that is repeated until the whole trajectory is generated. The first strategy has been developed using a standard CAD/CAM/CAE software, NX™, that provides an accurate model of the robot as well as precise computation of distances between the robot links and the obstacles in the environment. The second strategy has been completely implemented in Matlab, except for a final simulation carried out in NX™. The selection of one or another approach will depend on the trade-off between accuracy and efficiency, in terms of distance computation, required by the application.

INTRODUCTION
Nowadays, simulation in manufacturing processes has become an essential issue, mainly in the creation of new processes or in the redesign of the existing ones. The effort invested in simulation is inversely proportional to the associated costs of developing new processes, as it avoids the constant design changes that would take place at a later stage, with significant cost increment.

In this context, the utilization of a system that integrates CAD/CAM/CAE/Robotics platforms with simulation abilities is very interesting because it allows a direct information flow throw the different stages of the manufacturing process. In this sense, NX™ (NX web-page, 2008) is a product development system that provides a complete suite of integrated process automation tools to enable the capture and reuse of product and process knowledge. One of these tools, NX-Motion Simulation, is a complete solution for kinematics and dynamic motion analysis of rigid multi-body, where robots manipulators constitute a complex case.

In this work two strategies for manipulator trajectory planning tasks are described. The objective consists in simulating the trajectory of manipulator robots (KUKA KR 15/2 robot arm) in structured environments as a way of validating the movement of the robot. Robot trajectories have to be computed avoiding self collision as well as collision with the payload and the existing obstacles. The obstacle avoidance is based on the Artificial Potential Fields method (Khatib, 1986), that produces repulsive forces moving the robot away from the obstacles in the environment as well as attractive forces guiding it to the goal. It is also possible to introduce additional forces in order to avoid aspect such as singularities. These forces must be applied to the robot links exactly at the points that may produce collision. However, robot is controlled by forces and torques applied to the prismatic and revolution joints. The translation of forces into joint forces/torques is done by means of the kinematic and dynamic models of the robot arm, based on the modified Denavit-Hartenberg representation (Craig, 1989) and the Recursive Newton-Euler (RNE) formulation (Fu et. al. 1987), respectively. In this sense, the dynamic calculation is usually performed using the Lagrange-Euler formulation because of its conceptual simplicity, as in (Xie et. al., 1998) and (Chotiprayanakul et. al., 2007). However, the matrix computation involved makes it computationally expensive. Thus, we have modified the RNE formulation in order to tackle this problem. This method is conceptually more complex but much more computationally efficient.

The first approach is focused on the NX-CAD platform, where an accurate robot model is considered with the aim of computing precise distances to the obstacles. The trajectory planning and dynamic calculus is integrated as a Dynamic Link Library (DLL) into the NX™ platform. The resulting trajectory is afterwards simulated for validation using the NX Motion Simulation software.

The second approach is based on Matlab platform (Matlab web-page, 2008) using spherical bounding volumes to model the robot arm as well as the obstacles. Spherical volumes allow efficient distance computation which can be implemented under real time requirements (Tornero et. al., 1991). In this case, the trajectory planning and dynamics modules are Simulink blocks inside the Matlab application. However, simulation is still performed using NX Motion.

The paper is organized as follows: next section describes
the general overview of both implementation strategies; the following section introduces the NX™ platform features, where computation and simulations are performed; afterwards, robot modeling is addressed based on accurate NX model and spherical convex hulls; next, the path planning approach is described; kinematics and dynamics models of robot are then established while, finally, simulation results, conclusions and future work are stated.

GENERAL OVERVIEW

In this section the two strategies for manipulator path planning tasks are described. The first one is based on the NX-CAD platform, as shown in Figure 1. NX™ software is used to generate an accurate model of the robot and to compute minimum distances between the robot links and the obstacles in the environment. Later, these distances are used in the calculation of the Artificial Potential Fields (APF) that will produce attractive and repulsive forces on each link, depending on the configurations of the goal and the obstacles at each time. These forces are then introduced in the Recursive Newton-Euler module (RNE) to produce links displacements, by means of the inverse dynamics of the robot arm. This is an iterative process that is repeated until the whole trajectory is generated. Next, the computed trajectory is introduced in the NX-Motion for simulation.

The second strategy is implemented in Matlab where kinematics and dynamics as well as object modeling have been developed. Kinematics and spherical volumes are considered for distance computation while dynamics and APF are considered for trajectory planning, as depicted in Figure 2. NX Motion Simulation is also used for checking collision avoidance.

NX of Unigraphics™ is a digital product development system that helps companies in transforming the product lifecycle. It is a suite of integrated CAD/CAM/CAE applications that touches the full range of development processes in product design, manufacturing and simulation. NX™ offers a broad set of CAD solutions for the design of complex mechanical products that deliver higher efficiency and shorter design cycles at lower costs. One of the most interesting features is that offers an API (Application Programming Interface) for use with Microsoft’s .NET framework, which provides access to NX™ core application functionality, making it possible to create automation programs using any .NET-compliant language.

As a component of NX™ system, NX-Motion Simulation allows to understand, evaluate and optimize the complex motion behaviour of assemblies, products and systems. It provides a solution for kinematic and dynamic motion analysis of rigid multi-body. All geometric models are supported allowing top-down and bottom-up approaches. It is possible to evaluate multiple design alternatives at all phases during the design process.

ROBOT GEOMETRIC MODELING

Trajectory and path planning in robotics systems can be listed in the following hierarchical way: planning problem (higher level), collision avoidance, distance computation and object model representation (lower level). According to this classification, the object model representation plays an important role in the computational cost of the upper levels. In this sense, two different ways of modeling a robot arm are described here, CAD Modeling using NX™ for an accurate model and spherical modeling for an approximate model, generating different envelopes depending on the required accuracy. The advantage of the CAD Modeling is that distances between objects are accurately computed while the drawback is the high computational cost. On the contrary, the advantage of using approximate models is the low computational cost obtained reducing model precision.

CAD Modeling

The platform used to model the robot arm is NX™ software, described in the previous section. With this platform it is possible to accurately model a robot arm and, therefore, take into account every feature of the robot in the trajectory generation and path planning application. The model of a KR 15/2 KUKA robot is shown in Figure 4.

Figure 1: Flow Chart of NX Implementation.

Figure 2: Flow Chart of Matlab Implementation.

NX™ CAD/CAM/CAE PLATFORM
Spherical Bounding Volumes

A frequent option in geometric modeling of robotic systems consists in obtaining the minimum bounding volume of the system in order to reduce distance computational costs in path planning. In this sense, spherical models are a preferred option because of their simplicity.

The spherical volumes employed here are the spherically-extended polytopes (s-topes), widely used in robotics (Tornero et al. 1991), (Del Pobil, 1992), (Gilbert et. al., 1988), because of their robustness when representing robotic systems. An s-tope is the convex hull of a finite set of spheres \( s \), defined by \( s=(c,r) \), where \( c \) is the center and \( r \) is its radius. Given the set of spheres \( S=\{s_0, s_1, \ldots, s_n\} \), the convex hull of such a set, \( S_o \), contains an infinite set of swept spheres expressed by Equation (1).

\[
S_o = \left\{ s : s = s_0 + \sum_{i=1}^{n} \lambda_i (s_i-s_0), s_i \in S, \lambda_i \geq 0, \sum_{i=1}^{n} \lambda_i \leq 1 \right\}
\]  

(1)

Simplest s-topes (from one to four spheres) are called sphere, bi sphere, tri sphere and tetra sphere. In Figure 5, different s-tope models are used for modeling the KUKA KR15/2 robot arm. The selection of the s-tope employed depends on the level of accuracy required in the application.

PATH PLANNING APPROACH

The Artificial Potential Fields method (Khatib, 1986) has been widely used for the obstacle avoidance problem in both manipulators and mobile robots. This technique represents the robot as a particle under the influence of an artificial potential field, whose local variations reflect the structure of the free space.

The potential function in a position \( p \), \( U(p) \), is defined as the sum of an attractive potential \( U_{attract}(p) \), which stretches the robot towards the destination, and a repulsive potential function \( U_{rep}(p) \), which pushes the robot away from the obstacles, as stated in Equation (2).

\[
U(p) = U_{attract}(p) + U_{rep}(p)
\]  

(2)

This potential function generates a force field \( F(p) \), with attractive and repulsive components, which causes the robot motion in the negative direction of the gradient of potential, as indicated by the Equation (3).

\[
F_{attract}(p) = -\nabla U_{attract}(p) \quad F_{rep}(p) = -\nabla U_{rep}(p)
\]  

(3)

On the one hand, the end-effector of the manipulator is influenced by the attractive potential field generated by the goal and by the repulsive potential field generated by the obstacles nearby. On the other hand, all the manipulator links are affected only by the repulsive potential field.

The attractive potential function is defined in Equation (4), where \( \xi \) is a positive scaling factor, \( p_{ref} \) is the end-effector position vector and \( p_{goal} \) is the destination position vector, both defined in the global reference frame.

\[
U_{attract}(p) = \frac{1}{2} \|\xi (p_{ref} - p_{goal})\|^2
\]  

(4)

The repulsive potential function for the \( i \)-th manipulator link is defined in Equation (5), where \( \eta \) is a positive scaling factor, \( d_{obs} \) is the obstacle influence area and \( d_i \) is the minimum distance between the \( i \)-th link of the robot model and the nearest obstacle, \( d_i = \| p_{link} - p_{obstacle} \| \), being \( p_{link} \) the nearest point to the obstacle in the \( i \)-th link of the robot model and \( p_{obstacle} \), the corresponding obstacle point.

\[
U_{rep,i}(p) = \begin{cases} 
\frac{1}{2} \left( \frac{1}{d_i} - \frac{1}{d_{obs}} \right)^2 & \text{if } d_i \leq d_{obs} \\
0 & \text{if } d_i > d_{obs}
\end{cases}
\]  

(5)

In the NX approach, the \( d_i \) distance is provided by the software. In the Matlab approach, where s-tope models are used, this distance refers to bounding volumes and is named Minimum Translational Distance (MTD). The MTD is defined in (Bernabeu et al, 2001) as the shortest relative translation of two models to bring them in contact. Thus, when two models are not intersecting MTD represents the separation distance between them. Otherwise, if models are colliding MTD states the penetration distance.

ROBOT KINEMATICS AND DYNAMICS

The forces derived from the APF method must be converted into joint torques in order to get the robot to a non-collision configuration. For this purpose, the robot dynamic model is needed. In this sense, we propose a modified version of the recursive Newton-Euler (RNE) formulation with the aim of including the external forces obtained from the path planning module in the joint torques computation. The RNE algorithm requires the forward kinematic model of the manipulator and is described next.

Kinematic Model

The robot kinematic model refers to the geometric relation between the robot motion in the configuration space and the motion of the tool frame relative to the base frame. In this section, the forward kinematics using modified Denavit-Hartenberg (MDH) formulation is detailed.

The relative position and orientation of frame \( i \) referring to frame \( i-1 \) is described by an homogeneous transformation matrix, in (6), where \( p \) is the position vector of the origin of frame \( i \) expressed in frame \( i-1 \) and \( ^iR \) is the rotation matrix. The columns of \( ^iR \) represent the components of the unit vectors of frame \( i \) expressed in frame \( i-1 \).
\[ i^{-1} T_i = \begin{bmatrix} i^{-1} R_i & p_i \\ 0_{3 \times 1} & 1 \end{bmatrix} \]  

(6)

The forward kinematics transforms vectors expressed in the \( n \) frame into their components referred to the base coordinate frame, as indicated in Equation (7).

\[ ^0 T_n = ^0 T_1 T_2 \ldots n^{-1} T_n \]  

(7)

With the MDH conventions, displayed in Figure 6, the general joint transformations can be described as a composition of four elementary transformations:
1. Rotate an angle \( \alpha_{i-1} \) about the \( X_i \) axis.
2. Translate a distance \( a_{i-1} \) along the \( X_i \) axis.
3. Rotate an angle \( \theta_i \) about the \( Z_i \) axis.
4. Translate a distance \( d_i \) along the \( Z_i \) axis.

The resulting matrix transforms the coordinates of a point from frame \( i \) to frame \( i-1 \), as in Equation (8). This transformation is not constant, but varies as a function of the corresponding joint variable. In particular, the MDH matrix for the KUKA KR15/2 robot is shown in Table 1.

\[ ^i J_i = \text{Rot}(X_i, \alpha_{i-1}) \text{Trans}(X_i, a_{i-1}) \text{Rot}(Z_i, \theta_i) \text{Trans}(Z_i, d_i) = \begin{bmatrix} \cos \theta_i & -\sin \theta_i & 0 & a_{i-1} \\ \sin \theta_i \cos \alpha_{i-1} & \cos \theta_i \cos \alpha_{i-1} & \sin \alpha_{i-1} & -d_i \sin \alpha_{i-1} \\ \sin \theta_i \sin \alpha_{i-1} & \cos \theta_i \sin \alpha_{i-1} & \cos \alpha_{i-1} & d_i \cos \alpha_{i-1} \end{bmatrix} \]  

(8)

Table 1: MDH Matrix of the KUKA KR15/2 Robot

<table>
<thead>
<tr>
<th>Link</th>
<th>( \alpha_{i-1} )</th>
<th>( a_{i-1} )</th>
<th>( d_i )</th>
<th>( \theta_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Link-0</td>
<td>( \pi )</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Link-1</td>
<td>( \pi/2 )</td>
<td>0</td>
<td>675</td>
<td>( q_1 )</td>
</tr>
<tr>
<td>Link-2</td>
<td>( \pi/2 )</td>
<td>300</td>
<td>0</td>
<td>( q_2 )</td>
</tr>
<tr>
<td>Link-3</td>
<td>( \pi/2 )</td>
<td>0</td>
<td>650</td>
<td>( q_3 )</td>
</tr>
<tr>
<td>Link-4</td>
<td>( \pi/2 )</td>
<td>155</td>
<td>-600</td>
<td>( q_4 )</td>
</tr>
<tr>
<td>Link-5</td>
<td>-( \pi/2 )</td>
<td>0</td>
<td>0</td>
<td>( q_5 )</td>
</tr>
<tr>
<td>Link-6</td>
<td>-( \pi/2 )</td>
<td>0</td>
<td>0</td>
<td>( q_6 )</td>
</tr>
</tbody>
</table>

Modified Recursive Newton-Euler Formulation

The Newton-Euler formalism for a serial manipulator implies two recursive computations: a forward (kinematics) recursion from the base to the end-effector, which computes the velocity and acceleration of each link, and a backward (dynamics) recursion from the end-effector to the base, which computes forces and torques applied to each link. In the RNE algorithm, the following notation is employed:

- \( \sigma \): joint type (1-revolute joint, 0-prismatic joint).
- \( \rho \): unit vector along the \( i \)-th joint axis, in the \( i \)-th frame.
- \( R_{i-1} \): rotational transformation from frame \( i-1 \) to frame \( i \).
- \( p_{i-1} \): position vector of the origin and of the gravity center of the \( i \)-th link with respect to \( (i-1) \)-th link frame.
- \( q_i, \dot{q}_i, \ddot{q}_i \): joint position, velocity and acceleration in the \( i \)-th frame.

Forward (kinematics) iterations for \( i = 1, 2 \ldots n \)

The forward kinematics phase, expressed in Equation (9), recursively computes the angular velocity \( \omega_i \), angular acceleration \( \dot{\omega}_i \), linear velocity \( v_i \) and linear acceleration \( \ddot{v}_i \) of each link in terms of its preceding link. The initial conditions for the base frame are \( \omega_0 = 0_{3 \times 1}, \ \dot{\omega}_0 = 0_{3 \times 1}, \ v_0 = 0_{3 \times 1}, \ \ddot{v}_0 = g \cdot k \) where \( g \) represents the gravity.

\[ \begin{align*}
\omega_i &= i^{-1} R_i^T \omega_{i-1} + \sigma_i q_i e_i \\
\dot{\omega}_i &= i^{-1} R_i^T \dot{\omega}_{i-1} + \sigma_i (\dot{q}_i e_i + i^{-1} R_i^T \omega_{i-1} \times (q_i e_i)) \\
v_i &= i^{-1} R_i^T v_{i-1} + \omega_i \times p_i + (1-\sigma_i) i^{-1} R_i^T (q_i e_i) \\
\ddot{v}_i &= i^{-1} R_i^T [\ddot{v}_{i-1} + \omega_i \times (\omega_{i-1} \times p_i) + (\dot{\omega}_i \times p_i)] \\
&+ (1-\sigma_i) (q_i e_i + 2 \omega_i \times (q_i e_i))
\end{align*} \]  

(9)

The linear acceleration of the gravity center of the \( i \)-th link, \( \ddot{v}_i^C \), is calculated by Equation (10).

\[ \ddot{v}_i^C = v_i + \dot{\omega}_i \times c_i + \omega_i \times (\omega_i \times c_i) \]  

(10)

Backward (dynamics) iterations for \( i = n, n-1 \ldots 1 \)

Once the velocities and accelerations of the links are found, the joint forces \( f \) and torques \( t \) are computed in the backward computation, starting from the end-effector link and ending at the base link. In this sense, the dynamic parameters for the \( i \)-th link are indicated in Figure 7.

The modified RNE formulation including the obstacle repulsion forces is defined in Equation (11), where \( F_i \) and \( T_i \) are the inertial forces and torques acting on the \( i \)-th joint due to the dynamics of \( i \)-th link. \( F_{\text{rep},i} \) is the repulsion force, obtained in the APF method, applied in the global point \( p_i^{\text{link}} \). \( f_i \) and \( t_i \) are the reaction forces and torques acting on the \( i \)-th joint taking into account the dynamic effect of previous links. Finally, \( \tau_i \) is the \( i \)-th actuator force or torque required to move the \( i \)-th link.

Figure 7: Definition of Dynamic Parameters for the \( i \)-th Link with a Rotational Joint
\[
\begin{align*}
F &= m \ddot{y} \\
T &= \Gamma \omega + \alpha \times (\Gamma \omega) \\
f &= R \tau + F + (R)^{-1} F_{\text{rep}} \\
t &= T + R \tau + (R)^{-1} p \times (R)^{-1} F_{\text{rep}} \\
\tau &= \sigma f e + (1 - \sigma) f e \\
\end{align*}
\]  

For the end-effector, \( f_{n+1} \) is not a reaction force but the total force acting on it, induced by the artificial potential field, as indicated in Equation (12).

\[
f_{n+1} = F_{\text{art}}(p) + F_{\text{link}}(\vec{p}) 
\]  

**SIMULATION RESULTS**

In order to validate the simulation proposals described along the paper, some results are displayed in Figures 8 and 9. The trajectory followed by the robot to reach the goal (in red) avoiding static obstacles is shown. Figure 8 displays a top view while Figure 9 displays a front view. We found out that this trajectory is almost the same in both platforms with little variation based on the different models used. Obviously, the Matlab computation is much faster than NX computation but the performance is better in the last case when very accurate movements are needed.

**CONCLUSIONS & FUTURE WORK**

In this paper, two different strategies have been described for implementing robot trajectory planning using artificial potential fields. The standard recursive Newton-Euler algorithm for robot dynamics computation has been modified in order to consider the forces produced by the APF approach in an iterative process.

The first strategy has been implemented in the NX platform, a standard CAD/CAM/CAE software. An accurate robot model is generated and minimum distances between the robot links and the obstacles are computed. The APF and RNE modules are integrated inside the NX platform. The second strategy has been developed in Matlab, except for a final simulation carried out in NX. Kinematics and spherical volumes are considered for distance computation while dynamics and artificial potential fields are considered for trajectory planning. In both cases, computed trajectories are simulated using the NX-Motion Simulation software.

We have described two different approaches: the first focusing on distance computation while the other focusing in the calculus efficiency. Now we are working on an intermediate approach avoiding the existing trade-off between accuracy and efficiency.

**ACKNOWLEDGEMENT**

This research has been supported by the Spanish Government project BIA2005-09377-C03-02, as well as by the Regional Government Ref. AINFO/2007/048 and IMPIVA Ref. IMPVB/2007/10.

**REFERENCES**


IMPLEMENTATION OF THE SYMBOLIC SIMPLIFICATION FOR THE CALCULATION OF ACCELERATIONS OF MULTIBODIES

Dmitry Vlasenko, Roland Kasper
Institute of Mobile Systems (IMS), Otto-von-Guericke-University Magdeburg
Universitätsplatz 2, D-39016 Magdeburg, Germany
e-mails: Dmitri.Vlasenko@ovgu.de, Roland.Kasper@ovgu.de

KEYWORDS
Symbolic decomposition, multibody dynamics

ABSTRACT
This paper presents a method of simplification of multibody dynamics equations by preprocessing based on the symbolic decomposition and multiplication of sparse matrices. The method was implemented in the Virtual System Designer (VSD) software for the simulation of dynamics of CAD systems. Simulation tests show that the symbolic preprocessing greatly increases the numerical efficiency of the simulation.

INTRODUCTION
Nowadays there are many methods performing the simulation of multibody systems. If simulated models are described using absolute coordinates, the equations of motion include large sparse matrices. Decompositions and multiplications of the matrices are the most numerically costly procedures in the simulation process.

The numerical efficiency of simulation methods can be significantly reduced if the sparse structure of matrices is taken into account. In the last years we developed a method for symbolic simplification of equations of motion based on preprocessing, which performs the object-oriented multibodies with complex structures and redundant constraints (Vlasenko and Kasper 2007;2). The symbolic simplification of decompositions and multiplications of matrices has several advantages in comparison with standard sparse solvers:

- Sparse structure of matrices is used completely without any run time overhead.
- The numerical operations with numerical elements of matrices are performed already during the translation.
- Additional operations with arrays of indexes (like in usual sparse solvers) are not needed.

In this article we show the results of the implementation of the method for the calculation of accelerations of multibodies. We developed in Maple a preprocessing module, which performs the symbolical simplification, and integrated it with our tool Virtual System Designer (VSD) for the object-oriented simulation of dynamics of CAD systems. Tests show that the integration of the preprocessing module with VSD greatly reduces the simulation time and the number of computations.

SYMBOLICAL SIMPLIFICATION OF THE DECOMPOSITION OF MATRICES

Let us consider a multibody system, consisting of rigid bodies, connected by holonomical constrains. The equations of constraints can be written as:

\[ g(q) = 0 \]  \hspace{1cm} (1)

where \( q \) is the vector of coordinates.

Differentiating this equation, we get the equations of constraints on the velocity level:

\[ G(q)v = 0 \]  \hspace{1cm} (2)

where \( G \) is the constraint Jacobian matrix, \( v \) is the vector of velocity variables.

Differentiating (1) twice, we obtain the equations of constraints on the acceleration level:

\[ G(q)\dot{v} = u(q,v) \]  \hspace{1cm} (3)

where

\[ u = -\dot{G} \cdot v \]  \hspace{1cm} (4)

Combining (3) with the equations of motion in descriptor form:

\[ M\ddot{v} + G(q)^T \lambda = f(q) \]  \hspace{1cm} (5)

we get the index-one formulation of the equations of motion (Eich-Soellner and Führer 1998; von Schwerin 1999) which can be used for the calculation of \( \dot{v}, \lambda \)

\[ \dot{q} = v \]  \hspace{1cm} (6)

\[ \begin{bmatrix} \dot{v} \\ \dot{\lambda} \end{bmatrix} = \begin{bmatrix} M & G(q)^T \\ G(q) & 0 \end{bmatrix} \begin{bmatrix} v \\ \lambda \end{bmatrix} = \begin{bmatrix} f(q) \\ u(q,v) \end{bmatrix} \]  \hspace{1cm} (7)

where \( f \) is the vector of external forces, \( M \) is the mass matrix, \( \lambda \) is the vector of Lagrange multipliers.

This linear system can be efficiently solved by sparse solvers, exploiting a block-sparse structure of matrices \( M \) and \( G \) (e.g. null space methods, range space methods) (von Schwerin 1999; Lubich et al., 1995)

We developed an algorithm, based on the QR-decomposition of matrices, which can be used for the simulation of mechanical systems with complex structure, including closed loops and redundant constraints. Let us consider it more precisely.

From (2) we get the system of equations...
\[ \dot{v} = M^{-1}(f - G^T \lambda) \]  
\[ GM^{-1}G^T \lambda = GM^{-1}f - u \]

Computing the Choleski decomposition of \( M = LL^T \), we obtain from (7) the matrix equation for \( \lambda \):
\[ A^T \lambda = b \]
where
\[ A = L^{-1}G^T \]
\[ b = GM^{-1}f - u \]

If the matrix \( A \) is linearly independent (e.g. all rows of the Jacobian matrix \( G \) are independent), then, using the QR-decomposition of \( A = Q \begin{pmatrix} R & 0 \\ 0 & 0 \end{pmatrix} \), we can calculate the value of \( \lambda \) as
\[ \lambda = R^{-1} (R^T)^{-1} b \]

However, in the case of redundant constraints \( G \) has dependent rows! The presence of redundant constraints in CAD models is not unusual. In many cases design constraints in CAD models are independent, using more constraints than it is needful from the mechanical point of view. The redesign of CAD models and the elimination of redundant constraints by engineer is very costly procedure.

From (11) follows that if \( G \) has dependent rows, then \( A \) has dependent columns. Consider now the calculation of solution of (10) in this case. Clearly, if \( A \) has dependent columns then the product \( A^T A \) is singular and the solution of (10) is not unique. In our case we need only an arbitrary solution with limited norm Performing the QR-decomposition with pivoting of \( A \), we obtain (Golub and van Loan 1996):
\[ A\Pi = Q \begin{pmatrix} R_1 & R_2 \\ 0 & 0 \end{pmatrix} \]
\[ \lambda = \Pi_1 R_1^{-1} (R_1^T)^{-1} \Pi_1^T b \]

where \( \Pi_1 \) is a part of the permutation matrix \( \Pi \): \( \Pi = (\Pi_1, \Pi_2) \). From (15) follows that we do not need to calculate the matrix \( Q \), but only \( \Pi_1, \Pi_2 \). Substituting the value of \( \lambda \) in (8), we calculate the absolute accelerations \( \dot{v} \).

Since using absolute coordinates, the matrix \( A \) usually has a sparse structure and includes both numerical and symbolic elements, e.g. the elements that are constant during the simulation and elements, depending on the coordinates of bodies. Therefore, the QR-decomposition of \( A \) can be optimized. We have developed a preprocessing module, which symbolically simplifying the QR-decomposition of matrices and for each decomposition generates a corresponding C-code.

This approach has the advantages pointed out in introduction. Generating C-code directly not only avoids calculating with zero elements, but also allows to preprocess all numerical parts of expressions. Indexing of matrices is avoided completely as line code is generated.

It is well-known that the numerical complexity of the QR-decomposition depends on the order of columns. That is why we decompose not the matrix \( A \), but the matrix \( \tilde{A} = A\Pi \) which is obtained from \( \tilde{A} \) by the reordering of columns.

We do not identify dependent rows in the matrix \( \tilde{A} \) on the preprocessing level because some elements of \( \tilde{A} \) are not constant. That is why we can get the situation when, substituting in our C-procedure the numerical values of elements of \( \tilde{A} \) on an arbitrary time step, we obtain the matrix \( \tilde{R} \) having zero elements on the main diagonal. We propose the following algorithm of the solution of this problem:

1. Using a C-procedure, generated by the preprocessing module, we obtain from the numerical value of the matrix \( \tilde{A} \) the upper triangular matrix \( \tilde{R} \) having zero elements on the main diagonal.
2. We permutate rows and columns of \( R \) in order to maximize the size of the non-singular upper triangular submatrix \( R_{1,1} \) and to minimize the size of the lower submatrix \( R_{2,2} \):
\[ P_1 R P_2 = \begin{pmatrix} R_{1,1} & R_{1,2} \\ 0 & R_{2,2} \end{pmatrix} \]
where \( P_1 \) and \( P_2 \) are permutation matrices.
3. We perform the QR-decomposition of the submatrix \( R_{2,2} = Q_2 \begin{pmatrix} U_1 & U_2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \Pi_{2,1}^T \\ \Pi_{2,2}^T \end{pmatrix} \). Then the matrix \( \tilde{R}_1 \) from (13) can be calculated as
\[ \tilde{R}_1 = \begin{pmatrix} R_{1,1} & \tilde{R}_{1,2} \\ 0 & U_1 \end{pmatrix} \]
where \( \tilde{R}_{1,2} = R_{1,2} \Pi_{2,1} \). The matrix \( \Pi_{1,2} \) from (13) is a combination of submatrices of \( \Pi_1, \Pi_{2,1} \).

**DOUBLE INSULATOR CHAIN EXAMPLE**

Consider a double insulator chain example (Hagedorn et al. 1980; Lubich et al., 1995; Vlasenko and Kasper 2007:1), shown in Figure 1. Each chain consists of insulators, connected by revolute joints. The first end of each chain is coupled with the triangular distance holder; the second is coupled with the ground. The holder is connected with the high voltage line, which is modeled as a force \( f_n \) acting on the holder.
Moreover, not only the procedure of the decomposition of matrices can be simplified. If we consider precisely the equations (4), (11) and (12), we can see that the matrices $L^{-1}$, $G$, $M^{-1}$ and $G$ are sparse. Therefore, the calculations of the matrix $A$ and of the vectors $b$ and $u$ need the calculation of products of sparse matrices. Taking into account the sparsity of the matrices during the calculation of the products, we can significantly improve the numerical efficiency of the simulation.

Since the matrices $G$, $G$ etc. are sparse, we can write them down in a compressed form, including only the symbolical parts of matrices. The compressed form of matrices can be calculated, using the following procedure. Consider an arbitrary matrix $K$:

$$K = \begin{pmatrix} 0 & k_{1,2} & 5 \\ k_{2,1} & k_{2,2} & 0 \\ 0 & 0 & k_{3,3} \end{pmatrix}$$

Then the dense form $DK$ of the matrix $K$ can be calculated as union of rows of $K$, excluding all non-symbolical elements:

$$DK := (k_{1,2} \ k_{2,1} \ k_{2,2} \ k_{3,3})$$

The use of the dense form of matrices reduces the number of memory, needed for the simulation.

**Preprocessing module**

We developed in Maple a preprocessing module which performs the symbolic optimisation of calculations. Starting from the mechanical parameters of a simulated system (e.g. masses of bodies, types and places of connections, etc.), the module generates a set of optimised C-procedures, which are then compiled in a .dll library. The library is called from VSD during the simulation of the mechanical system.

The module generates the following procedures:

1. The constraint Jacobian matrix $DG$ in a compressed form as a function of coordinates $q$.
2. The matrix $A = L^{-1}G$ as a function of $q$ and $DG$.
3. The product $M^{-1}s$ as a function of $q$ and an arbitrary vector $s$ (we call this procedure during the calculation of $b$ from (12) and during the calculation of $\dot{v}$ from (8)).
4. The product $Gs$ as a function of $DG$ and an arbitrary vector $s$ (the procedure is called during the calculation of $b$ from (12)).
5. The vector $u = -G \cdot v$ as a function of coordinates $q$ and velocities $v$.
6. The matrix $R$ as a function of matrix $A$, where $R$ is an upper-triangular matrix, obtained by the QR-decomposition of $A$: $A = QR$.

The functions, generated by the preprocessing module, are much faster than the standard procedures of multiplications and decompositions and need less RAM because they use the sparse structure of matrices.
MANIPULATOR MODEL

We tested our preprocessing module, simulating the dynamics of a CAD model of a spatial manipulator, shown in Figure 3. Each stiff connection between bodies is defined by three plane-to-plane joints like it is usually defined by design engineers during the development of CAD models. This leads to the redundancy of constraints, which should be taken into account during the simulation of the manipulator. The complete model includes 8 bodies connected by 3 revolute joints and by 12 plane-plane joints.

Figure 3. Manipulator Model

The matrix $A$ matrix from (10), corresponding to the model, is shown in Figure 4, where symbolic elements are colored black, zero elements are blank and numerical elements are non-blank. Its length is 41 and the width is 50. The use of sparse methods for the decomposition of $A$ is very effective because its density is 17.6%. The QR-decomposition of $A$ using the preprocessing module needs 34916 flops, which is 83% less than the number of operations needed for the decomposition of $A$ using a standard dense solver.

Figure 4. The matrix $A$ of the model

The simulation of the dynamics of the model is carried out for the first 2 seconds using the Runge-Kutta integrator with the fixed timestep of 0.01 seconds. The use of the .dll library, generated by the preprocessing module, allows us to perform the simulation in 0.52 seconds, which is 5.5 times quicker than the simulation without preprocessing analysis (here we show the time of the simulation of the unstabilized model. In VSD is used also the second preprocessing module for the stabilization of constraints in the projection methods, but its description is very extensive and sophisticated and will be considered in other articles). This result shows that the use of the preprocessing module enables in future the real-time simulation of complex multibody systems.

CONCLUSION

The method of symbolical simplification of decomposition of sparse matrices can significantly increase the numerical efficiency of the calculation of multibodies’ accelerations. The proposed method can be used for the simulation of multibodies with complex structure and redundant constraints.

The method was tested during the simulation of double chain examples. The tests results of the decomposition of matrices shows linear increase of the numerical operations using the preprocessing module vs. a cubic dependency of the dense solver.

The preprocessing module, simplifying numerical procedures with sparse matrices, was developed in Maple and integrated with the simulation software VSD. The results of the simulation of a CAD model of a manipulator show the high efficiency of the method.

REFERENCES


BIOGRAPHY

Dmitry Vlasenko (born 1977) graduated from the Novosibirsk State University, Russia with a B.A. in 1998 and from St. Petersburg State University, Russia with M.A. in 1920. He obtained his PhD degree in Mechanical Engineering from Otto-von-Guericke-University Magdeburg, Germany in 2006. From 2006 he works as a research engineer at University of Magdeburg, Institute of Mobile Systems.
MEDICAL ROBOTICS
VIRTUAL PROTOTYPING AND MULTI-MODAL INTERFACES TO TEST THE CONTROL OF AN ORTHOSIS

Aron Pujana-Arrese
Joseba Landaluze
IKERLAN Research Centre
E-20500 Arrasate/Mondragon, Spain
E-mail: {APujanaJLandaluze}@ikerlan.es

Jesus Gimeno
Marcos Fernandez
ARTEC Group, Valencia University
E-46980 Paterna (Valencia), Spain
E-mail: {Jesus.Gimeno|Marcos.Fernandez}@uv.es

KEYWORDS
Augmented Reality, IAD, Optical Motion Capture, Orthosis, Virtual Prototype, Virtual Reality.

ABSTRACT

The test of wearable robots with users can be dangerous, especially initially, when the control algorithms and the controller are tested. The use of virtual prototypes, as opposed to creating real ones, whilst not eliminating the need to manufacture an ultimate prototype for the final tests, eliminates intermediate tests which could potentially be dangerous for a user. Multi-modal interaction and virtual prototyping are used to test and evaluate the controllers. The Ikerlan research centre has designed a 2 DoF arm orthosis to research the issue of the interaction of such devices with the environment. This orthosis was not physically constructed, but a virtual prototype was created for testing the performance of the different control algorithms implemented. The multi-modal interfaces considered are composed of traditional devices, one haptic device (Phantom) and a video-based motion capture and mark recognition system. As a result, two test platforms have been developed, where the orthosis is tested in Real-Time simulation based on the real movements of the user. Augmented Reality (AR), Virtual Reality (VR) and video-based optical motion capture techniques are used in order to evaluate a prototype.

INTRODUCTION

An exoskeleton or “human amplifier” is a mechatronic system worn by a person which offers the user the possibility of enhancing their physical skills. The term IAD (Intelligent Assist Device) is used for a special kind of human amplifier, a biomechatronic device, the purpose of which is to help the user perform a daily activity which requires a certain amount of effort. The word orthosis can be used in the same context. Ikerlan has designed and constructed an upper limb IAD of 5 DoF (Degrees of Freedom) (Martinez et al. 2007), an exoskeleton designed to help the user perform a routine activity requiring effort in the workplace.

As a preliminary step, and above all to research the issue of the interaction of such devices with the environment, a 2 DoF arm orthosis was designed. This orthosis was not physically constructed, but a virtual prototype was created for testing the performance of the different control algorithms designed. For this purpose multi-modal interaction techniques were used after integrating them with the Virtual Prototype. These multi-modal interfaces combine classic interaction devices (the mouse and the keyboard) and haptic interfaces (Phantom), as well as a video-based motion capture and mark recognition system. As a result, the orthosis is tested in Real-Time simulation based on the real movements of the user.

An active orthosis can be considered as a manipulator which connects directly with the human user and his position and force control should be considered. Normally, the user creates the movement set-point values, after detecting his “intention” and closes the loop helped by the sensors the human body posses. There are however applications, in rehabilitation for instance, where the set-point values may be generated in a different way. In principal it is not possible to control the position and force simultaneously, although strategies have been developed to allow the position and the force to be modulated.

A classic in this field is the Hybrid Controller strategy put forward by Raiber and Craig (1981). The controller carries out its action using selection matrices which establish some spatial directions where position control must be carried out and some others where force must be controlled. In this way the force and position control actions are uncoupled by using the appropriate treatment of the spatial geometry where the manipulation task is being carried out. Different variations appeared later, which did require knowledge about the restrictions of the environment. Another classic strategy is Impedance Control (Hogan 1985), which does not control the position or the force but the dynamic relation between the two. The concept of Impedance control is based on the assumption that during all manipulation tasks the environment contains kinetic inertias and restrictions, i.e. systems which accept force as input and respond through displacements. This type of system is known as admittances. In turn, the manipulator in contact with the environment must adopt the behaviour of an impedance, i.e. a system which responds to a displacement by exercising a certain force on the environment. If the manipulator presents an inescapable inertial element, the representation of the system can always be modified to include this element in the environment modelling. This type of control
strategy is deemed to be very suitable for IADs, although it needs to be adapted depending on the specific application.

The use of virtual prototypes, as opposed to creating real ones, whilst not eliminating the need to manufacture an ultimate prototype for the final tests, eliminates intermediate tests with real prototypes which could potentially be dangerous for a user. With the goal of getting a more realistic use of the virtual prototype, multi-modal interaction will be used. In this interaction, different input modalities have been used (not only traditional devices mouse and keyboard). On one side the haptic interfaces such as Phantom, supply a force feedback, thus facilitating interaction testing with virtual environments. But other techniques based on motion capture enable real user movements to be used as an input for the virtual world.

Motion capture can be considered as the computer’s sense of sight, through which it is able to observe the movements carried out by the user. Whilst this is a fairly new technology, a large number of approaches have been developed in order to create the capture process, using various technologies: acoustic, electro-magnetic, mechanical... But currently those most commonly used are optical capture systems for three main reasons: accuracy, speed, and freedom of movement for the user. Within this group two sub-types can be considered: infrared optics and video-based optics. Infrared optics is more accurate but requires rather expensive special hardware (infrared cameras). Video-based optics works very similar to the sense of human sight, where based on the image, an attempt is made to capture the position and movements of the object under observation. The work carried out and presented in this paper aimed to create an environment to evaluate the correct performance of a virtual prototype of the orthosis, but on a real user. This is possible by applying a new interaction paradigm: Augmented Reality (AR) (Azuma 1998). The difference between Virtual Reality (VR) and AR (Lores and Gimeno 2007) lies in the way they treat the real world. VR immerses the user into a world which completely replaces the real world, whilst with AR the user interacts with a mixture of the real world and virtual information.

This paper describes, firstly, the orthosis designed, as well as its modelling on Dymola/Modelica. Continuing with, the description of the position/force algorithms implemented and some results obtained in simulation are presented. Then, the multi-modal interfaces used, Phantom haptic device and AR paradigm, are stated, as well as the virtual prototype and the test platforms carried out based on them. Finally, the test procedures for the orthosis are explained.

ORTHOSIS DESCRIPTION AND MODELLING

Orthosis description

In order to make an initial approach towards the final objective, the IAD or orthosis of 5 DoF (Martinez et al. 2007), it was decided to simplify the system by reducing it to two DoF. An orthosis was therefore designed with three main parts: the two which will move jointly with the arm and forearm and a third fixed to the body. In order to achieve the two DoF of the moveable parts, two rotating joints were used: one for the shoulder and another for the elbow. As a result, the movement of the arm is limited to a vertical plane. The third main component is a kind of backpack which will be secured to the user’s back, as shown in Figure 1, and will withstand all the efforts made by the orthosis.

The main function of the exoskeleton, in so far as it has been designed as a support device, lies in amplifying the torque with respect to the elbow joint exercised by the human muscles when handling loads. In order to guarantee this functionality, the device consists of two DC motors to motorise the two joints with which it is equipped. Graphite brush motors by Maxon were chosen. More specifically, the RE 36 Ø36 model, which is capable of supplying power of up to 70 watts when operating axially. Each drive is fitted with a gearhead and a tachogenerator also by Maxon: GP 32C Ø32 1-6 Nm gear with 190:1 reduction, and DC Tacho Ø22 mm and 0.52 V tachogenerator. A servoamplifier, Maxon ADS 50/5 is used for current and speed control.

The abovementioned drives are placed near the rotating axes, hidden inside the cylindrical pieces as shown in Figure 1. The forearm part of the orthosis weighs 1.94 kg, and its moment of inertia with respect to the rotating axis is 261 kg·cm². The arm part weighs 3.5 kg and has a moment of inertia with respect to the rotating axis of 820 kg·cm².

Given that this first orthosis design was created to research the interaction with the environment, a force sensor was fitted to the designed orthosis at the tip of the forearm cylinder, at the level of the user’s hand. This sensor was designed by Ikerlan for robotic manipulators and supplies the force exerted at the tip in the 3 spatial coordinates.

Modelling of the orthosis

Although the design described has not been physically constructed, it has been created virtually. In other words, a virtual prototype has been developed consisting of the
dynamic model of the exoskeleton and upper human extremity, and of the controller which deals with the positioning of the hand in the work plane and the force exerted by the orthosis against the environment in the event of collision.

The dynamic part of the model was developed in the Dymola modelling and simulating environment, which bases its operation on the Modelica modelling language. Both elements from commercially available model libraries as well as objects specifically designed for this application have been used in the composition of the model.

The model has been designed considering two parts. On one side are the axes commanded by the actuators and on the other side the three-dimensional mechanical components. As has been mentioned in the previous point the drives selected are identical for both degrees of freedom of the system. Therefore the modelling of the two axes are the same.

Figure 2 shows the model of an axis which comprises a DC motor, a gear, a regulator and a tachogenerator to measure the rotation speed. Figure 3 shows the typical model of a DC motor. It consists of the electrical part and the mechanical part. The mechanical part is finished off by the component for motor load modelling. This subsystem has the working voltage of the motor as an input while its output port informs about the current value on the circuit. Although the two typical control loops (an internal current one and an external speed one) have been modelled on the regulator, when the impedance control is applied only the current loop is used, given that the control law generates the torque set-point values for the actuators.

Figure 4 shows the complete dynamic model, where the subsystem of the axes previously described has been included twice to control the angular position of each DoF. The model brings together the parts from the top half of the human body starting from the trunk, and moving onto the neck, head and the shoulders from where the upper limbs hang. For modelling the different parts of the body three-dimension components from the Dymola mechanical library, in the shape of spheres and cylinders, have been taken into consideration, where the inertia values and masses have been fixed in accordance with some average values. In the case of the right limb, the mass and inertial values correspond to the human arm in addition to the orthosis components. The mechanical model is completely rigid with the exception of the shoulder and elbow articulations, which have been reduced to one DoF each. The articulations have been modelled using rotary joints included in the mechanical library, although only the joints corresponding to the right limbs are actuated and controlled by the motors.

In order to carry out robustness tests with the implemented controllers, an external force has been added to the right hand whose value depends on the weight of the load fixed during each test.

Finally the model includes an additional three-dimensional component which serves to implement a vertical wall placed at a configurable position. The interaction force between the orthosis and the wall has been modelled in Modelica language within the “collision” element. The triangular icons correspond to the input and output interfaces of the model. These interfaces serve as a connection between the different controllers implemented using Matlab/Simulink blocks and the model described in Dymola/Modelica and exported to Simulink as an S-function.

POSITION/FORCE CONTROL ALGORITHMS

The basic control of an orthosis type device is based on position control, where the user creates the motion set-point value and closes the loop with the help of the sensors possessed by the human body. A key element is the detection of the “intention” of the user for creating the motion set-point values from it. This characteristic, of great interest to researchers, is a fundamental difference between an orthosis and robotic manipulators. Another very important factor to consider is the interaction with the environment, especially from the point of view of controlling the force exerted in order not to harm people.
who may be present in the field of action of the robotic device. In robotic manipulators there are two classic types of position/force control: hybrid control and impedance (or admittance) control. It has been considered that these types are valid in the case of an arm orthosis, although they have particular characteristics, different from the case of robotic manipulators.

Hybrid control

A hybrid position/force controller has been implemented in the orthosis, following the diagram shown in Figure 5. It basically consists of two independent controls, one for position and one for force, and a supervisor, which depending on the contact with the environment, switches between one type of control and the other. The supervision is based on the information supplied by a force sensor in the orthosis, in this case at the end of the arm. In a more general case, a system distributed with force sensors throughout the body/casing of the orthosis might be considered.

The supervisor, in addition to the set-point generator, is responsible for making the transition between controllers “smoothly”, to avoid bouncing and to ensure the stability of the system. In the case of the orthosis being described, for each one of the axes the control position is a proportional control and the force control is a PI control. The switching between position and force control is not carried out in a simultaneous fashion on the two axes, but when moving with interaction, and depending on the relative position with respect to the contact surface, one of the axes carries out the position control and the other the force control. With this system the position of the orthosis can be controlled very accurately as well as the force exerted on the environment.

Impedance control

Another classic system, and very interesting in the case of an orthosis, is impedance control. Impedance control does not need a supervisor and is capable of taking control of a compound task, with phases of free and restricted motion, maintaining the stability of the system without changing the control algorithm. It is based on the idea of controlling the dynamic relationship between the force and position variables of the physical system. It assumes that in every manipulation task the environment contains inertia and cinematic restrictions, i.e., systems that accept forces as input and respond through displacements (admittances). In turn, the manipulator in contact with the environment must adopt the behavior of impedance and responds with a certain force to the displacement of the environment. The general strategy can be established in terms of controlling a motion variable and, in turn, equipping the manipulator with a response to disturbances in the form of impedance.

In impedance control, the functional form of the torque of the actuators of a manipulator is well known:

\[
\tau_{\text{act}} = I(\theta)J^{-1}(\theta)M^{-1}K \left[ X_0 - L(\theta) \right] + S(\theta) \\
+ I(\theta)J^{-1}(\theta)M^{-1}B \left[ V_0 - J^{-1}(\theta)\omega \right] + V(\omega) \\
+ I(\theta)J^{-1}(\theta)M^{-1}F_{\text{int}} - J'(\theta)F_{\text{int}} \\
+ I(\theta)J^{-1}(\theta)G(\theta, \omega) + C(\theta, \omega)
\]

where each line of the second member represents a contribution to the total torque of a different nature: the first line corresponds to terms dependent on the position, the second to terms of speed, the third to terms of force and the fourth to terms of inertial coupling. This equation expresses, in the field of the actuators, the behavior that the controller must be capable of inducing in the manipulator, in the form of non-linear impedance. The input variables are the desired Cartesian positions and speeds, and the terms, linear or non-linear, which specify the required dynamic behavior, characterised by the magnitudes M, B and K. Figure 6 shows the typical structure of impedance control, where the feedback gains of the position, speed and force loops, \( K_p \), \( K_v \), and \( K_f \) respectively, depend on the inertia and reference mass tensors, and on the stiffness K and damping B (design parameters), and are deduced from the control law (1).

The \( F_{\text{int}} \) force feedback has the effect of changing the inertia apparent in the manipulator. But an interesting characteristic of the impedance control system is that it can be applied without having a force sensor. In this case the force is not explicitly controlled, but depending on the impedance values used in the design of the controller, the force exerted by the system on the environment is limited.

As mentioned previously, these position/force control systems have certain special characteristics when there is a user wearing the orthosis and modulating the resulting behavior in the interaction with the environment. The
different techniques applied in this research (motion capture system, VR, AR, etc.) help to evaluate the behaviour and to optimise these control systems in the case of an orthosis.

Simulation results

To test the capability of the implemented position/force controllers different tests were carried out in simulation, varying parameters such as the hand trajectory or the situation of the wall. It was considered a reference system, which was situated on the shoulder axis so that the YZ plane remains parallel to the simulated vertical plane, while the XZ plane remains parallel to the floor.

Figure 7 shows the evolution of the coordinates of the tip of the orthosis with the hybrid control algorithm when moving from the lowest possible point, $(0, -57)$ cm, to a higher point, $(35, 15)$ cm. In this case the desired trajectory penetrates into the wall located at 40 cm on the X axis. Figure 8 shows the force exerted by the orthosis when interacting with the wall during the movement with the force set-point value fixed at 100 N. The first significant perturbation appearing both on the position and force graph corresponds to the moment of impact with the wall. After the initial impact, the force exerted presents an even greater perturbation due to the force controller switching, so that the shoulder articulation goes on to be force controlled and the elbow, position controlled. The third discontinuity and the least significant for the force is due precisely to the contrary switching, with the elbow articulation being force controlled and the shoulder position controlled. Finally, there is a small jump in the position before the final set-point values are reached which corresponds to the moment of separation from the wall. From this moment on both articulations are position controlled until the movement finishes.

As has been mentioned in the previous point, the impedance control philosophy is completely different given that the aim is not to explicitly control the force nor the system position, but the dynamic interaction between the system and its environment. As Figure 9 shows the evolution of the position coordinates of the tip of the orthosis together with its set-point values. In a first stage the tip is driven to the lowest position $(0, -57)$ cm. After a couple of seconds a new trajectory is generated with the aim of positioning the arm in the point $(10, 50)$ cm. The tracking error is minimum until the impact with the walls situated at 40 cm on axis X occurs. At this moment the position error increases
coinciding with the increase in the interaction force graphically shown on Figure 10. Noteworthy is the fact that the results shown correspond to a simulation obtained with a controller implemented without force feedback. Supposing that a force sensor is available, the interaction force is explicitly controlled.

MULTI-MODAL INTERFACES AND AR

Multi-modal interfaces are those where more than one interaction channel is used with the computer. This research uses traditional devices such as interaction channels (keyboard and mouse), a haptic device (Phantom) and motion capture based on video (ARToolKit 2008).

An interaction paradigm is an abstraction of all the possible interaction models organised in groups with similar characteristics. There are at present 4 paradigms: the desktop computer, virtual reality (VR), ubiquitous computing and augmented reality (AR). The main difference between AR with respect to VR lies in the way they treat the real world. The VR systems aims to submerge the user in the virtual world, while the AR let the user see the real world around them and increase this vision adding virtual information, therefore, it brings virtual information to the real world.

In the case under study, both VR and AR techniques have been used in two different test platforms. The AR enables information to be taken from the virtual prototype to the real world and interaction in this virtual world or with virtual options. This is done using user movement as a data source for the movement of the orthosis virtual prototype, and by visualising the performance of the orthosis on the real user, adding a 3D model of the prototype to the video captured of the user in real time (non-immersive AR).

**Phantom haptic interface**

The Phantom is a haptic joystick made by the North American firm SensAble Technologies. More precisely, the Phantom is a product line of haptic devices makes it possible for users to touch and manipulate virtual objects. In the case of this research the Phantom Desktop device was used. It provides precision positioning input with 6 DoF positional sensing and high fidelity force-feedback output. Portable design and simple parallel port interface ensure quick installation and ease-of-use.

On the other hand, SensAble offers the OpenHaptics software toolkit. This toolkit enables to add haptics and 3D navigation to specific simulation and visualization applications. The OpenHaptics is patterned after the OpenGL API, making it familiar to graphics programmers. Using the OpenHaptics toolkit, it is possible to leverage existing OpenGL code for specifying geometry, and supplement it with OpenHaptics commands to simulate haptic material properties such as friction and stiffness.

Figure 11: Marks and angle calculation

The OpenHaptics toolkit includes the Haptic Device API (HDAPI), the Haptic Library (HLAPI), utilities, PHANTOM Device Drivers (PDD), and source code examples from which start developing new applications.

**Video based motion capture and AR techniques**

Video based motion capture consists of applying computer vision techniques, in order to recognise and obtain both the position and the orientation of the objects or people reflected in the images. This problem, which may at first seem simple, is very costly computer-wise and in certain cases remains unsolvable. In order to help resolve it, marks are often used, which tend to be flat with a specific shape (square, circular, etc.) with an image printed on them (Figure 11). With these markers, both the shape and the image are known in advance, thus enabling the vision problem to be assessed on recognising these marks, and not any object or person, obtaining a much simpler process of recognition and calculation (position and orientation). This type of capture system is perfect for use in AR systems, as it minimizes the problem of coherence (registration) between the virtual information and the observed images.

The main advantage of these kinds of capture systems is the user’s freedom of movement, as he does not have to wear wired sensors, but simply some small cards with the markers printed on them. However, these kinds of systems also have significant optical type disadvantages, being their main problem the concealment of the markers. If the user covers the markers with his movements, the capture system cannot see them. This problem is partially solved by calculating the possible position and orientation of the hidden marker by using its previous data and assuming that it will follow a similar trajectory. One commonly-used technique in this respect is the implementation of Kalman filters.

**VIRTUAL PROTOTYPE AND TEST PLATFORMS**

Based on the dynamic model and the controllers presented in previous sections, a Virtual Prototype has been developed consisting of the dynamic model of the exoskeleton and upper human extremity, and of the controller which manages the positioning of the hand in the work plane and the force exerted by the orthosis against the
Figure 12: Virtual Prototype and Phantom haptic device

environment in the event of collision. The prototype is completed with the integration of multi-modal interfaces.

A Simulink model comprising the controller and the physical part imported from Dymola/Modelica was modified to obtain a model suitable for execution in real-time. Some functions were added for the communications via UDP with the multi-modal interfaces.

Based on the aforementioned Simulink system, and using the RTW tool from MathWorks, an executable application was obtained which was downloaded onto an xPCTarget compatible PC/104 platform. The simulation is carried out with a fixed step integration algorithm of 1 ms, and communicates with the different multi-modal interfaces whose functions are, first, generate the position set-point values for the tip of the orthosis and, second, guarantee a sensorial feedback between the user and the prototype.

Using diverse aspects of the multi-modal interfaces two different platforms were implemented with the aim of validate all control features. The first one is based on a haptic interface and a virtual 3D representation of the orthosis and the environment. The second system uses motion capture and AR techniques in order to generate the position trajectory and visualize the resultant movement.

Phantom haptic interface based test platform

As has been mentioned before, the Phantom is a haptic device which boasts a total of six degrees of freedom, as well as force feedback. With the help of the OpenHaptics toolkit only two of the Cartesian directions and their corresponding movements are integrated into the virtual prototype of the orthosis as its movement is restricted to one plane. The Phantom takes charge of defining the position set-point values on each movement, allowing the user to control the arm movement at all times. On the other hand the force feedback enables the sensation of colliding with the virtual environment or that of supporting a weight at the hand. However the position set-point values generated with this device are not those natural to a human hand.

The system is completed with a graphical model that includes the orthosis, a human body and a vertical wall. Several two dimensional plots are also added to inform about the temporal evolution of the most significant variables. The link between the virtual world and the real-time simulation is established using the Matlab’s tool named Virtual Reality toolbox. In this way, every mobile part of the virtual scene shows a position and an orientation according to the data incoming from the dynamic simulation. Figure 12 shows an example of handling on the Virtual Prototype.

Motion capture based AR test platform

The movement capture used is the video based optical type. To this effect, ARToolKit (2008) meets the requirements stipulated for this study. ARToolKit works by detecting and capturing both the position and the orientation of flat square markers (Figure 11). These markers always adopt the form of a black square, on a white background, with an identifying image inside. Two markers are placed on the user’s arm to capture the information from it.

For the visualisation offers 3D drawing capacity, but they are very basic, simply enabling the designer to draw by using OpenGL (a low level graphics library) or VRML (a high level graphics library). It has therefore been considered that the OpenSceneGraph (OSG 2008) high level drawing library offers greater capacity, mainly due to the fact that it is based on the use of a scene graph. Consequently, the visualisation part has been developed by using this library. Although OSG has no software for AR applications, there are some approaches, such as the OSGART library, which combine the ARToolKit’s AR capacity with the visualisation power of OSG. In the case in question, a new OSG node has been created, JexAROSGTexture, which is responsible for showing the images captured by ARToolKit in a texture. This node is simply included in the OSG scene graph and is always painted in the background of the application. Finally, to make the virtual information (orthosis model) coincide with the real image (user), the OSG camera is configured so that it coincides with the camera values captured by ARToolKit. In this way, the problem of “registration” between the virtual information and the real information is reduced, as the system is ultimately drawing on the same image from which the calculations have been obtained.

TEST PROCEDURES AND RESULTS

Both described approaches are perfectly valid for testing both the position control and the force control ones. In any case, the system fitted with the Phantom has been applied mainly for the start up of the controller part related to the interaction with the environment. As can be seen in Figure 13, the behaviour of the orthosis with the position/force algorithms can be seen on different graphs, obtaining similar results to those shown in Figures 7, 8, 9 and 10.
From the moment of the impact the user can perceive on their hand the interaction force regulated by the controller.

For testing the AR system, a MFC based application on Windows platform has been developed, in which 5 views have been included using OSG, as well as some view configuration dialogues. Figure 14 shows the typical screen. In the first of the views, the view of the virtual prototype on the real user in the virtual environment (with or without wall) is shown, following the movements calculated by the control system. In the two right-hand views a virtual human is shown with the orthosis in one of these moving in accordance with the set-point values generated by the capture system, and in the second with the values calculated by the control and dynamic simulation subsystem, according to the interaction environment and the chosen position/force control algorithm. These two views enable the user to visually compare the performance of the control, as with this interface the feedback from the force exerted is replaced by the visual feedback. Finally, in the fourth and fifth views, graphs of the positions and of the force exerted in the case of interaction with the virtual wall are shown.

CONCLUSIONS

The virtual prototyping is a great help when testing new control systems related to safety and comfort. In the case of an orthosis or a wearable robot, the first controller test trials can be dangerous for the user as they are carried out on a real prototype. The integration of multimodal interaction and virtual prototyping, such as haptic devices and AR systems based on motion capture, enable testing of the controllers to be carried out in more natural conditions, similar to those of the user themselves interacting with the real system.

To investigate the interaction control techniques of an arm orthosis with the environment and as a simplified case, a two DoF orthosis has been designed, where classic position/force control strategies have been implemented. The integration of some multi-modal interfaces, such as a haptic device or a motion capture based AR system, enables controllers to be tested. Two test platforms have been made and evaluated: one based on the Virtual Reality toolbox from MathWorks and on the Phantom haptic device, and a second one which combines optical motion capture based Virtual Reality and Augmented Reality. With the first one precise force feedback is obtained, and with the second one, the AR system permits the generation of the movement set-point values by a real user, in natural movements.

ACKNOWLEDGMENT

The material used in this paper was partly supported by the Spanish Ministry of Education and Science and European FEDER Fund (research project DPI2006-14928-C02-01).

REFERENCES

MOVEMENT RECONSTRUCTION IN INDUSTRIAL AND MEDICAL DOMAINS BASED ON TOPOLOGICAL PETRI NETS

Markus Koch, Robert Beckebans and Jürgen Schrage
Cooperative Computing & Communication Laboratory
University of Paderborn, Siemens AG
Fuerstenallee 11, 33102 Paderborn,
Germany
E-mail: {mkoch,trebor,schrage}@c-lab.de

KEYWORDS
Movement analysis, sensor mapping, topological Petri nets, robotics.

ABSTRACT
This work shows a methodology of mapping real-world distributed sensor information to a simulation and evaluation framework. The methodology bases on Petri nets which are used to model the topology of different application examples scaled from industrial to medical domains. For an industrial example a robotic arm with optical sensor elements is shown. A medical application forms a human spine sensor wear. A third application forms a data glove. The different application domains and sensor topologies are mapped by a topological Petri net approach. The evaluation of this methodology and experimental results will show the feasibility of our approach.

I. INTRODUCTION

Movement detection or movement capturing is a vital requirement in different domains. Starting from movement detection in movies [1] for character animation up to applications in industrial [2,3,4,5] or even medical domains [6,7].

The different application domains use preferred sensor systems. The different sensor features require special sensor evaluation algorithms to reconstruct the sensor data to movements of the viewed body.

Optical based systems mainly rely on vision algorithms to reconstruct the posture or movements [1]. This solution for example requires an external camera apart from the tracked object. A solution which is attached closely to the moving object and follows the movements without the requirement of external laboratory setup would be desirable. In this work we show a sensory system which is suitable to different application domains and is based on fiber-optical cords [8].

The cords consist of different sensitive areas which need to be attached to the target system skin. The different application domains require different sensor configurations and so the topology and also the correlation between them depends on the physical objectives of the whole system. So an universal applicable description is needed for this special kind of sensory. The methodology should be able to specify the different sensor topologies and also be able to embed the different reconstruction and interaction functions.

II. RELATED WORK

To arrange this work in the related work we focus here on movement detection and body tracking solutions and on the Petri net basics needed to major our topological Petri net approach.

A. Movement detection and body tracking

Different solutions for motion capturing and movement detection are available. There are four major sensory principles: vision based, ultra-sonic-based, resistor-based and fiber-optical [1,3,4,5]. They all have their pros and cons. For example the vision-based solutions use marker on the moving body and recalculate the positions by evaluation of special oriented camera systems. This solution is successfully used in different animated videos and medical problems. The major problem here is the requirement of a special laboratory with high accurate adjusted cameras. So this solution is not suitable for mobile usage and requires marker on the object. The ultra-sonic-based solutions use ultra sonic sources which are distributed and mounted on the traced object. But this solution also requires a laboratory equipment to evaluate the different active elements to calculate the positions based on runtime differences. So also these solutions are not really suitable to mobile applications.

The third major principle bases on resistors where the bending of the sensor element results in a change of the resistance of the element. But unfortunately mostly only the bending can be measured but not the orientation of a
positive or negative bending. Despite the fact that resistor-based solutions can be used in mobile domains, they cannot be used in strong electromagnetic environments like e.g. MRT-scanner. The last solution is based on fiber-optical cords, which might be used like resistor-based solutions but is able to detect also the directions of bending. Because fiber-optical solutions are produced without any metallic components on the sensor side they are capable to be used in electro-magnetic environments without interferences. Also this solutions have a property which we have to deal with like shown in this work. The fiber-optical cords are not compressible and so if they are placed on the back of a person for example, they will slide on the back if the person is moving. So a mapping between the sensor element and the actual position on the back of the person needs to be done. Here topological Petri nets can be used to aim this goal.

B. Petri Nets

Since the first introduction of Petri nets by Carl Adam Petri [9] there are many different enhancements from simple uniform token based nets to high-level Petri nets. Petri nets are used mainly to model distributed and parallel systems. Different analyses have been applied to this methodology to ensure different features of a modeled net. So Petri nets are widely used in academics and also industrial applications. For example Petri nets are used to model the behavior of an railway crossing to analysis the optimal waiting time for cars in front of a crossing [9]. To long waiting time results in accidents caused by driver illegal cross the railway. And on the other hand too short waiting time will result in accidents caused by cars stopping or waiting on the railway. In this work we make use of the ability to model the topologies of the sensor elements with help of High-level Petri nets. The transitions of the net are than used to recalculate the mapping of the sensor elements to the traced object.

III. APPLICATION SCENARIOS

To motivate the design of a topological Petri net methodology here we introduce three different applications where the fiber-optical cords were applied. The applications are taken from industrial, medical and computing domains to show the transferability of the sensors and also the evaluation software suite shown in detail afterwards.

A. Robotic arm

We used a common robotic arm which did not have any sensory elements and applied fiber-optical sensors to track the real movements.

To visualize the reconstructed movements a robotic simulation is used. The arm has different fiber-optic sensors which consist of separated sensing elements. So the whole robotic sensory forms a complex topological network which needs to be evaluated to get all bending and rotation information. Furthermore the gripper is equipped by a newly developed tactile fiber-optic solution to be sensitive an object which is in the gripper and can sense the gripping forces [6]. The robotic application domain is shown in figure 1.

Figure 1: Robotic arm in real world and simulation

B. Spine sensor equipment

Another possible application of such movement detection is a human spine movement scenario [7]. In this case we place the sensory on the human back. Because even this placement is very difficult we developed a spine wear suite. Which enables us to place the sensor wires accurate on the back and the medical test person can easily dress the shirt like casual wear.

Figure 2: Spine sensory in real world and spine simulation

Caused by the incompressibility of the sensor wires the
sensors slip on the back during flexion of the person. Such a scenario is illustrated in figure 2.

C. Data glove

The third field of application is a device to sense hand and finger movements [4]. The fibers are placed on a finger shown on the left side of figure 3.

![Data glove in real world and in simulation](image)

So the sensory information can be used to reconstruct the movements in virtual reality shown on the left side of figure 3.

IV. TOPOLOGICAL PETRI NET APPROACH

Like mentioned in related work Petri nets provide the opportunity to model distributed systems like shown in [11,12]. Here we utilize a topological approach to map different topological distributed sensors to a consistent evaluation methodology.

A. Petri Net topology mapping for application examples

All of the shown application scenarios use fiber optical cords in different configurations. In the Data glove example the cords are placed on each finger separately. In this case the sensor segments of two different fingers do not interference. Only the slipping of the sensor segments needs to be considered in our mapping and evaluation methodology. For this configuration a mapping onto a topological Petri net is shown in figure 4.

![Data glove Petri net topology mapping](image)

The different traces are independent and only bending information are propagated in each trace to remind the slipping.

![Spine sensor Petri net topology mapping](image)

The spine sensor application needs another topological mapping. Torsion of the human back influences both sensor wires. So a kind of crosslink needs to be established. A layout for handling slipping and cross linking is illustrated in figure 5.

![Robotic arm Petri net topology mapping](image)

A combination of the two layouts is needed in the robotic application example. Here some sensor wires are mounted on the arm likewise the finger application and some sensor segments need crosslink computations. Or even more complex topologies for new invented optic-tactile gripper equipment. The resulting topology is shown in figure 6.

B. Workflow

Figure 7 shows the interaction workflow of the developed Petri net simulator for topological mapping and the spine software. The Simulator enhances the existing framework with the required topological mapping functionality. The framework is responsible for the sensor data acquisition, calibration, analysis and visualization.

The workflow starts with the activation of a sensor view and triggers the simulator to load the appropriate topological Petri net. The sensor values are read and placed onto the appropriate Petri net places of the instantiated net. The evaluation of the Petri net then calculates the mapping of slipping and cross linking if applicable. The results are send back to the framework. This is established by specialized result places. These results are than used to visualize the movements in this example.

The last step usually also includes an analysis and
evaluation step. To extract movement features. The human spine application for example uses this step to extract constrained movements which could possibly be caused by a disease.

![Workflow of petri net tool – Spine Framework interaction](image)

**C. Petri Net Simulator**

For the purpose of modeling sensor slipping and cross linking a Petri net simulator with the capabilities to be used in conjunction with our software suite is established. The simulator is instantiated and triggered from the main software framework. The simulator can create, load, modify and save the Petri net topologies for the different application examples. The saved nets can be used in more specialized Petri net analysis tools to ensure Petri net specific characteristics like circle-freeness. This analysis is not needed for the examples we focus on.

The simulation of the Petri net can be shown visual step-by-step for debugging purposes. To get a better performance the visualization of the net and the whole simulator can be switched off. This ensures the maximal possible throughput of the system.

**V. EXPERIMENTAL SETUP**

For the evaluation of the concepts a reduced experimental setup shown in figure 8 was used [6]. To proof the concepts we used one sensor cable with three segments.

![Experimental setup](image)

**VI. EVALUATION**

The experimental setup results are similar to the data glove example with just one finger. Snapshots of the simulation steps are illustrated in figure 9.

![Petri net evaluation](image)
Figure 9b: Petri net evaluation

The first picture is taken at beginning of the sequence and shows the initial Petri net with initialized input places. The place_0 contains a start trigger token. Place_1, place_2 and place_3 contain the sensor signals of three sensitive sensor areas. The initial marking is placed by the spine software framework into the simulator. The second snapshot shows the net after firing the transition_0. In this case transition_0 is responsible for the first slipping correction. Because the first segment is not affected by slipping the sensor data is handed over to the intermediate result Place_4. For the next sensor segment transition_1 calculates the slipping with respect to the intermediate result and the sensor segment data of the second segment hold on place_2. This continues until the results of every sensor slipping corrections are calculated.

VII. CONCLUSION

The different application examples show the wide range of possible fields for fiber-optic sensors. The different sensor configurations results in different mapping opportunities. The slipping and cross linking are two kinds of interferences shown in the three examples. For the required evaluation and mapping feature a topological Petri net methodology is used. The workflow shows the interaction with our evaluation software. Experiments show the functionality of the approach for a single fiber. The slipping correction is calculated by an applied Petri net topology.

APPENDIX

The products and trademarks of the mentioned related work underlay the trademarks and/or copyrights of the owning companies and are here only referenced.

REFERENCES

[2] M. Koch, R. Beckebans, J. Schrage, W. Richert: Real-time measurement, visualization and analysis of movements by fiber optical sensor applied to robotics, 47th IEEE Conference on Instrumentation, Control and Information Technology (SICE 2008), Tokyo, Japan, August 2008, accepted for publication
COMMUNICATIONS SYSTEMS
Expansion of OPNET Modeler's SIP Model for Performance Evaluation of Hierarchical Call Routing in NGN

Hajar Atrianfar
Zohreh Ayatollahi
Mohsen Sepasi
Communication Technology Faculty, Iran Telecommunication Research Center (ITRC)
Tehran, Iran
E-mail: hatrianfar@yahoo.com, z_ayat@itrc.ac.ir, m.sepasi@itrc.ac.ir

KEYWORDS
Performance Modelling, NGN, OPNET, SIP, Call Server, Call server, Gateway, Hierarchical Call Routing.

ABSTRACT
This paper introduces a hierarchical route management simulation model in OPNET Modeler to allow performance evaluation of different scenarios proposed for interconnections between different administrative levels of control segments - provincial and national- of Next Generation Network (NGN) in call routing procedure. The provided model is an enhancement over the original Session Initiation Protocol (SIP) model existing in the standard model library of OPNET Modeler to allow simulation based on the hierarchical routing, featuring a full implementation of provincial and national call servers task definition in routing various types of calls generated throughout the entire network. This model encompasses multi-domain routing capability that is not considered in existing SIP model. Additionally, this model is capable of implementing various numbering procedures legislated in the network. One of its key aspects is introduction to the concept of location server, which can be used as a feature to define various routing plans in different scenarios.

INTRODUCTION
However, it is an inevitable trend that the present situation of various networks running independently will converge on the Next Generation Network. In the first release of NGN architecture issued by the ITU, two architecture are proposed for service control support: call-server based and IMS based. Notwithstanding providing more enhanced services and ubiquitous applications, IMS hasn’t been deployed by network operators and telecom equipment vendors in large scale and its reference points hasn’t been fully standardized. So local service providers benefit call-server based one for the transition of their networks from the traditional Public Switched Telephone Network (PSTN) infrastructure to an NGN architecture. The main responsibility of call server -the control unit in NGN- is call routing. Accordingly, the technology of organizing networks and Call Routing become an important issue to be studied.

On the other hand, operators face so many challenges to the network planning nowadays. In particular, Quality of Service (QOS) management presents a big complexity to an efficient network operation, requiring real-time decisions to keep QOS commitments. In order to solve these problems and validate their decisions, simulation support for performance evaluation is needed. The main objective of modelling tools is to address the network providers towards the correct strategies and the most cost effective ones to be chosen when evolving to NGN. Therefore providing a detailed simulation model for NGN signalling protocols is a considerable requirement in NGN design, implementation and performance modelling. OPNET's integrated analysis environment is one of the most known network simulators intended for design, model and test the NGN architecture and innovate unique mechanisms. In this paper, we focus on the implemented SIP protocol in the existent OPNET modeler and propose an enhanced model to remove available limitations and adding more capabilities in accordance with routing management between different hierarchies of routing agents.

CALL ROUTING
According to ITU-T E.170, the call routing should be divided into two aspects, that is, routing architecture and routing plan. The routing architecture consists of hierarchy and nonhierarchy of call routing (ITU-T E.170 1992).

Within national networks it is often appropriate to adopt a hierarchy of switching units -e.g. local, area, trunk, regional trunk, international- with each level of the hierarchy performing different functions (Zhengguang et al. 2003).

Call routing on call server
A call server as the routing core of NGN have to support various signalling protocols. Today, call routing in TDM network is based on a prefix of the dialed number in E.164 format. However, it will take a long time for the present networks to evolve into NGN. Therefore, in the transition period, call server has to coexist with traditional circuit switching and E.164 numbering plan is used very widely not only for legacy PSTN switch, but also for call servers (Lim et al. 2007).

Call servers usually keep prefixes related to media gateways or phone numbers covered by access gateways in their routing tables depending on Class 4 or Class 5 switches have been replaced. The goal is to determine how the call can be routed towards the callee. If the dialed number is related to registered gateway in the same Call Server (CS), routing is simply done. Otherwise, originating CS should be aware of prefix allocation to the gateways under control of terminating CSs. Also, all the CSs should be kept informed about any changes in end-to-end routing information or policies in the network as quickly as possible. These operations must be encountering considerable difficulties in large networks (Ohrtman 2003; Rosenberg et al. 2002).
Various scenarios for call routing

There are two schemes considered for routing procedure in large NGN networks (Ohrman 2003):

- All the routing information should be built into each CS.
- Utilizing an individual server to share call routing information for limited number of carriers. So each CS needing routing information out of its region should refer to it. As shown in figure 1, the hierarchy of routing agents is very similar to that of circuit switching, so these locating servers are located at upper hierarchies. Routing Agent divided into two levels, that is, at provincial level and at national level. The Routing Agent at provincial level provides all routing information related to gateways of its own province. The Routing Agent at national level maintains and synchronizes Routing Agents under its coverage at provincial level.

The main benefit of the first scenario is existence of the whole information necessary for call routing in the originating CS. This scenario results in a lower call setup time. Also some disadvantages are enumerated as follows:

1. As the number of communicating regions increases, further information is needed to be provisioned in provincial CSs. So the dramatic increase in quantity of information greatly speeds up by growing in communication zones.

2. Whenever any changes occur in network routing specifications, it should be publicized in all CSs. This function is performed manually or different routing agents can broadcast by means of TRIP protocol to synchronize the newest routing information that has been added, deleted and updated. Existing delay in manual operation may cause errors in some call routings. On the other hand, TRIP is not fully supported by different vendors.

The advantage of second scenario is that routing information is centralized and therefore its amount has been lowered in CSs. This simplifies data updating and processing time in each server decreases. But as a disadvantage, exchanging messages between network servers at different hierarchies to find out destination and subsequent call route may affect the network performance and increase call setup time (Eyers and Schulzrinne 2002).

In a real network, we have CS at national and provincial control stages that are connected to Public Land Mobile Network (PLMN), International Switch Center (ISC) and PSTN network elements through provincial/national signalling/media gateways.

Figure 1: General Schematic of National and Provincial Control in NGN

In designing and optimizing number and interconnection of control nodes and interfaces among existing and new NGN network, we can find the best model by simulating proposed routing scenarios between CSs and analyzing obtained performance. So simulation support for hierarchical routing between routing agents is strongly needed. Furthermore, in any performance analysis, there should be a real implementation of signalling protocol in simulation tool. This is necessary to reach actual simulation results and measurements. Normally, the signalling protocol between CSs is SIP and CSs acts as SIP proxy server.

EXISTING SIP MODEL

In the case of Next Generation Networks, the modelling methodologies are much less developed, mainly because an inherent complexity of these networks. The OPNET Modeler software package is among the most popular and most comprehensive tools available for modelling new communication technologies and protocols (Hnatyshin et al. 2007).

OPNET Modeler obtains SIP as the signalling protocol for call routing in NGN network modelling. SIP is an application layer text based signalling protocol. It is used to establish and tear multimedia sessions (Rosenberg et al. 2002).

Currently, SIP support is so weak in OPNET as the other simulators. The main functional limitation existing in available SIP model is disability to define various domains which are managed independently through their own sip proxy servers. In other words, available SIP model does not allow the interaction among more than one SIP proxy between two user terminals and hence it is not possible to arrange SIP signalling flow between relating proxy servers and recreate multidomain hierarchical routes scenarios as it happens in real network. Moreover it is not possible to distribute routing information and policies in CSs of different hierarchies.

However, developing or expanding simulation models is a quite complex and challenging task. In this paper, we attempt to demystify the route of SIP signalling messages which leads to call establishment in OPNET as well as enumerate the key features implemented in new developed model.

PROPOSED MODEL

In this section, new approaches are proposed to eliminate the existing model limitations and synchronize it with the NGN configuration modelling requirements. In order to solve mentioned problem, 4 modifications are introduced as follows:

1. Modifying the Application Model existing in OPNET Modeler in order to set the Destination Preferences attribute in the initiating AG as the real phone number including appropriate pre-codes instead of IP address. Consequently, it differs the call initiator be in the same city as the call invitee or not. This change helps us to model the actual calling procedure occurring in the network, which is helpful especially in digit analysis of sip servers.

2. Introducing new model attributes such as Domain and Area for call agents and Domain and CSLevel for sip servers in order to assign a group of users to a certain domain and to
relate each server to a certain domain and level of server. So we can distinguish between national and provincial CSs introduced in different scenarios. The term Area in call agent attributes refers to the national CS that is serving the user. The new attributes of the model are shown in figure 2.

![Figure 2: SIP Attributes of Call Agent in Provided Model](image)

3- Introducing of Processing Delay and Diameter Delay concepts for sip servers equivalent to process delay control for each SIP message and location server queries delay control, respectively. The idea of this feature is the same as one introduced in (Hernandez et al. 2006).

4- Location Server attribute for SIP servers. This attribute is presented as a table with 3 columns of Call ID, Next Domain and CS Level. Call ID is the minimum digits of call number which is needed to be processed in national or provincial CSs in order to judge the next node in the signalling route. Next Domain and CS Level defines the next domain and level of CS(national or provincial) which call should be forwarded to. This attribute is set as PCS for Provincial Call Server and NCS for National Call Server. In cases which the next node is a Media Gateway (MG), CS Level explain the national or provincial level of MG and when the next node is callee, this attribute has no value. The importance of providing this attribute in addition to domain and area is the capability of modelling different routing scenarios, because routing is not always done only through provincial CSs and sometimes national CSs intervenes too. Secondly, incoming traffic to PSTN from NGN network could be from provincial or national MG depending on whether or not the call has originated inside the same province. Additionally digit analysis in CSs is modeled. As the other words, Location Server is added as a considerable feature to the SIP model in order to model the Hierarchical Routing through setting this attribute in all scenarios. The new attributes of model are shown in figure 3.

![Figure 3(a): Sip Server Attributes in Modified SIP Model](image)

![Figure 3(b): Location Server Attribute Introduced in Modified SIP Model](image)

In next section we concentrate on SIP-related process models which are invoked during call setup and enhance the functionality of them in order to support multi-domain hierarchical route management.

**OUR MODIFICATIONS TO SIP PROCESS MODELS**

The OPNET Modeler includes a hierarchical structure of models consisting of a network model, a node model, and a set of process models. The complete specification of an OPNET process model consists of a finite state machine, action statements expressed in C/C++, and configurable parameters (Hnatiyshin et al. 2007).

Like all other subsystems in OPNET models, processes are event-driven, driven by events. When an event is actually delivered to a process, it is termed an interrupt; the process is said to be interrupted which means that it is invoked to allow it to take some action in response to the interrupt. Considering this framework, we try to modify the contributed process models in OPNET.

When a voice application is running, sma_voice_calling_mngr is the process model that starts to directly interact with SIP process models. In fact, it keeps a state variable which points to a specific sip_UAC_mngr.

**Sip_UAC_mngr**

When the voice calling manager wants to setup a session with the remote host, it will call the function sip_request_invite(), trying to open and maintain an active connection with the remote side. With this event, it tries to arrange an interrupt to the sip_UAC related to the same call. the transfer of this interrupt should be realized via sip_UAC_mngr.

In sip_UAC_mngr, the client will enter the active “init” state in which some state variables will be defined. One of its actions is parsing the node attributes which were specified as SIP UAC Parameters. So we added the new parameters Domain Name, Current Domain and Current Area to the Model Attributes of sip_UAC_mngr process model instead of Proxy Server Specification attribute. then we changed the “init”
state related functions in a way that to read the new attributes values.

Sip_UAC_mgr creates a parent-to-child memory to store variables including above attributes which should be passed between the UAS mgr and UAC. Then it spawns a UAS process to accept requests.

Sip_UAC

In the modified sip_UAC, it tries to get the address of a provincial proxy server which located in the domain of initiating call agent. For this purpose, “init” state codes are changed in a manner to find the Server Address of a proxy which its Domain Name is the same as the domain name of the caller stored in parent-to-child memory. Additionally, the CS Level attribute of the proxy should be equal to “PCS”. If there is not a provincial proxy server defined for a domain, the provided model finds the relevant national proxy which supports that domain. The Domain Name of this server is the same as the Current Area of the caller.

Then the process opens a passive connection to the UAS of found server address so that it can connect to this UAC if it has a request.

After initiation, the client will enter the passive “listen” state to wait for the incoming request. Normally, it will process both request and response received from upper layer and remote host respectively. When receiving a REQ or RESP, the client will be pushed into the “fork” state to process a pending request or reponse.

In the case of request from sip_UAC_mgr to initiate a call, the process opens an active connection with sip_UAC_conn_open_passive() to the UAS address specified in “init” state.

Sip_UAS_mgr

In this process model, we added the new parameters Domain Name, CS Level, Diameter Delay, Processing Delay and Location Server attributes to the SIP Proxy Server Parameters. Then we changed the “init” state related functions in a way that to read the new attributes values.

Sip_UAS_mgr creates a parent-to-child memory to store variables including above attributes which should be passed between the UAS_mgr and UAS. Then it spawns a UAS process to accept requests.

Sip_UAS

If there is a remote interrupt to setup a session, the UAS will establish an active connection and at the same time inform the UAS manager to spawn a new UAS to listen the requests. After the connection is established, the process then sits in the “listen” state to wait for the further stream packets from the UAC side. If the packet is a sip packet, the UAS will get the message type first, and then take actions according to the type of the message. If the packet type shows an INVITE request, the UAS will call the function block sip_UAS_invoke_req_process to process the request. When the incoming request is destined to other destination UAC, the UAS will relay the packet to its destination. It is done without any respect to destination domain sip server. According to our project goal mentioned before, we need to allow the interaction among more than one SIP proxy between two user terminals. Also we want to distribute routing information in CSs of different hierarchies and recreate multidomain hierarchical routes scenarios as it happens in real network.

For this purpose, we modify “listen” state to get invitee_addr from the call information embedded in sip packet and calculate the Call ID. As we said before, Call ID is the minimum digits of call number which is needed to be processed in national or provincial CSs in order to judge the next node in the signalling route. Getting the Call ID, process refers to Location Server stored in parent-to-child memory passed from sip_UAS and assigns the Next Domain and CS Level (national or provincial) of call server which call should be forwarded to it. As mentioned before, Destination Preferences attribute in the AG initiating the call is the dialed phone number including appropriate pre-codes. Consequently, Call ID differs depending on the call initiator be in the same city as the call invitee or not. This results in assigning different sip server as the next node in signalling route and this is exactly what we wanted.

By the means of Location Server attribute, you can configure CSs with routing information as you want and define hierarchical routes between different levels of call servers. If the destination doesn’t belong to the gateways of a province, call information shouldn’t be available in the Location Server of its provincial CS. So Next Domain and CS Level would be related to a national CS at the upper level which the destination province is under its coverage.

This attribute also has the capability of assigning the next node as signalling gateway or media gateway which connects PSTN to NGN. “ack_delay” and “req_delay” states are added in order to implement process delay control for each SIP message and location server queries delay control, respectively. The idea of this feature is the same as one introduced in (Hernandez et al. 2006).

APPLICATION SCENARIOS

In order to verify the proposed SIP model capabilities, this section describes the simulation of scenarios involving a sample of call routing schemes discussed in section “Various scenarios for call routing”.

For this reason we choose a network topology that consists of three interconnected provinces as follows:

- The province of Tehran have 6 Access Gateways, 2 provincial CSs, 8 provincial SG/MG(Signalling Gateway/Media Gateway) and 1 national SG/MG.
The province of Tabriz have 4 Access Gateways, 1 provincial CSs, 3 provincial SG/MG.

The province of Yazd have 2 Access Gateways, 1 provincial SG/MG.

This configuration includes 2 national Call Servers in the entire network located in Tehran and Esfahan. Yazd subscribers are under supervision of Esfahan NCS and Tehran and Tabriz subscribers are in supervision of Tehran NCS. Esfahan NCS acts as the provincial Call Server for Yazd subscribers too. The imposing traffic of these provinces on each other can be handled through two different scenarios. Task definition of national and provincial CSs in call routing is as follows:

Supported interconnections of PCS in scenario 1:
1- Connections between NGN subscribers: intra-province and inter-province.
2- Connections between NGN and PSTN subscribers of the same province.

Supported interconnections of NCS in scenario 1:
1- NGN subscriber connecting with PLMN.
2- NGN subscriber connections with ISC.
3- NGN subscriber connections with PSTN of another province.

In second network scenario that has been modeled with our simulation model, all the connections to out of the province, either NGN or PSTN, should be processed via national CSs. As you see in first scenario, CSs at provincial level should have routing information of all NGN subscribers of all provinces. But in second scenario, all the out of province - NGN or PSTN- routing information is provided in national CSs.

SIMULATION RESULTS

Second scenarios has some preferences over the first one:
- In second scenario Location Server tables adjusted for PCSs of three provinces shows that data content is reduced by 36% on average.
- 3 digit analysis of numbers in provincial CSs leads to speed-up and ease in the management affairs.

In order to study the call setup time and other network performance parameters, we generated a combination of various traffic such as NGN, PSTN, mobile and international call traffics between provinces of the modeled network. We imported this traffic to AGs of Tehran, Tabriz and Yazd by the means of creating profiles in OPNET. The following pictures show some related analysis.

Figure 6: Call Setup Time for Originating Tehran AG1

Simulation results in figure 6 shows that two scenarios are not so different from the call setup time point of view. So according to mentioned advantages for second scenario in section “Various scenarios for call routing”, it can be selected for national and provincial control interconnections.

Figure 7: Active Calls in Tehran NCS

Figure 7 shows that the number of calls processed via Tehran NCS in second scenario is approximately twice what is done in first scenario. So it can be estimated that the Tehran NCS capacity in second scenario should be nearly two times what is in the first scenario.

CONCLUSION

According to the lack of simulation support in SIP, we proposed a modified SIP Model with hierarchical route management capability as a contribution to the OPNET Modeler. This model has eliminated major defects of the existing SIP model. By the means of this model we illustrated that the second scenario with centralized routing information has preference to the first one and doesn’t effect on call setup performance.

REFERENCES


IMPLEMENTATION OF DECISION SUPPORT SYSTEMS FOR COMMUNICATION SYSTEMS

Călin Munteanu,
Simona Caramihai,
Carmen Oudubășteanu,
Florian Antonescu
„Politehnica” University of Bucharest
Automatic Control and Computer Science Faculty
Splaiul Independentei 313, Sector 6, Bucharest, Romania
e-mail: mc_aurel@yahoo.com

KEYWORDS
Decision Support System, Hybrid Petri Nets, simulation.

ABSTRACT

The study of complex systems is one of the most challenging domains of the systems theory. It requires new techniques and approaches, including hybrid models and behavior analysis methodology. The present paper will focus on one of the most typical examples of complex systems – communication networks – and presents a decision support system software application, based on hybrid models, that allows analysis and performance evaluation for different behavioral scenarios. The modeling support is consisting in hybrid Petri Nets.

INTRODUCTION

A clear tendency in control engineering field is represented by the study of aggregated, highly non-linear and dynamical systems, as networks of systems – either communication, transport or enterprise networks. A main characteristic is that such networks, besides those already mentioned, is that every subsystem has several links with others and consequently can influence their behavior. It results that the global behavior of the overall system cannot result mathematically from the behavior of its subsystems.

Complex systems should be considered as rather probabilistic than deterministic and factors such as non-linearity can magnify apparently insignificant differences in initial conditions into important consequences. It means that the long term predictions for complex systems are not reliable. A reliable prediction procedure should be one based on iteration with small increments.

Moreover, there are usually severe time-constraints for every problem solved into the framework of complex system, and finding an acceptable solution with respect to those constraints is more important than finding the best solution.

Added to the fact that formal mathematical analysis of model is not possible and that the future behavior cannot be forecasted based on the past, these considerations imposed the extensive use of simulation tools combined with a reliable methodology for system specification.

The present paper will present a decision support system for highly dynamical communication networks, based on the simulation of hybrid models. The modeling framework was chosen in order to have an explicit difference between the physical structure of the modeled system and the initial conditions, as well as between the inner functioning of the processes and the desired functioning specifications. Moreover, as generally, the scaling factor of a non-linear system is highly important for the prediction accuracy, it was chosen a formalism that allows not only modularization but also scalability.

The next section of the paper will present the initial assumptions on the structure, behavior and functioning context of the considered complex system and, consequently, the modeling framework – Hybrid Petri Nets (Le Bail, et al., 1991; David and Alla, 2005).

Based on the analysis and performance evaluation methodologies of hybrid Petri Net models, the structure of the proposed Decision Support System (DSS).

Finally, the structure of the DSS software application for simulation (scenario analysis) is presented.

COMMUNICATION SYSTEMS ANALYSIS – PROBLEM CONTEXT

The paper is considering large communication networks with high dynamics of server load, where large amounts of information have to be transferred with given time constraints. It is assumed that there are several paths for information transfer between two given servers, usually the use of a single path being not sufficient for the transfer in due time.

The objective of the system analysis is to find an acceptable combination of paths, that used concurrently, will result into the accomplishment of the transfer problem.

Those assumptions imply that basically every problem has multiple solutions and that the modeling framework has to allow the modification of server load and path availability as initial conditions.
Communication, as well as transport networks, have in common the fact that they have to transfer a given flow of entities (informational or physical) from a source to a destination, via a net of possible paths, which will be called connections, usually with some time constraints.

The availability of connections can vary in time, as well as the amount of entities to be transferred, the server loads and the time limits, so as a problem will consist into a (physical) network structure and a set of (varying) constraints. The network structural model can be the same for many problems, but a problem solution will satisfy only a given set of constraints and will be unique. The problem solving time is usually as important as the solution, so as the DSS was conceived in order to find the first acceptable solution and not necessary the best one, even if some optimization is desirable.

Since usually large networks are constituted by aggregating different sub-networks, there was taken into account the possibility of model composition and scalability.

The information to be transferred, consisting of a given number of information packages, has to reach the destination into a given time interval. The packages can be sent by different routes, as the destination node has the possibility to organize them for reconstituting the initial information. Therefore, concurrent distribution of packages on different routes, starting from the source node could improve the overall transmission time, as a route will usually imply several servers, and the transmission time is more significant that the necessary time for expediting a package. The problem to be analyzed is how to distribute packages on different routes, according to connections availability, nodes (servers) load and transfer speed, in order to meet the overall time constraints. Obviously, a route will not consider twice the same node.

**MODELING FRAMEWORK**

Taking into account the above presented problem context, it is necessary for the modeling tool to reflect the amount of information to be transferred, the availability of every path and the transfer speed dynamics. Consequently, Hybrid Petri Nets (HPN) were considered as the most appropriate choice.

In a timed continuous PN, maximal speeds are associated with transitions. Other authors have proposed various results concerning these PNs (David and Caramihai, 2000).

The initial motivation leading to the concept of continuous PN was an approximate modeling of discrete systems with a large number of states, as a consequence of the management of a large number of entities by the modeled process. By this approach it was possible to analyze the flow of entities instead of following the evolution of every one. Another domain of application of this type of nets is for continuous system modeling.

Hybrid PNs (HPN) are formally defined (David and Alla, 2005) as a sextuple \( R = \langle P, T, \text{Pre}, \text{Post}, m_0, h \rangle \) fulfilling the following conditions:

- \( P = \{P_1, P_2, \ldots, P_n\} \) is a finite, not empty, set of places;
- \( T = \{T_1, T_2, \ldots, T_m\} \) is a finite, not empty, set of transitions;
- \( P \cap T = \emptyset \), i.e. the sets \( P \) and \( T \) are disjointed;
- \( h: P \cup T \to \{D, C\} \), called "hybrid function", indicates for every node whether it is a discrete node (sets \( P^D \) and \( T^D \)) or a continuous one (sets \( P^C \) and \( T^C \));
- \( \text{Pre}: P \times T \to Q^+ \) or \( N \) is the input incidence application;
- \( \text{Post}: P \times T \to Q^+ \) or \( N \) is the output incidence application;
- \( m_0: P \to R^+ \) or \( N \) is the initial marking.

In the definitions of \( \text{Pre} \), \( \text{Post} \), and \( m_0 \), \( N \) corresponds to the case where \( P \in P^D \), and \( Q^+ \) or \( R^+ \), corresponds to the case where \( P \in P^C \).

\( \text{Pre} \) and \( \text{Post} \) functions must meet the following criterion: if \( P_i \) and \( T_j \) are such that \( P_i \in P^D \) and \( T_j \in T^C \), then \( \text{Pre}(P_i,T_j) = \text{Post}(P_i,T_j) \) must be verified.

HPNs contain a discrete part (Petri, 1962) and a continuous part (David and Alla, 1987), usually interacting (Le Bail, et al., 1991; David and Alla, 2005). Given a timed hybrid PN defined as a pair \((R, \text{tempo})\) such that:

- \( R \) is a marked autonomous hybrid PN;
- \( \text{Tempo} \) is a function from the set \( T \) of transitions to the set of positive or zero rational numbers:
  - if \( T_j \in T^D \), \( d_j = \text{tempo}(T_j) = \text{timing associated with } T_j \);
  - if \( T_j \in T^C \), \( u_j = \frac{1}{\text{tempo}(T_j)} = \text{flow rate associated with } T_j \).

(Timings associated with discrete transitions and maximal speeds with continuous transitions), instantaneous behavior is analyzed in the following way: a stable marking of the discrete part is sought, then the instantaneous firing speeds of the continuous transitions are computed.

The semantics related to instantaneous firing speeds is relatively easy to define (a continuous transition fires if there is a positive marking in all its input places). There are two ways of enabling a transition: 1) having in all its input places a positive marking or 2) having in some input places positive marking and in at least one input place a zero marking - but all the input places with zero marking being fed by their input transitions). In the first case, the transition is called strongly enabled and it will fire at the associated maximum speed. In the second case, the transition is called weakly enabled and it will fire at a speed less or equal with the associated maximum speed. Iterative algorithms presented in (David and Alla, 1987) and (Alla and David, 1998) work very well if there are no conflicts but automatic calculation for the whole continuous PN in case of conflicts is difficult. Calculation by resolution of a linear programming problem (LPP) was used for some specific cases in (Balduzzi, et al., 2000). In case of conflicts
Stage two is characterized by the discrete marking [0 1], which means that T₂ is not valid at least until P₁ gets the marking back (5 time units). At the beginning of this stage P₁ will have 0 markings, P₂ -> 90-0.5*20=80 and P₃ -> 60-2*20=20. For the same reason as in first stage, T₃ will be strongly enabled and will fire with the speed of 2.5. So the balance of P₁ will be 2.5 and will allow the firing of T₁. The T₁ firing speed will be the minimum between 2.5 (balance of P₁) and 2 (it's maximal speed), meaning 2. Heating the speed for all transitions, we can compute the balances and evolution for each place corresponding to the second evolution stage. The values are: for P₁, balance of 2.5-2=0.5 and evolution given by 0.5*t; for P₂, the balance of 2-2.5=-0.5 and evolution 80-0.5*t and for P₃: balance of 0-2.5=-2.5 and evolution given by 20-2.5*t. The minimum time when a continuous place is emptied is 8 corresponding to place P₁. The discrete marking evolves after 5 time units from the beginning of the phase. So, the second phase lasts 5 time units (t.u.).

At T=25 t.u. Begins phase three of the HPN's evolution. The discrete marking is [1 0] and will remain unchanged for 20 t.u.

According to second phase's evolution, the continuous markings at the beginning of the third phase will be: for P₁ , 2.5; for P₂ , 77.5; for P₃ , 7.5. Having positive markings in both P₂ and P₃ produces firing of T₃ at maximal speed (2.5). Having positive marking in P₁ will produce firing of both T₁ and T₂ at their corresponding maximal speeds (2 and 1).

With the speed of each transition, the balances and evolution for each place can be computed for the third evolution stage. The values are: for P₁, balance of 2.5-(2+1)=0.5 and evolution given by 2.5-0.5*t; for P₂: the balance of 2-2.5=-0.5 and evolution 77.5-0.5*t and for P₃: balance of 1-2.5=-1.5 and evolution given by 7.5-1.5*t. The minimum time when a continuous place is emptied is 5 and it correspond to both P₁ and P₃ places. The discrete marking evolves after 10 time units from the beginning of the phase. So, the second phase lasts 5 time units (t.u.).

At T=30 t.u. begins fourth phase of the HPN's evolution. The discrete marking is still [1 0] and will remain unchanged for 15 t.u. Because P₁ is empty, T₁ will not be strongly enabled. Because P₁ is empty, neither T₂ or T₃ will be strongly enabled. So the continuous part will not evolve from this point forward. Only the discrete part will evolve in cycles of 25 t.u.

**DSS STRUCTURE**

The block architecture of the Decision Support System is presented in Fig. 2. Its structure was designed taking into account the specifications of the problems it has to resolve.

Consequently, the DSS will include a net-editor that will allow the user to design models of networks represented in HPN formalism.
The nets can be either directly analyzed or stored in a model database. With the models from the database, the composition block allows the building of larger global nets. Finally, the history module lets the user to either store analyzed scenarios or to verify and compare previous analyses.

As a further development direction, the history block will be used for a learning facility of a system, that will allow it to propose adequate initial sets of data for some classes of problems.

The user interface module allows the user to initialize the structural models with actual parameters and especially to simulate the net functioning.

A simple case study represented by a system with one source, one destination and two intermediate nodes is presented in (Munteanu et al., 2007), along with all possible routes, the modelling possibilities and different behaviors of the system.

**SYSTEM ANALYSIS (SCENARIO ANALYSIS)**

According to the DSS structure presented in Fig. 2, the developed software contains an user interface (Fig. 3) from which the editing (Fig. 4), viewing (Fig. 5) and composition (Fig. 6) modules can be called. In addition, the simulation algorithm can be chosen (using the simulation button from the main interface) and a history of simulations (regarding one PN) can be obtained using History button.

The editing module’s window is presented in Fig 4. The module’s window consist in 3 vertical areas. The left one is for adding new elements (discrete or continuous transitions or places and links between them). The middle area allows viewing of the net and interacting (placing, moving, deleting) with its elements. Placing and moving are realized using mouse, deleting using DELETE key. Selection of an element can be done using the mouse and is confirmed by changing the color (into red) of the selected element. In this module there are possibilities of selecting the drawing color for each type of element, opening and saving PN’s, exporting to an image format the current Petri Net.

The right area is for setting/viewing properties for each node or arc. According to node’s type there are different possible properties that can be selected.

In order to be more comprehensive for the user, there are graphical differences (except drawing color) between continuous and discrete elements. The one between discrete and continuous places consist in drawing one circle for discrete and two (concentric) for discrete place. The difference for transitions consist in a greater high for continuous ones.

The viewing module (Fig 5) allows viewing of a PN (as in editing module) but not also the editing function. Still, after an network is opened, there is the possibility of editing it (using Edit button) or creating a new one (using New button).
The composition module allows generating of a new Petri Net from two existing ones by selecting the nodes with the same meaning that appear in both networks (Fig. 6). After specifying the two PN that will be composed (using **Open first network** and **Open second network** buttons), the common nodes can be selected from the two Combo boxes below the buttons. The association of nodes is made by pressing the ADD button. Only nodes with the same type can be identical, so if different nodes are selected (place with transition or discrete with continuous), the association is not allowed. The composition operation is achieved when **Run composition** button is pressed.

![Figure 6: Composition module](image)

Different analysis scenarios could be obtained by setting the maximal speeds associated to either the transfer activities or other processing jobs (Munteanu et al., 2007). Its values could be constant a priori established, piecewise constant or stochastic (generated by the computer) on time intervals. Also by setting certain priorities/sharing for different possible transfer route various functioning scenario will be constructed. If a certain route is preferred, the associated modeling transition will be associated with higher priority than the other conflicted transitions modeling the other routes. Should no route is favorite (main), all the transitions will have the same weight in a common sharing group. The sharing situation must be carefully chosen, as in the RPC analysis algorithms there are only three priority/sharing combinations (David and Alla, 2005).

The analysis of the communication /transport system assumes the following steps:
- setting the configuration of a communication network (the user sets the initial discrete marking that corresponds to available connections);
- setting the package transmission speeds on each direct link. This could be done either by the user or by the system. In the last case stochastic values could be associated to the speeds.
- choosing the number of packages composing a message;
- setting the maximum time of message transmission.

Besides these initializations, before the simulation begins, a priority level has to be assigned to each transition. A numerical example presenting different resolution rules and their impact on system behavior is presented in (Munteanu et al, 2007).

**CONCLUSIONS**

The paper presents a decision support system designed on a modular approach, for modeling and analysis of the complex networks in communication/transport area. The partial models of intermediate nodes can be composed in order to obtain the overall system model. HPN formalism was chosen as modeling and analysis tool due to a significant modeling power appropriated for complex systems with hybrid nature.

The decision support system inspects the scenarios provided by the system analysis and proposes a time suited solution.

The future research trend is to search the optimal solution corresponding to the minimum transmission time. In this purpose, all the possible situations for priorities/sharing allocation will be analyzed.

**REFERENCES**


MODELING AND SIMULATION OF TRAFFIC WITH COMPRESSION AT MEDIA GATEWAYS FOR NEXT GENERATION NETWORKS

Dimitar Radev
University of Rousse
8 Studentska Street
7017 Rousse, Bulgaria
E-mail: dradev@abv.bg

Izabella Lokshina
SUNY Oneonta
324B Netzer Administration Bldg.
Oneonta, NY 13820, USA
E-mail: lokshiv@oneonta.edu

KEYWORDS
Modeling and Simulation in Communication Networks, Next Generation Networks, Media Gateways, Integrated Services, Traffic Compression, Markov Reward Model.

ABSTRACT

The paper is devoted to modeling and simulation of traffic with compression at media gateway nodes for the next generation networks based on Markov Reward Models (MRM). The bandwidth sharing policy with partial overlapped transmission link is considered. Calls arriving to the link that belong to VBR and ABR traffic classes, are presented as independent Poisson processes and Markov processes with constant intensity or random input stream, and exponential service delay time. Service delay time is defined according to MRM. Traffic compression is calculated using clustering and self-organizing maps. Numerical examples and simulation results are provided.

INTRODUCTION

Next generation networks (NGN) are expected to be packet-based networks that provide various services including telecommunications. The NGN uses broadband transport technologies that enable quality of service (QoS) management, in which service-related functions are independent from underlying transport-related technologies. The NGN offers unrestricted access to different service providers and supports generalized mobility that allows consistent and ubiquitous provision of services to users. The NGN considered here can be defined with the following fundamental characteristics:

- Packet-based transfer;
- Control functions that are separated from bearer capabilities, call/session, and application/service;
- Service provision that is independent from the network;
- Support for a wide range of services, applications and mechanisms, based on service building blocks that include real time/streaming/non-real time services, and multi-media;
- Open interfaces;
- Broadband capabilities that provide end-to-end QoS, and transparency;
- Ability to interconnect with legacy networks;
- Generalized mobility support;
- Converged services between Fixed/Mobile;
- Backward compatibility and support for IP based addressing schemes, including for a variety of IP address recognition schemes designed for routing in IP networks;
- Unrestricted access to different service providers. Such NGN networks support:
  - Unified service characteristics;
  - Convergence of broadcast and telecommunications.

The NGN architecture is layered including transport layer and service layer. The boundaries are strictly defined. Furthermore, three kinds of interfaces are available:

- User-to-Network Interface (UNI);
- Network-to-Network Interface (NNI);
- Application-to-Network Interface (ANI).

The transport layer provides connection between outer NGN elements, for example, user terminals, and elements located at the NGN servers, for example, databases and media gateways. Access depends on applied technology. For example, fixed access can be provided through DSL and wired LAN, and wireless access can be provided through WiFi, WiMAX, and CDMA. Media gateway (MGW) nodes present interfaces between the NGN and other networks. The service layer provides session and other services and delivery methods. This paper is devoted to modeling and simulation of traffic with compression at media gateways, developed with Markov reward model using learning vector quantification.

NEXT GENERATION NETWORKS AND MEDIA GATEWAYS

Various views on NGN have been expressed by operators, manufacturers and providers (Cochemec 2002; Fazekas et al 2002). The NGN concept is based on integration of currently divided voice and data networks into a simpler and more flexible IP-based network, where the transport, control and service layers are independent and interact via open interfaces. All IP networks allow different access options seamlessly integrated with an IP network layer. The NGNs contain both wired and wireless access networks. Important NGN requirements include simplicity to provide new services, portability and accessibility through different networks, and support for quality of services. The most popular access to the NGN is based on
media gateways with changing transfer and switching. The media gateway nodes can be implemented as independent devices, or can be integrated in another system. In traditional circuit-switched networks, the intelligence is observed in the core of the network (e.g., in central switches). In the NGN model, the intelligence for transfer and switching is expected to be decentralized at the edge of the network. The NGN architecture is conceived to achieve independence of applications and services from basic switching and transport technologies. The fundamental feature of the NGN architecture is independence of applications and control mechanisms from the access and transport layers. That can be possible achieved with migration of applications and call functions to open platforms, and introduction of common control protocols supporting communication between control functions and network resources. Particularly, that can be reached with the gateways providing conversion between different communication media and protocol adaptations. The NGNs with open architecture consist of three main layers:

- Connectivity layer;
- Control layer; and
- Application and service layer.

The connectivity layer consists of the following elements:

- Multi-Service Core: the IP-based transport backbone that carries multiple services over high-speed optical links. This part of the network acts as a long haul transport system providing connectivity among geographically distributed nodes. This network is shared by different services (e.g., phone calls, Web sessions, video-conferences, multi-player games, movies).
- Gateway Elements: they are needed to convert the information between different standards and representations.
- Access Segment: consists of various different broadband access technologies (xDSL, broadband wireless, optical technologies etc.).

The control layer (e.g., call control) is clearly independent from the transport (physical) layer, which provides open and programmable interfaces towards the independent application layer that seamlessly mediates between the signaling protocols of different interconnected networks. The access layer includes both wired and wireless network technologies. The core transport network will be built around Dense Wave Division Multiplexing (DWDM) transport system. Important elements of the NGN architecture are media gateways and soft-switches. Gateways will be employed to interconnect networks based on different representation of the same signal. Common elements of the control layer are multi-service soft-switches, which are able to operate regardless of different protocols they must mediate between. Soft-switch is designed as software application that runs on the server or switch that directs the MGW switching activities. The MGW nodes are located at the ends of the NGN network, and consist of the following elements:

- Interface with the networks with circuit switching (e.g., TDM network);
- Interface with the packet networks (e.g., LAN connection);
- Digital Signal Processor (DSP) for signal processing between circuit-switched networks and packet networks.

There are three categories of the MGW nodes in dependence of their size:

- Small Office/Home Office (SOHO) for small peripheral networks, including voice, VoIP, data and video devices;
- Office, for medium size peripheral networks;
- Provider or carrier grade with high capacity in terms of simultaneous sessions and aggregate bandwidth.

**BANDWIDTH SHARING MODELS WITH PARTIAL OVERLAPPED TRANSMISSION LINK FOR MEDIA GATEWAYS**

Partial overlap of the bandwidth sharing model is defined in the following way: traffic of service \( i \) obtains part of bandwidth equal to \( r_i m_i \) bandwidth units, and the rest of traffic classes concure for sharing the rest of the link capacity \( C-m_{\text{tot}} \) bandwidth units, where \( C \) is the whole capacity. Input traffic is described as traffic of service \( i \), if it has \( m_i \) existing items in reserved capacity of \( r_i m_i \) bandwidth units, or in sharing capacity of \( C-m_i \) bandwidth units; otherwise connection is blocking and lost (Balsamo et al, 2001). In this way, schemes for access in full sharing and full separating of traffic flows can be introduced as particular cases of the partial overlap scheme, where for \( r_1=r_2=0 \) full sharing is obtained, and for \( r_1 m_1+r_2 m_2=C \) full separating is obtained (Gross & Harris, 1998). The obtained function for retransmission with partial overlapped transmission link for \( (m_1, m_2) \in \Omega \) can be presented according to (1).

\[
\alpha_i(n_1, n_2) = \begin{cases} 
0, & \text{if } m_i > C - n_i m_i - \max \{n_2, r_i \} m_i \\
1, & \text{if } m_i \leq C - n_i m_i - \max \{n_2, r_i \} m_i
\end{cases} \\
\alpha_s(n_1, n_2) = \begin{cases} 
0, & \text{if } m_i > C - \max \{n_1, r_i \} m_i - n_2 m_2 \\
1, & \text{if } m_i \leq C - \max \{n_1, r_i \} m_i - n_2 m_2
\end{cases}
\]

The partial overlapped link (POL) is introduced in the paper as specific case of the partial overlap scheme. The basic idea consists in the fact that there is a traffic service class, for which part of bandwidth is reserved, e.g. \( r_i > 0 \) and \( r_2 = 0 \). At the same time, this class is not concuring for the rest of free capacity, e.g. \( m_1 = 0 \), \( m_2 = 0 \). For example, such service class can be represented by constant bit rate (CBR) service class at ATM, administrative channel at GSM channels, etc. In this way, a hybrid approach is developed for full separating scheme, such as CBR service class, and full sharing scheme, such as variable bit rate (VBR) service class. This approach helps to determine optimal capacity of broadband connection channels shared by two or more traffic classes and services, such as, for example, CBR traffic class for voice transfer, VBR traffic class for compressed video data, and ftp transfer.
Continuous Times Markov Chain (CTMC) is used for modeling a transmission link with integrated services, where single channel for bandwidth sharing policy is defined as Markov Reward Model (Rácz et al. 2003; Radev & Lokshina 2007b). Models of traffic classes with guarantee of bandwidth, such as conversational, adaptive and elastic traffic classes, are presented as stationary stochastic processes developed with the network of three parallel queues in the states space, as shown in Figure 1.

![Diagram of Bandwidth Sharing in Wireless System](image)

**Figure 1: Bandwidth Sharing in Wireless System**

If different traffic classes, such as conversational CBR traffic classes, adaptive VBR traffic classes, and elastic traffic classes with available bit rate (ABR) are observed at the same time, then there is no chance to obtain equivalent sharing of bandwidth capacity. Even if an assumption is made that it is possible to separate certain bandwidth capacity for VBR class, still there is no chance to make it possible for ABR class, because of the following reasons:

- ABR class doesn’t provide the same quality of service as VBR or CBR classes;
- Bandwidth offered for ABR calls has high variation in dependence of link overload.

When considering the bandwidth sharing policy of the link capacity, the approach of full sharing should be avoided, and an alternative method should be developed. As soon as ATM transmission link obtains traffic classes without strong restrictions on quality of service, the bandwidth sharing of the link capacity, which increases the throughput and decreases the blocking probability, should be properly used. The input parameters of the model consist of a set of arrival rates (λ1, λ2, λ3) and departure rates (μ1, μ2, μ3), the bandwidths (b1, b2, b3), and the throughput constraints (θmin, θmax). The state model is uniquely described by triple (n1, n2, n3), where n1 is the number of states in conversational (conversational) flows, n2 is the number of states in adaptive flows, and n3 is the number of states in elastic flows. In order to obtain the performance measurement, the CTMC’s generator matrix Q and the bandwidth sharing policy are defined, so that the link capacity C is divided into two parts: a common part and a part reserved for the only two traffic classes. As is shown in Figure 1, there are two cases in the POL bandwidth sharing policy: CASE A and CASE B.

CASE A demonstrates the partial overlapped transmission link between constant and adaptive traffic flows. The only case of the partial overlapped link between CBR traffic class and VBR traffic class is considered in the paper. However, the number of traffic classes considered in this model of sharing policy can vary. When the number of traffic classes is more then one, it leads to the system complexity increase, and the space of the possible states should increase as well. Let us assume that the system consists of ATM channel with the bandwidth capacity C, Mbps. The calls of two traffic classes arrive in the system as independent Poisson processes with exponential service times. The service times are defined according to MRM (Rácz et al., 2003). The following assumptions are made:

- VBR calls always use maximum possible bandwidth, which is a value that is less than or equal to bandwidth of b2. At the same time, this value is equal to free capacity for CBR classes;
- All VBR traffic flows share the bandwidth in equal parts, e.g. the newly arrived calls and in-progress calls are compressed to the same values, if they haven’t been assigned their peak bandwidth. If during the newly arrived call the bandwidth is less than b2, then the last call is not admitted to the system, next it is blocked and lost;
- VBR call management is ideal, e.g. the time for adapting the system to new widths of the bandwidth after the newly arrived calls is infinitesimal.

The actual residency time for VBR calls depends not only on quantity of transferred data, but also on obtained bandwidth for VBR. The following parameters are used to define this balance:

- The moment throughput of VBR call at the moment t is determined as discrete random variable \( \mathcal{A}(t) = \min[b_2, (C-n_1b_1)/n_2] \);
- The throughput of VBR call during the retranslation of x quantity of data for continuous random variables is determined as \( \mathcal{T} = \epsilon/\mathcal{T}_c \).

The following additional parameters are introduced for the proposed model of partial overlapped transmission link:

- \( b_1^{\text{max}} \) – maximal allowed blocking probability of CBR class;
- \( b_2^{\text{min}} \) – minimal allowed bandwidth for VBR class;
- \( \theta^{\text{min}} \) – minimal allowed throughput for VBR class;
- \( \epsilon \) – threshold, which determines the value of \( \theta^{\text{min}} \).

These assumptions let us develop a CTMC, which state is determined as \( i=(n_1, n_2) \), where n1 is the number of broadband CBR calls, and n2 is the number of VBR calls. The partial overlapped link capacity is divided into two parts: a common part capacity of \( C_{\text{COM}} \) and a separated part capacity of \( C_{\text{VBR}} \) only for VBR calls. The constraints that define the number of calls are presented in (2),

\[
\begin{align*}
n_i &\leq C_{\text{COM}}; \quad N_{\text{VBR}}b_1^{\text{max}} \leq C_{\text{VBR}}; \quad n_2 \leq N_{\text{VBR}},
\end{align*}
\]

where \( N_{\text{VBR}} \) is the maximal number of VBR calls. These constraints are guaranteed for VBR calls in difference with CBR calls, where the maximal number of VBR calls is limited, and in this way, the new arriving VBR calls get protected. If there are too many of the new
arriving VBR calls in the transmission link, the throughput $\theta$ is decreased to $\theta^{min}$, and must be regulated via $N_{ABR}$. The generator matrix $Q$ is built in such a way that only transitions between the neighboring states are allowed, where $q_{ij}$ is depicted transition from state $i$ to state $j$. There are four possible transitions between the states, therefore two equations describe the newly arrived calls, and other two describe their services. The compression between traffic classes can be obtained according to (3),

$$r_i = \min\left(n_i, \frac{C - b_i n_i}{b_i}\right)$$  \hspace{1cm} (3)

where $r_i$, $b_i$ is a common width of the bandwidth for VBR calls, when the system is at the state $i$. The partial overlapped transmission link model is completely defined with the following two input parameters: $N_{ABR}$ and $C_{COM}$. The purpose is to minimize the blocking probability of CBR calls and to determine minimal throughput of VBR calls, under criteria for required quality of service. CASE B demonstrates the partial overlap between adaptive and elastic traffic flows. This model of sharing policy is used in mobile communications as soon as modeling the media with integrated traffic flows with variable bit rate of the real-time working sources is required. There are two basic reasons to apply the proposed model in networks with integrated services. The first reason is related to quality of elastic traffic flows with constant width of the bandwidth, as the bandwidth that is busy with elastic flow depends on current load of the transmission link, and on management and control algorithms in the network nodes. The second reason is related to the blocking of elastic flows, where the service is complete even if available bandwidth is very limited during newly arrived calls. For many services the actual residency time of elastic flows depends on the throughput that flow obtains. For instance, a file transfer protocol (ftp) session would last longer if its throughput decreased. The traffic service class that is standard for GSM mobile communications doesn't change the bandwidth sharing policy model, because part of the link capacity $\delta$, which is obtained for this class of provider services is extracted from the link capacity dedicated to three basic service classes. In this way, the schema of full separating is used for the service class, e.g. the link capacity is left equal to $C - \delta$. The following possibilities can be considered for the proposed schema of the partial overlapped transmission link:

- If there is enough bandwidth, then all traffic flows will obtain their necessary peak bandwidth, and second and third traffic classes will obtain, respectively, $b_2$ and $b_3$ bandwidth units.

- If providing compression of the bandwidth is required, e.g. if $n_1 + b_1 + n_2 + b_2 + n_3 + b_3 > C - \delta$, then the bandwidth compression is organized in such a manner that the bandwidth is equally shared between adaptive and elastic traffic classes up to the moment when a constraint for minimal possible value of the one of two classes is achieved.

- If providing additional compression of the bandwidth is required for the newly arriving calls, then the class that can allow additional decrease of the width of the bandwidth decreases it up to the moment when minimal constraint for that class is achieved. After that, the newly arrived calls are rejected. The described rules demonstrate that not only adaptive but elastic flows always obtain their maximum possible bandwidth as well, which is less than:

- Requirements to the peak bandwidth for $b_2$ and $b_3$;

- Equal part of the bandwidth left from adaptive and elastic traffic flows for the constant traffic flows.

The compression of ABR calls is required to achieve better throughput. The particular feature of ABR classes is related with their bit rate that changes over time, e.g. vary in certain interval that makes them appropriate for compression. If the compression is applied with higher than maximum possible value, then it leads to the total loss of information. The case with VBR and ABR traffic service classes is used, with the calls arriving as independent Poisson processes and introduced to the model as Markov processes designed for pure birth with constant intensity or random input stream. The time for service delay is exponential. The time for service delay can be determined using MRM. All adaptive and elastic traffic flows share proportionally the available bandwidth among themselves, i.e. the newly arrived flow and in-progress flows will be squeezed to the same compression values. After that, if a newly arriving flow decreases the flow bandwidth below minimal accepted value, and is not admitted to the system, then it is blocked and lost. The compression of traffic classes can be presented via a common width of the bandwidth $r_i$, $b_i$ for ABR calls, as shown in (4).

$$r_i = \min\left(n_i, \frac{C - b_i n_i}{b_i}\right), \quad i = 2, 3$$  \hspace{1cm} (4)

The proposed model is based on MRM, in which only transitions between the neighbouring states are allowed, and possible state transitions are described with nonzero transition rates according to (5),

$$q(n_1, n_2, n_3 \rightarrow n_1 + 1, n_2, n_3) = \lambda_1 \cdot n_1 \cdot \mu_1$$
$$q(n_1, n_2, n_3 \rightarrow n_1, n_2 + 1, n_3) = \lambda_2 \cdot n_2 \cdot \mu_2$$
$$q(n_1, n_2, n_3 \rightarrow n_1, n_2, n_3 + 1) = \lambda_3 \cdot n_3 \cdot \mu_3$$
$$q(n_1, n_2, n_3 \rightarrow n_1, n_2, n_3 - 1) = n_1 \cdot \rho_1(n_1, n_2, n_3) \cdot \mu_1$$

where the first three equations represent the state transitions due to call arrivals, while the second three equations represent the transitions due to call departures. The $n_1 \rho_1(n_1, n_2, n_3) \cdot \mu_1$ quantity denotes the total bandwidth
of interactive flows when the system is in state \((n_1, n_2, n_3)\), and their compression is denoted as \(\rho_3\).

CALCULATIONS OF TRAFFIC COMPRESSION IN MODELS WITH PARTIAL OVERLAPPPED TRANSMISSION LINK

The bandwidth sharing policy is completely determined in the partial overlapped transmission link models with specifying the following parameters: the link capacity is divided into two parts: a common part of \(C_{COM}\) and a separated part of \(C_{VBR}\) only for VBR calls, where the total link capacity is equal to \(C = C_{COM} + C_{VBR}\). These two parameters guarantee a certain blocking probability for CBR service classes and throughput for VBR service classes. The blocking probability is guaranteed through proper determination of a common part of the link capacity \(C_{COM}\). If VBR load of the link changes, then adjusting the parameter \(N_{VBR}\) to maximal number of ABR calls is required to keep the same throughput. The input parameters are determined with the following two steps:

- Determine minimal required capacity for CBR traffic class, which guarantee the blocking probability \(B_1\) for CBR traffic class according to (6),

\[
\min \left\{ C_{COM} : B_1 \leq B_1^{min} \right\} \tag{6}
\]

- Determine maximal number of simultaneous VBR calls in the system with minimal blocking probability.

The partial overlapped single transmission link model is considered with the middle-size link capacity of \(C_1 = 60\) units and the small capacity of \(C_2 = 50\) units. The assumptions are made that the system has 3 traffic classes: adaptive, elastic and constant; and the last traffic class is not included in subject of future analysis. The partial overlapped transmission link is introduced as the bandwidth shearing policy, and part of the bandwidth is reserved for constant bit rate traffic class. At the same time, this traffic class doesn’t compete for the rest of free capacity. The policy is a hybrid between full bandwidth separating scheme considering CBR class, and full bandwidth sharing scheme considering VBR class. For the space of possible states for the link capacity of \(C_1 = 60\) maximal number of adaptive VBR traffic classes is \(N_1 = 9\), and maximal number of elastic ABR users is \(N_1 = 6\). The peak bandwidth requirements for traffic flows are \(b_2 = 8\) and \(b_3 = 5\), respectively. The flows are defined with minimum accepted bandwidth, which set to \(b_2^{min} = 5.6\) (e.g., 70% of the maximal), and \(b_3^{min} = 3\) (e.g., 60% of the maximal). According to assumptions above, the model has 70 feasible states, and for 27 of them the traffic flow is compressed below the peak bandwidth for adaptive and elastic traffic classes specified by \((b_2, b_3)\). More accurate results are obtained through the actual compression developed with \(\mu_2\) and \(\mu_3\). When defining the size of the generator matrix \(Q\), the calculation of each state capacity without compression is required: \(C = N_2b_2 + N_3b_3\). For the states with capacity obtained greater than present, the compression is applied in the following way:

- If \(C < N_2b_2 + N_3b_3\), then \(r_2 = r_3\) is a level of compression for both traffic flows, as soon as minimal constraint for adaptive and elastic traffic classes is applied \((b_2^{min}/b_2 \leq r_2 \leq 1\) and \(b_3^{min}/b_3 \leq r_3 \leq 1\).

- If more compression is required and minimal constraint still is not true for both traffic flows, then traffic class with newly arrived calls, which needs additional compression, is compressed up to the moment when minimal constraint becomes true.

The calculations of the service coefficients are made through multiplying the number of flows with the coefficients of compression for the corresponding states. For the link capacity of \(C = 50\) units 35 feasible states are obtained, and for 16 of them the traffic flow is compressed below the peak bandwidth for adaptive and elastic traffic classes, based on the proposed rule for minimal throughput. The clustering procedure is used with steady-state distribution calculated in stochastic node network with DTMC that provides capabilities to resolve the tasks described below, in the following conditions:

- The topology and structure of Markov chain are known in advance, e.g. the number and shape of clusters classes are determined. In that case the task of adjusting the target clusters classes size cannot be resolved with conventional stochastic methods. Neural structures with unsupervised learning and clustering algorithms such as Kohonen networks, K-means clustering and Gaussian Mixture models offer models with greatly reduced training time. These models, known collectively as Vector Quantifications (VQs), provide capability to present the winning node that represents the same class as a new training pattern (Buhmann 2002; Webb 1999). With two-layered learning neural structure one can successfully estimate probability density function, occupancy distribution, rare event probability of DTMCs and MRMs.

- The topology of Markov chain is unknown, e.g. the number of clusters classes only is determined in advance. The performance of Markov chain mapping should correspond to possible classes of event that generates the data.

CALCULATIONS OF TRAFFIC COMPRESSION WITH CLUSTERING AND SELF-ORGANIZING MAPS

Minimum-squared-error algorithm is used to solve the clustering problem. Let’s assume that a dataset \(x=(x_1, \ldots, x_n)\) of points is given in some Banach space, which partitions the data into \(k\) clusters (e.g., disjoint groups), so that some minimizing empirical loss function (Radev & Lokshina 2007a) can be written according to (7),

\[
D(x) = \frac{1}{n} \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left\| x_i - s_j \right\|^2 \tag{7}
\]

where

\[
x_i \in C_x, \quad x_i = x_{[\mu(x)]}, \quad \sum_{j=1}^{k} n_j = n
\]

and the dataset points belong to a \(d\)-dimensional Euclidean region \((d \geq 2)\), \(C_j\) denotes the \(j\)-th cluster, \(n_j\) denotes the
number of point \( x_i \) in \( C_j \). The centroid (with the same expected value) has been partitioned into \( d \) clusters with \( n_j \) elements; and the mean vectors \( s_j \) are given as (8).

\[
s_j = \frac{1}{n_j} \sum_{x_i \in C_j} x_i
\]  

(8)

On the other hand, as of (7) and (8), the patterns are moved from one cluster to another only if such move improves the criterion function \( D(x) \), which is known as (9),

\[
D(x) = \sum_{j=1}^{d} \sum_{y_{ij}} \| x_j - s_j \|^2
\]  

(9)

where \( y_{ij} \) is the indicator of \( \{ x_i \in C_j \} \) - \( y_{ij} = I_{\{ x_i \in C_j \}} \)

and \( n_j = \sum_{j=1}^{d} y_{ij} \).

The mean vectors and the criterion function are updated after each pattern move. Like hill-climbing algorithms in general, similar approaches can guarantee local (but yet not global) optimization. Different initial partitions and sequences of the training patterns can lead to different solutions. The goal of clustering is to partition the sample set of points into \( k \) (not necessarily equal) clusters \( C_j \), such that (7) is minimized. In other words, it is necessary to specify the set of centroids \( S^* = \{ s_1, \ldots, s_d \} \) and the corresponding partitions \( \{ C_j \} \), which minimize (7). The definition also combines both the encoding and decoding steps in vector quantization. Clustering techniques with the loss function are called minimum-variance methods. Most well-known clustering and vector quantization methods update the set of centroids \( S_0 \) starting from some initial set \( S_0 \) and using iterative, typically gradient-based procedures that are multi-extreme and depend on the initial value \( S_0 \) in the gradient-based procedures, then they converge to a local minimum rather than global minimum (Kohonen, 1997). We can associate with the clustering used with an \( n \) dimensional discrete distribution \( f(x; p) \) with independent marginal \( f_m(x_m; p_m) \), \( p_m = (p_{m1}, \ldots, p_{mk}) \), \( m = 1, \ldots, n \), and so that each \( f_m(x_m; p_m) \) represents a discrete \( k \)-parameter Probability Density Function (PDF) with masses at points \( x_m = 0,1, \ldots, k-1 \). Note that for \( k = 2 \), \( f_m(x_m; p_m) \) is reduced to Bernoulli PDF. It is crucial to recognize that each generation based on \( f(x; p) \) partitions the set samples into \( k \) clusters \( C_j \). The clustering procedure consists of advanced calculations of discriminating hyper-planes \( W_j \) for the pair-wise discrimination of \( k \) classes \( k^* = k(k-1)/2 \), and later, the prediction of the dataset sample \( x \), e.g., the classification procedure. The linear regression is applied that for to discriminate the \( k^* \)-dimensional vector \( V_{x,n} \) formed for each class. For example, we can assume that \( W_j(x) \) is a regression function, as given in (10),

\[
W_j = w_{x_1} + w_{x_2} x_1 + \ldots + w_{x_n} x_n
\]  

(10)

then, we can determine if the pattern \( x \) belongs to pairwise classes \( C_1 \) and \( C_2 \). In that case, the \( j \)-th component of the vector \( V_{x,j} \) is equal to 1 (\( x \in C_1 \)) only if \( W_j(x) = 0 \); and it is equal to -1 (\( x \in C_2 \)) only if \( W_j(x) > 0 \); and, finally, it is equal to 0 in all other cases. Using discriminating functions, vector function \( s_j(x) \) can be defined for each pattern \( x \), as it is shown in (11).

\[
sw: X_j \rightarrow \{0,1\}
\]

\[
sw(x)_j = sign(W_j(x))
\]  

(11)

For each class \( C_j \), function \( s_j(x) \) can be presented as (12).

\[
s_j(x) = C_j \sum_{j=1}^{k} V_{x,j} \cdot sw(x)
\]  

(12)

The pattern \( x \) is uniquely classified with discriminating hyper-planes \( W_j \) (\( j = 1, \ldots, k^* \)) into class \( C_j \) only if \( s_j(x) = k \), i.e., with respect to \( k-1 \) hyper-planes, which discriminate the class \( C_j \) from the other \( k-1 \) classes. Then, the pattern \( x \) is placed in its half-space that belong to class \( C_j \) (\( V_{x,j} \) and \( W_j(x) \) have the same sign for all \( V_{x,j} \neq 0 \)). If the pattern \( x \) is not uniquely classified, then the Euclidean distances of \( x \) to all these hyper-planes \( W_j \) are calculated, and \( x \) is assigned to the class with minimum distance. Kohonen’s network algorithm provides a tessellation of the input space into patches with corresponding code vectors (Kohonen, 1997). It has an additional feature that the centers are arranged in a low-dimensional structure (usually a string, or a square grid), such that nearby points in the topological structure (the string or grid) map to nearby points in the attribute space. The Kohonen learning rule is used when the winning node represents the same class as a new training pattern, while a difference in class between the winning node and a training pattern causes the node to move away from training pattern by the same distance. In training, the winning node of the network which is nearest node in the input space to a given training pattern, moves towards that training pattern, while dragging with its neighboring nodes in the network topology. This leads to a smooth distribution of the network topology in a non-linear subspace of the training data. The traffic compression example represents the link capacity of \( C = 50 \) units and 35 feasible states, where only for 16 of them the traffic flow is compressed below the peak bandwidth for adaptive and elastic traffic classes based on the proposed rule for minimum throughput; and the traffic compression is calculated with learning vector quantification (Radev et al 2006; Radev & Lokshina 2007). The distribution of arrivals is considered, where 700 independent and identically distributed stochastic values of traffic flows are generated on both inputs of the competitive layer of the neural network. These stochastic values simulate the behavior of the products \( N_1b_1 \) and \( N_2b_2 \), and they are generated with Gamma and lognormal distribution functions, respectively. Preliminary vector quantification is developed when independent arrivals are evenly distributed as 10 hits to 70 target classes.
The high concentration of arrivals in inner classes can explain the fact that there is no blocking obtained for the given capacities of $C_1=60$ and $C_2=50$ units. The main purpose of analysis is to achieve optimal throughput and reduce the blocking probability with compression for arrivals, which are out of the link capacity and, evidently, not admitted to the system. The developed compression technique with $(r_1,r_2)$ creates the new grouping of arrivals following rejecting the classes that don’t belong to compression.

Table 1: Simulation results: Losses in Arrivals Following Traffic Compression

<table>
<thead>
<tr>
<th>Cluster No</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>69</th>
<th>70</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>with</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>compression</td>
<td>9</td>
<td>7</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>7</td>
<td>29</td>
</tr>
<tr>
<td>without</td>
<td>10</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>9</td>
<td>12</td>
<td>74</td>
</tr>
<tr>
<td>compression</td>
<td>9</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>10</td>
<td>9</td>
<td>12</td>
<td>62</td>
</tr>
<tr>
<td>with</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

These classes with losses following compression are numbers 10, 20, 30, 40, 50, 60, 69 and 70, as soon as their probability mass function (PMF) has low values, as is demonstrated in Table 1. For the link capacity of $C_1=60$ units the blocking probability after compression is about 0.3%, which can be used for the high priority data transfers; while for the link capacity of $C_2=50$ units the blocking probability after compression is about 4.1%, which is a significant increase, and this link should be used for the low priority data transfers (e.g., e-mails, or ftp files).

CONCLUSIONS

The paper is devoted to modeling and simulation of traffic with integrated services at media gateway nodes for next generation networks based on Markov reward models. The bandwidth sharing policy with partial overlapped transmission link is considered. The traffic compression is calculated using learning vector quantification. Numerical examples and simulation results are shown for communication networks of various sizes.

REFERENCES


BIOGRAPHY

DIMITAR RADEV, D.Sc. is Associate Professor at Department of Communication Technique and Technologies at University of Rousse, Bulgaria. His main research interests are Teletraffic Theory, Simulation and Modeling of Communication Networks, and Performance Analysis of Queuing Systems.

IZABELLA LOKSHINA, PhD is Associate Professor of MIS and chair of Management, Marketing and Information Systems Department at SUNY Oneonta, USA. Her main research interests are Artificial Intelligence, Complex System Modeling and Simulation in Communication Networks; lokshiiLV@oneonta.edu
SIMULATION IN ENERGY
Protection System Refurbishment at Paks Nuclear Power Plant using Full-scope Simulator
János Sebestyén Jánosy
MTA KFKI Atomic Energy Research Institute
Head of Simulator Development Department
P.O.Box 49, H-1525 Budapest Hungary
E-mail: jjs@aeki.kfki.hu

Abstract

Replacing the aging instrumentation and control system (I&C) in a nuclear power plant (NPP) is always a difficult task, even without the more and more frequent plant-life-extension and power-up-rating projects. The NPP plant lifecycle tends to reach 50-60 years, while that of the I&C system is only 8-10 years. Not only the systems become obsolete; spare parts for the old system are practically non-existent.

Putting into operation reliable, thoroughly tested, verified and validated, approved by the authorities new systems could significantly increase the outage time of the units. However, about 90% of the work can be done off-line, by means of full scope replica simulators, and in the same time the operators become trained to these new systems, too. Taking this approach the refurbishment does not increase the outage time needed for refueling and repair, at least not significantly. The paper describes this methodology in detail.

1. Introduction

The Paks Nuclear Power Plant produces more than 40% of electrical power of the Hungarian energy network. The country has no significant fossil energy resources and practically all the rest is made from imported oil and gas.

The first unit of the NPP was put into operation in 1982 followed by three more units. Some years ago the old, relay-based reactor protection system became less and less reliable, that means unnecessary reactor trips happened without any reasonable cause. It has been decided to replace the old system with a new, computer-based, advanced one with automatic self-diagnostic features and full (100%) backup reserve.

This was the first occasion at our Paks NPP that the testing of the project was based mostly on the full-scope, replica simulator, and the replacement of the whole Reactor Protection System was accomplished in due time, without significant increase of the outage time.

Independently from this project, in the same time, the power of the units has been increased to 505 MWe from the original 440 MWe in several consecutive steps.

The original planned life-time of the units was 30 years, which for the first unit expires in 2012. It is rather common nowadays to extend the life time of the NPPs all over the world; it started in the USA, with the oldest NPPs and continues all over in Europe, too.

Based on the experience of successful NPP life extensions, similar project started in Hungary for the Paks NPP, too. No reason can be seen so far preventing the successful accomplishment of this project. Based on the successful Reactor Protection System Refurbishment Project, we prepare a proposal to set up several simulators in Paks for the planned I&C replacements.

2. The Structure of the Simulator

The basic structure of the original full-scope simulator is shown on Fig. 1. There are several processors executing the model programs in parallel. All state variables are stored in the common memory, accessible to all processors. The time step of the integration for the models is 0.2 sec, that means, in every 0.2 sec we reach a consistent new state of the simulated processes. The term 'state variable' means: variables, the previous values of which are necessary to calculate the new actual state.

The 0.2 sec time step is too large for some models, especially for the models of the neutron kinetics and the two-phase-flow thermo-hydraulics. These model are executed sometimes with significantly smaller time steps but the data exchange with other models takes place once every 0.2 sec.

The simulated measurements are calculated from the state variables with interpolation and added noise every 0.1 sec. All the Control Room actuators and switches are scanned with the same 0.1 sec time step in order to make the response in the Control Room acceptable for the operators.

There are several other parts of the simulator not shown on Fig. 1 because they are not important for the
Fig. 1. The structure of the of the I&C testing by means of the simulator

3. The Testing Procedure

During the verification and validation there were four stages of testing by the simulator:

**First stage:** the SW Switches are in (SW) and (Guided) mode

The functions of the new I&C are modeled by the 'New I&C model SW' box. This module is driven by the simulated measurements, the same way as the modules of the old system are. The real control is still evaluated by the old system. It is not shown at the picture, but the outputs of the new system are carefully recorded and compared with those of the new system.

This 'guided' mode is very convenient, because every transient is performed in the old, very well known way. If the new - and probably imperfect - actions of the new I&C would be entered into the technological models, due the strong feed-back the system would move to unknown, unrealistic and therefore meaningless states.
The investigation of these unrealistic states would be just loss of effort and time.
If the behavior and the control actions of the new systems are very similar to those of the old system, the second stage can be entered.

**Second stage:** the SW Switches are in (SW) and (Full) mode
The control of the technology is handed over to the model of the new I&C model software. This is the so called "software in the loop" configuration. The most important transients are to be repeated again, in order to make sure: the differences (caused by the slightly different behavior of the new I&C algorithms) are not leading the system to unacceptable state trajectories and final steady states.
This is the first occasion that the Authorities have to issue a permission to continue, because completing successfully the first two stages means: the projected new system is acceptable and can be purchased/manufactured.

The ready-made new I&C units are put into operation at the simulator first. New, temporary I/O system has to be constructed to connect the I&C HW with the simulator (red modules on Fig. 1.) Sometimes this can be omitted, if the reserve channels found in the stationary I/O system are satisfactory for the job. It is very important that the time resolution of the temporary I/O system should not influence the performance, therefore it has to be fine enough compared with the time response of the control system. Then the third stage commences.

**Third stage:** the SW Switches are in (SW) and (Full) mode
Generally, it is similar to the second stage: still the model of the new I&C governs the technological models. However, now both the model and the real HW of the new I&C are driven by the same measurements, and the outputs are compared. The differences in the HW and SW control signals can be originated partly by implementation errors, partly by time delays caused by the temporary I/O system - in each case it has to be carefully analyzed and clarified.
If the real HW seems to operate in the same way, the fourth and final testing-by-the-simulator stage can be entered.

Fourth stage: the SW Switches are in (HW) and (Full) mode
Now the full control is made by the external hardware of the new I&C set up next to the simulator. This mode is called as the "hardware in the loop" configuration. Every transient, every mode/state change can be carefully studied and evaluated while the power generating unit of the plant is still running. If some parameters are to be tuned, it has to be done on the software module first; the necessary transients of the previous stages should be repeated, too. Even accidental situations can be triggered and analyzed - this is not possible on the real unit. Ending the fourth stage means ending the testing by the simulator.

Fifth stage: Installation and training
Having the new I&C modules thoroughly tested, tuned and verified, the plant outage time can be minimized. During the installation of the new I&C on the real unit, the simulator is ready to train the personnel in the usage of the new system (SW modes, of course). During the start-up usually it is enough to test the new I&C only in the most critical modes of its operation: in general it is enough to show, that it is performing in the same way as on the simulator.

Summarizing: the replacement of the whole - very sophisticated - instrumentation, control and protection system of the four reactors in Paks NPP could be done in four consecutive years, using the normal refueling outage time. During this four years the simulator was able to provide training to the both (old and new) systems.

4. Basic Assumptions

A special study has been prepared to determine: what to do, how to make the I&C refurbishment if the license for another 20 or 30 years of operation will be granted by the Authorities. The experience of other NPPs - especially the Loviisa NPP in Finland - has been taken into account during this study [1]. The Loviisa NPP is very similar to the NPP in Paks - they have been constructed approximately in the same time and they are of the same type, using Soviet VVER-440 Pressurized Water Reactors (PWRs) fitted to Western standards.

According to the results of the feasibility study:

- We have to keep the functionality of the existing old I&C systems in full detail. The existing control and protection functions are approved by the Authorities, are well known for the staff and are proven as fully satisfactory during the last 20 years of operation. The last significant change has been made in the beginning of the nineties when our energy system became temporarily independent and we had to regulate the frequency in the network even with our NPP units, operating on full power whenever possible.
- together with the structure and functions, the existing cabling has to be preserved, too. These cables are laid in air-conditioned corridors and tunnels, separated by fire gates and equipped with automatic fire extinguishing features. The cables are covered by fire-protecting coating and are in good condition. The greatest challenge to them is to lay new cables, or to change something: in this case the fire gates are to be demolished temporarily and that goes with a lot of carving-out; this is extremely dangerous for the existing cables, therefore should be avoided or at least minimized.
- The refurbishment has to be divided into several stages, as independent from each other as possible. It is a lot easier to accomplish the replacement of only smaller parts of the existing I&C systems at a time.
- The testing, verification and validation has to be done by means of the simulator and the replacement of the old I&C with the tested and approved new one has to be done during outages for refueling, once in a year.
- The refurbishment in four or five steps will last therefore four or five years. The sequence started on one of the units has to be repeated three times, for the three other units, with a time delay of one year for each. The whole process - from starting with the first, up to the ending on the last - may take nine years, but a given system will be replaced only on one unit at a time, and the experience gained with the first can be used on the others.
- The training simulator should be able to train the staff for the any phase, any constellation of the I&C refurbishment state.
- The usage of the full-scale replica simulator should be the same as that used for the successfully accomplished Reactor Protection System Refurbishment project.

5. Simulator Requirements

The simulator is used already in two shifts even without any reconstruction project. There are over 20 staffs for the four units to be retrained - twice a year for a week - that makes over 40 weeks a year just for the retraining. That together with some repairs and development takes one full shift of the simulator.
Teaching newcomers and providing retraining due upgrading the qualification - e.g. turbine operator becomes shift supervisor - takes the second shift.

The Rector Protection System Refurbishment Project could be done only with significant cutting-back of the training on the simulator.

6. Structure of the Additional Simulators

The approach described above is not suitable for a new nine-year project. Additional simulators are to be set up for the I&C refurbishment projects. In order to unload the 'main' training simulator, several others are to be constructed to facilitate the I&C refurbishment projects at their different stages. These simulators differ basically from the existing one by the lack of the replica Control Room. Construction of these simulators is far less expensive than it seems.

First, the license provided by the simulator provider allows the usage of unlimited copies of the simulator software inside the fence of the NPP.

That means, duplicating the training simulator only a rather small software development is necessary, initiated by the lack of the replica Control Room. Practically two types of the simulators can be developed: the Engineering Simulator and the Development Simulator. The structure of both are shown on Fig. 3.

The black SW blocks are coming from the full-scope replica simulator; the blue ones are developed to model the new I&C system. The red HW blocks ensure the "hardware in the loop" testing, the same way as it was done during the Reactor Protection Refurbishment Project.

As it can be seen, the basic difference between these and the full-scope replica simulator is that the very costly Control Room and the real-time I/O system is missing. Both simulators are identical with the full-scope one concerning the model software and the real-time data base. This means, that even snapshots and steady states are easily transferable from one simulator to the others. The testing of different I&C systems does not mean a very great flux of information coming and going to and from the Control Room.
Fig 4. Instructor's system of the full-scope replica simulator

Usually just a small disturbance or malfunction is generated and after that the performance of the system is observed. Some "control quality" type variables (integrals of error signals, etc.) are calculated, logged and plotted together with some other characteristic state variables. There is a small amount of variables changing and few actions are to be taken manually: the Control Room can be represented by the standard Plant Computer screens, the usual Reactor Core Surveillance screens, the Instructor's screens (Fig. 4.) providing the possibility to control the simulator and a special "actuator screen" replacing the usual control knobs and switches.

7. The Engineering Simulator

The so called Engineering Simulator is just limited to the configuration described above. The basic tasks are to perform the model development, the Stage 1 and Stage 2 testing procedures (see Point 3.). These simulators are the cheapest and can be distributed in several copies.

8. The Development Simulator

The development simulator differs only by the software switch and the I/O software reserved for the "hardware in the loop" testing (Stage 3 and Stage 4 in Point 3.) After the hardware of the new I&C system is manufactured, it has to be delivered to the simulator together with the corresponding I/O system (at Paks we are using the VME industrial standard which is a modular, very flexible system). Reserve cards and units of the main replica simulators can be used.

We discussed the possibility to deliver similar simulators to the manufacturers of the I&C systems to their site in order to facilitate the in-house testing. It would be advantageous but in this case the usage of the software is not allowed by the original SW license: this solution could be expensive and therefore it has been rejected.

9. Conclusions

The upcoming I&C system refurbishment project is now in the pre-planning stage. The experience gained during the Reactor Protection Refurbishment Project shows that a lot of time and money can be saved with extensive use of on-line simulation provided by different additional simulators. Putting the models together allows the training of the operators to their future tasks well during the implementation phase. Moreover, the whole process converts the simulation and the simulator facilities into "every-day-tools", accepted and hopefully admired by large number of people in the plant, increasing the efficiency and reliability of future developments and plant operating practices.

10. References

New application fields of nuclear power plant simulator

MyeongSoo Lee
JinHyuk Hong
SeungHo Lee
JeongKwan Suh
DoHyun Hwang

Korea Electric Power Research Institute
65 Munji-Ro, Munji-Dong ,Yusung-Gu
Dajeon, 305-380, KOREA
E-mail: {fiatlux; jhhong; kepc099; jksuh; whitepeach }@kepri.re.kr

Keywords
Real-Time Simulation, Virtual Prototyping, simulator, nuclear power plant, IPVF

Abstract

The nuclear power plant (NPP) simulators have used for operator training and licensing. Increasing of computer performance and using the accurate simulation model give a good chance to open the new application fields of NPP simulator.

Some of immerring important application fields are as following:
- The verification and validation (V&V) of human factors and ergonomics (HFE) design of Human System Interaction (HSI).
- Integrated performance validation (IPV) and training of newly adapted control systems of NPPs.
- R&D and/or engineering for NPPs.
- Virtual Prototyping (VP) and simulation of NPPs.

This paper describes recent R&D activities in above fields using NPP simulator in Korea.

Introduction

Korea has developed nuclear power technology as ambitious during past 15 years, as a result total twenty (20) NPPs are in commercial operation in Korea in these days. The nuclear power plant (NPP) simulators have used for operator training and licensing since early 1980s, and almost all the operator training simulators are full scope simulators including full sized panels. Because the simulator training is requirements of standard (ANSI/ANS 3.5 1998), the main applications fields of NPP simulators are still operator training and operator licensing.

Nevertheless, by Increasing of computer performance and by using the accurate simulation model nuclear power plant simulator industry get more chance to expand the new application fields of NPP simulator.

The NPP simulator related application fields as following are showed in figure 1.
- The V&V of HFE design of HSI
- Integrated performance validation and training of newly adapted control systems of NPPs.

- R&D and/or engineering for NPPs.
- Virtual Prototyping and simulation of NPPs.

Figure 1 Application fields of NPP simulators based on best estimated codes

Korea Electric Power Research Institute (KEPRI), which is a R&D center of Korea Electric Power Corporation (KEPCO), has developed an Advanced Pressurized Reactor 1400 (APR1400) simulator, and developing KNICS integrated performance validation facility (IPVF) with Doosan Heavy Industry (DHI).

The simulators are the important base system for the projects.

HFE Design V&V of new MCR

The design of new generation nuclear power plant in Korea, APR1400, is a First Of A Kind Engineering (FOAKE), and its one of the most major change from the OPR1000, Korea Standard Optimized Pressurized reactor, to Apr1400 is advanced main control room (MCR) design adapting digital man machine interface system (MMIS) and I&C system.

The features of the MMIS system of APR1400 are as following.

- The interaction with ‘soft control’ via compact workstations
- Fully integrated information could be shared all of the operators for plant situation awareness via large displays (LDPs)
- Improvements of operator aids functions for plant operation via new digital alarm system and computerized
This neo-concept main control room design needs comprehensive verification and validation process to get the license for construction. (M.S. Lee et al., 2006)

**Integrated System Validation (ISV)**

U.S. NRC developed Human Factors Engineering Program Review Model (HFE PRM) to support the advanced reactor design certification reviews. The HFE PRM describes the twelve HFE program elements that are necessary and sufficient to develop an acceptable detailed design specification.

One of the review elements is verification & validation (V&V), and its role is to comprehensively determine that the design conforms to HFE design principles and it enables plant personnel to successfully perform their tasks to achieve plant safety and other operational goals. The integrated system for HFE V&V of APR 1400 MMI is developing to meet the requirements of NUREG/CR-6393, and ANSI/ANSI 3.5. and SKN 3 PSAR Chapter 18.

**Distributed Control System (DCS) Simulator Features**

The APR1400 MMIS is designed that the plant can be operated by the soft panel display control at the MCR for all anticipated operating conditions including design basis accidents. The DCS of the SKN 3 NPP, the reference plant of APR1400, is Westinghouse’s DCS, Ovation™.

The one of the major differences between old switch board type panel simulation and DCS simulation of NPP simulator is Backtrack and Replay function ability.

The simulator standard, ANSI/ANSI 3.5, requests simulator features to freeze, run, snapshot, backtrack, control room panel hardware override, and initial condition reset. Also other features, such as replay, slow time, fast time, component failure capabilities, operator performance monitoring, monitoring of parameters, and plotting capabilities are required to be included. The configuration of APR1400 simulator was specially designed to support those functions. A Global MMIS Control (GMC) Server manages the display changes of operator information stations (OISs).

The figure 3 shows the GMC server’s function

![Figure 3 Global MMIS Control Server for DCS simulation](image)

**Other features as for integrated HFE V&V system**

Wide range of human factors evaluation methods to investigate the operator behavior and performance were summarized and evaluated to develop the human performance evaluation support system (HUPESS) with the co-work of Korea Advanced Institute of Science and Technology (KAIST). Some of human factors evaluation areas were selected as follows.

- Plant performance,  
- Personnel task,  
- Situation awareness,  
- Workload,  
- Teamwork,  
- And Anthropometric/ Physiological factors

Because the conventional and subjective methods which are National Aeronautics and Space Administration-Task Load Index (NASA-TLX), Korean Situation Awareness index (KSAX) and check lists were selected and used, it is sometimes need to check, evaluate and review the operator’s comments with the simulator Backtrack and Replay functions. The GMC server make all of the OIS displays could work that functions.

![Figure 4 Real time OIS and AV monitoring display](image)

The other tasks to capture and sending their own displays to
HUPESS and Virtual MCR Design Verification Support System (VMDVS) in real time were developed.
Figure 4 shows the OIS real time monitoring displays.
HUPESS system is the first HFE V&V system which is fully integrated with an advanced MCR type full scope NPP simulator. Figure 5 shows the experiment of HFE using the HUPESS. (Jun Su Ha, 2007)

![Figure 5 APR1400 HFE V&V experiment test](image)

The full scope simulators (FSSs) as an Integrated Performance Validation Facility (IPVF)

The full scope simulators (FSSs) are used as an integration test bed to verify and validate (V&V) before applying new digitalized I&C technology for replacing and/or upgrading old control systems of operating NPPs, and for constructing of advanced MMIS of new constructing NPPs.
The Table-1 shows the some cases of the full scope simulator (FSS) use for validation test and/or operator training before application of the new digital control system to NPPs.
The full scope simulators (FSSs) are used as an integration test bed to verify and validate (V&V) before applying new digitalized I&C technology for replacing and/or upgrading old control systems of operating NPPs, and for constructing of advanced MMIS of new constructing NPPs.

<table>
<thead>
<tr>
<th>Unit</th>
<th>Type</th>
<th>Utility</th>
<th>I&amp;C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sizewell B</td>
<td>PWR</td>
<td>British Energy</td>
<td>Upgrading</td>
</tr>
<tr>
<td>Leningrad</td>
<td>RBMK</td>
<td>Rosenergoatom</td>
<td>Upgrading</td>
</tr>
<tr>
<td>Paks</td>
<td>VVER</td>
<td>VMV</td>
<td>Upgrading</td>
</tr>
<tr>
<td>Chooz B</td>
<td>PWR</td>
<td>EDF</td>
<td>Upgrading</td>
</tr>
<tr>
<td>Lungmen1&amp;2</td>
<td>ABWR</td>
<td>TPC</td>
<td>Constructing</td>
</tr>
<tr>
<td>SKN3</td>
<td>PWR</td>
<td>KHNP</td>
<td>Constructing</td>
</tr>
</tbody>
</table>

The full scope simulators (FSSs) as an Integrated Performance Validation Facility (IPVF)

<table>
<thead>
<tr>
<th>Unit</th>
<th>Type</th>
<th>Utility</th>
<th>I&amp;C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sizewell B</td>
<td>PWR</td>
<td>British Energy</td>
<td>Upgrading</td>
</tr>
<tr>
<td>Leningrad</td>
<td>RBMK</td>
<td>Rosenergoatom</td>
<td>Upgrading</td>
</tr>
<tr>
<td>Paks</td>
<td>VVER</td>
<td>VMV</td>
<td>Upgrading</td>
</tr>
<tr>
<td>Chooz B</td>
<td>PWR</td>
<td>EDF</td>
<td>Upgrading</td>
</tr>
<tr>
<td>Lungmen1&amp;2</td>
<td>ABWR</td>
<td>TPC</td>
<td>Constructing</td>
</tr>
<tr>
<td>SKN3</td>
<td>PWR</td>
<td>KHNP</td>
<td>Constructing</td>
</tr>
</tbody>
</table>

IPVF for KNICS

KEPRI and Doosan Heavy Industry (DHI) have developed together the KNICS Integrated Performance Validation Facility (IPVF) which could evaluate the performance of the newly developed control and protection system.
Even though KNICS I&C is developed and verified by the safety critical software design and V&V methods, safety analysis methods, and qualified hardware design technologies, it should be satisfied the proven technology requirements for using new developed digital I&C for real NPPs.
The user’s requirements documents of the EPRI, EPRI-URD, requires the ‘proven technology’ which should have at least three years experience of documented and satisfactory service as modules of subsystems in the NPP or other plant applications similar to that of NPP.
Because the newly developed system has some risks for safety and operation rather than the existing system that plant managers are reluctant to introduce it, KNICS took alternative option, EPRI-URD requires that it should be completed satisfactorily, a ‘well defined program of prototype testing’ to verify its performance by using the stimulated full scope simulator (H.B. Kim, 2005)
The stimulation method which uses the actual system hardware and software, modified to function properly in the simulator environment is used for the ISV of KNICS control system. Figure 6 shows the architecture of the KNICS, and the doted box shows the stimulated KNICS for test.

![Figure 6 The system architecture and stimulated KNICS system](image)

The Simulator Sever loaded with simulator model connects to Master Node which exchanges I/O signals with I/O system which is connected with stimulated KNICS system.
The integrated performance validation tests for the KNICS control systems, PCS, NPCS, PPS etc., are still ongoing with the Integrated Performance Validation Facility (IPVF).
The IPVF is consists of APR1400 simulator, I/O systems, and the KNICS test systems and MMIS systems for operators.
The compact field point (CFP) I/O system is used for the I/O communication with the APR1400 simulator and KNICS test systems. (M.S. Lee et al., 2007)
Figure 7 shows the configuration of the IPVF I/O system.
Virtual Prototyping (VP) and simulation of NPPs

So the Integrated system validation is a final step of Human Factors V&V, some of verification process are should be done to ensure that HFE principles and methods are appropriately incorporated into the design process. One of verifications is ‘HFE verification’ to verify to be consistent with accepted HFE principles, standards, and guidelines. Also all issues should be tracked and verified to be resolved adequately.

VR MCR Design Verification System (VMDVS)

VR MCR Design Verification (VMDVS) is a kind of a virtual prototyping to design and verify the digitalized MCR. It has not only basic functions for rapid prototyping, building room, layout the general and specific MCR furniture, measuring the distance, but also specific function for visualization of lightning effect by Radiance calculation. It could connect with the simulator and other human factor engineering data accumulation system to get the data to display MMI display of LDP and operator head movements.

Summary

Though the main purpose of using of nuclear power plant (NPP) simulators is still operator training and licensing, some of new application fields for nuclear power engineering become the one of the major area for nuclear industry. KEPRI has developed an APR1400 simulator referenced to SKN 3&4 NPPs. It shows some of immerging important application fields as following.
- The verification and validation (V&V) of human factors and ergonomics (HFE) design of Human System Interaction (HSI).
- Integrated performance validation (IPV) and training of newly adapted control systems, KNICS I&C, to satisfy the proven technology requirements.
- Virtual Prototyping (VP) and simulation of NPPs.

Acknowledgement

This study was supported by Ministry of Knowledge Economy (KME).

References


Biography

MYEONGSOO LEE, a Ph.D. candidate in the Mechanical Design department at ChungNam University. He is member of KSCS in Korea, and is interesting in modeling & simulation, HSI, and VR applications. fiatlux@kepri.re.kr
Optimisation System For A Cogenerative Power Plant

Florin Hartescu, Ion Miciu
National Institute for R&D in Informatics
8-10 Averescu Avenue, 73136 Bucharest 1, Romania
E-mail: flory@ici.ro
IPA SA - Automation Engineering
Calea Floreasca 169, 014495 Bucharest, Romania
E-mail: ionmiciu@ipa.ro

INTRODUCTION

Modern electric power plants are large complex systems with many processes whose operations need to be optimized. The electrical networks are designed for transmission, repartition and distribution of electrical energy, so that they present various structures. An optimization problem is usually a mathematical model, where we wish to minimize a numerical value that represents something like costs, energy losses, risks, errors etc or maximize something like profit, quality, efficiency. Today all the systems that manage the security are characterized by the complexity of their major functions like identification, authentication, access control and data protection. The implementation of these functions is usually and objectively based on a trusted model that uses a trusted architecture, which is the platform of the security architectures.

Automation systems use more and more communication systems to be able to carry out right their task. Therefore, today it’s already available a wide catalogue of products that allows the embedded systems used for the automation to interface to one or more field bus networks and to one or more wide range communication networks (telephone network, Internet, PROFIBUS etc). But currently the systems both for the industrial automation and for other kinds of control need several devices to be able to work within the network.

In particular, the main device containing the CPU foresees some interfaces for being able to be programmed, some interfaces for being able to talk with the I/O devices, some interfaces for being able to talk with the communication lines.

The concept of co-generation means the combined production of electrical energy and of thermal energy based on the same primary energy source. The thermal energy can be used for heating, or for cooling. The system is based on a RT-ARCH (Real Time ARCHitecture), an architecture of software tools used in process control. The system is composed by classical algorithms running on a network of PLC-s and controlling algorithms implemented in a process computer. Some of them are typical numerical algorithms, and the others are adaptive control algorithms.

The Open Source movement has developed new concepts of making process control algorithms based on transparent and co-operative ownership of software.

DESCRIPTION OF TECHNOLOGICAL PROCESSES

Most of the cogeneration systems can be characterized as topping systems or as bottoming systems. In the topping systems, a high temperature fluid (residual gas, steam) fuels an engine for producing electricity, while the low temperature agent is utilized in thermal processes for heating/cooling the extent.

In the bottoming systems, the high temperature agent is produced mainly for a process (for example, in a furnace, or a cement oven); after the end of this process, the hot gases are used directly for powering a gas turbine if the pressure is proper, or indirectly for producing steam (in a recuperation boiler), after which is used for powering a generator with steam turbine.

The temperature ranges for the two system types are presented in figure below:
steam turbine cogeneration system – 3 major components: a heat source, a steam turbine and heat tank; operates with Rankine cycle, or in a base form, or in improved versions (with reheating of steam and prior heating of the regenerative water)

- gas turbine cogeneration systems – operates open/close cycle;
- cogeneration systems with mutual internal combustion engines – operates with Otto or Diesel cycle;
- cogeneration systems with combined cycles – topping temperature cycle, eliminates the heat that is retrieved and utilized by the bottoming temperature cycle for producing supplementary electrical (or mechanical) energy; operates with Joule–Rankine or Diesel–Rankine combined cycle.
- fuel cells – electrochemical devices which converts the chemical energy of the fuel directly into electricity, without the intermediary states of combustion and mechanical work
- cogeneration systems with Sterling engine – operates in Sterling cycle.

The paper presents a system with large capacity about calculation and communication in real time but at low cost. The low cost, compared to the devices currently available on the market, is obtained thanks to the fact that the “Intelligent Cell” has an architecture very simple constituted by a microprocessor very powerful with only three single connectors; one for the power supply and two for the communications. The “Intelligent Cell” is produced by the factory only with communication abilities and it can immediately be installed where it will have to reside during its operation activity.

GENERAL DESCRIPTION OF A COGENERATIVE POWER PLANT SYSTEM

The paper presents the process for the realization of cogenerative supply of electricity and heat, and in given case for their combination with waste heat recovery, particularly in case of combined (gas-steam) cycle industrial power plants.

The essential feature of the system is that during the season without heat demand the condenser with heating capacity is operated by cooling with cooling tower, and during the heat-intensive season it is operated at a higher temperature required for the heat supplying operation mode, while the cooling tower is disconnected partially or completely, and the resultant reduction of the specific expansion and heating capacity of the steam turbine is compensated by increasing the steam parameters and steam output of the waste heat recovery boiler with supplementary firing.

The control of the electricity generation is executed by the control unit of the gas turbine unit (GTU). The key control devices are special configured PLCs that operate autonomous to any of the other control devices as the requested set point for the electricity generation is manually settable. Additionally to the manually setting of the requested power generation capacity value by the control devices of the gas turbine unit there is also the possibility of setting the requested power generation capacity value by the overall HMI-monitoring system related to the control unit. The required fuel (natural) gas for the GTU is prepared and delivered at a pressure level of some 26.0 bar g by an also autonomous controlled and operating gas compressor unit (GCU). For this fuel gas provision the gas compressor unit (GCU) is considered to be operated automatically and autonomous by its control unit or PLC-unit. The operation of the GCU is entirely and exclusively controlled by the PLC-unit and is not depending for starting and operating to any other control device / PLC. Nevertheless it is possibly to start the GCU remotely by the PLC-unit which is designated for the Hot-Water-System (HW-system) but contains also some superior control duties for the entire plant, as some essential signals are exchanged between units. For the heat extraction from the delivered exhaust gas into an hot water system (HW-system) a main HW-boiler is foreseen which is capable of a thermal heat extraction of up to 22 MW th at nominal power generation capacity. To obtain the contractually specified heat generation capacity of 25 MW th an auxiliary HW-boiler with an additional heat generation capacity of 3.25 MW th get established beside the main HW-boiler. Finally the heat extraction out from the exhaust gas of the running gas turbine and its transfer into the secondary circuit HW-network get controlled by a PLC-unit. Furthermore the PLC-unit is controlling some auxiliary systems as the instrument air provision units, some building ventilators, an emergency cooling equipment for the primary HW-circuits etc.

For the monitoring and remote operation of a new unit, two visualization computers or Human Machine Interface (HMI) - stations are foreseen and are located in
the central control room (CCR) of the power plant. The first HMI-station is exclusively considered for the monitoring and operation of the gas turbine (and is provided by the gas turbine manufacturer Turbomach) as the second HMI-station (with a visualization system of Siemens, Distributed Control System type Simatic PCS7) is considered for the monitoring and visualization of the HW-system, the gas compressor unit (GCU) and all side auxiliary equipment as well as for some superior operational demands of the entire unit. The CCR is located inside the main machinery building of first unit inside the power plant and which is some 300 meters away from the location of the new unit.

THE INFORMATIONAL STRUCTURE OF THE SYSTEM

The overall goal is to foster the usage of FLOSS backend platforms and services and to generate new business opportunities for the Open Source developer community. The open source-based set of tools is supposed to have a high socio-economic effect for both, the providers and users of FLOSS.

The informational system is designed first for data and technical parameter acquisition and then for operative supplying of information required for the process control, optimization, for exploitation and maintenance of the equipment. The entry information is received as digital or analog signals from transducers through the process interface. As a result of data processing, the computing system elaborates the output information, analog or numerical, that can be used for controlling, adjusting or for displaying on video terminals. The system has a database for data and programs on the mass memory (one or more magnetic hard drives).

TECHNIQUES DESCRIPTION

Typically, classical control techniques use analog PID (PI, filtered PID) controllers, or on/off controllers for slower processes. Controllers are used in structures like:

- Simple adjusting with imposed reference
- Cascade adjusting
- Combined adjusting by perturbation and reference

These types of controllers have the advantage that they can be easily implemented, but very often their parameters are fixed. That’s why classical control structures can be used only for those subsystems of technological installation that have very well defined models, or have modest control requirements.

From all the diverse modern control techniques we have chosen to implement some algorithms by computer programs. This implementation has a great flexibility: it is very easy to modify the dates of the program that contains the adjusting parameters.

HARDWARE AND SOFTWARE ARCHITECTURE

The hardware platform used as a support of the control programs is composed by several personal computers with a Pentium IV processor Dual Core, 1024 Mb RAM, 256 bits video graphic accelerator, SCSI 160Gb HDD, data acquisition cards, one multiplex for 64 serial ports and a network of PLC-s. On these machines we have used Unix operating system and the Distributed Control System Simatic PCS7. This system offers multitasking facilities for parallel managing aspects regarding data acquisition, data transmission, adjusting and data displaying. Using multitasking we could manage more control loops on the same computer even though some of the processes are controlled only by PLC-s.

The development platform used was Visual C++ because it can compile programs for Unix operating system and it has the following facilities:

- Permits creation of separate threads with different adjusting algorithms for each of the processes.
- Offers communication methods inter-threads for transferring data between data acquisition processes and controlling processes. Communications are made using message boxes and critical sections.
- Permits realization of communication modules in a TCP/IP network. That makes possible implementation of the hierarchical architecture.
- Offers the possibility to create a user-friendly interface for the product, in Unix environment.

The only disadvantage of this platform is that it hasn't any dedicated functions for complicated mathematical calculus needed in adjusting algorithms, so that we had to write them ourselves.

The process computer is placed in the control room of technological installation because it must be protected from the environment of the technological process.

PROCESS OPTIMIZATION

Many forecasting models have been proposed and implemented in this fields, based on time-series, regression, expert system models etc. A special attention has been paid to the use of Artificial Neural Networks (ANN) to model load. In the last years, control centers in distribution went through some drastic changes. Evolving from Supervisory Control and Data Acquisition (SCADA) systems, the concept of Distribution Management System (DMS) gained growing acceptance. A DMS must provide a set of functions, namely for switching decision and
operation, which rely on the basic tool of power flow calculus.
The mathematical problem posed by the general optimization formulation is often remote from that of a practical design problem.
The optimization problem is "given a dynamic system (S) that evolves on a finite time interval [T₀...Tₖ] and the performance index J(u) it is required a command that minimizes J". In this standard formulation there are no restrictions, but it follows a movement in the states space for minimum energy consumption. Using the computer we have chosen dynamic optimization, that relies on the idea of finding a procedure for generating a “relaxing” array defined by the following condition: X₁, X₂, X₃, …Xₙ, so that f (X₁) ≥ f (X₂) ≥ f (X₃) ≥… with the property that if Xₖ → X, then f (Xₖ) → f (X), where X is the extreme value of the objective function in a specific domain.
The relaxing arrays generating procedure has two components:
- Choosing the descending direction of the function with the modified Newton method. This is a gradient method that means successive approximations of the function.
- Determining the step on the descending direction. Newton-type method calculate the Hessian matrix H directly and proceed in a direction of descent using a line search method to locate the minimum after a number of iterations. Calculation H numerically involves a large amount of computation. Quasi-Newton methods avoid this, by using the observed behavior of f(X) to build up curvature information, in order to make an approximation to H using an appropriate updating technique.

CONTROL SYSTEM FUNCTIONS FOR THE COGENERATIVE POWER PLANT

The automation system works in a collecting-processing way, interfering with and conducting the controlling processes. The system does the following functions:
- Supervision of technological equipment and devices
- Displaying and printing any damage that may occur, showing the moment of appearance, the kind and the place of the damage
- Displaying and printing the alarms at the moment of the occurrence
- Showing the entries and outputs from the programmable logic controllers for debugging any anomaly or damage
- Controlling the measuring equipment and supervising the process
- On demand, displaying the parameters of the technological process

Some of the subsystems of the technological installation are variable in time due to action of stochastic perturbations. That determines large variations of the parameters of the attached mathematical model, but fortunately the structure of the model remains the same. The control takes place at the level of the central computer, and also at the level of the PLC-s. We have used a hybrid adjusting method, i.e. some adjusting loops are implemented with classical algorithms (PI, PID), and the others are adaptive.
The system is composed by:
- Classical algorithms running on the PLC-s that also have implemented data acquisition modules, and analog and digital command modules. Programs are written in the PLC’s language.
- Controlling algorithms implemented with the process computer. Some of them are typical numerical algorithms, and the others are adaptive control algorithms.

CONCLUSIONS

Implementation of this system in a Romanian cogenerative power plants has the following advantages: obtaining a high efficiency and combustible saving, limited efforts for developing a new application in a short period of time, and high performance of the system in solving the demands of applications.
The system uses secure socket connections (SSL) to transmit all sensitive information during the control process. The application has been tested in an integrated system, with several servers running Unix, connected in a network. The system was configured easily, and it has worked very fast because the communication protocol transmits just the information needed. The aim of the system is to realize a network of Intelligent Cells with large capacity about calculation and communication in real time but at low cost. The prominent advantages of the process control system are the following:
a) - favourable specific heat consumption
b) - optimal utilization of the waste heat during the winter-summer seasons
c) - saving the cost of investment by use of the supplementary firing
d) - stable electric power in the power plants supplying cogenerated electricity and heat.

REFERENCES

Kirillin V. A., Sicev, V. V. si Seindlin, A.E., Termodinamica, E. S. E. Bucuresti, 1985.
Vladea, I., Tratat de termodinamica tehnica si transmiterea caldurii, E.D.P.Buc.
Hahne, E., Grundlagen der Technischen Thermodynamik Band 1 und 2, Institut fur Thermodynamik und Wartemtechnik der Universitat Stuttgart
Cichoeki A., R. Unbehauen, Neural Networks for Optimization and Signal Processing, John Wiley & Sons, Ltd. 1993
Fleming P.J. Computer-Aided Control System Design of Regulators using a Multiobjective Optimization Approach, Proc. IFAC Control Applications of Nonlinear Optimization
Gill P.E., Murray W., Numerical Linear Algebra and Optimization, Vol.1, Addison Wesley
Selic B., G. Gulliksons, P.T. Ward: Real Time Object Oriented Modelling, John Wiley & Sons, 1994
W. Wang – Designing Secure Mechanisms for Online Processes
Tananbaum S. A., Computer Networks, Prentice Hall.
MODELLING AND SIMULATION OF FLUID POWER SYSTEMS IN AN INTELLIGENT PROGRAMMING ENVIRONMENT

Gunnar Grossschmidt
Institute of Machinery
Tallinn University of Technology
Ehitajate tee 5, 19086
Tallinn, Estonia
E-mail: gunnar.grossschmidt@ttu.ee

Mait Harf
Institute of Cybernetics
Tallinn University of Technology
Akadeemia tee 21, 12618
Tallinn, Estonia
E-mail: mait@cs.ioc.ee

KEYWORDS
Fluid power system, multi-pole modelling, intelligent programming environment, simulation.

INTRODUCTION
Most of modelling and simulation systems in existence such as SimHydraulics™, MATLAB/Simulink package, ITI®SIM and ITI SimulationX, DSHplus, Dymola, HOPSAN, VisSim, AmeSim, 20-Sim, DYNAST, MSIM™, etc. (International Journal of Fluid Power 2000 . . .2006) are object-oriented (systems are described as functional or component schemes) using equations with fixed causality or equations in non-causal form for each object. The obtained equation systems usually need to check and correct to guarantee the solvability. Solving large equation systems requires special methods, it is impossible to use logical conditions, it is difficult to debug such systems. The static and the steady-state conditions are usually calculated through transient responses, which take much time.

In this paper a new approach is proposed. The approach is based on using multi-pole models with different oriented causalities and signal-flow graphs of functional elements. This enables methodical, graphical representation of mathematical models of large and complicated systems (Grossschmidt et al. 1998, 2000).

A general purpose intelligent graphical environment CoCoViLa (Grigorenko et al. 2005) is used as a tool for modelling and simulation. The CoCoViLa is a platform independent, Java based programming environment with automatic program synthesis. Using the CoCoViLa environment enables one to build and use fast prototypes for modelling and simulation of different systems.

Modelling and simulation of dynamics of a hydraulic-mechanical load sensing system in the CoCoViLa environment is considered as an example. Modelling and simulation of components, subsystems and the whole load sensing system are discussed.

MULTI-POLE MODELS, CAUSALITIES
Generally, multi-pole models represent oriented mathematical relations between pairs of input and output potential and flow variables of functional elements. The multi-pole models take into account the signal propagation in both directions as it occurs in hydraulic and mechanical systems.

The multi-pole models are as models with oriented causality. Depending on the chosen input and output variables we can have various forms of models. The mostly used multi-pole models are the four-pole models. The four-pole models show the relations between two pairs of potential and flow variables. One of the variables in pair must be the input. Only in that way we can take into account the input and the output in the same port. Models of that kind express the physical content of processes with retroaction.

Four forms of such four-pole models, or otherwise – four forms of mathematical causality are used. Letters G, H, Y and Z as in electrical engineering denote them. Four-pole models are expressed by two relations.

The four-pole models of form G and H (causality) represent the orientation in both directions (transfer functions G12, G21, H12 and H21) with corrections through the cross dependencies (transfer functions G11, G22, H11 and H22 in Figure 1).

![Figure 1 – Oriented graphs of four-pole model of forms G, H, Y and Z (causalities)](image-url)

The four-pole models of forms G and H for the transient response calculation include the members with differentiation procedure in cross dependencies. Mostly the four-pole models of forms G and H are used for calculation of the statics, steady-state conditions and frequency characteristics.

The four-pole models of forms Y and Z change the type of variable, which proceed through other elements of the system. The four-pole model of form Y has potential variables as inputs and flow variables as outputs. The four-pole model of form Z has conversely flow variables as inputs and potential variables as outputs. The four-pole models of forms Y and Z for the transient response calculation include the transfer functions with integration procedure.

The four-pole model of form Y for mechanical and hydraulic inertia expresses only the dynamics. Hydraulic volume elasticity four-pole model of form Z is also used only for dynamics. The four-pole model of form Z does not exist for mechanical inertia and damping, and for hydraulic inertia and resistance. The four-pole model of form Y does
not exist for mechanical elasticity and for hydraulic volume elasticity.

For composing a model for the fluid power system, it is necessary to build multi-pole models of functional elements and connect them through poles. The models for static or steady state conditions and for dynamic transient responses may be different.

Figure 2 - Scheme of the modified hydraulic-mechanical load sensing system of Bosch GmbH

HYDRAULIC-MECHANICAL LOAD SENSING SYSTEM

Fluid power systems, in which working pressure (pressure in pump output) is kept proportional to load, are called hydraulic load sensing systems. Such systems are mainly used in mechanisms containing numerous drives to run with the purpose to save energy. The scheme of the modified hydraulic load sensing system of Bosch GmbH is shown in Figure 2, the scheme of hydraulic-mechanical controller in Figure 3.

In Figure 2, the variable displacement axial piston pump is driven by an electric motor M. Hydraulic-mechanical control of the pump volumetric flow is performed by control valve and positioning cylinder. The feeding chain of the hydraulic motor RVerbr contains tube R_{WPDE} measuring valve R_{VW} with pressure compensator R_{DRE} and check valve, meter-in throttle edge R_{SK, SP} and connection elements.

Figure 3 - Scheme of a hydraulic-mechanical controller

The output chain of the hydraulic motor R_{Vorh} contains a meter-out throttle edge R_{SK, t} and tube R_{L, eff}. The device contains load sensing pressure feedback. Feedback pressures have been taken directly from the measuring valve with pressure compensator R_{DRE} through tubes and resistors. The scheme of a hydraulic-mechanical controller (Figure 3) contains a spool valve (effective area A_{V}) with inflow and outflow slots, constant resistor (volumetric flow Q_{Th}), positioning cylinder (effective area A_{Z}), and swash plate with spring.

INTELLIGENT PROGRAMMING ENVIRONMENT

A new programming environment CoCoViLa is currently used as a tool in modelling and simulation of the fluid power systems. CoCoViLa is mainly based on the similar principles as the previously used programming environment NUT (Tyugu et al. 1997).

CoCoViLa is a programming environment, which supports declarative programming in a high-level language, automatic program synthesis and visual programming. CoCoViLa is elaborated in the Institute of Cybernetics at the Tallinn University of Technology, in 2005-2007. The CoCoViLa environment is Java based, free and platform-independent.

CoCoViLa supports a language designer in the definition of visual languages, including the specification of graphical objects, syntax and semantics of the language. CoCoViLa provides the user with a visual programming environment, which is automatically generated from the visual language definition.

When a visual scheme is composed by the user, the following steps – parsing, planning and code generation – are fully automatic. The compiled program then provides a solution for the problem specified in the scheme, and the result it provides can be fed back into the scheme, thus providing interactive properties.

Automatic synthesis of programs is a technique for the automatic construction of programs from the knowledge
available in specifications. Having a specification of a class, we are, in general, interested in solving following problems:

- Find an algorithm for computing the values of components y_1, ..., y_m from the given values of components x_1, ..., x_m.
- Find an algorithm for computing the values of all components that can be computed.

The automatic synthesis of programs is based on proof search in intuitionistic propositional logic.

From a user’s point of view the CoCoViLa framework consists of two components: Class Editor and Scheme Editor. The Class Editor is used for defining models of components of schemes as well as their visual and interactive aspects. The Scheme Editor is a tool for the language user. It is intended for developing schemes and for compiling (synthesizing) programs from the schemes according to the specified semantics of a particular domain. The Scheme Editor is implemented using Java Swing library. It provides an interface for visual programming, which enables one to compose a scheme from shapes of classes. The environment generated for a particular visual language allows the user to draw, edit and compile visual sentences (schemes) through language-specific menus and toolbars.

Having developed the visual language we are able to load it in the Scheme Editor and build schemes by putting visual objects on the drawing canvas and connecting them through ports.

The Scheme Editor is fully syntax directed in the sense that the correctness of the scheme is forced during editing: drawing syntactically incorrect diagrams is impossible. The way to handle large schemes in the Scheme Editor is to use hierarchical composition in building the scheme. Any part of a scheme can be encapsulated as a separate class, so a large scheme can consist of a hierarchy of schemes, where each scheme object can contain subschemes. This means that schemes can be viewed in several different levels of abstraction, in order to encapsulate and manipulate parts of the scheme which are relevant to a particular issue.

When the visual classes have been built by software developers who must understand the problem domain as well, the language user need not be a software expert, but can work on the level of visual programming, arranging and connecting objects to create a scheme. Manipulating the scheme – a visual representation of a problem, is the central part of the user's activities.

The combination of object-oriented and structural composition is useful in rapid development of visual languages for modelling and simulation.

**MODELLING AND SIMULATION OF FLUID POWER SYSTEMS**

Generally, simulation of fluid power systems includes three activities: simulation of statics, steady state conditions and dynamics. Usually, different simulations require different models, but in some cases the same model can be used for several simulations. A number of packages for various fluid power systems modelling and simulation have been implemented in the previously used NUT system:

- package for hydraulic elements and circuits,
- package for electro-hydraulic servo-systems.

Multi-pole models of functional elements have been described as classes together with their icons and images. Using visual specifications of described multi-pole models one can graphically compose models of various fluid power systems. When solving specific simulation problem, the model has to be adjusted by evaluating different parameters of the elements and adding sources to elements of the model that describe disturbances of the necessary shape and value. During the simulation, some elements of the model need parameters, values of which cannot be computed at the moment they are required. For computing values of such parameters, a special iteration method has been used. When starting the process, approximate values of critical parameters have been given as initials. At each step of the simulation process we try to refine initial approximate parameters. We use the programming environment to synthesize programs for recomputing some parameters and try to recompute them iteratively until precise values of the parameters have been attained.

It is possible to solve a great number of various computing tasks on each fluid power system evaluating some components as inputs and computing some other components as outputs.

At the simulation of dynamics transient responses caused by disturbances are calculated. Disturbances are usually considered as changes of pressures or volume flows in certain points of the hydraulic system. Disturbances of different shapes are considered, such as constant, step, impulse, sine, linear, etc. Calculations of dynamic responses are performed in time. The time step length and number of steps are to be specified.

Static, steady state and dynamic computing processes are organized by corresponding process classes. To follow the system behaviour in time, the concept of state is invoked. State variables are introduced for each functional element to characterize the element at the current simulation time step. The simulation process starts from the initial state and includes calculation of following state (nextstate) from previous states (from oldstate and state). Final state (finalstate) is computed as a result of simulation.

**COMPOSING MODEL OF THE LOAD SENSING SYSTEM**

Simulating load sensing system includes simulating steady state conditions and dynamics. Simulation of steady state conditions in the NUT environment has been considered in (Grossschmidt et al. 2006, 2007). Here, the dynamic simulation in the CoCoViLa environment is considered. In our case the same model can be used for steady state conditions.

To build up the multi-pole model it is necessary to decompose the scheme of the load sensing system into logical components and subsystems. Components of the load sensing system are described as multi-pole models and implemented as classes in CoCoViLa environment. Besides poles the component class usually contains inner variables and component parameters.
Figure 4 - Multi-pole model of the hydraulic-mechanical load sensing system

Load sensing system components: VP - Displacement of the control valve; resH, resG – Hydraulic resistors; RVP - Meter-in throttle edge of the control valve; IEH1-3, IEH2-2m - Hydraulic interface elements; ZV - Positioning cylinder; REL - Constant resistor; RVT - Meter-out throttle edge of the control valve; PV - Variable displacement pump; ME - Electric motor; tubeH, tubeG - Tubes; RIVVWin - Measuring valve with pressure compensator; RSKZ - Meter-in throttle edge for hydraulic motor; MH - Hydraulic motor; RSKA - Meter-out throttle edge for hydraulic motor; CJh, CJhM - Clutches.

Simulation task components: dynamic Process – Process organizer; constant Source, dynamic Source – Sources; Graph drawings; Clock.
Relations between variables are described as equations or as user-defined Java methods for representing more complicated ones. In relations state variables, inner iterations, logical conditions and CoCoViLa subtasks are used for describing the behaviour of the component. The multi-pole model (Figure 4) represents the scheme of the load sensing system (Figure 2).

Models of the following components of the load sensing system have been developed: electric motor, variable displacement axial piston pump, clutch connecting electric motor and pump, hydraulic-mechanical controller, valve block, hydraulic motor, tubes and multiple tube connection elements. Model of the hydraulic-mechanical controller includes models of control valve, meter-in throttle edge, meter-out throttle edge, constant resistor and positioning cylinder with swash plate. Model of the valve block includes a measuring valve with pressure compensator and check valve, meter-in and meter-out throttle edges of the hydraulic motor, clutch with drive mechanism.

Models of the components have been composed as multi-pole models with different number of poles. Nevertheless, the mostly used are the four-pole models, described in more detail above.

The multi-pole model of the whole load sensing system has been built up using the components models. First, necessary components have been connected through poles. Second, variables of connection poles have been defined as inputs or outputs for each component depending on required causalities (Grossschmidt et al. 1998).

SIMULATION STEPS

The most important stage in the design of hydraulic load sensing system is parameters specification. In order to help designer to find better solutions, computer simulation technology is proposed that includes following steps.

First, hydraulic motor and hydraulic pump parameters must be chosen.
Second, the fluid and its properties must be chosen. Cinematic viscosity, density and compressibility factor depending on the pressure are calculated at each step.
Third, initial approximate values of pressures and pressure drops for pump control must be set.
Fourth, maximum displacements of the valves must be set.
Fifth, all the models of components must be tested separately. For each component the simulation problem must be composed and input signals must be chosen. Behaviour of the component must be simulated. Initial approximate components parameters values (e.g. stiffness of springs, geometry of valves working slots, etc.) must be refined as a result of simulations.
Sixth, the separately tested components models must be connected into more complicated subsystems tested in behaviour and adjusted if necessary.
Seventh, the whole load sensing system must be built up and simulation tasks must be solved.

Steps S...7 must be supported by testing on pilot models and using the results for refining the parameters and models.

SIMULATING DYNAMICS OF A SUBSYSTEM

Components models used in the system must be previously tested separately. Then various subsystems must be built up and tested. The simulation task of variable displacement pump with electric motor as an example of a subsystem is shown in Figure 5.

![Figure 5 - Simulation task description of variable displacement pump with electric motor](image)

**Notations:** ME – Electric motor; CJh – Clutch with pump rotor; PV – Variable displacement pump; ZV – Positioning cylinder and pump swash plate; Dynamic process; Dynamic sources; Drawing graphs; Clock.

The model of the subsystem contains loop dependencies that cannot be resolved by the CoCoViLa algorithm synthesizer. To solve the problem we need to split three variables: angular velocity $\omega_{me}$ of ME, angular velocity $\omega_{2e}$ of CJh and reaction force $F_r$ of PV to the swash plate. Initial approximate values are assigned to these variables and algorithm for recomputing of them is found and constructed by the synthesizer. When performing the simulation the subtask solving program is iteratively performed until the values of required precision are calculated. Using the method enables to perform simulations without searching and solving large equation systems. Variables requiring iterative recomputing are pointed in each component model. When synthesizing the simulation algorithm these variables are detected automatically and handled in way described above. When simulating the whole load sensing system 17 variables are iteratively computed.

Parameters have been set up in each component model. Simulation parameters have been set up in “dynamic Process” and “dynamic Source” classes. Experimenting with various parameter values gives us different simulation results and enables to achieve optimal configuration of component models.

In Figure 6 and Figure 7 results of simulations are shown. Graphs of volumetric flow of the pump $Q_2$ (1), volumetric flow of the positioning cylinder $Q_1$ (2), moment of the electric motor $M_3$ (3) and angular velocity of the pump $\omega_{m}$ (4) in the case of different values of the torsion elasticity $C$ of the clutch are presented. As it can be seen, the dynamic responses of the moment of the electric motor and
the angular velocity of the pump are significantly affected by the value of the torsion elasticity of the clutch.

**Figure 6 - Simulated graphs (C = 1E-5 rad/Nm)**

**Figure 7 - Simulated graphs (C = 1E-3 rad/Nm)**

**SIMULATING DYNAMICS OF THE LOAD SENSING SYSTEM**

A simulation task for dynamics (calculating transient responses) of the load sensing system is shown in Figure 4. In our simulations, tasks of two types are considered. First, responses to a displacement of the directional valve (the load moment of the hydraulic motor is constant) are calculated. Second, responses to a change of load moment of the hydraulic motor (the displacement of the directional valve is constant) are calculated. Here we consider tasks of the first type.

In the following example, simulations of the dynamics have been carried out under the following operating conditions.

- Load moment of the drive mechanism is 65 Nm.
- Initial displacement of the directional valve is 0.0045 m. Step change -0.0005 m (during 0.01 s) has been applied to the initial displacement of the directional valve.
- Time step is 5 μs.
- Simulated time is 0.5 s (results have been calculated for 100000 points).

Automatic synthesizing time of the simulation algorithm on the PC was 0.435 s. Process computing time on the PC was 32.9 s.

In Figure 8 ... Figure 11 graphs of simulating the dynamic responses are shown.

In Figure 8 pump control valve takes a new position in 0.10 s (3). Pump regulating pressure increases from 1.25 to 1.34 MPa in 0.50 s (2). Position angle of the swash plate drops from 0.170 to 0.152 rad in 0.50 s (1).

In Figure 9 transient response of the angular velocity of the pump (2) damps in 0.20 s. Volumetric flow of the pump (1) drops asymptotically from 0.00076 to 0.00068 m³/s in 0.50 s. Output pressure of the pump (3) dampingly oscillates in 0.30 s.

**Figure 8 - Graphs considering hydraulic-mechanical controller**

**Figure 9 - Graphs considering hydraulic pump**

(t = 0.5 s)

**Figure 10 - Graphs considering hydraulic pump**

(t = 0.055 s)

**Figure 11 - Graph of angular velocity of the drive mechanism**

In Figure 10 graphs from Figure 9 are shown in more short time. As it can be seen, almost all the graphs of transient responses figure oscillations of different frequencies. Therefore, the simulations must be performed using
extremely short time step to get results of reasonable precision.
In Figure 11 graph of angular velocity of the drive mechanism (output of the whole system) is shown. Step displacement -0.0005 m applied to the directional valve causes the angular velocity of the drive mechanism drop from 120.0 to 102.0 rad/s in shape of damped oscillation.

SIZE AND COMPLEXITY
The package for modelling and simulation of dynamics of the load sensing system considered in the paper contains:
- 34 classes, including 24 load sensing system component classes;
- more than 1000 variables;
- 17 variables that have to be iterated during the computations;
- 72 links between system components.
Java code of the automatically constructed algorithm for solving the simulation task of the dynamics of the load sensing system presented above contains 4364 lines and includes 4 algorithms for solving subtasks.

CONCLUSIONS
The multi-pole model of the fluid-power system contains models of functional elements and carries the full information about connections of input/output variables, which express the considered mathematical causalities and guarantees the completeness of the model.
Using the CoCoViLa programming environment enables one in flexible way compose and experiment with various large and complicated models. Automatic program synthesis allows fast solving a great number of simulation tasks in order to find out better solutions to design problems.
Simulation is performed step by step, starting from simulation of components and moving to more complicated systems. Calculations are performed separately for each multi-pole model. Iteration methods are used in cases of loop dependencies that may appear between component models when they are connected together into more complicated ones.
As a result of performed simulations the following conclusions concerning the load sensing system can be pointed out.
- The hydraulic-mechanical controller must have independent feeding with constant pressure.
- All the system components must react fast with optimal damping.
- The clutches must have elasticities of optimal values.
- Diameters of the connecting tubes must be optimal and lengths must be different to avoid resonance.

As a result of the current research, a simulation system is proposed that enables one to perform computer experiments at the first stage of design. Results of simulations can be used as initial data while building of real fluid power systems.
Methodology considered in the paper can be used in modelling and simulation of technical chain systems with any kind of physical content.

ACKNOWLEDGEMENT
This research was in part supported by the Estonian Science Foundation (Grant No. 7091).

REFERENCES

AUTHOR BIOGRAPHIES
GUNNAR GROSSSCHMIDT received his Dipl. Eng. degree from Tallin University of Technology (1953) and Candidate of Technical Science degree (Ph.D) from the Kiev Polytechnic Institute (1959). His scientific interests are concentrated around modelling and simulation of fluid power systems for several decades. He has developed an original system for modelling of fluid power system.
His list of scientific publications contains over than 70 items. He has been lecturer in Tallin University of Technology 50 years, as assistant, lecturer, associated professor and head of the chair of Machine Design.
MAIT HARF received his Dipl. Eng. degree from Tallin University of Technology (1974), and Candidate of Technical Science degree (Ph.D) from the Institute of Cybernetics, Tallinn (1984).
His research interests are concentrated around intelligent software design. He worked on methods for automatic synthesis of programs and their applications to knowledge based programming systems such as PRIZ, C-Priz, ExpertPriz and NUT. His research on developing intelligent software based on using hybrid knowledge bases and structural synthesis of programs is continuing till now.
LOGISTICS AND SUPPLY CHAIN SIMULATION
CUSTOMER DEMAND
FUZZY SIMULATION OPTIMIZATION APPROACH FOR THE TRANSSHIPMENT PROBLEM WITH UNCERTAIN CUSTOMER DEMANDS

Mohamed HMIDEN, Lamjed BEN SAID, Khaled GHEDIRA

Unité de Recherche en Stratégie d’Optimisation des Informations et de la connaissance (SOIE), ISG de Tunis
41 avenue de la liberté cité Bouchoucha 2000 Tunis, Tunisia
Mohamed.hmiden@gmail.com
{lamjed.bensaid, khaled.ghedira}@isg.mnu.tn

ABSTRACT: Transshipment is the transfer of a product among N locations at the same echelon level. The transshipment problem consists to determine the replenishment quantities where a transfer policy is practiced. In this work, we consider the transshipment problem characterized by fuzzy customer demands. Our aim is to estimate the total inventory cost expected value in order to determine the optimal replenishment quantities. To realize this aim we propose: (1) Two fuzzy simulation processes, the first is to calculate the credibility measure and the second is to estimate the total inventory cost expected value, (2) a genetic algorithm including the above fuzzy simulation processes. We experiment several systems in order to study the effect of customer demands uncertainty on transshipment benefits. To realize this, we use fuzzy entropy quantifying the uncertainty level.

KEYWORDS: Transshipment problem, fuzzy customer demands, fuzzy simulation

1- INTRODUCTION

The inventory management is a crucial activity in the supply chain. It allows the regulation of the inventory level to face the unexpected customer demands fluctuation. In fact, a reduced inventory level causes the non satisfaction of customer requirements and an inventory excess generates a holding cost. This task becomes more complex when it concerns several locations. The transshipment is an inventory collaboration method consisting to transfer products between locations (Krishnan and Rao 1965). It has been widely used in practice to reduce inventory costs and to improve the customer fill-rates (Eppen 1979). It provides an effective mechanism for correcting discrepancies between the locations observed customers demands and their available inventories (Herer et al. 2006). The transshipment problem consists to determine the replenishment quantities where a transshipment policy is practiced. It has been extensively studied, where several configurations and parameters are considered. Two transshipment configurations could be distinguished (Tagaras 1989):
- Emergency transshipment: the product transfer is made at the end of the period after observing the customer demands. It is called the emergency transshipment. In this configuration, we can either apply a complete or a partial pooling. In the first case, the quantity to transfer is equal to the minimum between surplus quantity in the location in excess and shortage quantity in the location in need. In the second case, the quantity to transfer is determined by the location in excess.
- Preventive transshipment. The product transfer is made at beginning of a period.

We can identify three transshipment parameter types: (1) Replenishment parameters such as the replenishment lead times (Tagaras 1989) and the replenishment fixed costs (Herer and Rashit 1999). (2) Transfer parameters such as the transshipment mean capacity, (Ozdemir et al. 2006) and the transshipment lead times (Tagaras and Vlachos 2002). (3) Environment parameters such as the number of locations (Robinson 1990), (Tagaras and Cohen 1992), (Week and data 2005), the number of periods (Tagaras 1990) and the customer demands.

In all the above cited works uncertainty had only been considered in customer demands. Generally, it is modelled by probabilistic distributions (Tagaras 1989) derived from the analysis of the historical data. However, these data are not always available or reliable due to the market turbulence and products short cycle life. In these cases, the fuzzy sets are adopted to model the customer demands uncertainty (Petrovic et al. 1999).

Few studies had applied the fuzzy sets to model the uncertainty of the different inventory parameter types. We identify the following uncertain inventory parameter types:
- Supply parameters: the replenishment lead time (dey et al. 2005), the supplier reliability (Petrovic et al. 1998)
- Inventory cost parameters: unit ordering cost (Roy et al. 2008), the unit holding cost (Petrovic et al. 1996) and (Maity and Maiti 2007) and the shortage cost (Ishii and Konno 1998).
- Customer demands parameter: We can classify works interested in this parameter into two categories: (a) the works studied the inventory problem for a single location. These works are looking for the optimal inventory quantities when the customer demand is fuzzy, (b) the works studying the inventory problem for the whole supply chain. This category is interested in finding a coordination or cooperation policy between the supply chain entities that helps to reduce the inventory costs when the customer demands are fuzzy.

In this paper, we are interested in the transshipment problem where the location customer demands are fuzzy. Our
objectives are to model the transshipment problem with fuzzy customer demands, to design two fuzzy simulation processes to estimate the total inventory cost expected value, to derive the optimal replenishment quantities and to evaluate the customer demands uncertainty effects on the transshipment benefits.

The remainder of the paper is organised as follows: In the second section, we present the studied problem and the used notations. In the third section, we present the fuzzy simulation processes. In the fourth section we present the experimental results and we finish this paper by a conclusion and future works.

2- PROBLEM DESCRIPTION

2.1- The model

In our work, we consider the transshipment problem characterized by N locations having non-identical cost structures. These locations are supplied at the beginning of a period and they satisfy a fuzzy customer demands. During a period, the sequence of events represented by figure 1 is:

a- At the beginning of the period, for each location L_i the replenishment quantity S_i is received.

b- During the period, each location L_i observes its customer demands.

c- At the end of the period, the transshipment decision is made and executed then the customer demands are satisfied.

This sequence of events is adopted by ozdemir, (ozdemir et al. 2006) for multi-period configuration.

![Figure 1: Period sequence events](image)

Throughout this paper, we adopt the following notations and concepts:

- \( S = (S_1, S_2, ..., S_n) \) : a vector of replenishment quantities. Where, \( S_i \) represents the replenishment quantity at location \( L_i \).

- \( h_i \) : unit holding cost at location \( L_i \).

- \( p_i \) : unit shortage cost at location \( L_i \).

- \( C_0 \) : unit transshipment cost from location \( L_i \) to location \( L_j \). We assume that this cost is symmetric. Which means that \( C_0 = C_{ij} \) for all locations \( L_i \) and \( L_j \).

- \( \tilde{D} = (\tilde{D}_1, \tilde{D}_2, ..., \tilde{D}_n) \) : vector of customer demands. Where, \( D_i \) represents the customer demands in the location \( L_i \). The customer demands in each location can be specified based on managerial experience and subjective judgement. It could be expressed with linguistic expressions, for example \( D_i \) is in the interval \((b_i, c_i]\), this kind of expression is modelled by trapezoidal or triangular fuzzy sets defined respectively by the quadruplet \((a_i, b_i, c_i, d_i)\) or the triplet \((a_i, b_i, c_i)\) (Petrovic et al. 1999) and (Wang and Shu 2007) (Chiang and Wen-Kai 2002) presented in figure 2.

\[ \mu(\tilde{D}) = \min(\mu_1(\tilde{D}_1), \mu_2(\tilde{D}_2), ..., \mu_n(\tilde{D}_n)) \text{ : membership degree of the fuzzy customer demands vector.} \]

- \( \text{Pos}(\tilde{D}_i \geq r) = \sup_{x_i} \mu_i(x) \text{ : the possibility degree that the fuzzy customer demand in location } L_i \text{ is greater than a given value } r. \]

- \( N(\tilde{D}_i \geq r) = 1 - \sup_{x_i} \mu_i(x) \text{ : the necessity degree that the fuzzy demand at location } L_i \text{ is greater than a given value } r. \]

\[ \text{Cr}(\tilde{D}_i \geq r) = \frac{1}{2} \left( \text{Pos}(\tilde{D}_i \geq r) + N(\tilde{D}_i \geq r) \right) \text{ the credibility degree that the fuzzy demand at location } L_i \text{ is greater than a given value } r. \]

![Figure 2: Trapezoidal and triangular fuzzy demands](image)

2.2- The inventory costs

The transshipment problem consists to determine the replenishment quantities vector that minimizes the inventory costs function while a transshipment policy is practiced. At the end of the period, this function is given by:

\[ C(S, \tilde{D}) = \sum_{i=1}^{n} h_i |S_i - \tilde{D}_i| + \sum_{i=1}^{n} p_i |\tilde{D}_i - S_i| - k(S, \tilde{D}) \quad (0) \]

The first and the second terms represent respectively the holding and the shortage costs, evaluated before the transshipment realization. The third term represents the profit achieved by the transshipment as developed in the linear program (2).

\[ K(S, \tilde{D}) = \max_{x_i} \sum_{i=1}^{n} \left( h_i + p_i - C_0 \right) x_i \quad (2) \]

Subject to:

\[ \sum_{i=1}^{n} x_i \leq S_i - \tilde{D}_i, \forall i \in I \quad (3) \]

\[ \sum_{i=1}^{n} x_i \leq S_j - \tilde{D}_j, \forall j \in I \quad (4) \]

\[ x_i \geq 0 \quad (5) \]
This linear program, on the one hand shows that for each unit transshipped from location \( L_1 \) belonging to the set of locations in excess \( L^+ \), to location \( L_2 \) belonging to the set of locations in need \( L^− \), we decrease respectively the holding costs at \( L_1 \) by \( h \) and the shortage costs at \( L_2 \) by \( p \). On the other hand, the total inventory cost is increased by \( C \) as the transshipment cost.

The transshipment policy adopted here is the complete pooling proposed by Tagaras (Tagaras 1999). According to this policy, the transshipped quantity among locations is equal to the minimum between the surplus quantity at location in excess and the shortage one at location in need. In (2), \( X \) denotes the optimal quantity that should be transshipped from location \( L_1 \) in excess to location \( L_2 \) in need to satisfy its unmet demands. For each location \( L_i \) in excess, we have the constraint (3) that restricts the transshipped quantity to its surplus inventory. Whereas, the constraint (4) concerns each location \( L_j \) in need. It restricts the transshipped quantity to its shortage quantity. This linear program could be resolved by the simplex method.

### 2.3- Inventory costs properties

Given that the demand is fuzzy, the total inventory cost \( C(S, D) \) is imprecise. Thus, (1) could be modelled by fuzzy sets whose membership functions are unknown since they depend on the decision vector \( S \) of replenishment quantities and the fuzzy vector customer demands \( D \). However, we can compute its expected value. In this case, the transshipment problem consists to determine the replenishment quantities that minimize the expected value of the total inventory cost given by (6).

\[
\min \mathbb{E}(C(S, D)) \tag{6}
\]

The expected value of this fuzzy variable is given by:

\[
\mathbb{E}[C(S, D)] = \int \mathbb{E}[C(S, \tilde{D})] \, dr - \int \mathbb{E}[C(S, \tilde{D})] \, dr \tag{7}
\]

As the total inventory cost is positive thus the second term on the right hand of (7) is null. The expected value of the inventory costs is equal to the first term. The credibility measure of the fuzzy event indicating that the total inventory cost is greater than a predefined value \( r \) is given by (8)

\[
\tilde{C}(C(S, \tilde{D}) \geq r) = \frac{1}{2} \max(\mu(\tilde{D}), C(S, \tilde{D}) \leq r) + \frac{1}{2} \min(1-\mu(\tilde{D}), C(S, \tilde{D}) \leq r) \tag{8}
\]

Where \( \tilde{D} \) is a vector demand randomly generated.

\[
\tilde{D}_k = \tilde{D}_{1k} \wedge \tilde{D}_{2k} \wedge ... \wedge \tilde{D}_{nk}
\]

### 3- FUZZY SIMULATION

The fuzzy simulation technique is introduced by Liu and Iwamura (Liu and Iwamura 1998) to estimate the value of an uncertain function. We propose a simulation process to approximate the expected value of the total inventory costs for a particular replenishment vector \( S \). Since the expected value model is expressed with the credibility measure of fuzzy events, we begin by approximating the value of this measure.

#### 3.1- Process to estimate the credibility measure

To approximate the value of \( C(S, \tilde{D}_k) \), we propose the fuzzy simulation process. As shown in figure 3, the proposed credibility measure simulation process is composed mainly by three steps repeated \( M \) times. The first step consists to generate randomly the customer demands vector, apply the complete pooling transshipment policy and calculate the total inventory cost. The second step consists to compare the total inventory cost with a value \( r \) given as a parameter. According to the result of the third step, the possibility measure of the fuzzy event or the necessity measure of the same fuzzy event is updated. Once the number of iteration is reached the credibility measure is calculated.

#### 3.2- Process to estimate the expected value

To approximate the expected value of the total inventory cost \( \mathbb{E}[C(S, \tilde{D})] \), for a particular replenishment vector \( S \), we propose the fuzzy simulation process presented in the figure 4. This simulation process, approximating the total inventory cost expected value, is composed by principally two steps. The first step objective is to estimate the minimum and maximum values of the total inventory cost. This step is divided into two sub-steps. In the first sub-step the following events are executed: (a) the demand is randomly generated, (b) apply the complete pooling transshipment policy and (c) calculate the total inventory cost. In the second sub-step, the minimum or the maximum total inventory cost is updated. This step is repeated \( M_t \) times. The second step of the simulation process aims to approximate the total inventory cost expected value. It is
composed by two sub-steps repeated \( M_2 \) times. In the first sub-step, a real value \( r \) is randomly generated from an interval bounded by the minimum and maximum total inventory cost previously estimated. After that, the credibility measure of the fuzzy event indicating that the total inventory cost is greater than \( r \) is estimated by the credibility measure simulation process. The second sub-step consists of updating the total inventory cost expected value. The over of this step is iterated \( M_2 \) times.

![Figure 4: Expected value simulation diagram.](image)

This simulation process framework could be adapted in order to approximate the transshipment benefits for a particular replenishment vector \( S^i \). In fact, the same simulation process steps are used. However, instead of calculating the total inventory cost and its credibility measure, we calculate the transshipment benefits and its credibility measure.

### 4-GENETIC ALGORITHM

In order to derive the optimal replenishment vector that minimizes the total inventory cost, we use a genetic algorithm (Holland 1975). This optimization technique consists of deriving the optimal solution obtained by applying iteratively genetic operations to an initial population. It is characterized by the following steps:

- **Initialization:** this operation consists of initializing the initial population. We use a population of 30 individuals real coded. Each individual is formed by \( N \) genes, which represent a replenishment vector. The individuals are randomly generated.

- **Evaluation:** this operation consists of determining the fitness value for each individual. In our study, the fitness value represents the expected value of the total inventory costs, obtained by applying the expected value fuzzy simulation process.

- **Selection:** after evaluating the fitness of each individual and sorting the individual according to their fitness value, we select the fittest ones to reproduce and form the population of the next generation. In our case, the best individuals represent the replenishment vector \( \{ S^* \} \) that ensure low total inventory costs. We have chosen a deterministic selection procedure which consists of reproducing the best 10% of them to the mating pool.

- **Crossover:** this operation consists of combining genes from different parents to produce children. We have chosen the binary tournament selection to pick out parents for reproduction. Tournament selection runs a tournament between two randomly chosen individuals and selects the winner (individual with best fitness value).

- **Mutation:** this operation consists of mutating a selected individual by modifying one gene, which is selected randomly.

### 5-EXPERIMENTATION

In this section, we aim to analyse the effects of customer demand uncertainty, modelled by fuzzy sets, on the total inventory costs and the transshipment benefits. Since the sum of independent trapezoidal fuzzy sets is a trapezoidal fuzzy sets (Liu and Iwamura 1998), the system customer demands is modelled also by a trapezoidal fuzzy sets defined by the following quadruplet:

\[
(-a, a, b, c, d)
\]

In order to measure the location and the system customer demands uncertainty, we apply the fuzzy entropy (Liu 2007). The fuzzy entropy FE for the location \( L \) noted \( H(\tilde{D}) \) is given:

\[
H(\tilde{D}) = \frac{(d - a) + (\ln 2 - 0.5)(c - b)}{2}
\]

The fuzzy entropy for the system noted \( H(\tilde{D}) \) is given by:

\[
H(\tilde{D}) = \frac{(d - a)}{2} + (\ln 2 - 0.5)(c - b)
\]

To assess the effects of customer demands on transshipment benefits and total inventory cost, we identify four system configurations that differ by their customer demands. We assume that all the locations in the identified systems have a holding cost between 5 and 8 and a shortage cost equal to 25. For all experimentation, we look for the replenishment vector that achieves a high service level.
Once the optimal replenishment vector is determined, we compare between the systems transshipment benefits. The figure 6 shows that we obtain the maximum of transshipment benefits for the system $S_3$ and the minimum one for the system $S_1$. In fact, $S_4$ have the highest system customer demands uncertainty (FE=30.16) and $S_1$ have the lowest one (FE=15.08). However, the systems $S_3$ and $S_4$ have nearly the same system FE, the transshipment benefits is clearly important for the system $S_3$ having different location fuzzy entropy.

According to our experimentations we can conclude that the transshipment is more profitable for a system having high system FE and different location FE.

<table>
<thead>
<tr>
<th>Systems</th>
<th>Location number</th>
<th>Customer demand</th>
<th>Location FE</th>
<th>System FE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_1$</td>
<td>$L_1$</td>
<td>3, 4, 8, 9</td>
<td>3.77</td>
<td>15.08</td>
</tr>
<tr>
<td></td>
<td>$L_2$</td>
<td>4, 5, 9, 10</td>
<td>3.77</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$L_3$</td>
<td>5, 6, 10, 11</td>
<td>3.77</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$L_4$</td>
<td>6, 7, 11, 12</td>
<td>3.77</td>
<td></td>
</tr>
<tr>
<td>$S_2$</td>
<td>$L_1$</td>
<td>2, 3, 9, 10</td>
<td>5.15</td>
<td>25.60</td>
</tr>
<tr>
<td></td>
<td>$L_2$</td>
<td>3, 4, 10, 11</td>
<td>5.15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$L_3$</td>
<td>4, 5, 11, 12</td>
<td>5.15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$L_4$</td>
<td>5, 6, 12, 13</td>
<td>5.15</td>
<td></td>
</tr>
<tr>
<td>$S_3$</td>
<td>$L_1$</td>
<td>1, 4, 12, 16</td>
<td>9.05</td>
<td>25.52</td>
</tr>
<tr>
<td></td>
<td>$L_2$</td>
<td>2, 3, 9, 10</td>
<td>5.15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$L_3$</td>
<td>3, 4, 8, 9</td>
<td>3.77</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$L_4$</td>
<td>4, 6, 14, 16</td>
<td>7.54</td>
<td></td>
</tr>
<tr>
<td>$S_4$</td>
<td>$L_1$</td>
<td>1, 3, 11, 13</td>
<td>7.54</td>
<td>30.16</td>
</tr>
<tr>
<td></td>
<td>$L_2$</td>
<td>2, 4, 12, 14</td>
<td>7.54</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$L_3$</td>
<td>3, 5, 13, 15</td>
<td>7.54</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$L_4$</td>
<td>4, 6, 14, 16</td>
<td>7.54</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: System configurations
The four systems $S_1$, $S_2$, $S_3$ and $S_4$ have different customer demands uncertainty level, which is evaluated by their fuzzy entropy FE. We distinguish between two fuzzy entropies, the location customer demands FE and the system customer demands FE. The first one is calculated based on location customer demands and the second one is calculated based on system customer demands. In the systems $S_1$, $S_2$ and $S_3$ we have the same location FEs which are respectively equal to 3.77, 5.15 and 7.51 and in the system $S_4$ the location FEs are different. The systems $S_2$ and $S_3$ have nearly the same system FE.

In order to compare between the transshipment benefits for the four systems, firstly we derive the optimal replenishment vector for each system. Secondly, we estimate the transshipment benefits expected value for each system. As shown in figure 5, we have nearly the same total replenishment quantity for all the systems. This could be explained by the fact that the total replenishment quantity is influenced by the system customer demands expected value. In fact, all the systems have the same customer demands expected value which is equal to 30. The optimal total replenishment vector for each system is almost equal to the system customer demands expected value.

![Figure 5: Systems total replenishment quantity.](image)

![Figure 6: Systems transshipment benefits](image)

6- CONCLUSION

We have considered the transshipment problem characterized by a fuzzy customer demands. Our objectives were to determine the replenishment quantities and the expected transshipment benefits. We have proposed two simulation processes: the first is to approximate the value of the credibility measure for a fuzzy event. The second is to estimate the total inventory cost expected value. In order to derive the optimal replenishment vector, we have applied a genetic algorithm where we call the second simulation process to evaluate each individual.

Our experimentations showed that the optimal replenishment vector is influenced by the system customer demands expected value. That means, the total replenishment quantity is nearly equal to the system customer demands. The transshipment is more profitable for a system having a different and a high location customer demands.

Our work could be extended by considering other transshipment problem parameters uncertainty. We can consider for example the replenishment lead time and the inventory costs as fuzzy.

REFERENCES


Wang, J. and Y. F. Shu. 2007. A possibilistic decision model for new product supply chain design. European journal of operational research. 177, 1044–1061

PROPOSITION OF A SIMULATION TOOL FOR A COOPERATIVE-EFFECTIVE ATP FUNCTION

Mariem Siala, Khaled Ghédira*, Jean-Pierre Campagne**
SOIE/ISG/ Tunis University, Faculty of Sciences, Tunis, Tunisia
E-mail: mariem.siala@fst.rnu.tn
*SOIE/ISG/ENSI; ENSI/ Manouba University, 2010 Campus Universitaire la Manouba, Tunisia
**INSA Lyon, Industrial Engineering Department, Bat. Jules Verne 19,69621 VILLEURBANNE Cedex, France

KEYWORDS
Distribution network, ATP function, multi-agent system, cooperation, coordination

ABSTRACT
The Available to Promise (ATP) function presents one of the most important functions within the firm and the supply chain. According to its appellation, the first role of the ATP function is to provide customers’ requests with the products’ availability information. However, under the pressure of a highly competitive business environment, the companies need a decision support tool to satisfy at best the customers’ orders according to their available stocks to fit their supplies to a variable demand. Thus, the purpose of this paper is to improve this ATP function, within a multi-echelon distribution network, to be able to adapt to this dynamic environment and to the multiple stock localizations. We suggest the Cooperative ATP function (C-EATPE) which is based on a multi-agent approach, using the “Contract Net” protocol. In order to assess these functions, we have developed a simulation tool which allowed us to show the efficiency of the suggested functions as well as to study the impact of the system and the agents’ cooperation degree on the global and individual performances and so, its impact on the provided customer service.

THE ATP FUNCTION
The Available-to-Promise (ATP) function is an important element to manage the relationship between a company and its current and prospective customers. As its name suggests, the initial role of this function is to provide product availability information in order to promise customers' order requests. Traditionally ATP, called here «Conventional » ATP, represents nothing more than a simple consultation in the Master Production Schedule (MPS). It is unable neither to match supply with variable demand nor to operate on a short-term environment. Therefore, and under the competitiveness pressure, the ATP function becomes a critical component which requires more complex models. It has to reduce discrepancy between company supply or production and customers’ demand. Firms need smart decision support tools to face the supply-chain paradigm.

In literature, several articles emphasize the importance of ATP function. However, few works give quantitative models to support the ATP function.
In the supply chain, we can mainly distinguish a work of (Kügler and Schneeweiss 2000) which defines an ATP search procedure along three dimensions: the time dimension, the customer dimension and the product dimension. Yet, it is a central approach in which the research of the ATP quantities is made in the nodes without communication or cooperation between the different entities. (Jeong et al. 2001) suggested a batch AATP (Advance ATP) for the TFT Displays assembly. In this ATP system, only headquarter processes the data and gives a response to customers. This response is given in a batch mode, which means that orders are collected over a predefined length batching interval and processed together by a model or algorithm. Then, at the end of this interval, the response is given to customers. So, the ATP function is periodically executed (e.g. at the end of each day or week).
Several contributions for batch ATP have been achieved by (Chen et al. 2001), (Chen et al. 2002,) and (Ball et al. 2004). They developed mixed inter-programming models to allocate available components and production capacity to customer orders within a batching interval. (Zhao et al. 2005) adapted this model to the specific requirements involved by an electronic product supply chain within Toshiba Corporation. (Xiong and al. 2003a) proposed Web enhanced dynamic BOM-based ATP system then (Xiong and al. 2003b) extended this work by the “Web-based flexible available-to.promise computation system”.
In these works, the orders are received in the headquarters but not directly in the distribution centers which depend on these headquarters and can’t take decision. This doesn’t represent reality since, usually; the orders arrive directly in distribution centers.
If the order can not be fulfilled in the requested due date and quantity, all these works trigger new productions whereas in our work, we consider cooperation between the different entities to replenishment rather than trigger production. All these models don't handle demand variation and don't make some forecasting on the future demands but expect the receipt of the orders to find the solutions to the customers. We also note that these works present methods that can be applied to a particular company-specific setting only. Solely (Pibernik 2005) proposed two generic AATP models.
We don’t found any work that models the ATP function with Multi-Agents Systems. None of these works has also proceeded to the simulation of the proposed ATP function. Motivated by new challenges, the available to promise (ATP) function has migrated from a set of availability records in a master production schedule (MPS) toward a real-time decision support system and possibly even toward an automated intelligent-agent based negotiation infrastructure (Zweben 1996).

A GLOBAL APPROACH
A considered Supply Chain delivers a standard class of products. It's a network organization with autonomous or
semi-autonomous entities. These entities decide to coordinate their activities in order to deliver the products to the customers. Thus, our goal is not only to optimize the overall profit of the chain but also to improve the customer service and to maximize the individual profit of each entity.

Our framework only manages the finished products. So, the considered chain is assimilated to a hierarchic distribution network. More particularly, it is a divergent multi-echelon system, characterized by the property that each entity is supplied by exactly one other entity, and supplies one or more entities.

To answer to customer’s request, we propose an Advanced ATP function called C-EATP (Cooperative-Effective ATP function). This function is executed by each entity as soon as a new order arrives which implies that it is operated in a real-time mode. The other advantage of the C-EATP function that it has a multi-sites scope with a distributed execution. So, it is able to support the character naturally distributed of the supply chain since the orders are processed directly in the distribution centers, not in a centralized manner. This offers the centers an independent decision and a right data privatization.

We note also, that, as response flexibility, we can propose to customer back orders.

The given response to the customer is depended on order’s due date. Indeed, in case of an urgent order, the retailer would be unable to replenish its supplies and thus, a local solution must be found. If not, the retailer has to launch a new procedure in order to replenish the supplies from its superior distribution center. Hence, each entity, assimilated to a distribution center or to a retailer, divides its Master Plan on three horizons: The Short Term, the Mid Term and the Long Term.

For the Short term horizon, we proposed the EATP (Effective Available To Promise) function which ensures a better management of local stock. This function is based on booking of pseudo-orders which are determined with a new approach of estimating current demand. This approach was presented in (Siala et al. 2006a), (Siala et al. 2006b). Several experimentations has been made and the results demonstrated the effectiveness of the EATP function that gives an immediate response to the customers, privileges the rush orders and minimizes the number of backorders.

In the mid Term each Distribution Center can replenish from its direct supplier (its higher node) but where the latter can’t replenish. So, we suggest the Cooperative Effective ATP function (C-EATP) which is based on a multi-agent approach, using the “Contract Net” protocol, to assure the inventory redeployment between the different entities of the distribution network. We will detail this function below.

In the long term horizon, to satisfy the customer, we can trigger new productions. We are not concerned by this horizon in this article.

THE C-EATP FUNCTION

Before breaking down the multi-agent model suggested for the mid-term horizon, we break down the global progress of the C-EATP function.

When a retailer receives a customer order, he will try to provide him from his local stock by executing the Effective ATP Function. But if he does not have the required quantities, he will check if the order belongs to the mid term horizon. In that case, he can trigger a replenishment demand to the distribution center which is directly superior to him.

During this horizon, the distribution center can not stock up again from its superior node, so it has to meet the retailer’s demand from its local stock or from the stock or expected deliveries of its sons nodes. That is what we call a redeployment/replenishment process. During this phase, we redeploy the expected deliveries on the different retailers according to their current needs (figure 1). We also note that in order to minimize the logistical costs, the replenishment is held at delivery dates that are already fixed, all the distribution center has to do is to increase the required quantities during these periods.

![Figure 1: The redeployment process](image)

We note here that the retailer does not have to provide directly a retailer in stock shortage but to give him quantities that had been intended for him and that, in usual way, would had been delivered in the future by the distribution center agent. That is why there is no transshipment in this phase but only redeployment inventory.

Each unit constituting the distribution network will be assimilated to an agent. Thus, the model is made up of two types of agents: the Retailer agents and the Distribution Center agents. In this phase, we will only consider only one Distribution Center agent and its different Retailer agents.

**Multi-Agent Framework**

The model is composed of two agent classes: Retailer agent and Distribution Center (DC) agent.

A Retailer agent $\text{Ret}_i$ has no static knowledge but it has dynamic knowledge representing the master plan composed by: $R_{i,k}$ (Scheduled Receipts for period i), $O_{i,k}$ (promised orders for period i), $E_{i,k}$ (Estimated demand for period i); $B_{i,k}$ (Booking of pseudo-orders for period i), $EATP_{i,k}$ (Effective ATP for period i)

Its only contact is a Distribution center.

The table 1 gives an example of a master plan of each agent.

---

242
Table 1. The master plan of each agent

<table>
<thead>
<tr>
<th>Periods</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orders</td>
<td>70</td>
<td>36</td>
<td>44</td>
<td>25</td>
<td>27</td>
<td>30</td>
<td>30</td>
<td>16</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>Scheduled Receipts</td>
<td>190</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>80</td>
<td>150</td>
<td>0</td>
</tr>
<tr>
<td>Pseudo-orders</td>
<td>0</td>
<td>4</td>
<td>11</td>
<td>11</td>
<td>18</td>
<td>30</td>
<td>45</td>
<td>37</td>
<td>60</td>
<td>45</td>
</tr>
<tr>
<td>Effective ATP</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>22</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The Distribution Center (DC) agent has the same dynamic knowledge, concerning the master plan, that the Retailer agent. We note that this agent has no knowledge on the stock and commands of Retailer agents. It has for contact all Retailer agents.

The inventory redeployment is one of the key points of the C-EATP function. It is triggered as soon as a received order can not be satisfied in its due date by the stock on hand. Each order, in each retailer, is characterized by its due date \( d_{j,k} \), its quantity \( q_{j,k} \) and its realization date \( d'_{j,k} \). The role of this process is to minimize realization date \( d'_{j,k} \) of all orders.

We will detail below this process.

When an order \( O_{j,k} \) arrives in the retailer \( Ret_k \):

**a- Order satisfaction with a local stock:***

1. The Retailer Agent verify whether the enquiry order \( O_{i,j} \) can be satisfied by \( B_{j} \) (the Booking of pseudo-orders for period \( i \)). If \( B_{j} \) is less than \( q_{i,j} \) then the Retailer Agent searches \( (q_{i,j} - B_{j}) \) in EATP of preceding periods.
2. If \( O_{i,j} \) can be fulfilled then go to Step 3 else go to Step 4.
3. Give a positive response to customer inquiry.

**b- Order satisfaction with a Distribution Center local stock:***

In this phase, Retailer agent will ask the DC agent to replenish. It will introduce the emergency parameter in the request. Indeed, for satisfying the order, the Retailer agent has sought the required quantity in the Effective ATP but not in the booking of pseudo-orders. Thus, it will send the redeployment demand with a medium emergency.

4. Retailer Agent sends a message of replenishment to DC Agent with medium Urgency.
5. If DC Agent can fulfill the retailer demand, it sends it a positive response, go to Step 3. Otherwise, it triggers a replenishment process to the other Retailers Agents with the shortage.

**c- Redeployment demand based on Contract Net Protocol:***

6. Each Retailer Agent that receives an order from DC consults its EATP and proposes the requisite quantity or the maximum of quantity that it can propose. It has also to give a virtual price of the offer to DC Agent.
7. When DC Agent receives several positive responses, it selects the best offer.
8. If it receives all responses with a quantity lower than the request quantity, the DC collects the order quantity from different retailers.
9. If the research quantity is found in the network, the DC Agent sends a positive message to the Retailer else it proposes the maximum quantity that it found.
10. If the suggested quantity is equal to the required quantity, the Retailer gives a positive response to the customer else the retailer searches the remaining quantity in the reservations of pseudo-orders of the previous periods.

**d- Second redeployment demand***

10. If there is always a shortage, a Retailer Agent sends to the DC Agent a message of replenishment with a maximum urgency. This process will be identical to the previous, but instead of consulting the Effective ATP, the retailer must respond by consulting its pseudo-orders’ reservation.
11. If there is always a shortage, a Retailer suggests a backorder to the customer.

**The Determination Of Virtual Price**

The retailer selection’s process is based on CNP. But how to do it without using price? Indeed, pricing is not considered as a variable in this paper.

At several offers’ reception, the DC will choose a retailer according to a virtual price. This price depends on a function established according to EATP as well as pseudo-orders reservations.

\[
\hat{f} = \left( \sum_{i=0}^{N-1} EATP(i) \right) + 1/2 \left( \sum_{i=0}^{N-1} B(i) \right)
\]

We note that we multiply \( \sum_{i=0}^{N-1} B(i) \) by 1/2 to give more importance to the Effective ATP.

Therefore, when a DC request arrives to the retailers, each retailer calculates the virtual price then proposes its offer with the requisite quantity. Certainly, the DC Agent will choose the retailer Agent which offers the higher virtual price.

**SIMULATION TOOL**

**Features**

We have developed a multi-agent simulation tool in order to evaluate the performance of the C-EATP function. This tool has been developed so as to be parametrizable differently according to the situation to be studied. Thus, the tool’s aim is no longer only to study the effectiveness of our model but also to be used as a simulation platform which is able to study the best strategy to be adopted by the agents belonging to a distribution network and aiming to meet the customers’ demand facing different situations.

The simulation tool is developed in “Objective C” language in the multi-agent software platform “ Swarm”.

Swarm is a collection of object oriented software libraries which provide support for simulation programming. Users build simulations by incorporating Swarm library objects in their own programs. This platform has the following features [Stefansson 99]:

- A Swarm simulation proceeds in discrete time steps. So, agents are created and then interact according to a scheduling mechanism. This functionality allowed us to manage the periods within the master plan and to simulate the discreet passage over time from a period to another. It
also allowed the agents to react to the discreet events represented by the arrival of new orders.
- All agents and objects in Swarm can be probed. Probe attaches itself to an agent, can send a message, change a variable or retrieve values by calling agent or reading variable directly. This allowed us to enter the different variables related to the master plans of the different agents.
- Swarm provides a number of classes and protocols which generate a graphical user interface (GUI) to the user running a Swarm simulation. Thus, we managed to follow graphically the progress of the performance system indicators at each step of the simulation.
- It includes a generator of random numbers which allowed us to generate randomly the due date and the quantity of each new order.

**Determination Of The System’s Cooperation Degree**

The simulation tool can be used as a decision support system, which helps to find the best behavior to adopt by the agents facing different situations.
We can find, within the distribution network, a group (or a coalition) of cooperative agents that use the C-ETAP function and that consequently accept to use the redeployment process in case of shortages and another group made up of non cooperative agents.
The System’s cooperation degree depends on the number of cooperative agents (figure 2).

![Distribution Center Agent](image)

**Figure 2: The System’s Cooperation Degree**

According to this number, we define a system as:
- Totally cooperative: if 100% of agents are cooperative
- Very cooperative: if the rate of cooperative agents is included in the interval [75%, 100%]
- Cooperative: if the rate of cooperative agents is included in the interval [50%, 75%]
- Not very cooperative: if the rate of cooperative agents is included in the interval [0%, 50]
- Non Cooperative: if all agents are non cooperative
For example, in the figure, the system is cooperative since 4 agents among 8 are cooperative.

**SIMULATIONS**

**Evaluation Criterion**

Since the purpose of the C-ETAP function is to improve customer service, we want to minimize delays of customers’ promised orders. So we adopted the criterion for measuring the sum of orders’ delays. At each simulation step, the tool can measure the global performance i.e. the delays’ sum of all proposed orders in the distribution network and the individual agents’ performances.

**Simulations According The System’s Cooperation Degree**

In order to study the impact of the system's cooperation degree on the agents' global and individual performance, we have considered 6 identical agents, i.e agents which all have the same supply plan. Each agent can not meet alone the whole real demand. Thus, the expected ATPE quantities are lower than those of the customers' reel orders.
We carried out about fifteen tests for each combination and each type of system.
As showed by Figure 3, the global performance improved according to the cooperation degree.

![A global delays’ sum](image)

**Figure 3. Aglobal delays’ sum**

We thereafter compared the individual performances of a cooperative agent according to its belonging to a totally cooperative, very cooperative, cooperative and not very cooperative system. We could not conclude anything in these cases. In fact, we notice that the performance of a cooperative agent belonging to a non cooperative system (i.e the agent cooperates with just one other agent) is sometimes better than the one of a cooperative agent belonging to a very cooperative system (cooperating with three other agents) and vice-versa.

![delays’ sum for agent 1](image)

**Figure 4. Delays’ sum for agent 1**


**BIOGRAPHY**

Mariem Siala is PhD student and research assistant in the sciences university in Tunis. His mainly interested is the supply chain and Multi Agent systems.

Khaled Ghedira is a director of the ENSI (national school of computer sciences) and professor at the higher institute of management in Tunis. His main research interests are: Multi Agent Systems, Production management and constraint satisfaction problems.

Jean-Pierre Campagne is the head of the Industrial Engineering Department at INSA, Lyon. His main research interests are logistics systems and production management.
MULTI-PRODUCT SIMULATION
AN OVERVIEW OF MULTILEVEL AND HIERARCHICAL METHODS FOR DISCRETE-EVENT SIMULATION OF COMPLEX SYSTEMS

Esra E. Aleisa
Department of Industrial Engineering and Management Systems
Kuwait University
206, 8 KH, Khaldiya Campus
Kuwait
E-mail: aleisa@kuniv.edu

KEYWORDS
Discrete-event simulation, Multilevel, hierarchies, abstraction, Artificial Intelligence

ABSTRACT
Discrete-event simulation of complex systems contains numerous details of varying criticalities. Considering fundamental features of the model and inferior details alike will result in an extensive waste of time and effort. Details should be introduced gradually as appropriate, based on their significance, relevant to the problem context. This was the motivation that drove researchers to consider multilevel, hierarchical or abstraction methods to simplify complex and large-scale simulation projects. Methodologies developed in this regard are by large multidisciplinary, however, there is significant lack in the literature in analyzing and comparing these efforts to one another. The objective of this paper is threefold to: compare the literature of the most persistent approaches for simulation multilevel, hierarchical and abstract modeling, identify the properties that ensure effective development of these approaches, and accordingly, to establish guidelines for systematic development of these approaches.

INTRODUCTION
Complex simulation models result due to the extensive size or stochastic nature of the problem, as well as complications in the behaviors of the system. Hierarchies and abstraction were requisitely applied to make models more understandable, easier to implement, analyze, validate (Salt 1993; Ball 1998) and modify (Chwif et al. 2000; McGraw and MacDonald 2000). Moreover, hierarchies were applied to enable building simulation models before difficult details of the system are available, reduce simulation computational requirements and allow faster executions (Ball 1998). In addition, hierarchical modeling assists in developing intelligent decisions at an early stage of the simulation model, where interactions are less costly (Lu and Tcheng 1991). Moreover, hierarchical modeling has facilitated parallel programming (Chen and Ghosh 1997; Kiran et al. 2001) and model reusability (Zeigler 1987; Pidd and Castro 1998; McGraw and MacDonald 2000). Due to these advantages hierarchical, multilevel and abstract modeling (HMA) were repeatedly applied in the literature of discrete-event simulation. These terms were used interchangeably within the highly interdisciplinary research in this area. Despite the popularity of HMA methods, there is a lack in the literature in comparing and contrasting these efforts. In addition, the properties of effective HMA are rarely identified and well understood. Which resulted in the lack in developing systematic procedures to assess whether or not a specific HMA methodology genuinely simplifies complex models or otherwise, defeats its purpose. In this paper, we will compare persistent applications of simulation hierarchical modeling, in terms of their representation scheme, general approach and methods for integrating submodels. Next, we will associate abstraction methods to hierarchies and identify properties that will improve the efficiency of a hierarchical simulation study. Finally, we propose general guidelines to systematically structure and formalize HMA methodologies simulation. The established guidelines devise that effective HMA should consist of four phases: problem formulation, abstraction method selection, hierarchical model building and hierarchical model execution. Additionally, in this paper, we shall also show that the efficiency of the HMA methodologies is largely restrained by the efficiency of the underlying abstraction method used and the domain to which it is applied.

LITERATURE REVIEW
The majority of research on HMA modeling by large is interdisciplinary. The applications vary according to the purpose of the study, the behavior of the system being simulated, the source of the model complexity and the representation scheme used. These factors also affect the underlying abstraction process and the interface used to integrate sub-models.

Table 1 compares several HMA modeling applications with respect to their representation of primitive components, general approach and the interface used to manage submodels. From Table 1, one can observe that the most embraced approaches to simulation hierarchical modeling involve integration of concepts from object-oriented methodology (OOM) and artificial intelligence (AI). OOM and AI-based abstraction and hierarchical modeling will be discussed in greater depth in the following sections.
OBJECT-ORIENTED METHODOLOGY AS AN APPROACH FOR HMA MODELING

The popularity of Object-oriented simulation (OOS) has increased, this occurred as the deficiencies of procedural simulation software began to surface and lag behind. Specifically, in their lack of ability to expand simulation models hierarchically, to implement event routines on different levels, and in experiencing difficulty in model extension and maintenance (Sharma and Rose 1988). In addition, procedural simulation languages lack a unified conceptual and detailed design framework (Mize et al. 1992; Carr and Balci 2000), which requires model transformation that is error-prone (Lin et al. 1996). These shortcomings are rooted to the underlying basis procedural language used such as FORTRAN, which has propagated to limit the simulation languages upon which it is based, such as SLAM and SIMAN. Therefore, using OOS helps solve the root problems that prevent efficiency in hierarchical simulation methodology. Table 2 compares some OOS applications and tools.

In spite of the fact that OOM facilitates simulation hierarchical modeling, it requires considerable standardization across applications. Also it involves the associated overheads (Narayanan et al. 1998) of learning object-oriented programming principles and requires more set-up of supporting code (Joines and Roberts 1998; Kornecki 2001). Moreover, OOS compliers and software are not as mature as procedural simulation software (Reynders et al. 1995). In fact, we believe that although OOM has drastically improved simulation software and simulation software support, it has not contributed as much to increasing the efficiency of the simulation model obtained from a statistical perspective. In addition, the benefits of OOM shall be realized when the system is known to be reusable or extendable in the future, which is not always the case, as simulation projects are usually a one-time-type application.

Table 1: Comparison of Persistent Literature on Simulation Hierarchical and Multilevel Modeling

<table>
<thead>
<tr>
<th>Literature</th>
<th>Representation of Atomic components</th>
<th>Approach Description/Tools</th>
<th>AI</th>
<th>OOM</th>
<th>Other</th>
<th>Sub-model Interaction &amp; Interface</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zeigler 1987</td>
<td>System’s components</td>
<td>developed DEVS formalism</td>
<td>x</td>
<td>x</td>
<td></td>
<td>Coupling scheme</td>
</tr>
<tr>
<td>Fishwick 1988</td>
<td>Model dependent</td>
<td>Reasoning, database relations and OOM, developed and used developed HIRS a multilevel simulation language</td>
<td>x</td>
<td>x</td>
<td></td>
<td>Global database, control/primitives/production/inference rules</td>
</tr>
<tr>
<td>Ketcham et al. 1989</td>
<td>Data, knowledge</td>
<td>Information-based approach, Combines concepts of OOM and database</td>
<td>x</td>
<td>x</td>
<td></td>
<td>IBIS interfaces and IDEF0 and databases</td>
</tr>
<tr>
<td>Pooley 1991</td>
<td>Activities</td>
<td>Graphical, activity diagrams and activity cycle diagrams</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Čerić 1994</td>
<td>Static and dynamic, functional and informational aspects</td>
<td>OOM, and information processing object hierarchy (IPOH)</td>
<td>x</td>
<td></td>
<td>using OOM</td>
<td></td>
</tr>
<tr>
<td>Kamal &amp; Biles 1995</td>
<td>Model data and specifications</td>
<td>OOM, developed Visual Sim</td>
<td>x</td>
<td></td>
<td></td>
<td>Interface of Visual Sim</td>
</tr>
<tr>
<td>Karacal 1998</td>
<td>System activities, data primitive types</td>
<td>Separate logical/physical events/processes activities, state-space representation and OOM approach</td>
<td>x</td>
<td>x</td>
<td></td>
<td>SMALTALK and OOM software interface and database relations.</td>
</tr>
<tr>
<td>Barcio et al. 1997</td>
<td>Components, tasks</td>
<td>C++, SMOOCHES (State Machine for Object-Oriented, Concurrent, Hierarchical Engineering Specifications)</td>
<td>x</td>
<td></td>
<td></td>
<td>SMOOCHES interface</td>
</tr>
<tr>
<td>Pidd &amp; Castro 1998</td>
<td>Components of system</td>
<td>Simplifying model building by using selective external modularity, an extension to modularity to improve tractability</td>
<td>x</td>
<td></td>
<td></td>
<td>OOM interface</td>
</tr>
<tr>
<td>Benjamin et al. 1998</td>
<td>Information in a model</td>
<td>Knowledge-based heuristic to select appropriate abstraction level</td>
<td>x</td>
<td></td>
<td></td>
<td>IDEF5 to capture relations and KBSE component</td>
</tr>
<tr>
<td>Ball 1998</td>
<td>System components</td>
<td>OOM, AFS (Advanced factory simulator)</td>
<td>x</td>
<td></td>
<td></td>
<td>AFS user interface assists</td>
</tr>
<tr>
<td>McGraw &amp; MacDonald 2000</td>
<td>Parameters/variables</td>
<td>Input sensitization techniques, model abstraction techniques</td>
<td>x</td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Kiran et al. 2001</td>
<td>Sub-model components, specifications</td>
<td>Each sub-model is individually created and modified then interacted into an overall system</td>
<td>x</td>
<td></td>
<td></td>
<td>Sub-models interconnected via OLE PROMODEL’s capability and a visual basic interface to let user control level of detail</td>
</tr>
</tbody>
</table>
Table 2: Comparison among Several Object-Oriented Simulation Applications

<table>
<thead>
<tr>
<th>Author</th>
<th>Approach</th>
<th>Rational of Using OOM for Simulation</th>
<th>Tool(s)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zeigler 1987</td>
<td>OOM, KB-simulation, Coupling</td>
<td>x, x, x</td>
<td>DEVS scheme, SCOOPS</td>
<td>The core of the approach is the coupling scheme and the system entity structure</td>
</tr>
<tr>
<td>Sharma &amp; Rose</td>
<td>OOM/ hierarchical process-oriented simulation</td>
<td>x, x, x</td>
<td>HPSIM</td>
<td>Explains the deficiency of current simulation software from a software engineering perspective</td>
</tr>
<tr>
<td>1988</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kienbaum &amp;</td>
<td>OOM with emphasis on graphical interface</td>
<td>x, x, x, x</td>
<td>Activity cycle diagrams, SIMOJECT, MODSIMII</td>
<td>Describes the development of an OOM GUI for rapid discrete event simulation model building, emphases on manufacturing systems</td>
</tr>
<tr>
<td>Paul 1994</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kamal &amp; Biles</td>
<td>Develop a five layer approach for OOS</td>
<td>x, x, x, x</td>
<td>Visual Sim</td>
<td>Include Hardware, Kernel, presentation, specification, management, and application layers</td>
</tr>
<tr>
<td>1995</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lin et al. 1996</td>
<td>Context-based simulation</td>
<td>x, x, x</td>
<td>CSIM++, C++</td>
<td>Compares OOS with Procedural Simulation and emphasizes long and short term reusability</td>
</tr>
<tr>
<td>Barcio et al.</td>
<td>OOM, concurrent/hierarchical engineering</td>
<td>x, x, x</td>
<td>SMOOCHES C++</td>
<td>Applied to distributed systems of air defense system, focus on model portability, provides classification of simulation systems</td>
</tr>
<tr>
<td>1997</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Praehofer 1996</td>
<td>Used variable structure models an extension to</td>
<td>x, x, x</td>
<td>DEVS Scheme</td>
<td>Models in this approach have the ability to change their own structure and internal relations, compares DEVS and EMSY</td>
</tr>
<tr>
<td>DEVS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Joines &amp; Roberts</td>
<td>Provided fundamental of OOS</td>
<td>x, x, x, x</td>
<td>YANSL, C++</td>
<td>Compared the abilities of procedural and OOM simulation languages and compared Object based and Object oriented languages</td>
</tr>
<tr>
<td>1998</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ball 1998</td>
<td>Progressively grouping stations together and</td>
<td>x, x, x</td>
<td>AFS (Advanced factory</td>
<td>Emphasized the need for hierarchical techniques for manufacturing system and compares some of the available OOS software</td>
</tr>
<tr>
<td></td>
<td>comparing it with detailed model</td>
<td></td>
<td>simulator)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kilgore &amp; Burke</td>
<td>Discrete-event, process/object oriented</td>
<td>x, x, x</td>
<td>SILK classes of JAVA</td>
<td>Created a framework for distributed processors each of which represent a node in a network</td>
</tr>
<tr>
<td>2000</td>
<td>simulation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rossetti et al.</td>
<td>UML design framework</td>
<td>x, x, x</td>
<td>SIMFONE</td>
<td>aim to explain abstraction in a simulation framework and develop OOS libraries</td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buss 2000</td>
<td>Component-based simulation</td>
<td>x, x, x</td>
<td>Simkit</td>
<td>Possible to be implemented in a non OOM language, uses common interface rather than inheritance</td>
</tr>
<tr>
<td>Chen et al. 2001</td>
<td>Component-oriented simulation based on HLA</td>
<td>x, x, x</td>
<td>CORSA</td>
<td>Emphasized simulation interoperability and interchangeability</td>
</tr>
<tr>
<td></td>
<td>(high level architecture)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ARTIFICIAL INTELLIGENCE AS AN APPROACH FOR HIERARCHICAL, MULTILEVEL MODELING AND ABSTRACTION**

The most detailed research in abstraction and hierarchical modeling is conducted in the field of AI. AI-based abstraction is viewed as a learning activity (Lu and Tcheng 1991; Manfaat et al. 1998). Multiple learning algorithms such as regression analysis, inductive learning, neural networks and recursive decomposition are used to abstract detailed models or representations. Examples of such applications include forming empirical or general models from training examples (Reddy 1996; and Manfaat et al. 1998), and building layered abstract models from given detailed models (Lu and Tcheng 1991). When AI-based abstraction is used for planning or problem solving, abstraction models are constructed by dropping literals that are not relevant to a given abstract objective. One of the earliest semi-automatic AI-based abstraction systems was ABSTRIPS (Sacerdotti 1974). ABSTRIPS uses a state-space representation and is based on a STRIPS (Stanford University Research Institute Planning System) framework, which creates abstraction hierarchies by removing symbols from the formal language to form the definitions of the operators and space (Holte et al. 1996). The successors of ABSTRIPS are many, including PRODIGY/EBL (Minton 1988) and ALPINE (Knoblock 1994).
Integration of AI concepts to HMA modeling is most beneficial when simulation is used as a decision making tool. If this is the case, using AI-based abstraction under certain assumptions can result in an exponential speedup (Ellman 1993). However, in order to exploit AI-based abstractions to simplify simulation models, simulation components need to be represented using a state-space framework or knowledgebase. Yet, describing the parameters of an engineering design in terms of state-space representation, in many cases requires an explicit enumeration of all possible design states, which is cumbersome and complicated. Moreover, in state-space representation, there is often a need to identify goal states a priori, which are unknown in the case of simulation.

By reviewing the literature of HMA modeling, we conclude that object orientation and artificial intelligence are among the most frequently applied approaches for HMA modeling. Object orientation was embraced mainly because it is inherently hierarchical and supports interfaces, which is a challenge by its own to develop otherwise. Moreover, it has one-to-one correspondence to the simulation model components. On the other hand, AI concepts were integrated to simulation to produce abstract or general models from detailed training examples using machine learning advancements.

**EFFECTIVE ABSTRACTION METHODS AND HIERARCHIES**

Hierarchical models are a result of an iterative application of some abstraction methods. That is, an ordered sequence of abstraction spaces constitutes the skeleton of an abstraction hierarchy (Knoblock 1994). This is illustrated in Figure 1. Moving upward, the abstraction hierarchy results in compression, i.e. loss of information; while moving downward in an abstraction hierarchy results in refinement or the gaining of information (Fishwick 1995).

Since abstraction processes are the building blocks of hierarchical modeling, the lack in the development of systematic methodology for HMA modeling is inherited from the lack in systematic abstraction methods for simulation. Therefore, developing efficient simulation hierarchies requires efficient abstraction methods. Abstraction and properties of effective abstraction methods are discussed next.

**ABSTRACTION METHODS**

Research in abstraction is conducted in many areas, but lacks interaction, consistency and formalism across the research efforts by which it is addressed. However, the most detailed work in developing a general abstraction theory was that developed by Giunchiglia and Walsh (1992). They have defined abstraction as a mapping process that:

1. Maps the original problem representation onto an abstract representation, which
2. preserves certain desirable properties of the model, and is
3. simpler to handle as it ignores unnecessary model details.

Moreover, Sisti and Farr (1998) have defined abstractions as the process of capturing the “essence” of the behavior of the model without all the details of how this behavior is implemented.

![Figure 1: Abstraction Methods Derive Hierarchical Models.](image)

**Properties of Efficient Abstraction Methods**

As discussed earlier, efficient abstraction methods produce efficient hierarchical models. Despite that fact, analysis on the efficiency of hierarchical models and abstraction within simulation has not gained enough emphasis in simulation literature. Conversely, research in abstraction for planning and problem-solving in computer science and AI are forerunners for studying the anticipated efficiency and value of abstraction. From these areas we extracted properties to ensure effective abstraction processes and eventually effective hierarchical modeling. These properties restrict an abstraction method to:

1. Be formal and systematic.
2. Be complete and thus not result in loss of information when details need to be retrieved.
3. Produce simpler models.
4. Be computable and tractable.
5. Reduce cost of model building and execution.

Maintain the structure of solutions obtained at abstract levels while refining them at more detailed levels (Bacchus and Yang 1992; Knoblock 1994)

**Problem Domain and Abstraction Methods**

Having discussed the properties of efficient abstraction methods, the question that arises here is, does the domain to which abstraction hierarchies are applied affect the efficiency of the results obtained? According to Pidd and Castro (1998), whether a system can be hierarchically decomposed depends on
the perception of “how much” a system is a reductionist one, i.e., how meaningful it is when considered in part as opposed to as a whole (holism). This largely depends on the amount and type of interactions that exist among the different system’s components. In fact, Bacchus and Yang (1992) explain, that in problem domains which exhibit significant interaction, hierarchical abstraction is not expected to outperform nonhierarchical methods. Therefore, in addition to the efficiency of the underlying abstraction building process, the success of hierarchical methodologies also relies on the context of the problem to which HMA modeling is applied.

GUIDELINES FOR SYSTEMATIC HIERARCHICAL MODELING OF COMPLEX SIMULATION SYSTEMS

By reviewing the literature of hierarchical simulation modeling, we observe that research in this area is highly interdisciplinary, also the underlying methodologies are diverse to a great extent. However, we were unfortunate to witness integration among these approaches or have these approaches follow systematic and mutual guidelines. In this part of the research, we propose general guidelines upon which simulation hierarchical modeling need to be structured. These guidelines are extracted by comparing and contrasting literature of hierarchical modeling and abstraction for simulation and from forerunning research in this area from the field of AI and OOM. Following these guidelines will enhance the understanding of developed hierarchical methodologies and will hopefully expand their application, and facilitates their extendibility and automation. As a general guideline, we suggest that a methodology for simulation need to consist of four main phases: problem formulation, determining the appropriate abstraction method to be applied, hierarchical model building and hierarchical model execution. These are illustrated in Figure 2.

Phase I: Problem Formulation

The problem formulation phase for HMA modeling follows the classical simulation problem formulation which can be found in (Law and Kelton 2000). The simulation problem is formulated based on the data collected for the system to be simulated, the purpose of study and problem domain. The problem formulation should also indicate the representation scheme. Examples of representation schemes that have been used for hierarchical simulation were illustrated in Table 1.

Phase II: Determining the Appropriate Abstraction Process

The source of the system or model complexity determines the method of abstraction best to be applied to ensure efficient HMA modeling. Abstraction methods are classified into either elimination abstraction, estimation abstraction or grouping (clustering) abstraction (McGraw and Mac-Donald 2000). In an elimination abstraction, some of the model components, factors, literals or criteria are dropped for simplification (Knoblock 1994).

In an estimation abstraction, complex behaviors of the system may be replaced with simpler ones, i.e. replacing random variables with deterministic ones (Zeigler 1987; Coughlin and Sisti 1997). In grouping abstractions, different model components, factors or criteria are merged into an aggregate cluster. For example, merging of spaces or distances (Coughlin and Sisti 1997; Tam and Li 1991; Manfaat et al. 1998; Goldin and Klahr 1981), merging over time periods (Goldin and Klahr 1981; Kulkarni and Mohanty 1996), merging of product flow (Tang and Abdel-Malek 1996; Tam and Li 1991), and merging of criteria and factors (Partovi and Burton 1992).

Phase III: Hierarchical Model Building

The hierarchical model building phase is a bottom-up process of abstracting different simulation model components, factors or criteria into abstract or aggregate units. To construct the abstraction hierarchy, the abstraction process repeats iteratively until no further abstraction is necessary, i.e., the model has reached an acceptable simpler representation.

Figure 2: Phases of Simulation Hierarchical Modeling.
Phase IV: Hierarchical Model Execution

The hierarchical model execution phase is the top-bottom simulation abstract model’s dis-aggregation phase. Here abstract simulation models at the highest level (n) of the hierarchy are simulated first, then the ones one level below and so on. After each simulation, the solutions obtained are refined. The simulation process repeats progressively with refined solutions until statistically acceptable solutions are obtained or the detailed original model is reached (level=0).

Following the HMA modeling approach shown in Figure 2 has the advantage of gradually adding details to the simulation model which reduce modeling errors. Modeling errors may also be corrected at higher levels of the hierarchy where simulation is less costly. In addition, details that add cost to modeling and simulation time are added only when solutions at higher levels turn out to be unsatisfactory. More importantly, different simulation strategies can be simulated before difficult details of the system are available.

CONCLUSIONS

From this research, we conclude that object-oriented methodology and artificial intelligence are among the most frequently embraced approaches for simulation hierarchical modeling. This is because object-orientation is inherently hierarchical while artificial intelligence-based methods are capable of creating abstract models from detailed ones using machine learning techniques. Furthermore, in this research we have identified properties to ensure effective abstraction and eventually effective simulation hierarchical modeling and pinpointed the structural relationship between abstraction and hierarchies. Moreover, we presented how the problem domain could dictate the viability of applying hierarchical and abstraction methodology to simplify simulation complex systems. Finally, we used the broader literature of hierarchical modeling and abstraction to extract guidelines upon which simulation hierarchical methodologies need to be structured and formalized. These guidelines aim to create a common platform, to enhance the understanding of simulation hierarchical modeling, expand their application and encourage further research, extension and automation, and to make simulation hierarchical modeling an efficient technique that will genuinely simplify complex and large-scale simulation systems.

REFERENCES


http://doi.acm.org/10.1145/161465.161471.


AUTHOR BIOGRAPHY

ESRA ALEISA is an assistant professor of Industrial and Management Systems Engineering (IMSE) at Kuwait University. She has received her Masters and PhD in Industrial Engineering and production systems from the department of Industrial Engineering at the State University of New York at Buffalo in 2001 and 2005 respectively. In 1998, she has earned her B.S. degree in industrial engineering from Kuwait University.

Dr. Esra Aleisa research interests include, Planning and design of large scale facilities, simulation and improvement of manufacturing and service systems, multilevel planning and design of complex engineering design, group technology (GT), and design structured matrices (DSM). She is a member of Omega Rho, the international operations research honor society, IEEE, INFORMS, IIE, ASEE.
MODELLING AND SIMULATION FEATURES FOR PRODUCTS-SERVICES

Rinaldo C. Michelini and Roberto P. Razzoli
PMAR Lab, DIMEC
University of Genova
Via Opera Pia 15/a - 16145 - Genova, Italy
E-mail: michelini@dimec.unige.it; razzoli@dimec.unige.it

ABSTRACT

The product design methodologies and practices are expected to face in the short future important changes, to deal with the lifecycle requirements of the growth sustainability. These changes aim at granting the visibility over the points-of-service, in connection with the on-duty conformance-to-specification tests, and at the point-of-dismissal, to comply with the enacted recovery (reuse, recycle) rules. To that purposes, the major innovations develop along complementary lines: 1 to establish appropriate knowledge frames, which embed lifecycle economy of scope proper orientation; 2 to exploit modelling and simulation functions M&S, which provide the pertinent product-view specialisation.

The paper offers an introductory overview of these lines, to address the design practice management requirements, by a multi-disciplinary approach, with concern on available information technology CAD tools that embed algorithmic and heuristic blocks to offer mixed mode simulation-and-emulation aids, and PLM tools that explore Ambient Intelligence and net-concerns to tackle with the on-duty and end-of-life services provision.

KEYWORDS: Lifestyle Eco-Design; Modelling & Simulation; Net-Concern; Product-Service.

INTRODUCTION

Industrialised societies face greater challenges than ever, to keep the pace of the economical growth, due to globalisation with competition of low wage countries, and to sustainability with the exacting acts to preserve eco-conservativeness (Kleinert 2004). Now, the ICT aids enable to design, build, deliver and support enhanced goods, out of reach of past entrepreneurship; although the change is well underway, the efforts face apparent nonsense: enhancement is focused on redundancies and embellishments; manufacture technologies are addressed as end-processes, rather than enabling means of design-for-X issues; etc.; investment in proprietary solutions lead to deceitful results, quickly turned obsolete by new technologies; abstract problem solving capabilities happen to be illusory, when these are not linked to actual market onsets and environment needs, driven by emerging business tracks. Focus on these requires worthy attention.

The ongoing trend leads to expand suppliers’ responsibility, from the point-of-sale, to the point-of-service, to support the conformance assessment duties, for safety and environment protection acts, and to the end-of-life, to comply with the enacted recovery, recycling and reclamation incumbents. This means to deal with provision of products-services or extended artefacts, where extensions correspond to (mostly intangible) frames, going together with the (material) supply chains, to warranty the achievement of the desired tasks. Acknowledging this trend, the design steps will be critical: to optimise performance, to increase profitability, to enhance delivery quality, to respond to regulatory drivers, to satisfy customers, stakeholders and third parties, etc.; this requires reconsidering the design duties, by exploiting proper decision skills that cut across the manufacture business, to concentrate efforts on the extensions and to integrate traits of the extended artefacts lifecycle, addressing every views, still preserving the overall coherence.

Proper modelling and simulation (Tambe et al. 1995) is winning tool to expand the enterprise concern over products lifecycle, so that material, process, operation, maintenance, dismissal and recovery models are tested and assessed through virtual prototypes. As compared to earlier industrial habits, when competition was won by off-process presetting and optimal production plans, today an enhanced decision support is required to run on-process, with: - product assessment for lifelong performance and customised quality; - manufacture choice for process effectiveness; - service schedules for conformance-to-use rules; - dismissal provisions for eco-consistency (Weber and Hemmelskamp 2004). These prospects are consistent with: # real-time diagnosis and decision tools delivered as on-duty service engineering complements; # off-process presetting of the condition knowledge by the lifestyle design of the products. The latter option is discussed hereafter; then, the business widening, around the whole extended artefact, is sought with resort to enhanced PLMs, embedding a super-model, with federated architecture, assuring unified access to lifecycle data. The integration of every view (structure and function layout, fabrication process, quality certification, maintenance policy, dismissal incumbents, ...) into a super-model is critical competitive advantage to upgrade, with ICT aids, the production business. Thus, these aids are winning options, favouring the high-tech value-added offers.

LIFESTYLE DESIGN AIDS

The recalled perspectives show that the existing product data management, PDM, tools shall progressively turn into product lifecycle management, PLM, ones, with twofold upgrading: to improve the business native effectiveness; to widen the reach of all the engaged actors. The shared vision emerges, (Borg et al. 2004), (Lee et al. 1993), (Michelini et al. 1999), (Parsael et al. 1998), (Rautenstrauch et al. 2002), with key aspects:

* product-service unified data-frame; the extended artefacts delivery is primary business, and the lifecycle knowledge
is basic requirement from the early design steps;
- integrated data-flow management: hierarchic, cross-linked, parametric product/business models ensure that decisions are right away made, achieving enterprise-wide impact;
- distributed, flexible operability: robust communication and shared data-base in the process resources help establishing teaming relationships, to face every emerging requests;
- plug-and-play interoperability: all technical and business modules need be seamlessly compatible and self-adapting, to be immediately operational, with not integration cost;
- total connectedness: all acknowledged actors are linked by communication infrastructures that deliver the right data at the right time, whenever they are required;
- fully enabled extended artefact visibility: the PLM tools, linked to science-based and experience-driven knowledge, fully disclose decision schemes and achieved issues.

The unifying super-model, for the coherent arrangement of the whole conditioning knowledge, distinguishes because of a set specialised properties, needed to deal with extended artefacts. It is deemed by several development and research initiatives all over the world, at different level of complexity.

The IMTR Inc., e.g., is noteworthy example, with effective work aiming at joining the modelling and simulation functions, separately, used for the description of products, processes and enterprises into an unified frame, (IMTR 2000), (Vernadat 1996); the project shows the effective achievements at the range of product-process integration (by simultaneous engineering) and prospects lists of manufacture functions, deserving specific attention to reach appropriate enterprise integration. The IMTR 2015 vision is projected on the ten-twenty years span, and purposely focuses on manufacture functions, thus, mainly actions that belong to the conventional domain of the industrial companies.

<table>
<thead>
<tr>
<th>Craft-study</th>
<th>Industrial-company</th>
<th>Extended net-concern</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masterpiece</td>
<td>Mass deliverable</td>
<td>Products-service</td>
</tr>
<tr>
<td>Economy</td>
<td>Economy of scale</td>
<td>Economy of scope</td>
</tr>
<tr>
<td>Point-of-der</td>
<td>Point-of-sale</td>
<td>Point</td>
</tr>
<tr>
<td>Design with manufacturing</td>
<td>Off-process</td>
<td>On-process</td>
</tr>
<tr>
<td>Work organisation</td>
<td>Master leadership</td>
<td>Scientific job-allocation</td>
</tr>
<tr>
<td>Decision structure</td>
<td>Craftsmen</td>
<td>Hierarchical assignment</td>
</tr>
<tr>
<td>Motivation style</td>
<td>Individual</td>
<td>Cash</td>
</tr>
<tr>
<td>Knowledge features</td>
<td>Personal</td>
<td>Team</td>
</tr>
<tr>
<td>Competitiveness index</td>
<td>Quality</td>
<td>Manufacturing</td>
</tr>
</tbody>
</table>

Figure 1: Product-process vs. earlier supply chains

The need to look after at the enterprise level depends on the fact that earlier vertical flow-shop set-ups, more and more, are replaced by horizontal job-shop ones, incorporating non proprietary technologies and facilities, with extended resort to out-sourcing and productive break-up. The manufacture functions (strategic positioning, market assessment, risk management, resource management, quality deployment, work-flow schedule, factory management, operation facility planning, shop setting, performance evaluation, productivity monitoring, finance/cost management, etc.) are interlaced concern, with tasks and responsibilities distributed among several entities, and an optimal arrangement can only be achieved dynamically, with co-operative efforts and after reconfiguring. This is halfway step, towards new set-ups, when the suppliers responsibility will encompass product-service delivery, on the lifecycle span, recovery included. The approach comes to be business paradigm shift, similar to the one occurred, over two centuries ago, with the industrial revolution, Fig. 1, leading, now, to products-services supply chains, supported by extended enterprises.

The new set-ups are prospected outcome of sustainability (Michelini et al. 1997), (Steiger et al. 2005), as soon as governmental rules, to refrain from polluting and lower consumption, will require visibility on the supply chain (Petrovic 1998). The enterprise competitiveness will turn, from the capability of offering new products (fit-for-purpose to individual needs), to the ability of providing services, granting the expected functions to clients’ satisfaction and the enhanced tangibles effectiveness (fit-for-purpose to general benefits). The emerging business paradigms will profit of co-operative networked organisations, according to different styles:
- on one side, manufacturers could be spurred to keep in charge the all service: artefacts supply, lifecycle conformance and dismissal incumbents, so that the trade regulation would depend on single indentures;
- on the other side, independent enterprises could profit by safety rules and environment acts expansion, to become service dealers, with technology oriented qualification and infrastructure-based organisations.

Both styles require to focus on the design functions, moving the enterprise profitability to be critically dependent on the initial choices of the prospected supply chain. The new market leaders establish in this technical scientific frame, replacing the economy of scale, by the economy of scope, with, at least, two opportunities: # functions delivery and artefacts life-extension policy, with profitability in the business of supplying products-services; # increased recycling efficiency, with profitability in the new businesses aiming at reverse logistics (from waste, to ‘raw’ materials), (Blumberg 2004), (Dickhoff et al. 2004). The two lines develop outside the factories, and the knowledge contents of the required PLM aids shall, accordingly, expand to deal with the set of scopes, with critical relevance, aiming on on-duty performance or operation reliability, and at pollution rules or recovery effectiveness. In that frame, the PLM requisites (e.g., total connectedness) appear to be critical challenge, chiefly if the extended enterprise is built by independent companies that focus on their core competencies, and join the partners efforts in view to co-design, co-manufacture, co-market, co-maintain, co-service and co-recycle, and, there through, to fulfil the supply chain at clients’ benefit and environment safety, with on-duty and end-of-life provisions.

MODELLING AND SIMULATION TOOLS

The role played by the design steps, to develop the huge information outfit of the supply chain, explains the critical relevance of the ICT aids, since the business effectiveness, grounded on the economy of scope, needs, this time, to scan
through the whole delivery lifecycle, dismissal and recycling included, to take into account, both, structural and functional properties of the artefact, its parts and its materials. The intimate connection of the knowledge frame, Fig. 2, means iterating the ICT aids, design steps and betterment scopes, to obtain the self-comprehensive product-service description. The knowledge build-up has exacting relevance; the design of the product-service aims at sets of implicit scopes:

- close integration needs bind all actors on the lifecycle (e.g., for clear transmission of economic incentives);
- the users’ needs ought to be incorporated, pre-emptively designing duty-driven solutions for them;
- the artefacts characteristics must be attuned to the existing level of equipment (modular design could help);
- schemes for products temporary or final call-back should be focused on efficient transfers;
- the alternative options of items use could be foreseen, with included client’s information sheets;
- marketing strategies shall embrace the clients’ education about items use in the most appropriate way;
- advertising plans are done, promoting an environmentally more acceptable way of function fulfilment;
- the overlay of point-of-sale, -of-service and -of-recovery shall ensure conservative usage;
- the end-of-life and recycling duties have to devise optimal resources exploitation;

and the likes, so that the choices help achieving optimal performance, as for the operation technical performance and the eco-consistency rules.

![Diagram](image)

**Figure 2:** Lifecycle design information set-up

The information framework represents a noteworthy change by respect of the up-now existing manufacture practice, oriented on the product-process *internalities* of simultaneous engineering. From now on, new *scopes* add, directly driven by the *externalities* of the on-duty eco-consistency and of the end-of-life call-back and recovery according to the enacted sustainability regulations. As well known, the EU directives, presently, are oriented towards the extended producers responsibility, with already regulations of the free take-back of sets of widespread durables, such as end-of-life vehicles, ELV, and waste electrical and electronic equipment, WEEE. Accordingly, these directives request compulsory recovery (reuse-recycle) targets, incorporating the reverse logistic obligations, thus, conditioning the selling of the product, to the free take-back of the scrap. The productive organisation presetting, having the prescribed eco-consistency abilities, becomes quadruple achievement, covering:

- product specification, leading to the construction technical files by proper CAD, CAM, etc., to select producibility figures, operation constraints, dismissal duties, etc., as for the forecast performance;
- manufacturing process specification, leading to improve factory effectiveness by simultaneous engineering practice through product-and-process mutual betterment, by means of pace-wise up-grading:
  - delivery lifecycle specification, leading to regulation, maintenance, restoring, etc. plans, for conformance-to-use operation, and to preset dismissal, recovery, etc. plans, for enhanced reverse-logistics;
  - entrepreneurship specification, leading to define the business policy and the productive organisation as for the manufacture and service functions/facilities choice, incorporation and management.

The modelling and simulation tools help extrapolating the lifecycle, according to the chosen frames (causal orders and behavioural modes). By these aids, the logic relations are modelled on a judgmental basis that factually affects the parallel pre-set of deep knowledge schemas; the functional dependence, in the knowledge context, indicates guesses between entities without implying closed-form causality; therefore, the expert-modules provide human-like progress, as compared to just algorithmic blocks. The ensuing lifestyle design ICT aids will organise with a nested multiple-layer data-flow, Fig. 3, to combine structured information, leading to algorithmic models, with added empirical evidence, due to assessed expertise, given by databases and joint heuristics. The knowledge build-up progresses along different tracks. Every product-service view is given by expert-simulation, according to the said modelling. The link between the views comes out from model-federation. The super-model, once widened to investigate the all lifecycle, is exploited: directly, to complete the extended artefact design up to appropriate details; indirectly, to expand the (shared) reference database, for future expansions. The spot modelling, when embedded in the design ICT aids, brings to a module for virtual reality testing; a similar module can be developed for real-time operation, to be later exploited by service engineers, as diagnostic tool for monitoring maintenance.

For lifestyle design, the information vaulting and handling achieve high criticality, with falls-off dealing all over the extended enterprise, namely, covering the management of the internal processes and machineries and of the external interactions and fulfils, to keep full and responsible overseeing on the delivery at the points of service and call-back. The enterprise competitiveness, due to the economy of scope, shall, then, at once, comply with all lifecycle driving duties, to find out time-evolving set-ups, the manufacture and service plans, and the business governing policies.

![Diagram](image)

**Figure 3:** Layered simulation aid for lifecycle-design

The lifestyle design phase is always compelled to deal with integrating the four product-process-environment-enterprise, 2PE, scopes, based on the explicit coding of modelling and simulation features, M&S, encompassing the proper ranges:
- technology innovation, requiring enhanced focus on the ideation of new offers (products-services, etc.), in the lines of the lifestyle design (done with extended artefact M&S);
- increased production effectiveness by adaptivity (flexible automation, etc.) and leanness, with account of pertinent functions/facilities (by means of processes M&S);
- lawful compliance of the enacted environmental protection rules (according to the related M&S), suitably assessed and vaulted by (third-party) certifying bodies;
- appropriation of point-of-service and end-of-life duties, to rule the whole supply chain under unified responsibility (incorporating extended enterprise M&S).

The enhanced resort to integrated design approach exploits the super-model description, and empowered modelling and simulation features, M&S, with four ranges checks, Fig. 4.

**Figure 4:** The four ranges checks for lifestyle-design

These M&S ranges expand the PLM tools reaches, to cover the product, process, environment and enterprise spheres:
- the standard product data: physical shape; operation performance, quality, cost and affordability; producibility; lifecycle requirements; use and maintenance directions; dismissal instructions;
- the standard manufacturing data: materials procurement and processing, assembly and disassembly; quality testing and conformance checks; packaging; re-manufacturing;
- the standard environment data: voluntary agreements for low pollution and high remediation; mandatory targets to rule recovery (reuse, recycle) and safe dumping;
- the standard enterprise data: business functions: trading strategies, finance and resource management, etc.; operation functions: factory and facilities management, scheduling and planning, etc.

The fully devised modelling and simulation features, M&S, become the crucial qualifying element of the manufacturing business to come, (Michelini 2008), with the in-progress build-up of the knowledge entrepreneurship to face the sustainable growth demands. The focus will be moved on the information flow delivered in parallel with the product-service, leading to a series of scopes, such as:

- to adapt process-and-product, by simultaneous engineering procedures; to run intelligent manufacturing plans, satisfying the mix of clients, on the tactical horizon; to make use of normal standards archives, to explore parametric choices and modulations; to acknowledge the technology appropriateness, grounded on existing rules; to develop the resort to recycled materials, among the qualifying construction features; to fulfil the value chain analysis, and forecast the economic return of every variation;
- to monitor the eco-impact during lifecycle and at dismissal (recovery included); to reckon the eco-figures (resources yield per unit service, or similar assessments); to accomplish the productivity estimation of the (actually involved) natural capital; to recognise extended artefact liability, also in terms of eco-consistency; to pre-set monitoring-monitoring-maintenance schemes, to expand the on-duty conformance operation; to suggest new exploitation-modes, for improving the environmental protection;
- to understand the net-worked setting, as enabling aid of the product lifecycle management; to rationalise the procedures and to fix the intermediary thresholds; to code clients requirements, by full demands and checks transparency; to study the application demands, to orient the supply chain towards ‘function’ provision; to explore the buyers’ needs, for incorporating thriftiness along with enjoyment; to pre-plan the diagnostic and restoring tasks, to expand the delivery availability;
- to add any other scope, fit to improve the value chain with measurable falls-off into the product, process, environment and enterprise, 2P2E, functions and facilities.

At the design phase level, the super-model set-up shall incorporate the four series of scopes, with related modelling and simulation features, M&S, further granting the suited series of opportunities, Fig. 5.

**Figure 5:** Example super-model opportunities

The unifying super-model, for the coherent arrangement of the whole conditioning knowledge, distinguishes because of special properties, needed to deal with extended artefacts. The lowest level is given by the simultaneous engineering approach, with the product-process unified data-frame. At the mid-level, the integrated product-process-enterprise and/or product-process-environment are recently developed,
through consistent inclusion of the «knowledge», as «new economy» driving fact to deals with corporation policies in functions management, either with focus on lifestyle actions that belong to conventional supply chains. The balanced coverage of the 2P2E frame is the further challenge towards eco-sustainability, opening the business to externalities (with measurable falls-off in eco-consistency, to cover the whole supply chain, from material provisioning, along the product lifecycle, up to end-of-life withdrawal and recovery), used as new input to reorganize the enterprise internalities (facilities and functions, actually incorporated to fulfill the asked product-service provision). The externalities to internalities interplay has found way in the modelling and simulation features, M&S, methodology, and its driving motivation in the integrated 2P2E design. Accordingly, the full 2P2E design is the extension of the earlier simultaneous engineering, where the product-process, 2P, design made possible the cross-adaptation of products, with client-driven quality, to processes, with issue-adapted characteristics. The methodology can move at different conditional patterns, so that, at the end, it will be useless to distinguish whichever, the product or the process, comes out as the leading inspirer of the improvement. Similarly, the 2E expansion can develop to encompass the environmental externalities, directly referring to enterprise embedded functions/facilities, so that the lifecycle service provision is achieved by small or no company up-dating of the productive organisation, exactly as the earlier client-bent of the product quality was reached by the already offered set-up flexibility. Then, the full 2P2E design offers benefits, by the interplay transparency between the environment data (mandatory lifecycle eco-targets) and the enterprise data (improved productive set-up incorporation). A considerably different situation is faced, if the productive organisation does not embed all the facilities and functions required by the on-going service-delivering or (mandatory) end-of-life recovery business. The useful alternative, looks after the virtual-enterprise set-ups, namely, the networked concerns, gathering autonomous partners, to join the requested facilities and functions that make effective the requested product-service provision. The opportunity is easily implemented on condition that the properly wide facility/function market exists, and the suitable brokerage helps choosing and integrating the scope-driven virtual-enterprise (Cunha and Putnik 2006), (Michelini 2008). The alternative is not further discussed here, because our interest focuses on the design phases, and related modelling and simulation functions M&S, method. In the framework of the present discussion, we might point out that the virtual-enterprise arrangement is preliminary answer in front of compulsory regulations, (Acaccia et al. 2006a, 2006b), enacted to impose extended manufacturers’ responsibility, with the free take back of the end-of-life products. For the enterprise profitability, dependence on the strategic positioning in the market is standard admission, and the supply chain concept has, also, been modified into value chain, to join parts and materials delivery, with the related intangibles flow (value web), supporting a main contractor with vital complements. Today, the expansion of off-process design tasks and connected data-bases, (Lambert and Gupta 2004), (Lee et al. 1993), (Qian and Pagello 1994), is compelled by enacted product lifecycle regulation constraints, and is made possible by existing IT aids, once all three M&S are acknowledged. A relevant issue in M&S deployment is to link the views, developed for different purposes or by different teams, in order to achieve model federation, making easy to create high-level representations, allowing reuse of existing frames and data, and propagating the knowledge environments by seamless continuity. Designers will be able to assess the candidate prototypal deliveries, from virtual factories, for virtual point-of-use set-ups, up to virtual point-of-dismissal situations, to evaluate and to improve productivity, function performance, operation reliability, maintainability, eco-impacts, end-of-life falls-off, etc., of every supply chain. They will be able to fulfil these checks beforehand and with full completeness, to quickly reach effective hints, with the ability to zoom in at critical details and compare alternatives. In the future, proper exploitation of M&S description will be the way to improve the manufacture business efficiency, to reach the best balance of all constraints in designing, developing, producing, servicing, recovering and supporting the products. Cost and time decrease, customer satisfaction, lifecycle responsibility, environmental safety and resources conservativeness will be part of relational models, with algorithmic and heuristic modules, balanced by captured knowledge, virtual tests, on-line analyses and designer’s decision making. The M&S aids will support best practice, from concept creation, through artefacts use and call-back, up to waste handling and remediation; every duty will be known by the users and transparent to the authorities, to assess the resource productivity. The federated modelling approach will be dynamic, with learning a bit from every accomplishment and training the design teams to combine views and integrate functions.

CONCLUSION

The turning of PLM tools to cover all accomplishments at the point-of-service and at the end-of-life means the huge restructuring of the design tasks, since most of the product information are totally outside the original manufacturing enterprise. Indeed, the old industrial revolution, first, introduced the off-process design tasks, leading to optimal choices, according to scale economy, by the most effective arrangement of the manufacture shop and process planning. The design is off-process critical phase, with comparatively overall effects on final unit products. The recent industrial evolution has, already, modified its business paradigms, aiming at buyer quality, with diversified offers supported by flexible automation: product-process design by simultaneous engineering is current practice, according to scope economy. The scenarios to come look after further widening the enterprise intervention areas, with responsibility moved to lifecycle operations, to comply with pressing environmental acts for resource sparing and pollution avoiding. This quite directly leads to similar expansion in the design tasks, to cover product-service delivery, with inclusion of on-duty conformance to specification and end-of-life recovery and remediation incumbrances. The problem is discussed by the paper, with focus on the ICT aids that make possible these new design tasks and efficient this overall innovation.
The attributes of these aids are organised, in keeping with the modelling and simulation features, M&S that are the critical means to implement the appropriate PLM tools. The joint product-process design is suitably accepted practice, and has already contributed to apply economy of scope by simultaneous engineering techniques. Novel PLM tools distinguish by few additions, such as the re-manufacturing M&S, say, the process features that support call-back and reprocessing of products upon completion of originally intended use, to take profit from reverse logistics. Actually, the design-for-manufacture, DF/M, design-for-assemblage, D/A, etc. scopes move to design-for-disassemblage, DjD, design-for-recycling, DjR, ones, since earlier product ideation steps. The business paradigms will include remanufacture, recycle and reuse of products, parts and materials, to minimise the tangible resources consumption and to maximise the overall efficiency of manufacture business. The M&S recognition shall enable designers to explore and analyse the reverse logistics as worthy option, to enhance the value chain, for effectiveness, profitability and environmental sensitivity.

The joint product-process-environment-enterprise 2P2E aim shall be basic practice, once the lifecycle, recovery included, accomplishments become standard supplier’s responsibility. This brings to extended enterprises; the M&S description shall evolve to include business and operation management areas. The challenges are here more problematic, since the physic-based frames of engineering and manufacture are replaced by economic transactions, human and intellectual activities, social and legal constraints. Objects and events are specified by texts, frames, spreadsheets or graphic trends; model validation and simulation testing quickly lead at data reduction in terms of cost propagation and due dates. This provides effective means to acknowledge business and operation functions, and to help providing the visibility for reliable prediction and on-process steering and control. From now on, the incorporation of appropriate extended enterprise M&S is domain open to new options, where the ICT aids shall play the strategic role.

REFERENCES


BIOGRAPHY

RINALDO C. MICHELINI, professor at the school of Engineers of the University of Genova, Italy. ASME Fellow, Member of IFIP WG 5.1 (CIME) and WG 5.5 (COVE), SCS (Senior), Russian Academy of Engineering (Correspondent), etc. Author/co-author of over than 500 technical papers/chapters/books, in the areas of integrated design, robotics, expert automation, industrial diagnostics.

ROBERTO P. RAZZOLI researcher at the PMAR Lab of the University of Genova, Italy. His scientific activity is about eco-sustainable design, mobile and modular robots, industrial automation. He is author/co-author of over ninety publications.
SUPPLY CHAIN SIMULATION
USING SIMULATION TO SUPPORT SUSTAINABLE LOGISTICAL DECISION MAKING IN COMPLEX SUPPLY CHAINS

P.J. Byrne
Dublin City University Business School
Dublin City University
Ireland
E-mail: peter.byrne@dcu.ie

Paul Ryan
Cathal Heavey
Enterprise Research Centre
Department of Manufacturing and
Operations Engineering
University of Limerick
Ireland
E-mail: paul.ryan@ul.ie
E-mail: cathal.heavey@ul.ie

KEYWORDS
Sustainable, Logistics, Decision Support System, Supply Chain

ABSTRACT
Across the European Economic Area (EEA) more goods are being transported over longer distances more frequently than ever before. As a result, Greenhouse Gas (GHG) emissions from the transport sector increased by 24% between 1990 and 2003 for the EEA – 32 countries. The objective of this paper is to highlight the current situation for freight transport by road and proposes a simulation based tool to assist more environmental sustainable supply chain decision making. GHG emissions, energy demand and costs in freight transport are increasing. Across the EEA, GHG emissions from most sectors (energy supply, industry, agriculture, waste management) decreased between 1990 and 2004, however emissions from transport increased. In Ireland, there has been a 160% increase in GHG emissions from the transport sector in the period 1990 to 2005. With energy use in transport now 99% dependant on imported oil, sustainability in logistical operations is a concern for the Irish economy. The paper will focus on inefficiencies in modern Supply Chain (SC) design and show how using simulation can aid decision makers in choosing more environmentally friendly SC options. A simulation tool to aid decision makers to choose more sustainable options will be proposed, and preliminary results discussed.

INTRODUCTION
The 2006 Stern report on climate change highlights how the earth’s climate is warming. As the climate alters, the “basic elements of life” will change for people around the world. Access to water, food production, health, and the biological environment are at risk. Hundreds of millions of people could suffer hunger and coastal flooding as the world warms (Stern 2006). The greenhouse effect is attributed to this warming. If no action is taken in reducing GHG emissions, and in particular CO₂ emissions then based on current trends, a 2-3°C increase in average global temperature is predicted in the next 50 years. In the EEA, GHG emissions from most sectors, (energy supply, industry, agriculture, and waste management) have decreased between 1990 and 2004, however emissions from transport increased, (EEA 2007).

Over recent years, the increase in popularity of Supply Chain Management (SCM) techniques such as lean manufacturing, outsourcing and in particular, Continuous Replenishment (CR) policies for materials and stock levels has meant a redesign of traditional SC’s. Transport has become a major factor in these supply chains. The ability to get the right materials, to the right place, at the right time, is central to modern day manufacturing. The practice of CR has increased the frequency of material delivery, using transportation with high GHG emissions (i.e. smaller trucks, partially filled). Globalisation and the outsourcing of products and services has de-centralised production facilities creating large distances between interacting companies that now contribute to the production of a finished consumer product. This makes efficient transportation economically vital to all in the supply chain, while also causing enormous CO₂ emissions in the transportation stage of the supply chain, (Archibald, Karabakal et al. 1999), (Edgington 1993).

GHG emissions (excluding marine and aviation) from transport grew by 24% between 1990 and 2003 for the EEA – 32 countries. Emission increased most in Ireland and Luxembourg by 130% and 122% respectively for the same period (EEA 2006c) Logistically, Ireland’s geographical situation is unique. The relatively short distances and lack of suitable infrastructure in Ireland offset any advantages which could be gained by using intermodal freight transport, therefore, 90% of freight traffic in Ireland occurs on the road (Ryan, Dundon et al. 2003). Using simulation to identify areas of potential environmental wastes in the SC will provide planners with decision support to choose more environmentally friendly options. Potential economic verses environmental gains may also be identified and or highlighted for discussion when choosing a manufacturing contract.

This paper focuses on the environmental effects of freight transport and gives potential software solutions for
environmental compliance. Section 2 assesses the negative environmental impacts of increased road transport, documents Ireland’s poor environmental performance verses other EEA countries and introduces the new practice of sustainable logistics. This developing strategy in logistics management improves efficiency for the mutual benefit of the environment and company profit. Section 3 then highlights how recent environmental analysis can help identify potential environmental and economic savings in SC design. Section 4 explores the advantages of using Discrete Event Simulation in SC decision making. A Discrete Event Simulation (DES) based tool to aid decision makers choose more sustainable SC options will be proposed, and preliminary results discussed.

**FREIGHT TRANSPORT AND THE ENVIRONMENT**

In the transport sector GHG emissions are the result of burning petrol, diesel and kerosene in internal combustion engines. Across the EEA more goods are being transported over longer distances more frequently than ever before. As a result, GHG emissions from transport increased by 24% between 1990 and 2003 for the EEA – 32 countries (EEA 2006c) With countries now complying to Kyoto regulations, public opinion resistant to heavy freight traffic, efficient freight transport is a growing concern, (Joseph 2006).

![Figure 1: Rise in transport emission 1990 – 2003 (EEA 2006c)](image1.png)

As of 2006, only 8 of the EU-15 Member states are predicted to reach their Kyoto targets. The remaining 7 states are forced to implement “Kyoto Mechanisms” (i.e. carbon trading, forestation etc.) to reach their commitment targets (EEA 2006a). Transport is the fastest growing consumer of energy in the EEA and energy demand particularly in road freight is predicted to grow (EEA 2006b). For the EEA-32 countries, freight transport volumes have grown by 34% from 1993 to 2003, with the economy only growing 26% in the same period. 21% of the EU-15 total GHG emissions now are attributed to transport. Freight transport by road now holds a 77% market share over rail and water freight with road transport contributing 93% of the total of all transport emissions., (EEA 2007), (EEA 2006c).

**Freight Transport Growth in Ireland**

According to (Mason and Lalwani 2004), the dominant mode of freight transport in many economies is by road, due mainly to the flexibility and speed that the movement of freight by road offers when compared to railway, inland waterway or sea transport This is particularly true for Ireland as 90% of freight traffic occurs on the road. (Ryan, Dundon et al. 2003) comments in a report on road transport in Ireland that; “In essence, what is in evidence is a congested road network, with regions that are poorly served, clogged-up metropolitan areas, blocked access points, lack of traffic information provision and inexplicably and incomprehensibly slow planning and implementation of schemes”. Dramatic increases in goods vehicle activity on Irish roads has been documented by the Central Statistics Office (CSO) (CSO 2006) in Ireland. The percentage increases from 1995 to 2005 are shown below;

- Tonnes-Kilometres carried (output / quantity of work done) – 230% increase
- Tonnes carried (total weight of goods carried) – 259% increase
- Total distances travelled (by Irish registered goods vehicles) - 166% increase
- Average no. of goods vehicles (unladen weight of 2 tonnes and over) - 190% increase

A good benchmark for efficiency and vehicle utilisation in freight transport is the level of empty haulage, i.e. the percentage of the journey that a vehicle is running empty. For 2001, (Mason and Lalwani 2004) highlight the UK statistic for empty haulage as 26.4%, in Ireland for 2003 Forfas, quotes a figure of 38% (Forfas 2007). Forfas, (Ireland’s national policy and advisory board for enterprise, trade, science, technology and innovation), released a document in 2007 called “Perspectives on Irish Productivity” (Forfas 2007). Chapter 8 of this report measures the productive performance of the Irish road freight sector, and rates it as only 70% as efficient as the UK. It states “Irish goods vehicles have smaller carrying capacities, are emptier for longer, travel fewer kilometres and have lower average laden factors”.

As part of the research for this project a survey to assess the “Current State of Logistics in Ireland” is currently being administered. What has become evident from this study is that companies who outsource their logistics have no integration with the transport companies with whom they work. Goods are shipped solely on the basis of the lowest cost transport provider, with minimal levels of collaboration or information sharing. With transport companies highlighting consumer pressure, manufacturing deadlines and insurance policies on carrying low bulk high
value goods a major factor, preliminary results show that the empty haulage figure for Ireland could be as high as 60%.

Methods for Reducing GHG Emissions Freight Transportation

The use of bio-fuels is an alternative to reducing emissions in freight transport. However, bio-fuels are as yet unproven, and larger infrastructure development is required for implementation. Highlighted by (Roj 2007) “diesel will probably remain the most dominant fuel for commercial vehicles for at least two decades”.

In modern supply chains, it is not possible for transport managers to rely on having complete control of full truck loads, running on pre-defined timeframes. Therefore, flexibility in transportation management is central to efficient supply chain execution. (Potter 2005) identifies two main themes in the importance of integrating transport in the supply chain, both of which lead to overall increases in efficiency. Minimise cost and where possible integrate transport into the SC.

Table 1: Ten ways to Increase Sustainability in Logistics

<table>
<thead>
<tr>
<th>Five ways to reduce costs</th>
<th>Five ways to better integrate transport into SCs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduce the number of haulers</td>
<td>Co-ordinated distribution networks</td>
</tr>
<tr>
<td>Transport cost visibility</td>
<td>Transport cost visibility</td>
</tr>
<tr>
<td>Reduce administration costs</td>
<td>Exploitation of ICT</td>
</tr>
<tr>
<td>Maximise equipment use</td>
<td>Collaborative relationships</td>
</tr>
<tr>
<td>Consolidate shipments</td>
<td>Information feedback</td>
</tr>
</tbody>
</table>

ENVIRONMENTAL ANALYSIS OF SUPPLY CHAINS

When looking at the carbon footprint of a specific product it is important to consider the total carbon output from the product. This includes raw material extraction, manufacturing, distribution, retailing, consumption, disposal and recycling (Carbon Trust 2006a). In many instances emissions from logistics make up a large part of total product emissions. This paper is focused on reducing emissions in the logistical stage of the SC. The Carbon Trust identifies “integrated supply chain analysis” as the next step for companies to reduce carbon emissions, and cite three reasons for reduction of emissions in the SC;

- Increases in energy costs,
- Legislation that penalises high energy consumers, and
- Changing consumer attitudes towards low carbon products. (Carbon Trust 2006b)

A company’s full carbon footprint covers 3 main areas. 1) Direct emissions from activities the company controls, such as heating and energy use used in manufacturing or processing. 2) Emissions from electricity use, and 3) indirect emissions from products or services, (i.e. the sourcing of products and their eventual disposal or reuse). The Carbon Trust (http://www.carbontrust.co.uk/) defines two main reasons why a company would calculate its carbon footprint 1) To manage the footprint and reduce emissions over time; 2) To report the footprint accurately to a third party. In a recent report on Carbon Footprinting the (Carbon Trust 2007) highlights the following five steps for accurate calculation of a carbon footprint:

- Define the methodology – use consistent conversion factors, and a reliable method of calculation
- Specify the boundary and Scope of the coverage – which and what emissions are included, how far up and down the supply chain to calculate
- Collect emissions data and calculate the footprint – accuracy relies on calculating the correct data
- Verify the results – verification from a third party to add credibility
- Disclose the footprint – make all the above information available for review.

Calculating GHG Emissions Using Static Models

Recent studies focusing on elimination of air emissions have shown that some lean techniques (i.e. Continuous Replenishment, Just in Time) may not be the most environmentally conscientious method of manufacturing. By using CO₂ emissions as a Key Performance Indicator (KPI), (Venkat 2006) analyses supply chains by emissions and categorises them as being highly sensitive to the frequency and mode of delivery of goods. The type and amount of inventory stored at each company is also a factor, though the main impedance to CO₂ minimisation is distance.

Static simulation studies carried out by Venkat have shown that by varying the order size at each point of the supply chain, the optimal order size at each stage (that minimises emissions), can be calculated. In short, Venkat claims that lean supply chains do not necessarily reduce CO₂ emissions, particularly when distances along the supply chain are significant, adding that larger deliveries at less frequent intervals all along the supply chain generally lead to the lowest emissions. An example of a generic supply chain used in the analysis is shown in figure 2.

The generic supply chain example has 4 process and 4 transport stages. The distances between the process stages are fixed. Transport to the customer varies on order quantity and destination and requires further analysis of sustainable distribution and vehicle routing methods for the emissions calculations, Venkat assumes;
SIMULATION TO ANALYSE SUPPLY CHAIN DYNAMICS

The work in this paper is part of an on-going research project to evaluate the use of simulation as a decision support for designing environmentally sustainable SC’s. Globalisation has now made SC’s more complex, with material moving in smaller quantities, more often, over longer distances than ever before. This paper proposes DES to help decision makers evaluate and design SC’s in a more environmentally friendly fashion. DES permits the evaluation of operating performance prior to the implementation of a system. It also enables companies to perform powerful what-if analyses leading them to better planning decisions. DES also allows the comparison of various operational alternatives without interrupting the real system; it permits time compression so that timely policy decisions can be made. (Chang 2001)

Simulation to Support Supply Chain Design

Most simulation tools are designed as interactive tools to be used by a human planner not as real time decision-making tools, which are directly linked to control systems to dispatch tasks. It is envisaged that the tool completed in this project would be of that type. The use of simulation tools can aid the human planner to make the right decision by providing relevant information to the current system. However, in order for the correct use of the tool, the human planner should be able to interpret and modify the plan in order to achieve better supply chain performances, (Chang 2001). When studying a supply chain model and analysing the data requirements Chang highlights some of the questions the users might have:

- Which supplier policy is achieving best delivery performance under given demand pattern?
- Which supplier policy is most robust under demand fluctuation?
- Which is the most cost saving inventory policy under given demand pattern?
- How would profit be impacted by adding X% more capacity?
- What is the trade-off between delivery performance and inventory cost when building more inventory?
- What is the impact of information accuracy on the manufacturing performance (e.g. cycle time, order fill rate)

All of the above questions have the capacity to be addressed by an SC simulation model; however it is the addition of another four questions that forms the basis of the environmental SC simulation model undertaken for this project. The questions are;

- What are the environmental and economic trade-offs in altering supply chain design?
- How is the carbon footprint of a supply chain altered by increasing or decreasing suppliers, and or focusing on supplier locations and delivery modes?
• How is the carbon out put of a SC affected if there is a spike in demand, and how to find the best resultant action, environmentally?
• What is the optimum environmental SC design of interacting companies using different inventory policies economic order quantities and production systems etc?

Environmental Decision Support System - Model Basics

The following section introduces the participants in a typical SC for the electronics sector and gives an example of their interaction. Although the model found in figure 3 is not based directly on an individual supply chain, it is representative of currently used supply chains in the electronics component manufacturing sector in Ireland. It is first useful to describe the players in the supply chain of which there are six;

![Figure 3: typical supply chain](image)

OEM - Original Equipment Manufacturers. (DLH Services 2003) refers to an OEM as “a company that builds products using components from other companies. An OEM sells these products under its own name. Many OEMs differentiate themselves by building solutions rather than just products for their customers”.

CM - Contract Manufacturers. (Hassig 1995), outlines that as the electronics industry changed, OEM’s realised the industry was moving to quickly for them to compete. They were purchasing new machines that were becoming obsolete, and were forced to lay off workers that were hired during production peaks. Therefore as the industry changed, contract manufacturing became the norm. Hassig claims, “World Class CM’s have manufacturing capabilities that rival all but the very largest OEM’s”, with “CM’s assembling components and materials often owned and supplied by an OEM”. CM’s have become a popular management incentive as they, “help reduce personnel costs, inventory management headaches, and capital expenditures”.

Suppliers - A contract manufacturer may work with up to 50 suppliers at any one time (Teravainen 2002). Suppliers form an integral part of the Contract manufacturer’s success and form a “crucial” part of the supply chain. Teravainen adds, in order for contract manufacturers to remain competitive in the marketplace, suppliers are expected to provide flexible, fast service and maintain competitive prices.

4PL - 4th Party Logistics Providers. “The term 4PL was actually coined by the consulting group Accenture, who also own the trademark. Accenture defines a 4PL as; “An Integrator that assembles the resource capability, and technology, of its own organisation and other organisations to design, build and run compressive supply chain solutions”. With (IT Toolbox 2003) noting that “A 4PL organisation (i.e. consultancy firm) would then build a set of activities focused around a specific set of supply chain initiatives goals and services, such as warehousing, IT implementation and administration”.

3PL - 3rd Party Logistics Providers (Piasecki 2003) describes 3PL’s as “businesses that provide one or many of a variety of logistics related services. Types of services include;
- Public Warehousing
- Contract Warehousing
- Transportation Management
- Distribution Management
- Freight Consolidation

Customers – Any party buying products or services

Depending on the amount of goods sourced, produced and distributed the SC in figure 3 can become quite complex and very intricate. Though its main operations can be characterised in the following 7 points;

• Orders come in from customers to the OEM
• The OEM then issues work orders to their CM’s.
• If the CM’s cannot fulfill the order straight away, a request for material is sent to their suppliers who ship whatever stock possible
• If the supplier stock cannot be meet. The work is placed on back order and shipped when available
• The once the CM has received the stock it is manufactured and sent to the OEM
• On leaving the OEM the goods may take any number of routes before it reaches its intended customer. Distribution centers, and retailing outlets around the world may be used
• The transportation of the products are facilitated by a 4PL and carried out by a 3PL

For environmental calculations in the model, conversion factors from the department of Environment, Food and Rural Affairs U.K. (DEFRA) were used. DEFRA released Guidelines for Company reporting on Greenhouse Gas Emissions (DEFRA 2005). The emissions calculations are based on the National Atmospheric Emissions Inventory for 2003 and the UK GHG Inventory for 2003 developed by Necten and are available at. (http://www.naei.org.uk/). The NAEI (National Atmospheric Emission Inventory) has developed this information for the UK. The factors are given in the form of “per Kg unit of CO2”, released. The report includes
• Fuel Conversion Factors – (Natural Gas, Grid electricity etc.)
• Heat and Power conversions
• Electricity Conversion Factors from 1990 – 2003
• Process Emissions (Metal production, chemical etc.)
• Conversion factors for emissions from Refrigerants and Air Conditioning Systems, and most importantly
• Transport Conversion Factors – which consist
• Standard road transport fuel conversion factors – (Petrol, Diesel etc)
• Passenger Road Transport Conversion Factors – (according to engine size, for Petrol and Diesel cars)
• Rail and Air Passenger Conversion Factors
• Rail Air and Sea Freight Conversion Factors
• A Breakdown of Diesel Road Freight Conversion factors

Using the conversion factors available at DEFRA, the emissions from goods vehicle used and the percentage fill can be calculated. It is worth noting that these emissions values change on an ongoing basis. However the nature of the tool allows the user to openly define the factors for use. DSS described in the next section uses the “Transport Conversion Factors” described above. However an expansion of the simulation tool to allow for all other necessary conversions is planned.

DSS Model Evaluation

The simulation software package used for building the model is eM-Plant. eM-Plant is an object-oriented discrete event simulation package developed by Tecnomatix (www.tecnomatix.com). The eM-Plant package was chosen to develop this model due to existing licensing availability and existing expertise in the University of Limerick. Models in eM-plant can be quickly built due to development techniques such as polymorphism and the inheritance of component structures. With these development methods coupled with its powerful internal programming language called SimTalk. (Su and Shih 2002) highlight 5 advantages of using eM-plant as;

• They can build hierarchical models either bottom up or top down
• They can reuse sub models or parts of models stored in the engineering database
• They can use several modelling techniques for different levels to verify the results of each level and dramatically increase productivity
• They can centralise and define changes of all relevant system parameters in one table, and
• They can integrate in real time with other software.

Initial model development has been carried out by the authors using the eM-Plant package. Figure 4 shows a screenshot of the five different “frames” used. The frames contain the specific and characteristic information for the Product Demand, OEM, Manufacturer, Supplier(s) and a Results frame for analysis. They are all contained in the Carbon Model frame and constitute the primary elements of the model. The different variables for this for each of the elements are outlined below;

![Figure 4: Screenshot of Simulation Model](image)

**Demand**
- Profiling annual and periodic (weekly)
- Demand patterns can be profiled

**OEM**
- CM’s Flexibility Constraints
- Product Quantities
- Purchase Orders

**Manufacturer**
- Bill of Materials (BOM) for the product
- Set up and Production times
- Production Costs
- Capacity Constraints i.e. Shift Times
- Demand Fluctuation Planning abilities
- Component Stock Levels
- Specific Material Order Policies (i.e. (Q, r ) (s; S))

**Supplier**
- Component Part Numbers
- Unit Prices
- Leatimes
- Batch Sizes
- Material Order Quantities
- Finished Stock Levels
- Back Order Strategies
- Time(s) to Replenish Stock

**Results**
Results for the model include the % of demand satisfied by the CM and Suppliers on time while also a definitive breakdown of the costs involved.

**- Costs**
The following costs have been verified with industrial partners and added to the model to accurately describe any Economic verses Environmental tradeoffs.
- Administration
- Storage
- The unit cost per product, and
- Logistics costs
- Kg of CO₂ released
On starting the simulation sequence the model works as follows;

- Product demand from the OEM is received by the CM where it is referenced against the CM's bill of materials (BOM) data and corresponding raw material inventory levels held at the CM.
- Where necessary, components are ordered from suppliers in order to maintain levels of stock adequate to satisfy forthcoming product demand.
- Purchasing constraints (i.e. minimum order quantities and batch sizes) for products on the supplier side will have influence on the size of these orders.
- The suppliers then ship the requested amount of their respective components to the CM, if they are unable to ship the entire requested amount then the remainder is placed on backorder and shipped when available.
- The CM then determines its maximum production capacity (based on physical capacity and material constraints) and produces the outsourced product accordingly. Any available or required buffers of finished goods are also considered at this point.
- Once manufactured, the products are tested and then shipped to the OEM where they are recorded into stock.

When the model has finished the total Kg emissions from the scenario used is collected and correlated to the cost and % of product demand satisfied. The model has shown the potential to analyse SC’s and give an accurate account of the carbon equivalent outputted by SC activities. Further work on the model is required to answer the all four questions outlined at the start of the section. Experimentation on finding the optimum design for a SC has begun. Further analysis in the coming months will create more complexity in the model in order to give a broader and more detailed account of the emissions from a SC.

CONCLUSIONS

The logistics associated with supply chains is becoming increasingly, complex and difficult to manage as trends in globalization expand. This paper highlights the impact that this logistical activity is having on the environment, particularly with respect to CO2 emissions related to the freight sector. As is clear from this paper, the emissions associated with freight transportation is increasing rapidly, particularly in the Irish context. In an attempt to reduce these emissions this paper examines the impact of supply chain design on the environment and attempts to quantify its impacts.

The paper reviews present modeling techniques to measure this, and proposes the use of discrete event simulation as a support tool that could be used to aid decision making with respect to supply chain design. Such a tool would capture the inherent variability associated with supply chains and combined with logistical emission conversion factors would outline the logistical environmental impact from that supply chain design. Such a tool could then be used to experiment with alternative supply chain designs. This tool would thus enhance the decision makers ability to easily identify emission saving activities and identify the financial repercussions (positive or negative) of such supply chain reconfigurations.

As this is an ongoing research project, definitive models have not been completed, but initial findings have great potential with preliminary results outlined in this paper. However, further work is required.

REFERENCES


Mason, R. and C. Lalwani (2004). Integrating Transportation into the Supply Chain to Improve Supply Chain
modelling. Research interests are: Simulation Modelling of Discrete Event Systems; Modelling and Analysis of Supply Chains and Manufacturing Systems; Process modelling; Component-based simulation; Decision support systems. His email address is <cathal.heavey@ul.ie>.

ACKNOWLEDGEMENTS

This work is funded by the Environmental Protection Agency (EPA) of Ireland

BIOGRAPHY

PJ BYRNE is a Lecturer of Operations/Supply Chain Management in Dublin City University Business School. He has a B.Eng. degree in Production Engineering, from the Manufacturing and Operations Engineering department in the University of Limerick. He has also received a PhD from the same university in the area of supply chain simulation. His research interests are Supply chain design, analysis and optimisation. Environmental impacts of supply chain construction. Company outsourcing, decision-making and costing and Simulation applications in the semiconductor industry. His e-mail address is <peter.byrne@dcu.ie>.

PAUL RYAN is a currently pursuing a Ph.D degree in the Enterprise Research Centre in the University of Limerick, Ireland. He received a B’Sc degree in Manufacturing Systems in 2006 from the Manufacturing and Operations Engineering Department in the University of Limerick, where he held work experience in the Semiconductor Industry. His work is funded by the Environmental Protection Agency with research interests including Environmental Technologies, Sustainable Logistics, Supply Chain Design and Operation. His e-mail address is <paul.ryan@ul.ie>.

CATHAL HEAVEY is a Senior Lecturer of Operations Management in the Department of Manufacturing and Operations Engineering at the University of Limerick. He is an Industrial Engineering graduate of the National University of Ireland (University College Galway) and holds a M. Eng.Sc. and Ph.D. from the same University. He has published in the areas of queuing and simulation
A CONCEPTUAL MODEL FOR SME MECHATRONICS SUPPLY CHAIN

Jihene Tounsi  
Université de Savoie  
SYMME-Polytech’Savoie  
12, chemin de Bellevue,  
BP80439  
74940, Annecy-Le-Vieux  
Jihene.tounsi@univ-savoie.fr

Julien Boissière  
Université de Savoie  
LISTIC-Polytech’Savoie  
12, chemin de Bellevue,  
BP80439  
74940, Annecy-Le-Vieux  
Julien.boissiere@univ-savoie.fr

Georges Habchi  
Université de Savoie  
SYMME-Polytech’Savoie  
12, chemin de Bellevue,  
BP80439  
74940, Annecy-Le-Vieux  
Georges.habchi@univ-savoie.fr

KEYWORDS
Small and Medium Enterprises (SME), mechatronic, supply chain (SC), modeling methodology, conceptual modeling, simulation

ABSTRACT
Nowadays, in order to meet global dynamic market and its perturbations, the supply chain and its integration across the different organizations has become a competitive business advantage along the production processes. In particular, the mechatronic industry is very sensitive to these requirements. Our research deals with modeling the supply chain in order to manage and implement a simulation platform. Basically, we develop a modeling solution for an industrial environment composed by mechatronic Small and Medium Enterprises located in Savoy (France). The study is based on elements from real applications and the SME network organization. An analysis of the domain helped us to develop concepts and a modeling approach adapted to these companies. We present in this work the studied area, the methodology used to construct the conceptual model of the supply chain, its structure and management rules. Then, we present the model obtained. And finally, we apply this conceptual model on a case study.

INTRODUCTION
The integration of the supply chain between the various actors is a major strategic challenge and a strong competitive advantage. Indeed, the supply chain is a complex macro-system, first due to the variety of implicated organizational structures and the relationships between them, and secondly due to the strategic decisions it involves.
Small and Medium Enterprises (SME) evolve in an unstable and complex economic environment. In order to survive and to be able to support supply chain’s requirements (internal factors: influent actors, delay, consumer satisfaction, etc.) and market requirements (external factors: competition, unpredictable mutation, etc.), SME have to collaborate altogether to achieve their goals without losing their autonomy and identity (Villarreal Lizarraga et al., 2005; Julien, 1997).
The industrial environment of Savoy, corresponding to our research field, consists of numerous productive SME of the mechatronic industry working together to achieve a common goal in a complex global supply chain. We are interested in mechatronic SME as it represents a quite new concept that emerged in several industrial areas that need to integrate mechanical, electronics and computer knowledge simultaneously into production. These SME are an economic and a competitive advantage for the survival and the prosperity of region’s economy.
Due to the complexity and heterogeneity of a supply chain composed of SME and the lack of work in this area, we have focused our research in this direction. For all these reasons, we propose to study the structure and the SME dynamic behavior in their global supply chain.
In this paper, we present an analysis of the industrial environment as well as our modeling approach in the context of mechatronic SME. Thus, this work is divided into three main parts. The first part presents a study of the existing situation concerning theories discussed in literature. In the second part, we define the modeling methodology inspired from research in literature. Then in a third part, we develop a conceptual model representing our supply chain vision. The last section concludes the paper by developing the model based on an actual industrial case.

LITERATURE OVERVIEW
In this section, we define the place of SME in the supply chain and in supply chain management (SCM) based on the literature. We also present the problem which encouraged us to try to reduce the complexity of supply chain in our context. Moreover, in the last part we present the different modeling visions in the supply chain and SCM field.

Supply Chain and SCM
The supply chain concept was born in the 90’s when management techniques in the business world had evolved from separate logistics to collaborate logistics. According to Christopher (Christopher, 1992), the supply chain can be seen as a network of companies involved, both upstream and downstream, to the different processes and activities that create value in the form of goods and services brought to the ultimate consumer.
As such, a company’s success lies in its ability to integrate managerial processes of the supply chain but also to coordinate with other actors (Drucker, 1998; Lambert and Cooper, 2000).
According to Bagchi and Skjoett-Larsen (Bagchi and Skjoett-Larsen, 2005), the integration can be defined as a complete collaboration between decision-making levels: strategic, tactical and operational level. It is based on 5 interrelated key points:
• Sharing of information and communication
• Collaboration among the actors of the supply chain
• Collaboration leading to sharing risks, costs and benefits
• Sharing of technical expertise, ideas and institutional cultures
• Organization

In fact, to be able to talk about a successful SCM tool, we must consider the supply chain structure, the processes that add value and the management components that define how to manage the entire supply chain (Cooper et al., 1997).

SME Role in the Supply Chain

On the one hand, we have found in literature a lot of work dealing with the integration of the supply chain from a big company’s point of view. In this case, the company is considered as the heart of the supply chain structure. This point of view mainly focuses on the producing company performance and pushes other companies to fulfill requirements of this major company or to synchronize with its manufacturing system.

On the other hand, SME are embedded in a complex and unpredictable system and subject to economic and social mutations. According to Julien (Julien, 1997), to find equilibrium between big companies and SME, SME must maintain and even improve their competitiveness in an increasingly competitive market. This competitiveness can be achieved by the control of information and processes but also thanks to a network structure.

Indeed, a SME in a supply chain becomes more efficient when it evolves into a network; Mallidi et al. (Mallidi et al., 1999) explains the advantages for SME to cluster in an SME network as it provides:
• Access to new markets by manufacturing new products that are out of reach of a single SME
• Increased productivity by grouping and aligning the productive capacity of each actor of the network
• Improved responsiveness thanks to common responses to the disturbances which are unbearable by a single actor
• Improved use of resources by avoiding functions duplication in the cluster.

The theory of SME network, which is based on collaboration, cooperation and coordination, is a crucial roundabout for its management strategy. These alliances will reduce their vulnerability and this stresses SCM integration in a global supply chain context. This philosophy is explained by Assens (Assens, 1994): “the enterprises network is based on an autonomous and independent SME network which interacts in the context of mutual trust relationships to reduce the uncertainty of their environment. During these local interactions, a global organization form emerges in the absence of a central regulatory entity”.

Modeling Approach

Modeling is a mechanism that reflects the actual system which, when coupled to simulation, provides a very powerful decision-making tool. This tool can consider the dynamics of the system and the behavior of different autonomous actors.

The literature is unanimous on the positive role of modeling and simulation (M&S) in complex systems study, analysis and performance evaluation. For example some authors (Lee et al., 2002; Longo and Mirabelli, 2008; Ingalls, 1998) highlight the features and advantages of a decision-making tool based on the modeling and discrete event simulation. The M&S approach has been adopted in several works in order to reduce the SCM complexity and evaluate its performance (Bagchi et al., 1998; Labarthe et al., 2007).

Indeed, the simulation translates the supply chain conceptual model and recreates the complexity and highly stochastic environment of a real supply chain. The conceptual model defines concepts (implicated entities) and parameters that give the possibility to a supply chain manager to analyze different scenarios by changing input parameters (Longo and Mirabelli, 2008).

Several orientations of supply chain modeling arise when studying literature. On one hand, Thierry (Thierry, 2003) provides two orthogonal visions:
• Product Vision: this vision considers the supply chain dedicated to the product (or product line) from the raw materials to the final customer. It focuses on the product flow to define the starting point and end point of the supply chain. Rota et al. (Rota et al., 2002) define the supply chain dedicated to a product (or family of products) as “all companies involved in the processes of manufacturing, distribution and sales of the product, from the first of suppliers to the ultimate customer”.
• Company Vision: this vision is centered on the company rather than the product. In this case, the object is the supply chain of the company. This vision focuses on the company modeling, as an organization within a supply chain with all the products it manufactures and all the relationships it might have with other actors in the supply chain. Many models have been proposed in this perspective among them the famous SCOR model, or the model proposed by Keaney (Kearney, 1994).

These two visions are disjointed. They cannot be handled in the same time due to the number and the complexity of the separate supply chains crossovers.

On the other hand, two other visions can be found in the literature:
• Process Vision: it is based on the process classification according to decision-making level (Chopra and Meindl, 2001; Stevens, 1989): strategic, tactical and operational level. This point of view deals with identifying processes that operate in the supply chain and their decision-making levels.
• Structure Vision: it has been proposed by Cooper et al. (Cooper et al., 1997). The structure vision considers the structure of the supply chain made of:
  o Actor’s type: in a model, the actors who have an important decision-making role (primary actors) should be identified as well as those who synchronize activities without taking part in the decision-making (secondary actors).
  o Network structure: the structure is defined in two dimensions. The horizontal dimension defines the different roles that actors can have in the supply chain (suppliers, producers, etc.). The vertical dimension defines the number of actors for each role.
Relationship characteristics: actors share several processes that define their behavior characteristics in the supply chain (process management, control and monitoring or outsourcing). These two visions are complementary: the choice of one does not exclude the use of the other at the same time. The first defines the supply chain from a process point of view and the second from an organizational point of view.

OUR RESEARCH SCOPE

According to a survey that we conducted in mechatronic companies in the region and some other former investigations carried out, we have found out that the industrial environment is mostly composed of subcontracting SME specialized in mechatronic. We have drawn from a questionnaire, several features of mechatronic supply chain in the region:

- Complexity: the supply chain is composed of several autonomous and independent actors. The relationships between actors differ depending on the objectives and geographical location. In fact, there are network companies working together to achieve a given process. Each actor may be seen as an expert actor for a specific activity.
- Decentralization/outourcing: it is a strategic decision due to globalization and the desire to reduce costs. The studied SME are mostly subcontractors, which are not necessarily located in the same region or even in the same country.
- Lack of visibility in the global supply chain: SME collaborate by linking their activities to achieve an objective or a process. Or, as a result of the previous characteristics (decentralization and outsourcing), the supply chain is divided in many sites spread over several geographic locations depending on their purpose or activity in the global supply chain (distribution, manufacture, supply, etc.). These sites only have a local visibility but are coordinated with other sites through products flows.

In the following section, we discuss the scope of our research, namely a supply chain in SME mechatronic. Modeling and simulation of the supply chain behavior in the SME mechatronic context, is a challenge and a growing need expressed by professionals. Our choice for M&S approach is justified, on the one hand, by the lack of analytical models to model the dynamic of the global supply chain with all its complexity and on the other hand, by the power of simulation to manage stochastic behaviors of the supply chain (Lee et al., 2002).

MODELING METHODOLOGY

In the previous part, we have defined different points of view adopted to model a supply chain. In this section, we present the modeling methodology we followed. Our approach combines in the same time the “structure vision” with the “product vision”. These two visions are based on the “product vision” which represents our modeling framework. To generate the conceptual model, we propose an incremental methodology based on these 3 visions recalled in figure 1: product vision, structure vision and process vision. In each step, a vision is applied to construct and to refine the conceptual model. The result of each step (intermediary model) will be the input of the next one. Hence, at the end of the 3 steps, the architecture of the conceptual model is obtained.

![Methodology Framework](image)

**Figure 1. Methodology Framework**

Product Vision

This vision leads us to define a first abstract model of the supply chain. This first model follows the product life cycle. So by adopting a product point of view, we identify the frame of the supply chain which consists in:

- The environment: it is characterized by the flow circulation and the different steps of the product’s transformation as well as related disturbances.
- The organizations: to assure the physical flow management, we need to define the organizations which carry out one or several product’s transformation stages. Knowing that we are in a supply chain context essentially composed of SME, the implicated organizations can be a network of firms that collaborate to accomplish one or several transformation stages.

Structure Vision

On the basis of the abstract model provided by the previous step, the “Structure Vision” details the organizations that are involved and also the physical environment. This vision is based on the identification of 3 essential elements: implicated entities, vertical and horizontal hierarchy as well as organizational relationships between them.

- “Structure vision” of the environment: this is the part containing the physical flow. So, we need to model the product flow and the resources needed to achieve its transformation.
- “Structure vision” of organizations: this consists in identifying the actors who take part in the network. Actors will be prioritized in the organization according to their involvement in decision-making level and tasks that will be awarded. The information flow management will be different depending on the decision-making level.
At the end of this step, we obtain a more detailed intermediate model.

Process Vision

While implementing “Process Vision”, various categories of processes that may occur can be identified. This can be done according to the decision-making level, but also depending on the actors’ relationships. These relationships can be classified in two categories:
- Management and control: contains the processes that ensure proper decision implementation in the perspective of a continuous improvement of processes in terms of added value.
- Synchronization: contains processes for exchanging information and physical flows according to a processes scheme developed and already predefined by the decision-making layers.

At the end of this stage, we get a refined conceptual model of the supply chain.

CONCEPTUAL MODEL GENERATION

In this section, we present the generation of the conceptual model as applied to SME context following the different phases of the proposed methodology. The objectives are to identify the concepts of the model, its architecture and processes.

Applying “Product Vision”

By applying the “Product vision” method, we get a first abstract model of the global supply chain (Figure 2).

![Figure 2. Abstract Model of Supply chain](image)

This model consists of:
- An environment defined by the space allotted to the product flow and management between the elements of the supply chain known by internal resources as well as all the external elements that can influence the chain activity.
- Sub Supply Chains (SSC): each SSC represents a group of SME which collaborate to achieve an internal aim and/or the overall objective of the global supply chain. The SSC is responsible for the processing of the product in a certain stage of its life cycle.

- A perimeter (boundary) of influence, representing the visible environment for the SSC and on which it can act by internal conferring (if the action does not disturb the environment located at the outside of its visibility) or by conferring with other SSC.
- Shared perimeter of influence: it is the perimeter representing the area of flow transfer between two SSC. It’s a shared zone where two SSC coordinate their activities to allow the flow transfer.

This abstract model is the first result of applying “Product vision”. In the following paragraphs, we detail this model by applying the other two visions to structure the architecture and the processes of the model...

“Structure vision” application

In this stage of modeling, we detail the internal architecture of the SSC and the environment which is visible through its perimeter of influence.

SSC and Environment Structure

In the conceptual model, a SSC (Sub Supply Chain) defines a network of companies that collaborate to achieve common goals (internal and/or external). It consists of three layers representing the different decision-making levels (Figure 3):

![Figure 3. Layers of the SSC](image)

Each layer is composed of several concepts and plays a specific role in the SSC.
- “The Monitoring System”: it is the intelligent layer of the SSC that ensures the monitoring and the control of the other SSC’s layers as well as the coordination with the other SSC of a global supply chain. Indeed, it analyses the information coming from the “Execution System” and establishes the necessary metrics. Based on these metrics, it measures the SSC performance and accordingly acts on the other layers. This system must also coordinate decision-making activities with the other SSCs of the supply chain in order to optimize the product flow and to achieve the global objectives. At the “Monitoring layer”, the flows are managed by different components. These components are the units responsible of the decision-making and control.

As presented in figure 4, the “Monitoring System” layer is composed of Monitoring Actors (MA) which can be defined as follows:
- Monitoring Actor (MA): it is the concept that models an intelligent actor. It represents the actors involved in all SSC layers and particularly in the “Monitoring System” for decision-making activities...
and the control of the other SSC layers and the global supply chain. In some cases, some MA involved in the “Execution System” could have dual roles or objectives (synchronization / monitoring).

**Figure 4. Monitoring System**

- **“Execution System”**: it corresponds to the reactive layer of the SSC. Firstly, it synchronizes the operations of the “Physical System” according to the information flow gathered from this system. Secondly, it is a reactive system; it acts on the Physical System by observing it and its perimeter of influence. In exceptional situations, the “Execution System” refers to the “Monitoring System” for coordination and decision-making.

**Figure 5. Execution System**

The “Execution system” layer (Figure 5) is composed of two types of actors: the monitoring actors (MA) and the executive actors (EA). These actors apply a reactive control method on the “physical system” layer.

- **Executive Actor** (EA): the EA act on the “physical system” changes by applying a reactive control. They observe the physical environment and from the captured perception they synchronize the operations of the “physical system” in order to correct perturbations.
- **Monitoring Actor** (MA): a MA of the Execution System is identical to a MA of the Monitoring System. But in this case, the control activity of a MA is the synchronization of the physical flow.

- **“Physical System”**: it is the visible part of the environment for a SSC corresponding to its “perimeter of influence”. It contains the elements that model the physical flow and the resources needed to achieve the SSC’s goals or to process an operation.

**Figure 6. Physical System**

As it is presented on figure 6, the “Physical System” is composed of concepts that are passive towards themselves and are controlled by the other layers of a SSC:

- **Moving Entity (ME)**: it models the physical flow (or product) in circulation (motion) on the supply chain.
- **Resource (R)**: this concept models the resources needed for actors to carry out their tasks or activities.

**“Process Vision” application**

During the previous steps of modeling, we built up an intermediary supply chain model by identifying the organizations structures and the environment in which they operate. In the following section, we refine the model by integrating different types of processes. This integration depends on the role and the decision-making of the different systems.

- **“Monitoring System”**: at this layer, we identify two kinds of processes that could be described as “Monitoring and Control Processes” (MCP) and “Strategic Processes” (SP). The MCP is short and medium terms management processes: their role is to control the system according to the perceived perturbations, but also to drive and evaluate the SSC’s local performance and in the global supply chain. The SP, which are the management processes in the long term, appear in the global supply chain. The decision at that level must be done in coordination and interaction among all the SSC integrating the supply chain.
- **“Execution System”**: at this layer, we identify the management processes in the short term and they can be described as “Operational Control Processes” (OCP). These processes allow the actors to synchronize and control the flow circulation in the physical layer.
- **“Physical System”**: at this layer, there are “Physical Processes” (PPs) circulating between resources. They define the transformation stages of the product. A physical process can be supported by one or many SME.

**The Final Conceptual Model**

At this step of the methodology, the final conceptual model is built. This model is divided in two main sub models: the first being the sub model of the global supply chain and the second is the sub model of network’s organizations involved in the supply chain.

**Figure 7. Global Supply Chain Model**
The global supply chain model focuses on organizations and processes that coordinate the activities. This model is presented in figure 7.

The SSC model corresponds to the structure of the SME network as well as the supported processes. This model is presented in figure 8.

![Figure 8. SSC Model](image)

In our model, we can distinguish two possible relationships. The first one corresponds to the coordination of the activities between several SSC of the global supply chain. The second one corresponding to collaboration is an internal communication mechanism between the actors involved in the same SSC. We have not chosen the cooperation as an integration mechanism because it does not meet our needs. Indeed, we have based our choice on resources, benefits and risks sharing, common decision-making and objectives.

**Coordination in the Global Supply Chain**

SSCs will coordinate their activities in order to achieve the global supply chain objectives. At the global level, each SSC has a limited perimeter of influence. For this reason, a SSC needs to communicate with other SSC to face the environment disturbance by coordinating their activities and local objectives in order to achieve the global objective. Coordination will be done through the exchange of physical flow between “Physical Systems” of SSC. The physical flow exchange is unidirectional. Coordination also occurs between the “Monitoring System” of SSCs through information flows and performance indicators. In this case, the information exchange is bidirectional. This coordination takes place in the following scenarios:

- **Initial configuration of the system**: SSCs interact in order to set and coordinate their goals in the global supply chain.
- **Environment perturbation affecting all goals**: if a perturbation occurred in a “Physical System” of a SSC which requires a modification of its local objective and which affects other SSC objectives, coordination between SSCs is necessary in order to find an appropriate solution to the problem.
- **Reconfiguration of the system**: if SSCs recognize that the current system configuration can not achieve the global objective, the interaction between SSC is required to set new objectives according to their experience.

**Collaboration in the SSC**

The SSC represents a SME network collaborating to reach a local objective. The collaborative process takes place in the following scenarios:

- **Physical system synchronization**: in the “Execution System” of the SSC, actors will synchronize their information and their actions. This collaboration is frequent in the SSC in commonly encountered situations.
- **Monitoring SSC**: monitoring and control of the SSC will be based on the performance evaluation and the resolution of new perturbations on the physical environment which do not disturb the other SSC’s objectives.

**INDUSTRIAL CASE**

The case study presented here deals with a SME group that acts as a Mechatronic providers in the French territory and the Benelux countries. Its products are manufactured in
China and Thailand factories. Then, they are distributed worldwide by providers such as the one studied. The production process is invisible to providers. Indeed, in this example, customer needs (forecasts and demands) are gathered annually by decision-making actors in the network. After that, the forecasts are sent to other organizations involved in the production planning. The visibility in this supply chain is defined by the geographical location. In consistency to the conceptual model formerly presented, we consider the example of this supply chain as a node between several sub-supply chains (SSC). Let us call this SME network “SSC-France” and the other networks “SSC-related country”. Thus, we get a model of the global supply chain which is reflected in figure 9.

![Global Supply Chain Model](image)

**Figure 9. Global supply chain of SSC-France**

In this model, we have identified the different networks in our supply chain example. Then, we identify the structure of SSC-France network and its processes. We have studied the actors of the SSC-France depending on their role and their decision-making level according to the conceptual model. We have also studied the processes that are supported by SSC-France and have sorted them in the previously identified categories of the SSC conceptual model.

**SSC-France Actors**

In this part, we will categorize SSC-France actors according to their role. The different roles as well as the number of actors for each role are listed below:

- **Logisticians**: they are actors who supervise the flow of cartons from their arrival at Marseille’s docks until their delivery. They coordinate activities of other actors in the SSC, plan SSC’s requirements and monitor the evolution of the physical environment and conditions delivery. They also manage reverse logistics in litigation case. For this role we identify 4 SME. Each of them has a specific monitoring role according to processes evolution. We will designate these SME L1, L2 and L3.

- **Storehouse**: these are organizations that support the products storage and orders preparation. In this kind of role, we identify two actors which we denote S1 and S2.

- **Carrier**: the SSC deals with 2 transport companies according to the travel mode, destination and delivery type. We will designate this 2 SME by C1 and C2. C1 handles transportation from the warehouse to customers or to another carrier in the case of some Benelux countries. C2 is responsible for transporting cargo from China or Thailand. This transport is done in the procurement case to meet customer’s orders.

**SSC-France Processes**

In this industrial case, we identify many processes which are managed by the SSC-France. These processes can be classified in five macro processes (in reference to SCOR model): plan, distribution, outsourcing, deliver and return management. Each macro process is divided in many processes carried out by one or more actors.

So, we will consider in our example four physical processes: PP1, PP2, PP3 and PP4.

**The conceptual model of SSC-France**

In this section, we refer to the modeling concepts, an identify entity from the industrial case.

**Actors**

L1, L2, and L3: Monitoring actors (MA).
S1: Monitoring actor involved in the “Execution System” as well (MA).
S2, C1 and C2: Execution actors (EA).

**Resources**

In our case, we need 3 resources to achieve all processes.

**Processes**

PPh1, PPh2 and PPh3: physical processes. In this case, physical processes can be supported by only one resource.

According to this correspondence, the conceptual model of the SSC-France is shown on figure 10.

![SSC-France Conceptual Model](image)

**Figure 10. SSC-France Conceptual Model**
CONCLUSION AND PERSPECTIVES

Mechatronic Small and Medium Enterprises integrate complex supply chains whose overall scopes are sometimes not visible. SME collaborate to achieve one or more processes in order to respond quickly to the supply chain demands and to survive to the environment perturbations. In this paper, we have analyzed SME in a mechatronic environment, developed a modeling methodology and proposed concepts dedicated for supply chain modeling and simulation. This methodology and these concepts were applied to an actual supply chain which is mainly composed of SME. We have also identified the communication mechanisms to manage and monitor the SME network and the supply chain.

A major perspective for this work that is being carried out at the moment is to study the dynamic behavior of each concept. The final aim of such a research is to implement a simulation platform for supply chains that mostly involve SME.

THANKS

We thank the APS (Assemblée des Pays de Savoie) and the Laboratory SYMME (Systèmes et Matériaux pour la Méchatronique) who fund this work through a research grant.

BIBLIOGRAPHY

LOGISTICS SIMULATION
TRANSSIM-NODE – A SIMULATION TOOL FOR LOGISTICS NODES

Larissa Neumann
Simon Deymann
Chair of Transportation Systems and Logistics
Department of Mechanical Engineering
Dortmund University of Technology
D 44227 Dortmund, Deutschland
E-mail: neumann@vsl.mb.uni-dortmund.de

KEYWORDS
Logistics, traffic simulation, yard management, forwarding agency, freight traffic

ABSTRACT
Strategic planning and operational decisions affect the efficiency of the business of a forwarding agency. The simulating-tool TransSim-Node is one of the first simulation tools for the purposes of forwarding agencies. It permits reproduction of internal and external material and information flows in logistics nodes of road haulage in detail. In addition, it enables to evaluate the effects of various strategies on forwarding agencies.

SYSTEM DESCRIPTION
The structure of a LTL terminal can be described by two subsystems: the yard and the terminal building. These subsystems are linked with each other by ramps. The basic task in the operation of a LTL terminal consists of routing vehicles in the yard, unloading and loading shipments on the trucks as well as sorting and buffering these shipments inside the terminal.

INTRODUCTION
Forwarding agencies dealing with LTL (less than truckload) offer transportation services for often heterogeneous general cargo including the collection of the advised consignments at the consignor and the transport of the shipments in appropriate time and quality to the indicated consignee. The general cargo is usually handled as groupage freight by the shippers. I.e. the transport of several small consignments, which are first collected by medium-weight trucks in the local range of a forwarder at different consignors (forerun), then consolidated and transported by line haul as collective consignment to another forwarder, and finally distributed to the consignees by the receipt carrier.

In order to provide these transports, the shipments collected at local traffic routes in one region first have to be sorted at a terminal for the specific long-distant relations. Due to divergent mass flows in long-distance traffic, semi-trailer or motor vehicles with trailers are used, whereas the local traffic is usually done with medium-weight solo trucks.

The provided services of a LTL forwarder can be divided into “transport processes” and “handling processes in the terminal”. The investment and operating cost of the terminal, accounting for approx. 35 % of total costs, are determined by the layout and the arrangement decisions of the operating forwarding agency. Consequently, the performance of a terminal has significant influence on the cost structure of a forwarder.

Especially the heterogeneous character of the goods and the flexibility with respect to costs and throughput reduces the application ranges of mechanized sorting and unloading technologies and make cost-intensive handling personnel necessary (DHL 2005). Hence, the challenge operating a LTL terminal is the effective personnel management.

When an empty or full truck arrives at the terminal, it has to be registered by the porter in most of the cases. Depending on its task (unloading/loading), an assigned or free door is attributed to the vehicle, or it is sent to a waiting place. Regarding the handling of outbound groupage freight, the processes are as follows: After the docking process of a medium-weight truck, the shipments are unloaded to an unloading buffer, serving only for temporary storage of unloaded consignments, by the freight carrier or personnel of the terminal. During this process, an alignment of the physical consignments according to the specifications on the loading papers takes place. Additionally, the shipments are labelled with barcodes where required. After having completed the unloading process, the truck leaves the strip door and the yard area as well.

In the course of the process, the consignments are consolidated within the terminal. They are removed from the unloading buffer, usually by fork lifts, and normally brought to the loading buffer for a specific route. Partly, universal buffer areas are used, but a relation-specific buffer enables the systematic pre-assembling for
dispatching and thus a faster loading of the long-distance vehicles.

After registration at the doorman, the long haul truck docks at the assigned load door, usually a fixed one or one allocated close to the relation-specific buffer in order to minimize the internal material flows. The loading process in general is done via terminal staff. Upon completion of loading, the vehicles leave the yard for driving to another terminal overnight (main carriage), being located near by the desired destinations of the shipments.

The process chain for inbound groupage freight is quite similar to the outbound processes. The long-haul trucks arrive at the LTL-terminal in the early morning, the shipments are unloaded at the assigned strip door and transported to the buffer areas for the different short-haul traffic routes. Following the load door allocation, the medium–weight trucks are loaded with the related shipments and distribute them to the consignees during the morning.

By analyzing the processes at an individual LTL terminal with its outbound and inbound traffic, normally four temporal shafts during the day can be identified (approx. 12 a.m. – 6 p.m., 8 p.m. – 12 p.m., 1 a.m. – 6 a.m., 6 a.m. – 9 a.m.) (see figure 2) (Lublow 1995).

**Terminal-layout**

Strategic decisions, which are made within the scope of new construction or the enlargement of a terminal, are described by the term layout planning.

One of the first decisions concerning the design which must be made for the terminal layout is how many doors a LTL terminal should own. The terminals offer two types of doors, strip doors and load doors. Therefore you have to distinguish between strip and load doors. The number of load doors can be determined easily because to every relation (daily truck-tour) not less than one load door is assigned. To avoid LTL terminals with a huge quantity of doors, the relations within the local range and main carriage are usually combined at one door.

The number of strip doors is not as easy to calculate. But according to Little’s law, the first step estimating the number of receiving doors is to multiply the daily throughput of trailers by the average time to unload a trailer and divide this time by the allocated time slot for unloading the trucks (Gue and Bartholdi, 2000).

Besides this task of layout design, it is important that the doors can be used flexible as strip or load door, optimized for the tactical freight mix. Therefore the dimensions of the functional areas in front of the doors must be the same all-over the terminal. These areas are important to build tightly packed loads and place fragile loads on top. They must be built up rectangular like the whole terminal, adjusted to the prevalent handling unit pallet (Deymann, 2006). Finally the layout designer has to choose between large storage areas in front of a door and a short distance from the door to the intralogistics path network. In a LTL terminal the throughput depends significantly on this distance because every shipment has to pass this distance two times, from the strip door to the intralogistics path network and back to the load door. The fourth challenge designing a LTL terminal is the temporary function as a short-time warehouse to fulfill the customers’ needs. The design of the warehousing areas inside the terminal depends on the quantity of stored pallets and the direct access on each pallet in percent.

For the improvement of such terminal-layouts complex process-chains have to be modelled, taking varying amounts of shipments, congestion and the path capacity for forklifts under consideration. But the most terminal-layouts have been constructed based on intuition, considering the key performance indicators or perhaps with the help of some spreadsheet calculations (Clausen and Goerke 2004). The consequences are specialized layouts that are poorly compatible for the varying daily business and expansions of LTL terminals.

**Assignment of trailers to load doors**

A ‘good’ dispatching and/or planning of resources has large effects on the costs of a terminal, because an optimization of the terminal by the change of technology is often not possible. Assignment decisions determine e.g. the path length between strip and load doors and therefore have an effect on handling and personal costs. So far, these decisions are made manually or based on the experience of the dispatcher, which are feasible and yield to the demanded performance (throughput). To optimize the
assignment of trailers to load doors on the tactical level, Chmielewski developed a helpful doorassigner, using a branch and bound algorithm. By using the doorassigner, reductions in transport afford [km] of about 15-20% are feasible (Chmielewski 2007). But a systematic optimization of both tactical planning and operational processes rarely takes place, especially the varying amount of shipments and the inclusion of layout restrictions are still problems for this complex mathematical optimization. For this reason a large potential to reduce costs by using simulation can be assumed.

Allocation of trucks to strip doors

Several Authors have proposed solutions to layout problems for freight terminals (PECK, 1983; TSUI and CHANG, 1990, 1992), but in every case, freight flows from strip doors to destinations were assumed to be known and independent of the layout. Under such conditions, the supervisor’s policy is equivalent to a First Come First Serve policy (GUE, 1995).

If the supervisor assigns incoming trailers to strip doors and manages forklifts between the functional areas based on the contents of the trailers waiting in the queue and the location of the doors, material flows depend on the layout. Normally the FIFO principle (first-in-first-out) is being used for the allocation of the vehicles to strip doors, but other strategies can be used either in the yard management. They are thereby not only affecting the workflow in the yard, because internal processes are influenced as well. Problems which occur in day-to-day business, e.g. delayed vehicles or daily fluctuations in volume, are remedied merely by using adequate operational strategies, for instance by modified strategies in the yard management.

APPLICATION OF SIMULATION

As there are not yet developed any explicit theories concerning the mentioned strategic and operational tasks in LTL terminals, simulation seems to be the most efficient way to reach a well-founded basis decisions to be based on. Therefore, the Chair of Transportation Systems and Logistics, Dortmund University of Technology, developed a simulation tool for logistics nodes called TransSim-Node in cooperation with Incontrol Enterprise Dynamics GmbH. TransSim-Node, based on the simulation software Enterprise Dynamics. It permits the illustration and simulation of material and information flows in logistics nodes (e.g. terminals, distribution centres) of road haulage.

The tool is used to test the interaction and the effects of strategic planning (e.g. layout decisions) on tactical and operational procedures and to derive rules for internal and external logistics processes. TransSim-Node enables to evaluate strategic planning and operational decisions for forwarders. Due to the fact that there is hardly any rationalisation potential in using different technical items if state of the art equipment is used, e.g. handling technology, the improvement of strategies and operations is the most important task to increase the economical performance of these agencies.

MODELING FORWARDING AGENCIES USING INDIVIDUAL ATOMS

The modelling of logistics nodes with TransSim-Node is done on a microscopic scale by using several atoms, which enable to simulate each individual shipment and thereby the connected internal and external processes such as vehicle movements and transports with fork-lifts. Every process on the yard area and within the terminal, starting and ending with the registration of vehicles at the doorman of the terminal, is modelled.

For TransSim-Node, some preconfigured logistical atoms of Enterprise Dynamics could be used (e.g. cranes, forklifts), but in addition, forwarding specific atoms had to be designed.

Most of the atoms which had to be developed are layout atoms, internal as well as external. Based on the processing of vehicles, which is not limited to traffic infrastructure, new area atoms were needed, e.g. waiting and shunting areas as well as areas at the doors or other loading points of the terminals. Concerning the internal areas, paths for forklifts, different buffers (loading, unloading, general), docks, ramps and processing and storage locations have to be modelled.

In addition, strategic atoms for routing the vehicles (‘Yard Management’) and for administrating buffer areas were developed.

Strategic Atoms for buffer areas

The clever utilization of internal areas is a substantial factor of success for forwarding agencies due to the high percentage of manual activities, especially in transportation and sorting of consignments. For implementing the strategies concerning the usage of buffer areas, a strategic atom was developed. Based on a grid for each buffer area, this modelling permits the calculation of the current used areas with consideration of the different sizes of the consignments (euro pallets, industrial pallets, half pallets, boxes, parcels and bar material).

The administration of buffer areas in TransSim-Node can be controlled by external algorithms, which determine the position of each shipment as well as the time required for the retrieval. For this calculation, not only the dimensions of the area and the consignments have to be considered, but also the orientation of the shipments on the fork lift and therefore the storage direction, the accessibility of the pallets as well as reloading if necessary.

Yard Management

In the simulation model different vehicle types and transport containers can be applied for the fulfilment of internal and external transport requirements, e.g. solo trucks, vehicles with trailers and semi-trailers, each with various transport containers. The term ‘transport container’ comprises fixed superstructure, swap bodies and containers, which can be used both in road and railway traffic. While fixed superstructures are connected to a vehicle durably and thus clearly assigned, swap bodies are without reference to a vehicle and hence realized as autonomous units in the model.
Due to the complex fleet and the various permissible combinations concerning superstructure and transport containers – which lead to 36 vehicle combinations – in combination with the task of each vehicle (loading/unloading/both), the routing of vehicles on the yard is extraordinarily complex, although the model distinguishes only between three areas at the yard (waiting area, shunting area, loading area):
- At shunting areas the various vehicle combinations are separated or combined. In addition they serve as buffers for trailers, swap bodies or containers, which can be moved on the yard not only by vehicles but also by transfer cars or reach stackers.
- At loading areas, consignments are handled additionally to the functions described above for the shunting area. This applies both to the motorized vehicles and to swap bodies or semi-trailers. Some vehicles can only be loaded/unloaded at specific loading areas, e.g. for side or yard discharge. Each loading area is equipped with a door and a dock leveller for balancing different heights of loading edges.
- Waiting areas only serve for temporarily buffering of vehicles, e.g. if no strip door can be assigned by the porter.

For controlling the traffic on the yard, a strategic yard management atom is used, which supervises routing of the individual units of the fleet to the correct place of destination. The yard management atom conducts a list of all permissible transport chains depending on the situation and the task of the vehicles entering and leaving the yard, based on 250 identified transport chains. In addition, it administers the states of all external areas and the destinations of each transport container. This atom e.g. supports the doorman with the assignment of vehicles to loading areas or a solo vehicle with coupling the correct trailer at the shunting area.

EXPERIMENT AND ANALYSIS

As a first application of model building with TransSim-Node, a medium-sized forwarding agency in North Rhine-Westphalia, Germany, was modelled. The scope of the model is as follows:
- gate with one entrance and one exit, operated by one doorman
- yard with 15 waiting areas as well as 20 shunting areas
- terminal with 15 strip doors, 52 load doors for national shipments and 20 load doors for special and/or international shipments
- 1,500 shipments per day for outbound groupage, and 2,000 shipments per day for inbound groupage
- approximately 75% of palletized shipments (65% euro pallets), 13% of parcels, 12% of bar material
- 150 medium-weight solo trucks for local range transports and 52 semi-trailers for main haulage

Due to the fact that the real terminal requires massive data of terminal operation and planning, which were not available, a reduced model of the forwarding agency was used, focusing on outbound groupage. Part of the input data were generated by using a complex data generator, which enables the detailed modelling of the system load for forwarding agencies based on statistic distributions and real data. Therefore, the model data concerning the shipments (e.g. consignment type, consignor, consignee, assigned local range vehicle, relation, assigned main haulage vehicle) and concerning vehicles (e.g. type of vehicle, arrival time, cut-off-time) are artificially generated.

The terminal described above was simulated concerning the outbound groupage of the terminal to obtain several performance indicators. These enable to compare the basis scenario (T15, consisting of 15 strip doors) with different scenarios concerning the number of strip doors.

**Figure 3: Model of the LTL-Terminal**

The utilisation of the yard and doors intraday for outbound groupage for two scenarios concerning the variation of strip doors is shown in Figure 4 (herein T15 is the scenario with 15 strip doors respectively T10 the one with 10 strip doors). Two points are noticeable: in the T10-scenario the overall yard time and the waiting time of the medium-weight trucks is longer (comparing the surfaces of the lines), and the loading of the long distance vehicles is finished later.

**Figure 4: Number of vehicles on the yard intraday (comparison of scenarios T15 and T10)**

In addition, some other observations were made regarding the scenarios
- The average yard time for unloading (medium-weight trucks) raises from approx. 28 minutes up to 40 minutes
- The number of medium-weight trucks waiting increases by 30 % from 68 to 89; their average waiting time from approx. 7 minutes up to 25 minutes
- The last shipment in the T10-scenario is transported to the loading buffer more than one hour later than in T15 – an increase of total internal processing time of 18 %
- The last long distance truck in T10 leaves the yard 22 minutes later than in T15
- The average utilization rate of strip doors between 12:30 and 18:30 increases from 52.6 % (T15) up to 72.8 % (T10)
- The reduction of strip doors without new assignment of doors leads an increase of transport distances (+ 3.8 %) and transport times (+3.4 %) for the forklifts

![Graph showing average process and waiting times of vehicles for scenarios T15 to T10](image)

Figure 5: Average process and waiting times of vehicles for scenarios T15 to T10

The comparison of the performance indicators for the different scenarios shows the consequences of changes in the layout. As an example, the layout of a terminal causes the basic distances between the doors, and if no re-assignment of strip or load doors is done, the transport performance inside the terminal decreases. In addition, strategies on the operational level are required to obtain a better performance, e.g. a changed strategy for unloading the trucks by allocating them to time-slots. The used scenario planning in simulation is an ideal tool to obtain a well-founded understanding of the whole system and its interrelation.

**SUMMARY**

This paper introduces a simulation tool for logistic nodes, e.g. forwarding agencies, which is developed based on Enterprise Dynamics. To prove that TransSim-Node is efficient and effective, a medium-sized forwarder was modelled and the performance of the reference system as well as of a scenario with variation of the number of strip doors for outbound groupage was analyzed.

TransSim-Node offers various applications for simulating forwarding agencies: First, strategic decisions which are made within the scope of new construction or the enlargement of an terminal can be analyzed; the term layout planning also includes the determination of shape and size of the terminal or the number of docks. In addition, tactical tasks - the assignment of trailer to load doors, the allocation of trucks to strip doors, or decisions concerning size and spatial arrangement of internal areas (e.g. buffering) and of travel paths – can be modelled and improved. Furthermore, TransSim Node can be used for improving the operational sector. Especially questions concerning the prioritisation of loading and unloading actions, allocation of resources like technical equipment and personnel as well as the course of action during the transshipping processes can be solved. These aspects are strongly depending on the strategic decisions made before, but they are also influencing them on the other hand. So the need of an integrated solving tool is quite obvious.

In order to solve these problems and obtain valid information, the developed simulation tool TransSim-Node will be useful. For further study, it is planned to devise various layouts (rectangle, U-shape, L-shape), with fixed parameters for comparing and analyzing their key performance indicators based on several standardised system loads.

**REFERENCES**


Deymann, S. 2006. „Visionen zum Neubau von Logistikimmobilien“, VDI-FML Fachausschuss Transportlogistik – Workshop Stiftung Industrie- Forschung, Dortmund University of Technology


A DESIGN APPROACH FOR ASSET SUPPLY LOGISTICS

Sebastiaan A. Röben¹, Jaap A. Ottjes², Arnold van den Dool¹, Gabriël Lodewijks²

¹ Shell International Exploration and Production B.V.  Arnold.VandenDool@shell.com
² Delft University of Technology Faculty of Mechanical, Maritime and Materials Engineering
J.A.Ottjes@tudelft.nl

KEYWORDS
Asset supply logistics, transportation safety, oil & gas industry, process-interaction simulation.

ABSTRACT
In order to provide adequate logistics support for an oil & gas producing asset, onshore as well as offshore, a logistics support concept needs to be developed at an early stage in the opportunity realisation process. This logistic concept is eventually to define all personnel and material sources, discrete supply points including routings and hence infrastructure, and transportation equipment to address the requirements over the full lifetime of an asset. Typically, due to a vast amount of options for infrastructure (roads, ports, bases), the stochastic nature of impacting parameters, a distinct set of requirements for different phases in the lifetime - i.e. from development to abandonment - and the potentially changing business environment over time, finding an optimum integrated logistics solution is a difficult task without any system(atic) and/or software support.

This subject matter led TU Delft and Shell E&P (Exploration & Production) to take a joint approach to develop a software tool that is to determine the optimum logistic concept for new developments in a new (greenfield) or existing business environment (brownfield). Objectives defined included to develop a generic tool for generating a number of alternative, but potentially optimum concepts for any given business environment. Secondly, objectives included the tool to evaluate all generated concepts using simulation techniques, assess performance by defined Key Performance Indicators (KPIs) and advise robustness of the concept to external stochastic variables (e.g. weather, delays in material supplies and people movements).

This paper presents the tool development and how to interactively generate of a set of feasible concepts and assess concepts’ robustness by KPIs, which include operational cost, availability, operational safety and environmental impact.

INTRODUCTION
On a high level, the development of a new oil & gas producing asset follows three phases: design phase, construction phase and an operational phase. During the first phase, an in dept analysis of the opportunity results in a comprehensive strategy for production and a design for associated infrastructure and processing plants. During the second phase, infrastructure and plants are constructed while an organisation is put into place to support both the project and the operations phase. During the third phase, the installations will start up production. The construction and operational phases both require logistic support, but the type of support may differ considerably: the construction phase generally is a relatively short period with high and irregular transportation volumes, whereas the logistic demand for the (typically much longer) operational phase shows a more constant pattern with lower logistics needs.

The logistic concept, providing logistic support during the construction and operational phase, is developed in the design phase of the opportunity development. It defines personnel and material sources, discrete supply points including routings and transportation equipment to address the requirements over the full lifetime of an asset. It also defines infrastructure, i.e. the locations of marine supply bases, air support bases and other land based facilities such as warehouses/yards. Finally, the transport modes, types and sizes of transport equipment are determined and the routings and schedules for transportation are established.

Figure 1 shows an example of a brownfield development. Two offshore platforms are planned at predefined locations with a third one already producing. Existing and optional supply bases onshore and sources of material and personnel are indicated, as well as all possible transport connections. Obviously, choosing and opening any new base will have an effect in the cost evaluation whereby Capital expenditures (CAPEX) is discounted in time. Possible transport modes include road, air (rotary/ fixed wing), rail and sea.
A logistic concept is defined as a sub set of the possible sources and basis. The quality of any concept is assessed using four KPI’s being cost, robustness/availability, transportation safety risk exposure and CO₂ emissions from transportation activities. Different transport modes have different contributions to KPI’s. For example, helicopter transport and road transport incur higher transportation safety exposure than other modes. As the KPI’s are essential for the development of the tool, they are discussed in more detail in the next section.

KEY PERFORMANCE INDICATORS

Cost

Each concept has a number of factors that add to the total cost of executing logistics. First, and in most cases main contributor, is cost associated with transport resources. Costs may be split up in Operating Expenditures (OPEX), a yearly contracting fee and a variable cost component, the latter depending on actual usage. A second contributor may be investment costs (CAPEX) to build or to bring existing facilities up to standards and the operational cost for providing services from the facility.

Robustness/availability

Robustness indicates to what extend the supply of personnel and materials is resistant to disturbances in the logistic chain and actually reveals availability. Disturbances are for example bad weather conditions, equipment break down or irregularities in the external supply chains, consumption of materials or people movements. Robustness is expressed by a service rate. The service rate, at its turn, relates to the probability that an asset temporarily runs out of stock of one or more of its main supplies.

Safety risk exposure in transportation

Around 50% of all incidents and fatalities in the Oil & Gas industry are logistics related, (Association of Oil and Gas Producers, 2005). In order to combine the risk of all transport movements that arise from a certain logistic concept, a unified risk measurement model has been used. Shell has recently conducted a study to scientifically determine the transport risk exposure and build a mathematical model to compare different transport scenarios for any part of the world (ref. Koornstra 2007).

Subject study uses a mathematical model that relates transport safety data from a number of countries with well-maintained transport accident records to their Gross National Income (GNI) to calculate a ‘relative risk factor’ for any other level of GNI.

\[
Risk = R_m \times M_i \times M_j \times M_d
\]

With:
\[
R_m : \text{Modal reference risk}
\]
\[
M_i : \text{Risk Multiplication factor GNI}
\]
\[
M_j : \text{Country GNI correction factors (wealthy but unsafe, or poor but safe countries)}
\]
\[
M_d : \text{Additional, mode specific correction factors}
\]

CO₂ Emission

This KPI addresses one of the environmental aspects of an opportunity development and its operation. Currently, there are no global legal requirements for monitoring or control of CO₂ emissions or other (greenhouse) gasses such as SO₂, NOₓ. For European energy producing and industrial companies, some control is in place by the recently created CO₂ emission ‘rights’. The right to discharge CO₂ has become as a tradable ‘commodity’, with a real and changing price. This legislation however does not include any vehicle, vessel or aircraft emission. As CO₂ emissions will very likely become an important issue for any new future E&P development, the tool also includes a KPI on emissions of transport movements.

CO₂ emission can be calculated for all vehicle, vessel and aircraft movements. The calculation is based on the total travelled distance multiplied by the specific fuel consumption for the mode of the transport. This results in the total fuel consumption per transport mode and will then be converted to an estimated CO₂ emission in kilograms. Fuel consumption data for land transportation are widely available. If the fuel consumption rate of a vehicle is known, CO₂ emissions can be calculated with a fair degree of accuracy by applying a fuel/CO₂ conversion factor. For marine transportation, average emission factors can be derived from ‘Service Contract on Ship Emissions’, 2005, while for aviation, figures are used from the IPCC guidelines (2006).

DEVELOPMENT APPROACH

Cost, emission and safety risk can all be calculated for a static scenario. Robustness however, and hence
availability, is influenced by stochastic parameters over time. For this reason, a second pass of evaluating concepts is done using simulation.

The starting point of any design case include locations of assets, personnel and material sources, supply bases and data available with respect to expected consumption personnel movement patterns. Further data needed is related to cost, transport risk and CO₂, and statistics for patterns of weather, material consumption and people movement.

Once data is available or estimated (‘User Input’), steps to be made for an initial analysis are shown in Figure 2.

**Figure 2: Steps in the logistic concept design process**

**Case data**

The tool uses a common data base both for concept generation and for simulation. Figure 3 shows what type of data is required for the common data base; figure 4 zooms in on marine related data.

**Figure 3: General data needed for concept design**

**Generating feasible logistic concepts**

An interactive design approach has been selected for generating the initial set of potential concepts. The prime reasons for choosing an interactive approach are that 1) the professional logistics user can work with a set of relevant local data and logic other than having a tool generate non-feasible options and 2) the tool cannot be designed such that it will converge to feasible options without placing too many constraints. As such, the user support consists of an object oriented database and tools to easily define and evaluate concepts and subsequently simulate them to take stochastic influences into account.

To choose the type and number of resources needed for a certain supply scenario at this stage, the user must at least get a preliminary insight in the degree to which the assigned resources are suitable for the task. The main parameters needed are the resources’ capacity, the frequencies and quantities of the material orders and the travelling distances in the region. In practice these parameters are all interconnected: the total supply capacity of a resource over time is the result of the number of trips possible multiplied by loading capacity per trip. The number of trips is a function of the length of a (set of) routes, travelling speed and turnaround times. The eventual occupation rate is determined by the ordered quantities and frequencies of the calls, which in turn are dependant on the material requirement and the storage capacity of the assets. Because of the interconnection, without any constraints the matter becomes a circular reference and unsolvable.

Different techniques are known to get by with this issue. The closest resolution in operations research terms of a
‘Resource Routing Problem’ is the ‘Travelling Salesman Problem’ with vehicle capacity constraints (Charikar et al. 1998). This problem however has the limitation that it calculates the solution for a static problem and also does not include the assets’ capacity. Another technique to solve the issue is also known as the unbalanced transhipment problem (Wayne Winston, 2003).

Although both techniques would provide a part of the solution, it has been concluded that these types of mathematical approaches would, in this particular development, constrain the objective of developing a generic tool or would require unacceptable simplifications of the system.

Therefore, it is argued here that a simulation based approach will provide best functionality and flexibility to the user. A method has been developed as an extension of a ‘relay based’ approach based on simulation of traffic in a network (Veeke, H.P.M., Ottjes, J.A., 1999). To aid the user in the process of selection a set of feasible logistic concepts, for each initial concept, the algorithm can determine the optimum routes, while respecting all capacity constraints. The expert user must determine the call-off frequency of all Assets in the field, combined with the Assets’ stock capacity and demand over time, which results in the quantity needed per call. The algorithm uses the call quantities and the limited capacity of the resource to find the shortest route(s) to serve all Assets. A solution may comprise of a set of routes, with multiple intermediate visits to the source point to re-supply the resource. The user can now review and accept or decline the proposed routes, but also put in routes manually.

At this stage, with the definition of routes, the call-off frequencies and quantities, it is now possible to statically compare effectiveness of the assigned resources.

Figure 5 illustrates the step of generating logistic concepts, all having a sub set of all possible road, air and sea connections.

The generation of each concept consists of the following steps:

- **Infrastructure:**
  - Choose location of Marine bases
  - Choose location of Helicopter bases
  - This has to be done ‘manually’ by the expert user

- **Routes over land, sea and air:**
  - Find routes for Personnel
  - Find routes for Materials
  - Optimising routes is supported by the tool

- **Resources:**
  - Find Number & Type of Vessel needed
  - Find Number & Type of Helicopter needed
  - This is an iterative process using the results of the routing step

**Selecting the final concept using simulation**

To this stage, all calculations have been carried out for a static situation, so without any stochastic influences. One of the important features of the simulation routine of the tool however is the ability to model stochastic influences in order to assess the concept’s robustness. This simulation routine, which incorporates reorder policies, safety stocks and priority ordering, is used to find the final concept, determined by its KPIs.

To demonstrate working with the tool, figure 6 and 7 show the results of simulation runs of a (simple) logistic concept in which stock level fluctuations of an asset are plotted against time. Figure 6 shows output whereby the simulations were performed on irregularities in the consumption pattern of the platform. Initially, intolerable out of stock situations occur in the platform stock levels (‘initial scenario’), but by varying the set points for reorder stock levels of both asset and supply base, the situation is reached which satisfies the asset requirements (‘best case scenario’).
By performing experiments with several concepts, the design team gains ‘experience’ with this specific case. The model can now be used to simulate and improve concepts, and to explore the dynamics and sensitivity of the particular case. The final concept is selected on the base of a weighed combination of the KPI’s. The weighing factors are case specific.

CONCLUSIONS

In this paper a design approach and prototype tool are discussed that support logistics concepts generation, at an early stage, for different stages of an E&P development. The best concept can be selected from a number of initial concepts on the basis of four key performance indicators. The objective to automatically generate potentially optimum concepts has not been met yet as the initial concepts still have to be build by expert users. The tool however does support that task with a comprehensive case data base and functionality to optimise supply operations. The desired optimisation aspect of the objective needs further attention. The essence of the work up till now is that the initial static models of the concepts can be automatically converted into simulation models to determine the robustness of each concept as a function of internal and external stochastic influences. A simple example is given to explain the way of working with the tool.

REFERENCES

- Association of Oil and Gas Producers, Safety performance indicators - 2005 data.
- IPCC Guidelines for National Greenhouse Gas Inventories, 2006
USING CONTAINER CALL TIME INFORMATION FOR RESTACKING REDUCTION

Wing Sum Lee, Jaap A. Ottjes, Hans P.M. Veeke and Joan C. Rijsenbrij
Delft University of Technology
Faculty of Mechanical, Maritime and Materials Engineering
Transport and Logistic Technology Group,
Mekelweg 2, 2628 CD Delft The Netherlands.
E-mail: J.A.Ottjes@tudelft.nl

KEY WORDS
container, stacking, discrete event simulation, call time information

SUMMARY
Eight heuristic stacking methods for containers are tested using simulation in a realistic setting of a container stacking lane. As the main performance indicator the ‘performance loss factor’ is used, defined as the number of restacking moves divided by the number of retrieved containers. Special attention has been paid to the possibility of utilizing uncertain call time information of (a part of) the containers to be processed. Depending on the percentage of containers with call time information, the uncertainty of this information and the filling rate of the stack the reduction of the performance loss factor is determined. Especially methods that evaluate the ‘remaining stacking capacity’ of all ground slots will yield reductions of the performance loss factor up to 85%.

INTRODUCTION
In deep sea container port operations the main issue is to serve the large container carriers within the contract period. A typical turn around time of carriers is 24 hours. After arrival and berthing of the vessel along the quay, the import containers are unloaded and stacked in the terminal main stack. In general this stack is situated near the deep sea quay wall. The export containers, available in (another part of) the main stack are then loaded into the vessel. After that the vessel is ready to depart.

The container main stack serves as a buffer between deep sea ship and other transport modes. The main stack of a container terminal requires quite a large area because it has to accommodate both export loads and import loads of large container carriers. A container stack consists of a number of ground slots on which containers are stacked in piles with a certain maximum height. Depending of the type of stacking equipment used, a stack may be composed of a number of “modules” each with own stacking/retrieving equipment, for example a portal type stacking crane.

Each module has three dimensions: width, length and (maximum) height, see Figure 1.

Figure 1 stacking module as used in the simulation experiments with 6 x 40 = 240 ground slots

The stacked import containers have different final destinations and retrieval times. The retrieval time of a container is the time it has to be retrieved from the stack for further transportation. After having been called, for example if a truck has arrived to pick up a specific container, this container has to be retrieved from the stack. If the pile with the container is stacked randomly and is n containers high, the probability that the right one is on top is only 1/n. In the other cases one or more containers of the pile have to be removed and put somewhere else in order to have access to the desired container. We call this “restacking”. If during the stacking of an import batch, knowledge of the retrieval time of the containers is taken into account, the number of restacking operations can likely be reduced.

Kim (1997) provides a method to estimate the number of restacking actions to pick up an arbitrary container. Also the total amount of restacking actions to pick up all containers in a bay can be estimated by this method. In later research (Kim and Hong 2006), two restacking methods, branch-and-bound and a heuristic rule, are compared to estimate an expected number of additional restacks for a stack. Narasimhan and Palekar (2002) proved the problem is NP-Complete and also investigated both a branch-and-bound method and a
heuristic solution. These articles all consider a stacking area which is divided in bays. Exchange of containers is only done within one bay. Duinkerken, Evers et al. (2001) developed a simulation model of a quay transport system to determine optimal stacking heights, numbers of AGVs and other variables. A stacking method based on categories was developed and tested. Category stacking is already applied at various terminals operated by straddle carriers, rubber tired and rail mounted gantry cranes.

**Research Question**
The research question in this work is to determine, for a number of (re)stacking methods, the benefit of a relationship between extra call time information of individual import containers and possible reduction of restacking effort.

**MODELING APPROACH**
The scope of work has been restricted to one stacking module with typical dimensions for a deep sea terminal. The dimensions of the test module are set to the ones shown in Figure 1. Width: 6 containers length: 40 containers and a maximum stacking height of 4 containers. All containers are considered to have the same dimensions. The stack will be filled at time=0 up to a predefined filling rate. The arrival order of the containers is random. The containers are put into the stack according the stacking method applied. After its dwell time, the container is retrieved from the stack and the number of restacks is recorded. After the first fill, the filling rate of the stack is maintained during the simulation run by replenishing the retrieved containers. These new containers are also stacked according the stacking method applied.

All containers are collected within 10 days after arrival. As a rule of thumb, taken from experience of some Rotterdam based terminals, we assume that 60% of the total is collected the first 4 days and 40% the remaining 6 days. This gives a mean time in stack (dwell time) of 4 days. The dwell time pattern is shown in Figure 2.

![Dwell time pattern](image)

**Figure 2** simplified dwell time distribution (days)

Each container is assigned an “estimated retrieval time” (ERT) and a confidence range of b time units. b is a parameter in the experiments. The estimated retrieval time is used in the (re)stacking algorithms. The estimated retrieval time is sampled for each container using the dwell time distribution of Figure 3. It is used in some of the stacking algorithms. The real retrieval time (RRT) is used to generate the actual retrieve action. In the model the real retrieval time of each container is sampled from a uniform distribution at (ERT- b and ERT + b). We distinguish containers with extra call time information and containers without extra call time information. For containers without call time information the ERT is not used for (re)stacking purposes.

![Figure 3 Estimated Retrieval Time (ERT), Real Retrieval Time (RRT) and time window (-b,+b) around the ERT in which RRT falls.](image)

**STACKING METHODS**
Eight different stacking methods are tested. In some methods use is made of the expected retrieve time of containers. If a fraction, say x%, of the containers has an ERT and consequently 100-x% have not, the available slots are divided proportionally.

If a container has to be (re)stacked, a “cost” factor, determined according the actual (re)stacking method, is assigned to each ground slot. The slot with the lowest cost factor will be chosen as the destination of the container to be (re)stacked. If multiple slots have the lowest cost factor the nearest position with respect to the origin of the container to be handled is selected. In some stacking methods the time axis is split up into a number of discrete time intervals of length c hours. Each container is then assigned the serial number of the time interval corresponding with its ERT. Next all eight methods will be discussed.

**Method 1: Random**
The method “random” does not use the ERT. For (re)stacking the target slot is randomly sampled from all slots that are not full.

**Method 2: Leveling**
The method “leveling” does not use the ERT. It takes the pile height of each slot as the cost factor.

In Table 1 an example of the each of the remaining methods 3 through 8 is shown. A square represents a container. The number in the square indicates the container’s ERT. After the table each method is explained.
Table 1 example of stacking methods 3 through 8. 10 time windows are available.

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
<th>Example</th>
<th>Priorities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>5</td>
<td>7, 2, 4, 1, 8, 3</td>
</tr>
<tr>
<td>2</td>
<td>Container to be (re)stacked. 5 means ERT = 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Minimum Time Difference</td>
<td>Cost factor: 2, -1, -2, 10</td>
<td>Priorities: 1, 3, 4, 2</td>
</tr>
<tr>
<td>4</td>
<td>Remaining Stack Capacity (RSC)</td>
<td>Cost factor: 2, -2, -6, 40</td>
<td>Priorities: 1, 2, 3, 4</td>
</tr>
<tr>
<td>5</td>
<td>RSC Improved</td>
<td>Cost factor: 2, -2, -6, 40</td>
<td>Priorities: 1, 2, 3, 4</td>
</tr>
<tr>
<td>6</td>
<td>RSC Improved with continuous time</td>
<td>Cost factor: 2, -2, -6, 40</td>
<td>Priorities: 1, 2, 3, 4</td>
</tr>
<tr>
<td>7</td>
<td>RSC Pile Search</td>
<td>Cost factor: -12, -2, -6, 40</td>
<td>Priorities: 4, 2, 3, 1</td>
</tr>
<tr>
<td>8</td>
<td>Restack Chance</td>
<td>Cost factor: 5, 2, 2, 1</td>
<td>Priorities: 4, 2, 2, 1</td>
</tr>
</tbody>
</table>

RSC = (maximum possible height – height of the top container) * (n_{top container} – n_{container-to-be-stacked}).

If a slot is empty, the maximum possible time window will be used. In the example of Table 1 the maximum time window is 10 time units. After calculation of the RSC values, the slot with the smallest RSC >= 0 is preferred. If no slot is available, the slot with the least negative RSC value is selected. If there are still no choices available, a third loop is performed. This third loop will search through slots that are actually destined for containers without ERT information.

Method 5: RSC Improved

The only difference with method 4 is that now, after calculation of the RSC values, the slot with the smallest RSC >0 is preferred. An extra demand is that (n_{top container} – n_{container-to-be-stacked}) > 0. In Table 1 this leads to the same answer as method 4, however in case the container to be stacked would have ERT=7, the empty slot will get preference.

Method 6: RSC Improved with continuous time axis.

This method is the same as method 5, RSC Improved except the partitioning of the time axis. Instead of discrete time windows, it uses a continuous time axis for RSC calculations.

Method 7: RSC Pile Search

This method is derived from method 5, RSC Improved. Instead of only looking at the top container of a slot, the container in the slot that gives the lowest RSC is normative for the score of the corresponding slot. RSC values lower than zero are taken into account too. Looking at Table 1, the cost factor of the first slot referring to the bottom container becomes: RSC = (4-1)*(1-5) = -12 and for the next slot: RSC = (4-2)*(4-5) = -2.

Table 2 Example of cost factor calculation for method 8.

<table>
<thead>
<tr>
<th>Pile with ERT's</th>
<th>#restacks</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

Cost Factor for Table 1 | 5
Method 8: Expected Restacks
This method uses a measure for the expected number of restacks per slot after the container to be stacked would have been placed on top. For each container in a pile the number of restacks is determined assuming that this particular container is called according its ERT. The numbers are added up for all containers in the pile, forming the cost factor of that pile. An empty slot gets a restack number of 1 to promote the usage of available stacks that do not require restacks.
In Table 2, the score is calculated for the first slot of the example of Table 1: after placing the new container on top, retrieving each container individually in the pile will cause a total number of restacks of 5.

EXPERIMENTAL RESULTS
In Figure 4 a part of the simulation screen is shown. The model is build in ‘TOMAS’ and is configurable for stacking lanes of any dimension. The eight stacking methods can be selected for each run. It’s possible to make a distinction between restacking method and stacking method. Further all varied parameters can be adjusted. The container dwell time distribution can be selected from predefined distributions or entered manually.

The model has been applied on a stack module with typical dimensions, according Figure 1. The tests are performed always for all stacking methods. The parameters varied for the consecutive sets of runs are:
- Run time, expressed in numbers of removed containers. Default value: 100,000 containers
- Percentage of Container with Departing Information: CDI factor (in %). Default value: 100%
- The pursued average Occupation rate of the stack (in %). Default value: 90%
- The possible deviation of the real retrieving time (=estimated retrieving time ± b). Default: b=1 h.
- Time step in case the time axis is split up in discrete time intervals of length c (hours). Default: c=1 h.

Figure 5 shows the performance loss factor for different run lengths. It shows that all outcomes stabilize from about a run length corresponding with 10000 removed containers. All runs are performed with a run length corresponding with 100000 removed containers. The relative standard deviation of the Performance Loss Factor proved to be less than 5%. It is concluded that both methods 5 and 6 (RSC Improved and RSC Improved with continuous time axis) outperform the other methods. The number of restacks is reduced with a factor 15 compared to random stacking.

![Figure 4 screen shot of the model](image)

![Figure 5 Performance loss factor for different run lengths](image)

Figure 5 Performance loss factor for different run lengths CDI = 100%, Occupation rate = 90%, b = 1 hr, c = 1 h.

Figure 6 shows the results of a set of runs varying the percentage of containers with call time information. All methods, except of course those not using ERT information, perform better if more information is available. It can be concluded that, in case of 50% containers with information, the number of restacks can be halved.

296
In Figure 7 the influence of the average stack filling rate is analyzed. It can be expected that at low occupation and thus low piles, restacking will be reduced. Still it is salient that both improved RSC methods hold on giving a large reduction in restacking until 90% filling rate. At filling rates close to 100%, the Performance Loss Factor becomes 1.5 for the first two methods. This corresponds to the value obtained theoretically, assuming infinite number of slots and 100% occupation and a maximum height of 4 containers.

In the last set of runs the accuracy of the estimated call time is varied by varying the value of b. Increasing b will introduce larger deviations between the real call time and the estimated one. One would expect the performance to be worse with increasing b. That expectation is confirmed in Figure 8. Still it is concluded that all methods using ERT give a reduction in the Performance Loss Factor of at least 50 % in case of an uncertainty of RRT of ± 6 h. Method 5, RSC improved even shows a reduction of 70% in that case.

CONCLUSIONS

In this work the results are shown of simulation experiments with eight heuristic (re)stacking methods to be used in container terminal operation.

If use is made of incomplete and, to certain extend, unreliable information of retrieval times of containers in (re)stacking operations, the number of restacking activities can be reduced considerably. Compared to the methods not using information (method 1 and 2), stacking methods using extra information perform significantly better. Methods using “improved minimal remaining stacking capacity” as “cost” criterion perform best. Some quantitative conclusions are:

- The number of restacks can be reduced with 50% if for half of import containers the retrieval times are know with a possible deviation of ± 1 hour and the stack average filling rate is 90%.
- When for all containers the departing time information is known within a bandwidth of 2 hours, the number of restacks in the considered case can be reduced with a factor 15.
- If the uncertainty of the real retrieving time in relation to the estimated retrieval time is ± 6 hours, all methods using departing information give a reduction in the Performance Loss Factor of at least 50 %. Method 5, RSC improved even shows a reduction of 70% in that case.

In general the simulation results clearly indicate that pre-information of container retrieval times can be very helpful to diminish retrieval times and thus to improve service times.
FUTURE WORK
The model is only applied on a reference case of one stacking module with fixed dimensions. The model will be further employed by investigating the influence of other interesting parameters like call patterns, call batches for example for loading a train or barge. The model is ready for realistic simulation as a part of a container terminal model with a multi-module stack. In that case the cycle times of stacking cranes and assignment of containers to stacking lanes will be taken into account as well as developing an extension for serving land side modalities.

REFERENCES


LIFTING AND STACKING
LIFT SYSTEM OPTIMIZATION BY SIMULATION

Jaroslav Sklenar
Michal Touš
Pavel Popela
Department of Statistics and Operations Research
University of Malta
Msida MSD 2080, Malta
Web: http://staff.um.edu.mt/jskl1/
Faculty of Mechanical Engineering
Brno University of Technology
Technická 2, Brno, Czech Republic
E-mail: jaroslav.sklenar@um.edu.mt, tous@fme.utb.cz, popela@fme.vutb.cz

KEYWORDS

ABSTRACT
Simulation methods are a very powerful tool in operations research. This paper deals with simulation and optimization of a lift system with respect to the average time spent in the system. Due to the complex and random nature of such a system, it is impossible to find an analytical form of the objective function. That is why the simulation is used for its evaluation. After a brief description of the particular lift system, a simulation model of this system is introduced. Then the way of obtaining input data is discussed as well as verification and validation of the simulation model. Results of some experiments and of an optimization based on the local search method are shown and interpreted in the last part of the paper.

LIFT SYSTEM
We consider the lift system in the building A1 of the Faculty of Mechanical Engineering of the Brno University of Technology. The building has 19 floors (1 – 19). There are five electric self-service lifts divided into two groups:
- triplex 3 lifts, floors: 1 – 18
- duplex 2 lifts, floors: 1 – 19
Table 1 contains the basic technical data of the lifts.

<table>
<thead>
<tr>
<th>lift</th>
<th>max speed</th>
<th>acceleration</th>
<th>capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>triplex lifts</td>
<td>2.5</td>
<td>1</td>
<td>15 each</td>
</tr>
<tr>
<td>duplex lift 1</td>
<td>2.5</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>duplex lift 2</td>
<td>2.5</td>
<td>1</td>
<td>16</td>
</tr>
</tbody>
</table>

Table 1: Lifts parameters

It is important that the lifts in each group are dependent on each other because they “cooperate”. Let’s illustrate this feature by two examples:
1. All lifts of one group are idle when a customer arrives and places a request for service. Then just the nearest one of the lifts serves this request.
2. A lift is staying on a floor and it is serving requests (customers are getting in or out the lift). Another lift from the same group going via this floor does not stop on this floor as far as there is no inner request on getting off the lift.
Each group has its own control panel with two buttons - so called outer requests - for direction up and down in every floor except for the ground floor (only up) and the top floor (only down). In every lift there is a control panel with buttons for every single floor - so called inner requests.

Although lifts run according to deterministic rules, the entire system is stochastic due to presence of random variables such as times of customer arrivals.

The described system is a complex service system and it is impossible to use analytical methods in order to obtain average time spent in the system and other characteristics; therefore simulation is used. The heaviest load of the system is during lunch time so this period is simulated.

SIMULATION MODEL
The simulation model was written in Java with the use of the SSJ (Stochastic Simulation in Java) tool. SSJ is a Java library for stochastic simulation, developed under the direction of Pierre L’Ecuyer, in the Département d’Informatique et de Recherche Opérationnelle (DIRO), at the Université de Montréal, see [2]. It provides facilities for generating uniform and nonuniform random variates, computing different measures related to probability distributions, performing goodness-of-fit tests, applying quasi-Monte Carlo methods, collecting (elementary) statistics, and programming discrete-event simulations with both events and processes paradigm.
At first we created an abstract model of the real system.

Abstract model
The state of the system is changed when an event occurs, for example a customer arrival. This change takes no time and the state does not change between two adjacent events; obviously the system is discrete.

As mentioned above, stochastic variables occur in this system and thus the system is also stochastic.

Situations which either happen rarely or have no significant influence on the results were not included in the model. The list of simplifications is as follows:
- Customers cannot get into the lift which has stopped on the same floor as the served customer is on, but is going in the opposite direction.
• First customer in the queue in every floor presses the button on both control panels (triplex, duplex).
• Customers are not able to place requests for service wrongly. Thus a waiting customer cannot press button for service in opposite direction and a customer inside the cab box cannot press button for a wrong floor.
• Customers queue for service in sequence in which they have been generated. FIFO is the way of organizing the queue.
• Breakdown of a lift is not possible as well as other situations which cause slowdown or stop of the system operation.

**Customer generation**

Customers are generated in each floor, thus each floor has its own random number generator. Customers can arrive either individually or in groups. All customers in a group have the same destination floor (usually it is a group of students going together). There are two queues in every floor (except for the top and the ground floors), one for customers going upwards and one for those who are going downwards. When a customer is generated he/she either places a request for service if he/she is the first one in a queue or just queues and waits for service. Destination floor of a customer is generated simultaneously with the arrival.

**Input data**

**Customer data**

In addition to lifts data, inputs include customers’ data which deal with arrivals and destination floors. It is assumed that arrival interval is a random variable with the exponential distribution for both individuals and groups. The parameter $\lambda$ is the intensity of arrivals computed as the number of arrivals per unit time. The random variable representing choice of the destination floor has an empirical distribution function obtained from statistical data. The standard method of inverse transformation is used for its generation. Another property of groups is the number of group members. It is assumed that these random variables have a uniform distribution so maximal and minimal numbers of members have to be set.

**Lift data**

As far as lift data are concerned the user may set:

- number of floors
- number of lifts
- number of groups
- time of getting in/getting of a lift
- height of a floor
- assignment of lifts to groups
- time of opening/closing lift door
- the lowest and the highest floor served by each group of lifts (these are the result of optimization discussed later on)

• the strategy of loading customers and the mode of lift stopping according to customer’s requests under which the system is running.

Then mode and strategy have to be chosen. Next we describe them in more details.

**Modes**

- none - lifts stop for every request.
- mode 1 - lifts don’t stop for outer requests in case that the last stop was in an adjacent floor. The lifts stop for every inner request.
- mode 2 - lifts don’t stop for outer requests in case that the last stop was either in adjacent or in floor following the adjacent one. The lifts stop for every inner request.

The lifts cannot accelerate to their maximal speed in case that they stop in adjacent floors. That’s why the mode 1 is included. Mode 2 is included because the lift remains in maximal speed for longer time.

**Strategies**

- Full container load (FCL) - the principle of FCL is shown in the example in Figure 1:

![Figure 1: The FCL strategy](image)

Briefly, the lift loads customers only in the current direction of its running. If there is no request in the current direction the lift direction is changed to the opposite which becomes current.

• First - when a lift is empty (the last customer has got off) it locates the customer’s request which is waiting for the longest time and moves to floor where the request is placed.

**Discrete-event simulation**

The lift system is simulated by events which are objects of Java classes. There are four basic event classes:

- Arrival – is the event of customer arrival.
- InOff – is the event of getting in/getting off the lift.
- Shift – this event occurs when the customer places his/her request and at least one of the lifts is waiting for requests. The second condition is
that customer goes in opposite direction than in which the lift comes. In this case the lift goes to the floor where the request was placed without any stop.

- FirstFloor – the first condition is the same as for the Shift event. Unlike Shift, the customer goes in the same direction in which the lift comes in the second condition. If there is the request in the same direction as the lift goes in and the lift will go via this floor during its way to request which made it moving, then the lift will serve the new request at first and then it will serve the older one.
- Floor – this event represents the movement of an occupied lift from the current to the next floor.

In addition there are two classes which create customers (class Customer) and lifts (class Lift). Links between events in the model are shown in Figure 2.

![Arrival into the system](image)

Arrival into the system

![Departure from the system](image)

Departure from the system

Figure 2: Links among events in the model

Lifts (Lift class objects) are created at the beginning of the simulation. Then an arrival event occurs and the customer (Customer class object) is created. Customer either queues as described in the “Customer generation” paragraph or enters the lift. If the customer places his/her request when at least one of lifts is staying on another floor, then, according to conditions, one of the events FirstFloor and Shift is planned to occur. If the lift is staying on the same floor as the customer has arrived, the event InOff (InOff class object) occurs almost immediately. If there is no staying lift, the customer waits in the queue for the service. After the customer gets in the lift (InOff event), the InOff event occurs again to check whether there is another customer waiting for service. If not, the lift moves from the current to the next floor in the requested direction, thus the Floor event occurs. This event is planned to occur in every next floor in order to check whether the request is placed there. If it is and if it can be served with respect to the mode of operation, the InOff event occurs. When the lift is empty (all customers got off the lift) it checks if there is any request for service. If not, the lift waits for a new request.

**Verification and validation of the model**

Process visualization, which is a part of the program for simulation, is useful for validation. According to visualization the simulation model corresponds to the abstract model.

Several measurements of the time spent in the system (time in the lift plus waiting) were taken in both directions and with different starting and destination floors. These results were obtained:

- average time spent in the system = 99.08 s
- standard deviation = 48.58 s

Simulation results are:

- average time spent in the system = 94.29 s
- standard deviation = 43.56 s

Thus we believe that the model corresponds to reality.

**SIMULATION RESULTS**

The load of the system was obtained from statistical data:

![Table 2: Customers data](image)

<table>
<thead>
<tr>
<th>floor</th>
<th>intensity</th>
<th>size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.00420</td>
<td>2-5</td>
</tr>
<tr>
<td>2</td>
<td>0.00119</td>
<td>2-4</td>
</tr>
<tr>
<td>3</td>
<td>0.00167</td>
<td>2-6</td>
</tr>
<tr>
<td>4</td>
<td>0.00071</td>
<td>2-3</td>
</tr>
<tr>
<td>5</td>
<td>0.00429</td>
<td>2-4</td>
</tr>
<tr>
<td>6</td>
<td>0.00214</td>
<td>2-5</td>
</tr>
<tr>
<td>7</td>
<td>0.000143</td>
<td>0-5</td>
</tr>
<tr>
<td>8</td>
<td>0.00071</td>
<td>2-6</td>
</tr>
<tr>
<td>9</td>
<td>0.00095</td>
<td>3-6</td>
</tr>
<tr>
<td>10</td>
<td>0.00143</td>
<td>2-4</td>
</tr>
<tr>
<td>11</td>
<td>0.00143</td>
<td>2-4</td>
</tr>
<tr>
<td>12</td>
<td>0.00214</td>
<td>2-5</td>
</tr>
<tr>
<td>13</td>
<td>0.00452</td>
<td>2-7</td>
</tr>
<tr>
<td>14</td>
<td>0.00214</td>
<td>2-8</td>
</tr>
<tr>
<td>15</td>
<td>0.00119</td>
<td>2-5</td>
</tr>
<tr>
<td>16</td>
<td>0.00262</td>
<td>2-6</td>
</tr>
<tr>
<td>17</td>
<td>0.000357</td>
<td>2-8</td>
</tr>
<tr>
<td>18</td>
<td>0.00071</td>
<td>2-3</td>
</tr>
</tbody>
</table>

![Table 2: Customers data](image)

<table>
<thead>
<tr>
<th>floor</th>
<th>intensity</th>
<th>size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.00420</td>
<td>2-5</td>
</tr>
<tr>
<td>2</td>
<td>0.00119</td>
<td>2-4</td>
</tr>
<tr>
<td>3</td>
<td>0.00167</td>
<td>2-6</td>
</tr>
<tr>
<td>4</td>
<td>0.00071</td>
<td>2-3</td>
</tr>
<tr>
<td>5</td>
<td>0.00429</td>
<td>2-4</td>
</tr>
<tr>
<td>6</td>
<td>0.00214</td>
<td>2-5</td>
</tr>
<tr>
<td>7</td>
<td>0.000143</td>
<td>0-5</td>
</tr>
<tr>
<td>8</td>
<td>0.00071</td>
<td>2-6</td>
</tr>
<tr>
<td>9</td>
<td>0.00095</td>
<td>3-6</td>
</tr>
<tr>
<td>10</td>
<td>0.00143</td>
<td>2-4</td>
</tr>
<tr>
<td>11</td>
<td>0.00143</td>
<td>2-4</td>
</tr>
<tr>
<td>12</td>
<td>0.00214</td>
<td>2-5</td>
</tr>
<tr>
<td>13</td>
<td>0.00452</td>
<td>2-7</td>
</tr>
<tr>
<td>14</td>
<td>0.00214</td>
<td>2-8</td>
</tr>
<tr>
<td>15</td>
<td>0.00119</td>
<td>2-5</td>
</tr>
<tr>
<td>16</td>
<td>0.00262</td>
<td>2-6</td>
</tr>
<tr>
<td>17</td>
<td>0.000357</td>
<td>2-8</td>
</tr>
<tr>
<td>18</td>
<td>0.00071</td>
<td>2-3</td>
</tr>
</tbody>
</table>

Intensities of arrivals of individuals (number of arrivals per second) in every floor are in the left table and intensities of arrivals of groups are in the right table which contains size of groups as well. Choosing of destination can be characterized as follows: customers going from first floor choose almost uniformly all floors. Customers going from other floors choose mainly first floor.

**The division in groups**

This experiment deals with how the division of lifts in groups changes the system characteristics (average time, st. dev.). Each of them has none mode and ECL strategy:

- comb. 1: first group 3 lifts, second group 2 lifts
- comb. 2: first group 2 lifts, second group 3 lifts

The difference between these two combinations shown in Figure 3 is probably caused by stopping of one more lift in the floor 19 for comb. 2. Customers going from 18 to 19 and from 19 down (32000 customers together) wait 20 sec.
shorter, while the others (1290000 customers) wait only about 5 sec. longer. But if one compares the numbers of customers, it is obvious that shorter waiting time mentioned above cannot significantly influence the total waiting time. While 5 sec longer waiting time of the bigger group of customers makes the difference between combinations being compared.

![Figure 3: Division of lifts in groups](image)

**Number of lifts**
Results of this experiment are in Figure 4. There is only one group and all lifts have the same parameters (floors 1-19, max. speed 2.5, acceleration 1 and capacity 15).

![Figure 4: Number of lifts](image)

Zeros in the bar chart represent an unstable system that doesn’t cope with the load. We see that the time spent in the system declines approximately exponentially with growing number of lifts.

**Joined and disjoined lifts**
The purpose of this experiment was to compare two systems – one with lifts running under triplex and duplex model (joined) and one with lifts running absolutely independently (disjoined). The result in Figure 5 shows clearly the advantage of the joined mode of operation.

![Figure 5: Joined vs. disjoined lifts](image)

**OPTIMIZATION**
The aim was to find such a setting which results in the best (i.e. the shortest) average time spent in the system. It is impossible to find an analytical form of the objective function $f$. That is why the simulation was used for evaluation of $f$.

The variables are the minimal and the maximal floor served by each group. So we have two variables for the lift system in the A1 building: maximal floor served by the triplex and minimal floor served by the duplex. We note that the triplex serves the lowest floor and the duplex serves the highest floor. Generally there are $2n-2$ decision variables for $n$ groups. The optimization problem is:

$$\min f(x_1, \ldots, x_{2n-2}),$$

where $f$ is the average time spent in the system. As mentioned above in the motivation paragraph, the local search algorithm was applied for optimization.

**Local search algorithm (hill - climbing)**
The solution space of the problem is depicted in Figure 6.

![Figure 6: solution space and vertex neighborhood](image)

The neighborhood of a vertex is selected as shown in Fig. 6 (the arrows show the neighborhood). Thus for $n$ groups each vertex has at most $2n$ neighbors. The neighborhood is not very large even for bigger $n$ which allows use of the steepest descent method in which the entire neighborhood is searched and the neighbor with the best objective value is selected.

The well known problem of the local search algorithm is its possible stop in local optima. This problem can be partially solved by repeated run of the algorithm from another initial vertex. Multiple runs increase the
probability of reaching the global optimum. It is possible to set the number of runs in the program.

Exhaustive search
The exhaustive search is feasible only for systems with few groups and small number of floors. For example for the system in the A1 building, on which the simulation was built, it is better to find the optimum by the local search method because the building has 19 floors which is a rather high number. The user of the simulation program may select either the local or the exhaustive search method.

Optimization results
The optimization of the lift system in the A1 building was performed for all feasible combinations of strategies and modes (see Table 3) with results in Figure 7.

<table>
<thead>
<tr>
<th>Combination</th>
<th>Mode</th>
<th>Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>none</td>
<td>FCL</td>
</tr>
<tr>
<td>2</td>
<td>none</td>
<td>first</td>
</tr>
<tr>
<td>3</td>
<td>mode 1</td>
<td>FCL</td>
</tr>
<tr>
<td>4</td>
<td>mode 1</td>
<td>first</td>
</tr>
<tr>
<td>5</td>
<td>mode 2</td>
<td>FCL</td>
</tr>
<tr>
<td>6</td>
<td>mode 2</td>
<td>first</td>
</tr>
</tbody>
</table>

Table 3: Combinations of modes and strategies

Note that the optimal values are max = 18 and min = 1 for all combinations. Thus both triplex and duplex serve all floors which they can serve. This is probably caused by uniformly distributed load to all floors. For more variable results we tried load which is not uniformly distributed, as shown in Table 4, and described in the following. In comparison with statistically obtained load the main difference is in intensities of arrivals of groups, for floors 2 – 10 in particular. Customers going from first floor go mainly to floors 11 – 19. Customers going from floors 2 – 10 choose mostly floor 1 as their destination while those going from floors 11 – 19 choose only floor 1. The results of optimization of the system with this load are shown in Figure 8. Note that each of the combinations has different values of maximum and minimum floors.

<table>
<thead>
<tr>
<th>floor</th>
<th>intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.364</td>
</tr>
<tr>
<td>2</td>
<td>0.0001</td>
</tr>
<tr>
<td>3</td>
<td>0.001</td>
</tr>
<tr>
<td>4</td>
<td>0.01</td>
</tr>
<tr>
<td>5</td>
<td>0.01</td>
</tr>
<tr>
<td>6</td>
<td>0.0001</td>
</tr>
<tr>
<td>7</td>
<td>0.001</td>
</tr>
<tr>
<td>8</td>
<td>0.01</td>
</tr>
<tr>
<td>9</td>
<td>0.001</td>
</tr>
<tr>
<td>10</td>
<td>0.01</td>
</tr>
<tr>
<td>11</td>
<td>0.0086</td>
</tr>
<tr>
<td>12</td>
<td>0.03</td>
</tr>
<tr>
<td>13</td>
<td>0.0126</td>
</tr>
<tr>
<td>14</td>
<td>0.0136</td>
</tr>
<tr>
<td>15</td>
<td>0.03</td>
</tr>
<tr>
<td>16</td>
<td>0.0214</td>
</tr>
<tr>
<td>17</td>
<td>0.0152</td>
</tr>
<tr>
<td>18</td>
<td>0.03</td>
</tr>
<tr>
<td>19</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Table 4: Customers data for non-uniform load

![Figure 7: Optimization results (uniform load)](image)

![Figure 8: Optimization results (non-uniform load)](image)

SUMMARY
A lift system can be very complex and random. The only way of obtaining information about such a system is simulation. The simulation model presented in this paper was built in Java program language with use of SSJ. The user of this simulation model may set a lot of data and therefore it is possible to compare various settings; few of them are presented in this paper. It is also an example of application of simulation methods in operation research. Simulation is used for evaluation of the objective function (average time spent in the system) because analytical methods cannot be applied due to the complex and random nature. The objective function is optimized by local search (hill - climbing) or by exhaustive search (when preferable) algorithm. The optimization of two different loads was performed and results are shown at the end of the paper.

REFERENCES
DYNAMIC CONTROL OF ANIMATION SCHEMES
FOR THE EFFICIENT 3D-VISUALIZATION OF MATERIAL FLOW SIMULATIONS

Wilhelm Dangelmaier
Benjamin Eikel
Matthias Fischer
Christoph Laroque
Heinz Nixdorf Institute, University of Paderborn
Fuerstenallee 11
D-33102 Paderborn, Germany
E-Mail: whd@hni.upb.de, eikel@upb.de, maf@upb.de, laro@hni.upb.de

ABSTRACT
This paper describes a method for the animation of a large number of objects within a dynamic 3D visualization of a material flow simulation model. It uses key-frame based animation. The number of animated objects may grow constantly in complex simulation models, which might lead to an amount of animations that is too big to be computed in real-time. By the use of a dynamic adjustment, the presented algorithm prefers important animations. Less relevant animations are updated rarely, whereby the selection itself is taken by multiple indicators, e.g. the visible size of the animated object on the screen, in order to keep a good optical impression. Dependent on the computing power of the computer, the algorithm controls the animations in such a way, that the fluid visualization of a large number of objects is still possible. Though the algorithm is to be used within a material flow simulator, it is moreover implemented in a specific animation editor, which allows the design and control of animation schemes. It supports the use of grouping to allow the creation of hierarchical structures for complex animations in a fast and easy manner. The evaluation of the algorithm is proven by a test scene, consisting of tens of thousands animated objects.

MOTIVATION
“Change happens”, and therefore the flexible adaptation of all processes close to production is a continuous challenge for companies in the global competition. Products as well as their manufacturing processes are to be reviewed critically at regular intervals. One well established tool for the design and safeguarding of those processes is the material flow simulation. During the last years, trends go more and more to the digital factory [7], which allows among multiple other things the most realistic 3D visualization of the designed processes [1]. Although the capabilities of computers are constantly growing, out of this approach the need for a realistic 3D view on the designed processes demands new techniques not only for the rendering of complex static scenes, but also for the dynamic animation of the material flow processes. Especially, the dynamic animations consume a lot of the resources of the computer. For example the animation of a conveyer where each coil is moving will lead to a problem even for small distribution centers. In order to be able to efficiently animate the dynamic behavior of such a factory, new methods for animation design and control are needed.

The presented method aims at the calculation and dynamic control of animation in such a 3D environment for even more complex scenarios. All represented animations can be implemented via key-frames, where single animation states can be defined by the designer at specific points in time. These states describe the position and orientation of an object. All states for times between the key-frames can be calculated by the computer by interpolation. Moreover the presented method can regulate the amount of time used for animation calculations within a complex simulation scene based on the speed of the specific computer. If the computing power is large enough, all animations can be calculated and presented to the user. If not, the calculations for some animated objects are skipped in order to achieve a good responsiveness of the system. The selection of objects tries to skip objects which are not important to the user. It uses an approximation of the visible size of the animated object from the user’s point of view in the 3D environment. Animations of small objects or those that are far away from the user are skipped first. The implemented method itself is, of course, not consuming lots of resources, in order to allow an overall improvement by its usage.

The next chapter will give a short overview over the implemented functionalities of the animation editor, whereas the next chapters will describe in more detail the implemented approach for the animation control. Afterwards, the testing of the implemented approach based on a specific test scene is described. The paper closes with a summary of the results and gives a short outlook on additional research topics.

ANIMATION EDITOR
For the design of the animations, needed within the material flow simulation, a cross-platform animation editor was implemented. 3D geometry models can be loaded from the file system and be moved or rotated in a virtual 3D scene. By giving a defined time period for an animation, a specific state can be saved as a key-frame. After the definition of several key-frames for different objects, the animation editor is able to show a dynamic view on the constructed animation in real-time. The animations can be saved in a file for later use. Additionally, the user can group several objects or groups together and manipulate the resulting group. Hereby, several hierarchical animations can be created.

Figure 1 shows a screenshot of the animation editor, where the 3D model of a forklift was loaded. By creating key-frames, it is possible to animate the up- and down-raising of the fork first and then let the entire forklift drive through a factory, where the tires of the forklift can rotate continuously.
Within the animation editor, an algorithm for the automated generation of test scenes was included. The generation of measurement protocols for evaluation, which cover calculation times and resulting performance key indicators, is also possible.

![Animation editor](image1)

**Figure 1: Animation editor**

The pre-defined animation itself can take place on different levels. At first, an object itself can be animated, as described above. Moreover, an animation can be applied to an entire group, which again can be a part of another group of a higher level. To give an example, again the forklift should be taken into consideration (cp. figure 3). At first, the fork itself is to be animated up- and downwards (A). Additionally, the driver’s head is animated via a rotation (B). Several additional animations follow (C, D). All animations are independent from each other.

![Conveyor with packet and KUKA robotics](image2)

**Figure 2: Conveyor with packet and KUKA robotics**

Fork, head and the rest of the forklift are then grouped to the overall object. Afterwards, this group can be triggered by the material flow simulator and thereby send around in a factory. It is also possible to store multiple animation schemes to a specific group, so that during the execution of a simulation model, it is possible to select a specific animation out of a pre-defined set, and thereby again accent the realistic aspect within the visualization.

![Forklift](image3)

**Figure 3: Forklift**

**ALGORITHM FOR THE ANIMATION CONTROL**

This chapter presents the algorithm for the control of the animation calculations in large, virtual scenes as they are typical for the material flow simulation, e.g. in the automotive industry. The first clause describes in detail the general approach as well as the algorithms used, whereas the next clause focuses on the implementation by the use of a scene graph.

**Control Algorithm**

With an interpolation between the defined key-frames it is possible to calculate single steps of an animation in the execution of a simulation run. In a typical material flow model, there are multiple animated objects at the same time, so that a more complex calculation of all intended animations is needed. The intention of the presented algorithm is to control this amount of animations. It fixes time and amount, if and when an object is to be animated. A cycle below is seen as a time period for the calculation of the animations before the rendering of the picture itself. In figure 4 “Animation” describes a time period needed for the calculation of the accumulated animations. The top row shows a typical execution, where every animation is calculated before each rendering step (“Image.”).

![Animation calculations delay the render engine. Bottom row: Control of the animation time allows a higher frame rate.](image4)

**Figure 4: Top row: Long animation calculations delay the render engine. Bottom row: Control of the animation time allows a higher frame rate.**

To reduce this amount of calculation for the animations, the entire duration is to be limited as it is shown in the bottom row. In each of the cycles, the animation of single objects is therefore skipped. In order to keep the visual quality high for the viewer in front of the screen, those animations which are of less importance are to be found. Typically, these animations are very small or in a long distance from the user’s point of view. Visible size of an animation here means the size of the geometry which is animated, projected onto screen. Here, the algorithm uses a ratio - the radius of the bounding sphere of the object divided by the distance to the viewer - to approximate the visible size of an object. The animations are
then divided into different levels of detail. Each level of detail consists of animations with a similar visible size. For instance with three levels existing, there is one level with large, one with medium and one with small animations (cp. figure 5).

![Division of animations into different levels of detail.](image)

**Figure 5: Division of animations into different levels of detail.**

In order to prefer some animations, the first animation levels, covering the large animation schemes, are calculated more often than those of the smaller animation schemes. The method itself is described in algorithm 1, where the preference of lager animations is induced by the loops. By moving through the 3D visualization, the calculated visible size of single animation schemes may vary. Therefore, the levels of detail are adapted in each step by selecting them out of a loose octree [3], which is a special form of the spatial data structure octree [2]. At the beginning of the visualization, the loose octree is created dependent on the object geometry of the entire scene. All animated objects are inserted into the loose octree according to their visible size.

**Algorithm 1** Control of animation calculations

```plaintext
procedure AnimationControl
    repeat
        for allowed from numberLevels to 1 do
            for level from 1 to allowed do
                Search the animations for the current level level from an octree
                Calculate the animations that were found
                Render the scene
            end for
        end for
        Adjust the number of levels numberLevels
    end for
    until Program terminated
end procedure
```

The number of animations which are covered by a specific level of detail is the quotient of all animations and the number of levels. With a growing number of levels, the number of animations within one level is reduced. The control of the animation time can be adapted by the number of levels of detail thereby. Therefore, the time needed for calculation of the animations is measured and compared to a pre-defined threshold. If the calculation time exceeds the threshold, the number of levels is increased which decreases the number of animations in each level and thus the animation time.

The search for objects to be animated is sped up by not calculating the visible size of each animated object, but instead using the size of the nodes of the loose octree, which contain multiple objects. The result gets a little more inaccurate, but the calculation speed is enhanced.

**Algorithm 2** Search quantity animations from an octree

```plaintext
procedure OctreeSearch(octree, quantity) maxR — minimal ratio of the search for the previous level
    repeat
        minR — minimal ratio of the elements in the list rejected
        Remove the element with the minimal ratio from the list rejected
        TreeWalk(octree, minR, maxR)
    until |found| ≥ quantity ∨ octree traversed completely
end procedure
```

Algorithm 2 describes the search for animations for a level of detail. It returns the animations which have a visible size in the given range of minR and maxR, where minR is the maximum ratio of the lower level of detail and maxR is the minimum value of the higher level of detail. So every animation is contained in exact one level of detail.

**Algorithm 3** Tree walk through an octree for searching certain animations

```plaintext
procedure TreeWalk(octree, minR, maxR)
    ratio — radius of octree / distance from viewer
    if ratio ≥ minR then
        if ratio ≤ maxR then
            Insert animations in octree into list found
        end if
        for all Children C of octree do
            TreeWalk(C, minR, maxR)
        end for
    else
        Insert pair (octree, ratio) into list rejected
    end if
end procedure
```

The visible size itself is calculated in each step via algorithm 3. Here, the loose octree is actually traversed and for each octree node the visible size is calculated dependent on the user’s point of view in the virtual scene. The ratio is then compared to the given ratio range.

**Implementation of Hierarchical Animations**

The hierarchy of key-frame animations was implemented with OpenSceneGraph [5], [6], which is a concrete scene graph library. In such a scene graph there are nodes that apply geometry transformations to all child nodes. These transformations (translation, rotation) are represented by a transformation matrix. Such a transformation matrix is calculated for each animated object in every time step by interpolating between the key-frames. It is then used by the rendering engine to transform the objects for calculating an image (cp. figure 6). Before the rendering of the scene, the transformation matrices for the animations are updated by algorithm 1. In such a scene graph the arrangement of the animation nodes is possible, such that only single objects or whole groups are animated.

![Hierarchy of key-frame animations in a scene graph.](image)

**Figure 6: Hierarchy of key-frame animations in a scene graph.**
EVALUATION
During the implementation of the intended and presented approach a test scene was created, which covers typical animation schemes from a material flow simulation, such as packets moving over a conveyor, forklifts moving through the entire scene and robotics moving at a specific place in the factory.

Test Scene
The entire test scene was build from two different animation schemes, which were created by the animation editor. The first animation scheme consists of a circle of conveyors and a KUKA robotics [4]. In Figure 2 and 3, some of the animations to be calculated are shown by the arrows. Each roll of the conveyor is moving (A), also the packet is moving over the conveyor (B). The different parts of the robotics are also moving in such a way, that handling the packet can be animated (C).

The second animation scheme is the moving forklift, already described in the earlier chapter about the animation editor. Both animation schemes cover hierarchical animations, since the described parts move independent from each other and nevertheless, all are moving as a group (e.g. the forklift is driving through the factory).

Figure 7 shows a screenshot of the test scene. The animation schemes with the KUKA robotics and the conveyor are placed in 21 rows as well as columns, which leads to 441 animation groups just for the conveyor sets. Between the columns, 20 animated forklift animation schemes are moving, each starting at a different point in simulation time (compare blue arrows in figure 7). Each object is animated independently, as described before. Additionally, the camera, which replaces the user’s point of view, is moving through the scene.

![Figure 7: Test scene with more than 50,000 animated objects.](image)

Analysis
Figure 8 shows a diagram of six measurements, which are taken over 30 seconds moving the camera through the test scene introduced before. All blue measurements show the time for calculation of all animations without using the animation control algorithm. It can be seen that the animation calculations need about 120 ms in each cycle.

The three red curves show the calculation duration, including the time for searching inside the loose octree, needed by the presented approach. The threshold for the calculation time was set to 30 ms. As can be seen, the calculation time is in a range centered on that value. If the required calculation duration is lower, the number of levels of detail is reduced. If the calculation duration exceeds this amount, the number of levels of detail is increased. The animation time moves around the desired 30 ms. Test users could identify a difference in the 3D environment of the material flow simulator, if the presented approach was used.

![Figure 8: Diagram of measurements in the test scene](image)
CONCLUSION
In this paper an algorithm was presented which is able to control the calculation duration for animations in large material flow simulation scenes, in order to allow a most realistic 3D visualization of the dynamic behavior of a simulation model. The measurements done with the implementation show that the control system is working correct and allows the control of complex scenes with more than 50,000 animated objects on a standard PC.

The control algorithm itself is kept simple, in order to keep the amount of additional calculation resources needed for the control itself small. As the fundamental data structure a loose octree is used, which holds and supports the search of the animations. Important animations can be distinguished from less important animations, which are updated less often.

Subsequent some aspects, which are not covered by the presented solution shall be explained and possible upcoming research projects are presented. By talking about interpolation, just a simple interpolation between key-frames was implemented because the control mechanism was in the focus. A meaningful enlargement of the presented solution would be the use of more complex interpolation techniques, e.g. adherence of velocity and acceleration continuity. Moreover, techniques could be interesting, which allow the dynamic adjustment of the animation during the execution of the simulation run, e.g. during the validation and verification step of a simulation study.

In addition, an implementation of a combination of rendering and animation calculation could benefit both sides. Animations could be eliminated if they are not in the user’s field of view, so that traditional culling algorithms could be taken into account. Typically they have to be adapted, e.g. for the case that an animation moves an object from outside the viewing frustum to the inside.

REFERENCES


loads/cadfiles/ File: kr60l30_3.stl. Access date: 2008-02-25. Model was modified.


BIOGRAPHY

WILHELM DANGELEMAIER was director and head of the Department for Cooperative Planning and Control at the Fraunhofer-Institute for Manufacturing. In 1990 he became Professor for Facility Planning and Production Scheduling at the University of Stuttgart. In 1991, Prof. Dangelmaier has become Professor for Business computing at the HEINZ NIXDORF INSTITUTE; University of Paderborn, Germany. 1996, he founded the Fraunhofer-Anwendszentrum für Logistikorientierte Betriebswirtschaft (ALB).

BENJAMIN EIKEL studies computer science since end of 2004 at the University of Paderborn. At the end of 2007 he finished his Bachelor’s thesis about the control of animation calculations. As group leader of the project group “3D-Rendering and Modeling of simulation-controlled manufacturing systems” he is part of the development of a system for parallel rendering of complex factories in order to allow the 3D visualization of complex material flow models.

MATTHIAS FISCHER studied computer science at the University of Paderborn, Germany. Since 1995 he has been a research assistant at the Heinz Nixdorf Institute. In 2005, he received a Ph.D. for his work on distributed virtual environments. His research interests are computer graphics, real-time rendering algorithms, and distributed computing. His e-mail address is <mafi@upb.de> and his Web address is <http://www.upb.de/cs/mafi>.

CHRISTOPH LAROQUE studied business computing at the University of Paderborn, Germany. Since 2003 he has been a Ph.D. student at the graduate school of dynamic intelligent systems and, in 2007, received his Ph.D. for his work on multi-user simulation. Since then he is research assistant in the group of Prof. Dangelmaier, Business Computing, esp. CIM. He is mainly interested in material flow simulation models and the “digital factory”. His e-mail address is <laro@hni.upb.de> and his Web address is <http://www.hni.upb.de/cim>.
RFID SIMULATION
DECISION SUPPORT SYSTEM FOR RFID

Gabriël Lodewijks, Adriana M. López De La Cruz, Hans P.M. Veeke
Department of Transport Engineering and Logistics
Delft University of Technology

e-mail: G.Lodewijks@tudelft.nl, A.M.Lopez delaCruz@tudelft.nl, H.P.M.Veeke@tudelft.nl

Key words: RFID, Decision support, fuzzy logic.

ABSTRACT

Testing the applicability of RFID for a specific application requires knowledge on fundamental performance parameters. Having this knowledge an optimum advice to apply RFID depends on the relative importance of each of these parameters. This paper combines the data from basic performance tests, the company restrictions, and demands to provide advice on the applicability of RFID. We use fuzzy logic to standardize the decisions and provide an objective assessment.

I. INTRODUCTION

Radio frequency identification (RFID) is an automatic identification (Auto-id) technology that enables data capture and identification without contact or line of sight.

Companies usually run into unnecessary costs for the promotion of RFID as an ultimate solution, ensuring its applicability for any case scenario; without having complete understanding of the technology or a truthfully assessment.

We developed already a testing methodology that includes three stages mainly designed to evaluate the technology, the impact of tagged products in the technology, and the impact of the overall application environment (Lodewijks, Veeke, and López, 2006). The testing methodology was designed following these parameters to evaluate the performance of RFID. The data obtained from every experiment are quantified into performance parameters (read rate, read area, antenna tag orientation, tag to tag distance and impact of materials) to evaluate different RFID equipment, with the same measure.

This paper proposes a decision support system based on fuzzy sets. The motivation for using fuzzy logic lies on the nature of the problem. When a company wants to apply RFID there are multiple constraints that RFID should meet in order to be integrated in the company processes. The testing methodology examines these constraints and produces results in terms of performance parameters. The performance parameters are a measure of the behavior of the system, however, every application is different, and the parameter’s importance depends on the company demands. Fuzzy logic provides the possibility to deal with the subjectivity of different objectives and constraints in a rational way (Yager, 1977). This methodology deals with constraints and objectives in a weighted manner. The importance of each parameter with respect to each other is determined using the company objectives. By combining the system constraints with the weighted performance parameters, the decision support system is capable of providing an objective advice based on company requirements.

II. DECISION SUPPORT SYSTEM

In order to provide an optimum advice while observing a simultaneous set of constraints and goals we propose the use of fuzzy logic. Fuzzy theory has been used widely to help represent mathematically a class of decision problems called multiple objective decisions. This class of problems usually involves many vague and ambiguous goals and constraints (O’Hagan, 1993).

The concept of fuzzy logic was introduced by Zadeh (1965). Fuzzy sets assess different propositions using membership functions to represent numerically the degree to which they are true or false. Fuzzy sets are an extension of classical set theory. A membership function in classical notion, maps all elements to either 1 or 0, as an indicator of an element belonging or not belonging to a set (true or false). A fuzzy set $A$ in $X$ is defined as:

$$A = \{(x, \mu_A(x)) | x \in X \},$$

where $\mu_A(x)$ is the degree of membership of the element $x$ in the fuzzy set $A$. The value 0 denotes complete no-membership, the value 1 denotes complete membership, and values in between denote intermediate degrees of membership (Zahed, 1965).

A fuzzy system consists of a number of rules that transform different variables into a "fuzzy" result. In order to understand the meaning of such a result, the values need to be defuzzified. A defuzzification technique adds the values related to a membership function and transforms it into a unique result that indicates a course of action (Zahed, 1965).

A. Applicability of RFID

We have collected information about the fundamental behavior of RFID. At this point we can not produce a statement on whether or not RFID is suitable for the application, but we can decide on whether or not RFID is functioning satisfactorily enough to continue with the next phase of experiments.
At this stage of testing we want to decide:

1. to continue with the experiments.
2. to continue with the experiments, but pay attention to certain parameters and conditions i.e. place the tags in a certain place of the product, leave more distance between products, stack the boxes in a given direction, etc.
3. Do not continue with the experiments.

Using the performance parameters as an input:

1. Read rate
2. Read area
3. Antenna tag orientation
4. Tag to tag distance
5. Impact of materials

Under the company requirements, for example:

1. The inbound/outbound area where RFID is planned to be installed is utmost 4m²
2. The dimensions of the product vary, but the minimum width of the boxes is 6cm
3. The products are mobile phones and they are packed in carton boxes
4. There is no stacking technique, the cartons are randomly filled with products. Around 20 to 25 cartons per container
5. The main requirement is to achieve 100% readability under any condition
6. Read rate is not an issue at the moment as long as 100% of the products are identified

In this case, the membership functions of the fuzzy set are going to be the performance parameters mentioned, with 100% being 1. Using these membership functions, it is possible to evaluate the degree of contribution of each parameter in the company requirements. For example, if we use the utmost read area required and we look for the membership of this parameter on figure 1, we see that for an area of 4m² the membership is 1 in case of a dual dipole tag. This means that for the required read rate the readability of RFID is still 100%.

After getting the membership value for each one of the performance parameters, we need to combine the results into a single application indicator that is an aggregated measure of performance parameters at different levels of implementation. This application indicator is based on the impact of RFID in the business process. The application indicator can be used to determine whether or not to continue with the tests. The application indicator (AI) can be calculated as follows:

\[
AI = \frac{\sum_{i=1}^{n} w_i P_i}{\sum_{i=1}^{n} w_i}, \quad 0 \leq AI \leq 1,
\]

where \( w_i \) is the weight of the \( i^{th} \) performance parameter and \( P_i \) the \( i^{th} \) performance parameter.

The weights assigned to the performance parameters are based on Saaty’s priority principle (Saaty, 1977), and are summarized in table 1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>read area</th>
<th>ant-Tag</th>
<th>Tag-to-tag</th>
<th>Material</th>
<th>read rate</th>
<th>Row sum</th>
<th>normalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>read area</td>
<td>1</td>
<td>0.4</td>
<td>0.33</td>
<td>0.5</td>
<td>0.67</td>
<td>2.90</td>
<td>0.10</td>
</tr>
<tr>
<td>ant-tag</td>
<td>2.5</td>
<td>1</td>
<td>0.83</td>
<td>1.2</td>
<td>1.5</td>
<td>7.03</td>
<td>0.24</td>
</tr>
<tr>
<td>Tag-to-tag</td>
<td>3</td>
<td>1.2</td>
<td>1</td>
<td>1.5</td>
<td>2.5</td>
<td>9.19</td>
<td>0.31</td>
</tr>
<tr>
<td>Material</td>
<td>2</td>
<td>0.83</td>
<td>0.67</td>
<td>1</td>
<td>1.2</td>
<td>5.70</td>
<td>0.20</td>
</tr>
<tr>
<td>read rate</td>
<td>1.5</td>
<td>0.67</td>
<td>0.4</td>
<td>0.83</td>
<td>1</td>
<td>4.40</td>
<td>0.15</td>
</tr>
<tr>
<td>Col. sum</td>
<td>10</td>
<td>4.10</td>
<td>3.233</td>
<td>5.03</td>
<td>6.86</td>
<td>29.2</td>
<td></td>
</tr>
</tbody>
</table>

| Inverted  | 0.10      | 0.24    | 0.31       | 0.20     | 0.15      |

Table 1: Weight List

The weights are determined by an assessment of the relative influences that a certain performance parameter has in the application relative to each other. By clustering applications and analyzing the typical characteristics of each situation, the weights for a given application can be obtained. The advantage of this method is that the weights only need to be determined once per set of application. This is a typical warehouse situation.

In order to find the decision, the application indicator needs to be defuzzified. Analyzing several possible cases that can be presented and combining the results and the influence of each parameter in the overall performance of RFID, the application indicator can be defuzzified using the function shown in figure 2.

![Figure 1: Read Area performance parameter](image-url)
Figure 2: Application indicator. Fundamental test TNT logistics

This function is used to determine whether or not to continue with the application tests, and if it is necessary to do some tuning, i.e. antenna configuration or tag orientation, to improve the performance for the next phase. At the end of the study a similar figure is derived to give the final recommendation on the applicability of RFID.

Using this approach, the case study for the application of RFID in the business process of TNT logistics is going to be developed.

B. Case study: TNT Logistics

The TUDelft RFID Research and Realization Laboratory (TUD3R Lab) executed readability tests for TNT warehouse and distribution of KPN mobile phones.

TNT objectives were to achieve 100% readability at item and container level in the inbound/outbound area and packaging table. In order to study the feasibility of these objectives, the readability test was performed using the designed methodology. At the end of the study a recommendation was made to TNT logistics.

Using the performance parameters found in the fundamental tests, the application indicator of this phase is calculated. The TNT application characteristics and the corresponding performance parameter values are shown in table 2.

<table>
<thead>
<tr>
<th>PP</th>
<th>TNT Requirement</th>
<th>Membership</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read area</td>
<td>≈ 4 m²</td>
<td>1 (100%)</td>
</tr>
<tr>
<td>Antenna tag</td>
<td>Not specified</td>
<td>1 (100%)</td>
</tr>
<tr>
<td>orientation</td>
<td>(desired 100%</td>
<td></td>
</tr>
<tr>
<td>Tag to tag distance</td>
<td>6 cm</td>
<td>0.9 (90%)</td>
</tr>
<tr>
<td>Impact materials of</td>
<td>carton</td>
<td>1 (100%)</td>
</tr>
<tr>
<td>Read rate</td>
<td>Not specified</td>
<td>1.6 Tag/sec</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 (100%)</td>
</tr>
</tbody>
</table>

Table 2: Application values

Table 2 shows the performance parameters values using the specified 4 m² area. The read rate corresponds to the case of 200 tags read in 120 seconds.

Using the PP values from table 1 and the weights from table 2; the application indicator is calculated as follows:

\[
\frac{\sum w_i PP_i}{\sum w_i} = \frac{w_{RA} + w_{AT} + w_{TT} + w_{IM} + w_{RR}}{w_{RA} + w_{AT} + w_{TT} + w_{IM} + w_{RR}}
\]

\[
A_i = 0.10 \times 1 + 0.24 \times 1 + 0.31 \times 0.9
+ 0.20 \times 1 + 0.15 \times 1 = 0.97
\]

Defuzzifying the AI value, 0.97, we can determine that it is suitable to continue with the next set of experiments.

The recommendations given using this method are very useful when trying to compare RFID equipment performance, since it allows the vendor and client identify strengths and weaknesses regarding the technology and the application.

III. CONCLUSIONS

It is important to have a testing plan at the beginning of any readability tests. This facilitates the process of determining what the best RFID configuration for a specific application is.

We also recommend the use of a decision making tool to provide advice on whether or not RFID is applicable. We propose fuzzy logic as a decision making tool because it provides an optimum advice while observing company demands, and restrictions. Though, the presented system is application dependent, it gives a mean to provide straightforward advice without the bother of having to analyze vast amount of data without knowing what it means or what to do with it. Furthermore, in (López, Veeke and Lodewijks 2006), it was mentioned the need to integrate an information system and a decision support system to diagnose and predict early failures and constantly monitor processes and systems. We think that RFID is a good candidate to apply as a data capture system in such a infrastructure.

IV. REFERENCES


V. ACKNOWLEDGEMENTS

This research was sponsored by the Transumo research program.
PERFORMANCE MEASUREMENT OF RFID
Adriana M. López De La Cruz, Hans P.M. Veeke, Gabriel Lodewijks
Department of Transport Engineering and Logistics
Delft University of Technology
e-mail: A.M.LopezdelaCruz@tudelft.nl, H.P.M.Veeke@tudelft.nl, G.Lodewijks@tudelft.nl

Key words: RFID, performance measurement.

ABSTRACT
This paper presents the minimum required information to determine whether or not RFID is appropriate for a specific application. This paper focuses on gathering specific information about the performance of RFID in a fundamental, in a semi-operational and in a fully operational stage.

I. INTRODUCTION
Radio frequency identification (RFID) is an automatic identification (Auto-id) technology that enables data capture and identification without contact or line of sight. This promising technology has been documented as being suitable for plenty of applications. However, the lack of unbiased, reliable, and standard information has made it difficult for companies to determine whether or not RFID is suitable for their operation.

Companies usually run into unnecessary costs for the promotion of RFID as an ultimate solution, ensuring its applicability for any case scenario; without having complete understanding of the technology or a truthfully assessment.

Here, we discuss the design of a testing methodology to provide information on the performance of RFID in a specific application.

The reliable implementation of RFID should include three stages mainly designed to evaluate the technology, the impact of tagged products in the technology, and the impact of the overall application environment (Lodewijks, Veeke, and López, 2006). The testing methodology was designed following these parameters to evaluate the performance of RFID. The data obtained from every experiment is quantified into performance parameters to evaluate, with the same measure, different RFID equipment.

II. METHODOLOGY
The reliability of a RFID system depends on the quality of the logistic visibility of an object, which is partly determined by the readability of the tags attached to them. (Lodewijks, Veeke, and López, 2006). In order to determine the readability of RFID tags in a certain application, a series of tests was designed taking into account the characteristics of radio frequency communication systems (Want, 2006).

Choosing the best technology requires precise knowledge of the desired solution’s requirements (Thiesse et al., 2006). The tests designed investigate the reliability of the application conditions to determine the relevance of RFID in a given situation.

The readability of a RFID system is influenced by the radio frequency RF air interface and the application. The first two parts of the tests focus on evaluating the RF air interface performance by determining functioning conditions like the read field of the tags, the effect of tag orientation and read rate, among others. The second part of the test assesses the influence of the application’s environment in the readability of the tags. In this part, test conditions like the packaging of the products, impact of the operational equipment, total number of boxes that can be segregated and identified, accuracy of the system, interaction with the physical environment, and more is analyzed. Fout! Verwijzingbron niet gevonden. shows the testing methodology. At the end of each test, the data obtained is analyzed and quantified in performance parameters. With this structure, the proposed methodology is a way of identifying key parameters and finding their individual influence on readability performance.

A. Fundamental test.

The fundamental test aims to research the technology, the limitations and the performance in a controlled environment. This part of the readability test is done in a RFID laboratory, where precautions have been taken to restrict possible causes of interference.

This part of the assessment is made by testing the RFID equipment in different conditions. In (ISO/IEC, 2005) a summary of the testing conditions is available. This assessment is only done once for every type of equipment under test, after having characterized the RFID equipments, the results can be used to compare which one gives better performance for a given application.

B. Single product test.

The single product test helps to find the proper location of the tag in the product. For this test, the RF properties of the packaging are studied. Analyzing the packaging composition of the products, it is possible to choose a number of positions to test the readability of the tags.

Another possibility to study the tag placement is being developed by the auto-id lab of Cambridge University in the UK. The Center of Distribution Automation and Control has developed a tool that simulates the interaction between tags,
C. Box tagging test.

The box tagging test is the first part of the tests designed to determine the influence of the operational environment in the performance of RFID. For these tests the information about tag antenna orientation is very important since it determines whether or not it is necessary to implement a specific stacking technique and tagging position.

D. Pallet tagging test.

The pallet tagging test is the second part designed to test the operational environment. The performance parameters determine the functioning level of RFID when it is integrated to the business processes. During the implementation of these tests, the reduction in readability is measured by isolating different causes of noise and error, and determining their influence on the overall readability.

III. PERFORMANCE PARAMETERS

Performance parameters are a standard measure to evaluate a RF technology in a given application. The performance parameters in the case of RF technology are created in order to have a measured system to compare RF technology providers, and to quantitatively determine the feasibility of RFID for an application environment. The testing results are estimated into performance parameters to obtain standardized and objective information from the collected data.

As shown in Fout! Verwijzingsbron niet gevonden., every test is quantified in one or more performance parameters. The fundamental tests have five performance parameters expressed in terms of readability [%] vs. area [m²] or time [s]: read area, antenna tag orientation, tag to tag distance, impact of materials, and read rate. These performance parameters are used to determine RF characteristics. Since this is the main part of the readability test, only the performance parameters of the fundamental tests will be discussed in this paper.

In order to find the performance parameters mentioned, a series of tests were performed at the RFID research and realization laboratory at Delft University of Technology. The data collected from the tests was quantified using a rectangular window with constant length (maximum reading distance in the z-plane as defined in the ISO/IEC TR 18046) and variable width to calculate the readability percentage per window area. The window’s width is increased by steps of 25 cm in the x-plane (y-plane kept constant) until it reaches the maximum reading point in the x-plane, as defined in ISO/IEC TR 18046. The tests were performed using a fixed structure with a circular polarized patch antenna, and two different types of tags.
A. Read Area

The read area is one of the most important parameters since it can be used to determine the main characteristics of the RF air interface. Documenting how the read area increases or decreases according a number of variables, is a good indicator to evaluate the degree of robustness of the RF system. Furthermore, it is possible to compare the performance of the reader when it is exposed to different types of tags and vice-versa.

Figure 2 shows the read area of a dual dipole tag and a single dipole tag. A dual dipole tag has two dipole antennas and is less sensitive to the direction of the antenna’s radiating field. The single dipole tag only has one dipole antenna. This makes the tag sensitive to the direction of the antenna’s radiation field, and will only work if it is place parallel to one of the radiating fields of the circular polarized antenna.

![Figure 2: Read Area performance parameter](image)

B. Antenna Tag Orientation

This second parameter is used to determine optimum orientation between the tag and the antenna configuration. After having a first impression on the size of the reading area, this performance parameter allows optimizing the reading area by using the antenna tag orientation that gives better performance.

Figure 3 shows the performance parameter of the antenna tag orientation when a dual dipole tag is place parallel to the antenna (YX plane), and rotated clockwise.

![Figure 3: Antenna tag orientation YX plane. Rotation angle θ](image)

C. Tag to tag distance

The third performance parameter is tag to tag distance. With this parameter it is possible to identify the capacity of the RF technology to read multiple tags and to overcome tag to tag interference problems.

D. Impact of materials

The impact of materials determines the influence of packaging materials in the performance of RFID. Figure 5 shows an example of this parameter.

![Figure 5: Impact of materials in RFID readability](image)

E. Read Rate

The read rate determines how many tags can be identified and read in a period of time. This parameter helps the application to determine how many read points are necessary in order to have 100% readability of products. Figure 11 shows an example of a bulk load of mobile phones.

![Figure 6: Read rate of mobile phones](image)

IV. Conclusions

It is important to have a testing plan at the beginning of any readability tests. This facilitates the process of determining what the best RFID configuration for a specific application is.
RFID should be tested at four levels: fundamental, single product, box and pallet level. The parameters to be measured are: read area, antenna tag orientation, tag to tag distance, impact of materials and the read rate.

Measuring these parameters presents an objective view of the parameters that will influence the performance of the RFID application. These parameters however are only the start of deciding to implement RFID. Currently our research focuses on how to decide, because each application may give different weights to each of the parameters presented in this paper.

V. REFERENCES


VI. ACKNOWLEDGEMENTS

This research was sponsored by the Transumo research program
EMERGENCY ROOM SIMULATION PROTOTYPES FOR INCIDENT MANAGEMENT TRAINING

Guodong Shao
Charles McLean
Manufacturing Simulation and Modeling Group
National Institute of Standards and Technology (NIST)
100 Bureau Drive, MS 8260
Gaithersburg, Maryland 20899-8260, U.S.A.
E-mail: guodong.shao@nist.gov

KEYWORDS

ABSTRACT
First responders and incident management personnel need better training resources to prepare for future disasters. Live training exercises while valuable are often very expensive to organize and conduct. Training using modeling, simulation, and gaming technologies could help to prepare for a more diverse range of scenarios than live exercises, as well as support individual, team, or multi-organizational training needs at lower cost. Effective, technically sound, and commercially available standards-based solutions are needed. A concept demonstration has been developed at the National Institute of Standards and Technology (NIST) to facilitate the identification, development, and deployment of standards that enable the integrated use of modeling, simulation, and gaming technology for incident management training, mission planning, and operational support. One of the simulation tools selected in the concept demonstration was ProModel. It was used to model a hospital Emergency Room (ER) in Washington DC. The concept demonstration was also used to identify use case scenarios, integration requirements, interoperability issues, standards needs, and available solutions.

DISCLAIMER
No approval or endorsement of any commercial product by the National Institute of Standards and Technology is intended or implied.

INTRODUCTION
There is a growing need for preparedness for emergency response both for man-made and natural disaster events. Effective emergency response presents a number of challenges to the responsible agencies. One major challenge is the lack of opportunities to train the emergency responders and the decision makers in dealing with the emergencies. The responsible agencies have tried to meet the need through the organization of live exercises, but such events are often difficult to organize and expensive to conduct. Modeling, simulation, and visualization techniques can help address many of these challenges. The limitations of live exercises can be overcome through use of integrated gaming and simulation models that allow emergency response personnel across multiple levels in multiple agencies to be exposed to the same scenario. Use of integrated gaming and simulation over a distributed network can allow people to participate from different locations and thus provide some flexibility in scheduling the resources. A preliminary investigation (Jain and McLean 2003a) indicates that a number of modeling and simulation applications for analyzing various disaster events exist. These need to be brought together for studying the impact of disaster events as a whole.

In the following sections, the NIST approach of integrating gaming and simulation systems for training of decision makers and responders on the scenarios preparing them to work together as a team is described. A proposed integrated systems architecture is briefly reviewed (Jain and McLean 2005). A concept demonstration use case scenario is introduced. A simulation model of a hospital emergency room, issues, and standards needs are also discussed.

EMERGENCY RESPONSE TRAINING SYSTEM ARCHITECTURE
Simulation and gaming-based technologies can together provide a highly effective means for incident management training if integrated correctly using an appropriate architecture. Simulation involves defining the rules of operation, the probabilities of paths of action, time durations, and letting the events unfold. Gaming relies on a trainee’s actions to determine the course of events under defined rules and probabilities. In the context of incident management, simulations are suitable for training emergency managers and decision makers of involved agencies. Gaming is suitable for training first responders. As illustrated in Figure 1, the architecture proposed by Jain and McLean has two major subsystems – simulation and gaming. Simulation and gaming subsystems could have individual communication integration infrastructures. This is based on the fact that currently the gaming and simulation worlds favor different integration mechanisms. The two infrastructures will be linked through a data synchronization and transfer

The simulation subsystem will contain a number of modules within each of the groups. The individual modules will model an aspect of the incident or response. They will interact with other modules based on the scenario. All the interactions will go through the simulation communications integration infrastructure. The gaming communications integration infrastructure may be based on the Massively
Multi-player Online Gaming (MMOG) architecture. The MMOG architecture needs to be enhanced to meet the requirements of incident management training and the need to synchronize with the simulation subsystem modules.

The proposed architecture will allow the training environment to be highly configurable. Simulation and gaming components can be selected and integrated based on a defined scenario. A scenario involving a terrorist attack using a dirty bomb can be modeled using components of the proposed system. The simulation modules employed for such a scenario may include crowd, traffic, explosion, plume, weather, fire, law enforcement, health care, transportation, and communications. The gaming modules for the scenario may include victims, general public, terrorists, fire, police, emergency medical technicians (EMTs), hazardous material teams, hospitals, shelters and public transportation. A natural emergency event such as a hurricane would require a different set of modules. The available modules in the proposed architecture can thus be configured to train incident management personnel across a range of scenarios.

SIMULATION SUBSYSTEM

The architecture description in the previous section showed two major subsystems - simulation and gaming, each with a number of major groupings. In this paper, we focus on the modeling of simulation subsystem as shown in Figure 2.

The simulation subsystem includes simulators that model the major capabilities and phenomena involved in emergency response. The simulation subsystem will create the emergency incidents in the virtual world. The ability to represent the incident in the virtual world together with the associated major aspects and the creation of unanticipated interactions will provide a valuable training environment. The description of the simulators below provides a brief description of their capabilities. The Social Behavior Simulators will simulate collective behavior phenomenon created through actions of multiple individuals. These include modeling of crowd, traffic, epidemic and consumer behavior. Physical Phenomena Simulators will model the physical phenomena involved in the creation and growth of the emergency incident. These may include such physical phenomena as earthquakes, explosions, fires, chemical, biological, or radiological plumes. Environmental Simulators will model the environmental phenomena that may affect the growth or containment of the emergency incident, its impact on the population or on the efforts by responding agencies. Such environmental phenomena include weather, watershed, indoor climate, and ecology. Organizational Simulators will model the actions of the organizations involved in any aspect associated with the incident. The organizations modeled may include fire departments, law enforcement, health care, other government agencies, and even terrorist organizations. The Infrastructure System Simulators will model the behavior of the infrastructure systems following the occurrence of an emergency incident. They will model the propagation of the impact of damage throughout the infrastructure system based on the damage to one part due to the emergency incident. In this paper, a transportation model of a hospital emergency room will be discussed in more detail.

CONCEPT DEMONSTRATION USE CASE SCENARIOS

The architecture for simulation and gaming for incident management training allows for the creation of a federation of simulation and gaming modules to represent the defined incident management scenario. The goal is to establish an effective, technically sound, and commercially viable standards-based solution for the development and integration of a simulation-based incident management training system. It is also to facilitate the identification, development, and deployment of standards that enable the integrated use of modeling, simulation, and gaming technology for incident management training, mission planning, and operational support. A concept demonstration has been developed to illustrate the potential benefits of integrated modeling and simulation tools for emergency response and incident management training. It is also useful in helping to identify use case scenarios, integration requirements, interoperability issues, standards needs, and available solutions. It will serve to help government, industry, and academia develop, test, and deploy required data standards in key mission areas.

A scenario of a terrorist attack using a dirty bomb in Washington DC on Fourth of July was defined for the concept demonstration. The concept demonstration system elements include simulation modules, gaming modules, and data server and information flow. A serial of simulators and gaming engines were selected to model the following simulations, a database system was also developed to provide the simulation with required data.

Figure 1: Architecture concept for simulation and gaming emergency response training system
- Incident management strategy gaming
- Plume simulation
- Crowd simulation
- Emergency vehicles response simulation
- Incident area traffic simulation
- Triage gaming
- Metropolitan train system simulation
- Hospital emergency room simulation

In this paper, the hospital emergency room simulation is discussed in detail.

**HOSPITAL EMERGENCY ROOM SIMULATION MODEL AND INTEGRATION REQUIREMENTS**

The emergency department simulator is a discrete event simulation model of an emergency patients flow in a hospital. The purpose of this simulation is to provide a small but realistic model of resources and patients flow and congestion in the ER of the hospital in response to an emergency incident including the deployment of resources and actions for triage and treatment of injured, movement of casualties to hospitals, and treatment at the hospitals. The model demonstrates how the incident affects: dispatch of ambulance to transport of injured to the hospital, as well as the waiting time in different areas, and evaluates the resources needed according to different scenario. The simulation will allow hospital management teams to train by responding in real-time to crises that affect ER flow and evaluate the impact of their decisions. Although it is important to understand the types of data required and how they should be represented, precise validated models are not critical to the demonstration scenario.

The area to be modeled include the whole hospital although we will concentrate the emergency department, patients location and off shift site and alternate hospital. The primary entities are patients, medical records, and soiled linen, resources are medical staffs, and specialists, emergency vehicles, triage and exam rooms, test lab, and other beddings.

Patients are modeled as first in – first out (FIFO) queues. The model uses the PROMPT statement, allowing the user to make modifications to selected model parameters. The user can change the number of patient arrival quantity and the average number of trauma and average number of cardiac patients per day. There are trauma rooms, cardiac rooms and specialty treatment rooms. Ambulatory and ambulance entrances exist as patients arrival points. The arrival of a cardiac or trauma patient, who will use of more resources, will cause the backlog of regular patients.

**Model Inputs**

The inputs of the simulation model are listed as follows:

- Number, location and type of casualties
- Availability of staff at work and off (on-call)
- Availability of resources
- Time and resources required for attending to each casualty type
- Probabilities of death from different casualty types over time.
- Hospital location
- Layout of the hospital
- Process stations
- Station capacities
- Processing times
- Patient arrivals rate
- Hospital shifts
- Medical resources
- Symptom-treatment profiles

**Model Outputs**

The outputs of the simulation model may include the operation of the health care system over time such as:

- System utilization
- Utilizations of process stations and resources

---

Figure 2: Simulation subsystem and its modules
• Updates of the status of the patients and medical staff
• Number of people treated and released, admitted, dead, waiting for treatment over time
• State of the staff and facilities (to determine their capability to deal with another incident)
• Run Time Interactions
• Simulated clock time – from Execution Control Supervisor
• Number of EMTs and ambulances dispatched over time to Traffic Simulation
• Number of ambulances and casualties arrivals over time from Traffic Simulation

Model Logic

The model logic includes relevant policies and procedures for emergencies including calling in medical staff, using temporary accommodations for the injured, acquiring needed supplies and equipment. The simulation was developed using ProModel simulation software. Figure 3 shows the model overview.

![ER Model Overview](image)

There are two kinds of patients as arrival entities of the model: Ambulance and General. Ambulance patients are those patients who are in critical situation, such as trauma and cardiac patients. There are limited rooms and beds for ambulance patients, if all the rooms are occupied at the time; the patient has to be redirected to an alternate facility. After a patient is taken into the room, Tech and RN will treat the patient right away, create medical record, and take to the MD for review, MD will make the decision, and the patient will be moved to the Nurse unit when the necessary procedures are done. General patients are ambulatory patients who can walk in to the hospital and waiting for exam and treatment. They have to go through the triage process first, if all seats are taken, triage-waiting area is provided. After the triage, patients are sent to the main waiting area waiting for calls to the different exam rooms based on their categories. Exam rooms include general exam, orthopedic exam, OB/Gyn exam, pediatric exam, and critical exam rooms. If it is not critical, the patient can be discharged. If further test or X ray is needed, patients have to be in the queue for these procedures.

ISSUES AND STANDARDS NEEDS

The interoperability between incident management Modeling and Simulation applications is currently extremely limited and the cost of transferring data is often very high. The development of simulation models often requires significant expertise and special skills. The process is often very time-consuming. Much effort is also needed to verify and validate the simulation models that are developed. The analysis of the simulation output data is also a complicated process. Incident management organizations often lack the resources for simulation model development. Standard interfaces, good reference models, validated scenarios, or scenario data in standards formats, interoperable tools to allow rapid integration and reusable data sets could help to reduce the costs associated with simulation model construction and data exchange. This could make the technology more affordable and accessible to a wide range of potential incident management organizations.

The major challenge for the realization of the scenarios according to the architecture is the development of mechanisms for communications and time synchronization between and among simulation modules and game modules. Major issues associated with distributed multi-player games are how and when players receive information on fellow player’s actions. Time lags may occur between when a player initiates an action and when other players see the action. This latency causes problems in the execution of distributed games. The High Level Architecture (HLA) has been used for integrating distributed simulation models in the manufacturing domain. A neutral reference architecture was developed for integrating distributed manufacturing simulation systems with each other, with other manufacturing software applications, and with manufacturing data repositories (McLean, Leong and Riddick 2000). The need for standardization of interfaces was highlighted. The HLA Run Time Infrastructure (RTI) technology does not require the use of servers for centralized management of game data but uses time synchronization mechanisms that may be unacceptable in a game environment. In the HLA RTI world, simulators publish and subscribe to data objects to communicate. Simulations may be time regulating or time-constrained, i.e., control the advancement of time or have their time advancement controlled by other simulations progress.

An associated challenge is managing the training of people at different levels of the incident management hierarchy. The best mode for training the first responders using a game client is to execute in real-time (i.e., time progress in a game environment is the same as wall clock time). The best mode for training the incident management managers and other personnel operating in Emergency Operation Centers (EOCs) may be segments of real-time execution interspersed with accelerated time (i.e., time progresses in the simulation and gaming environment faster than wall clock time) and fast forwards (i.e., simulated time jumping to a few hours or a day later). This mode will allow the EOC team to train in decision making over few simulated days of an unfolding
emergency event while spending only a day in wall clock time. Combined training of first responders and EOC teams would require careful orchestration of time segments and fast-forwards.

Distributed simulation architectures have come a long way. HLA is currently a standard architecture for integrating distributed simulations. However, HLA has grown from the background of war games executed in real-time. While HLA has been used for accelerated time execution to some extent, it may be a challenge to achieve speedups of 10 times or more with the large number of modules envisaged for full scenario training.

Software licenses for game development systems and game distribution are often quite expensive. Pervasive use of this technology will require that many incident management organizations get access to licenses to develop training applications.

FUTURE WORK

Incident management must deal with many different scenarios with multiple, interacting objectives:

- Assist the victims
- Prevent further casualties
- Preserve law and order
- Mitigate further damage to infrastructure and private property
- Preserve forensic evidence, etc.

The solution is an integrated suite of software applications that enable simulation-based, training, mission planning, and operational support that is:

- Based upon commercially-developed and supported software
- Affordable to acquire and maintain
- Configurable to meet regional needs
- Easy to integrate with independently developed modules
- Based on standard data formats that enable access to local, state, and federal data sources
- Evolvable to support technological changes over time

Simulation-based training systems and exercises will need to integrate modules and data sources from different developers and user organizations. The proposed framework, system concept definition, and architecture are steps towards addressing the challenge. A coordinated effort is required to achieve the vision of an interoperable simulation-based training system; standards are needed to enable the integration. NIST has the expertise and experience to provide the technical support to accelerate the standards development process. More effect needed in the following aspects:

- Use simulation technology to provide technically correct models, behaviors, and data for various phenomena that affect training, mission planning, and operational support.
- Use gaming technology to provide an immersive, engaging graphical, audio, and haptic (force feedback) environment that offers high quality realistic experiences.
- Enable time synchronization, data sharing, checkpointing, time warp, rollback, replay, and logging functions between various simulation and gaming applications.
- Allow for centralized distribution and management of updates to software and data sets.
- Provide security features that prevent unauthorized access to, or modification of, computer systems, software, and data.

CONCLUSION

First responders and incident management personnel need better training resources to prepare for future disasters. Live training exercises while valuable are often very expensive to organize and conduct. Training using modeling, simulation, and gaming technologies could help to prepare for a more diverse range of scenarios than live exercises, as well as support individual, team, or multi-organizational training needs at lower cost. Effective, technically sound, and commercially available standards-based solutions are needed. A concept demonstration has been developed at NIST to facilitate the identification, development, and deployment of standards that enable the integrated use of modeling, simulation, and gaming technology for incident management training, mission planning, and operational support. ProModel was used to model the hospital emergency room for the concept demonstration. The concept demonstration of a dirty bomb explosion was used to identify use case scenarios, integration requirements, interoperability issues, standards needs, and available solutions.

REFERENCES


Modelling Causality in Medical Scenarios

Peter F. Summons
The University of Newcastle
University Drive, Callaghan, NSW 2308, Australia
Peter.Summons@newcastle.edu.au

KEYWORDS

ABSTRACT

This paper outlines formalisms for the representation of state and causality as mechanisms in an Object-Oriented approach to modelling medical knowledge. Some of the temporal denotations used by the MDS, an Object-Oriented based medical knowledge acquisition methodology and modelling architecture, to model causality are described.

1. INTRODUCTION

The MDS is both a knowledge acquisition and modelling methodology and an Object-Oriented (OO) architecture that is suitable for modelling medical domain knowledge (Summons, 2005). The architecture is based on object classes that reflect a medical scenario (Summons and Villanueva, 1996) and its application in a typical medical scenario was outlined in Summons and Colloc (2007). Modelling temporal information is an integral requirement in medical systems and must be catered for in any medical ontology. The MDS architecture provides causal mechanisms between instantiations of its classes. Objects in the MDS have formalised behaviour and support formal definitions of temporal aspects and relationships that are normally required for modelling time in medical knowledge–based systems (Summons and Colloc, 2006).

This paper describes some of the temporal axioms and denotations created for the MDS architecture. These denotations enable the MDS to model temporally associated information, and to model knowledge for Knowledge Based Systems (KBS) that may utilise temporal reasoning. The particular temporal concept of causality is examined.

Section 2 provides a brief overview of the object Classes used in the MDS architecture and Section 3 introduces Allen’s temporal Interval Algebra (Allen 1984) that has been used in the MDS. Section 4 covers general causality and its temporal considerations for medical systems, as well as Allen’s approach to causality. Section 5 introduces some basic proposals that the MDS adopts for its treatment of time in a system to reflect some of the natural terminology occurring in medical practice. Section 6 then outlines the MDS formalisation of causality before the paper concludes in Section 7.

2. MDS ARCHITECTURE OVERVIEW

The MDS architecture can be broadly interpreted as having three major meta-classes (Summons and Colloc 2007):

- the Temporal Object class, described to clinicians as being Patient Properties consisting of physical things in the system that exist through time (such as Patient, Right Lung, Circulatory System);
- the Action Object class, described to clinicians as being Staff Interventions consisting of things or actions that staff would either perform, or have others (ward nurses etc) perform in relation to the system under consideration (such as tests, measurements, observations or signs);
- the Association Object class, described to clinicians as being Diagnoses consisting of labels given to the possibility of a Disease or a Pathological Process being present in an instantiation of a Temporal Object (such as Hypertension or Pain).

3. ALLEN’S TEMPORAL INTERVAL ALGEBRA

Allen’s interval algebra is based on a primitive operator $meets(x, y)$, and describes the set of 13 mutually exclusive relationships that can hold between two intervals $x$ and $y$. Allen’s interval relations are denoted by:

$$\begin{align*}
AR-1: & \quad before(x, y), after(y, x) \\
& \quad meets(x, y), met-by(y, x) \\
& \quad overlaps(x, y), overlapped-by(y, x) \\
& \quad starts(x, y), started-by(y, x) \\
& \quad during(x, y), contains(y, x) \\
& \quad finishes(x, y), finished-by(y, x) \\
& \quad equals(x, y), equals(y, x)
\end{align*}$$

Allen’s logic supports three ontological categories that are associated with the time intervals over which they hold, or during which they occur. The first is a Property, the second is an Event and the third is a Process.

A Property, which can exist throughout, or persist over, an interval by means of the $holds$ predicate is defined by Allen’s axioms H.1 and H.2:

$$H.1 \quad holds(p, i) \iff \forall i'(\text{in}(i', i) \Rightarrow holds(p, i'))$$

Where $\text{in}(i', i) \equiv starts(i', i) \lor during(i', i) \lor finishes(i', i)$

$$H.2 \quad holds(p, i) \iff \forall i'(\text{in}(i', i) \Rightarrow \exists i''(\text{in}(i'', i') \land holds(p, i'')))$$
The holds predicate defines the continuity of a property \( p \) during a time interval \( i \), so \( \text{holds}(p, i) \) indicates that the property \( p \) holds over the interval \( i \), and that there is no sub-interval \( i' \) of \( i \) over which \( p \) does not hold. For example, if the property \( p \) was “headache” and the interval \( i \) was “the last two hours” (a temporal reference that is relative to the anchor point “now”), then \( \text{holds} (“ \text{headache}, “ the last two hours”) \) indicates that a headache was experienced continuously over the last two hours, that is, at every sub-interval during the last two hours.

The relationship between Allen’s holds predicate and the definition of the MDS exists predicate for objects was presented in Summons and Colloc (2006). The basic definition is presented as OP-2 below. In the MDS architecture, the knowledge that an instantiation of an object exists over a time interval \( i \) in a system is implied if it can be established that a property of the object is held over that interval. The object has to exist for the property to hold.

\[
\text{OP-2: } \forall \text{Obj}, \forall P, \forall i: \text{holds}(\text{P(Obj), i}) \\
\Leftrightarrow \text{exists}(\text{Obj, P, i}) \\
\text{The Object Obj exists with Property P over the interval i}
\]

4. CAUSALITY IN MEDICAL KBS

Causality has been applied in many medical KBS, notably in the causal links between diseases and symptoms in diagnostic systems (Mora et al, 1993).

The MDS knowledge model uses Production Rules involving instantiations of its meta-classes as antecedents and/or consequents to show causal associations between symptoms and diseases, in a similar manner to that of the medical diagnostic Knowledge-Based System QMR (Lemaire et al, 1999). Some of these Production Rules are acquired from medical domain experts using the knowledge acquisition processes of the MDS methodology and some of the rules are inherent in the causal relationships existing between class structures of the MDS architecture.

4.1 Temporal Considerations of Causality

Causal inference requires two factors: the ability to have an exclusive explanation so that a consequence is not attributable to many possibilities, and a consideration of time, so that the cause usually occurs before the effect (although sometimes the cause remains hidden and is discovered afterward). Causal systems therefore employ temporal knowledge, as considerations involving the time dimension occur in the reasoning processes operating on the system ontology.

The need to include time in causal explanation and inference implies that a temporal reasoning method must be employed in a medical decision support system (O’Connor et al, 2001; Weng et al, 2002). This involves an instrument capable of modelling temporal relationships, and also a mechanism that supports temporal inference. Time units and conversion functions must be defined so as to be able to automatically express the different clinical pictures of a disease in different scales of time in a consistent manner. Events need to be recognised and the associated interval duration of stages tracked. Deduction operators must also be implemented in order to express the ordering of activities such as previous to, or next to, according to domain time scales.

Keravou (1996) argued that general theories of time do not adequately cover common situations and requirements found in the medical domain. He holds that an adequate temporal ontology for “…medical diagnostic problem-solving… must support durations, temporal uncertainty and multiple temporal granularities” (ibid p236). Keravou identifies objects of importance in the medical domain, such as disorders or patient states, as being compound objects having a temporal existence component. He also suggests that causal temporal interactions between objects may involve modelling a time object as a central primitive of the medical ontology. This time object can have its existence defined in several time-axes, which may correspond to definitions over several domains. However, there may be additional definitions for the object established between these time-axes where links can be formed between the domains. The specific time-axis definition of the time object fixes its granularity within the particular time-axis domain.

The MDS architecture does not identify a separate time object as in Keravou’s approach. However it does integrate temporal interval characteristics within all of its objects. The MDS also adopts a minimal time interval that is dependent on the domain under consideration (Summons and Colloc, 2006). The minimum duration interval depends on the domain of interest and is formally defined in Summons (2005). Any property that an object possesses is assumed to persist for at least a length of time equal to a minimum duration interval as it is not possible in the domain for the MDS to recognise or measure a change in the property within this interval. In this way the MDS recognises Keravou’s concept of different time-axes that can be representative of several temporal dimensions within a domain, or of several sub-domains.

4.2 Events

One of the fundamental considerations of causal systems is the existence and recognition of events. Dojat and Sayetatt (1996) viewed events as being instantaneous objects, or objects whose duration was without interest, but which could modify the state of the world. This view of an event has some similarity to the MDS Action Object meta-class concept. MDS Action Objects also have some similarity to Fishwick’s (1995) definition of discrete event objects, in that they can represent either a function, with implicit sub-events, or time-point occurrences. The MDS views an event as an activity that is modelled, as far as its effect on the system is concerned, as being instantaneous. An event can be caused by factors external to the system, or by MDS production rules affecting the system internally.
Allen’s axiom O.1 defines the predicate _occur_ so that his logic can support the occurrence of _events_ (his second ontological category).

O.1 \[ \text{occur}(e, i) \land \text{in}(i', i) \Rightarrow \neg \text{occur}(e, i') \]

Allen defines events as being limited to the time interval over which they occur so that an event _e_ occurs wholly over the interval _i_ and there is no proper sub-interval of _i_ over which the event occurs. For example, _occur(e, i)_ indicates that an event _e_ occurs exactly over the time interval _i_, and could not occur over any sub-interval _i’_ of _i_. This, of course, also implies that the interval _i_ is the smallest interval over which the event occurs, and that there is no larger interval (of which _i_ is a proper sub-interval) over which the event occurs.

While the concept of an event is recognised in the MDS architecture, it is modelled differently to the explicit form used by Allen. The MDS architecture represents a specific form of event, which has the potential to cause change in the system, as an instantiation of the MDS Action Object meta-class. This allows an event of this type to be represented as an object instantiation that can be scheduled (created), that can have attributes such as measurements, and that can have properties such as a “result”.

Allen’s third category of _Process_ supported the concept of an extended event. His Axiom O.2 below relates _occurrences_ to _processes_ through the _occurring_ predicate. Occurrences were not strictly considered as events as they extend over the interval in which they are defined, but do not necessarily persist throughout the interval.

O.2 \[ \text{occurring}(p, i) \Rightarrow \exists i' \left( \text{in}(i', i) \land \text{occurring}(p, i') \right) \]

The example of a _process_ given by Allen (1984) is: “If I am walking during an interval, I must be walking during its first half. I could have stopped walking for a brief period within the interval, however, and still have been walking during it”. In their discussion of Allen’s logic, Ma and Knight (2001) point out that Allen thought this intuitive definition was too difficult to formalise and so he used the weaker axiom O.2 to indicate that, via the _occurring_ predicate, a _process_ _p_ can exist over an interval _i_ if there is some proper sub-interval _i’_ of _i_ over which the process is occurring. Allen had to introduce stronger axioms in an ad hoc fashion to cater for special classes of processes, such as his treatment of “an object falling” in O.3 below:

O.3 \[ \text{occurring}(\text{falling}(\text{object}), i) \Rightarrow \forall i' \left( \text{in}(i', i) \Rightarrow \text{occurring}(\text{falling}(\text{object}), i') \right) \]

Ma and Knight claim that these _processes_ may simply be modelled as one of Allen’s _properties_ using Allen’s own axiom H.1. In a similar manner to Ma and Knight, the MDS architecture views a process that can be modelled as a property as a _state_ of an object which the activity of _walking_ would be one of the possible state values, where, for example, a _person_ object could be in a state of _walking, sitting, standing, jumping, skipping, or running_.

There is a class of extended events, however, that the MDS treats as _processes_ that are _occurring_ over an extended duration as defined by Allen’s Axiom O.2. Some medical conditions and treatments associated with a patient are treated as _processes_ by the MDS. They have a duration over which they occur, but they are not confined to occur at all proper sub-intervals of that duration. For example, a person may experience pain for a day but have brief periods during that day in which no pain is felt, and a treatment may consist of many individual activities that are not contiguous.

### 4.4 Causality in Allen’s Temporal Interval Logic

Allen’s causal model is presented in this section. Allen proposed a predicate _ecause_ to indicate a causal relationship between two events _e_ and _e’,_ such that _ecause(e, t_1, _e_2, _t_3) was true if and only if the occurrence of _event_ _e_ at time _t_1 causes _event_ _e_2 to occur at time _t_3.

Allen defined _ecause_ by axioms O.4 and O.5 below. Allen’s axiom O.4 denotes the causal relationship between event _e_ and event _e’_ and O.5 provides the assertion that an event cannot cause other events prior to its occurrence (effects cannot precede their causes)

\[
O.4 \quad \text{occur}(e, t) \land \text{ecause}(e, t, e’, t’) \\
\Rightarrow \text{occur}(e’, t’)
\]

\[
O.5 \quad \text{ecause}(e, t, e’, t’) \\
\Rightarrow \text{in}(t’, t) \lor \text{before}(t, t’) \lor \text{meets}(t, t’) \\
\lor \text{overlaps}(t, t’) \lor \text{equals}(t, t’)
\]

### 5. MDS PROPOSALS

One of the problems of many decision systems involves tracking all the possible states that may arise due to change in the system. Tracking states of objects that experience change as well as those objects whose states do not change introduces an extreme degree of complexity called the Frame Problem. Allen’s solution to the Frame Problem was to employ _Closure Axioms_, which state that no properties change unless explicitly changed by an event, and no events occur except as the result of actions.

The approach used by the MDS to cope with the Frame Problem is to adopt the principle of _Minimum Change_ as a pragmatic assumption of the medical domain. The principle prescribes that the value of the property, or the state of an object, is maintained until it is required to be changed – in the case of the MDS this means that it is recognised as having changed. So, once observed to hold, an object state, or a property value, holds for an indefinitely open interval _i_, unless another interval _i’_ is found in which the object state or property value has changed and the relation _meets(i, i’)_ is true. The fact that a change in the object state or property value is established actually defines the start of the interval _i’_ and the finish of the interval _i_. The formalisation of the _Minimum Change Principle_ is given by MCP-1 below:

The MDS also proposes the _Normality Assumption_ as another pragmatic assumption of the medical domain. This states that in the absence of any initial knowledge, an
object’s state is considered to have a value of normal. This imposes default initial conditions for state and the Minimum Change assumption then maintains these conditions.

MCP-1: \(\forall \text{Obj}, \forall P, \forall S \in P, \forall i_1, \forall i_2: \exists \text{exists(Obj, P, i_1)} \wedge \neg \exists \text{exists(Obj, S, i_2)} \wedge (\text{meets(i_1, i_2)} \lor \text{after(i_2, i_1)}) \Rightarrow \exists i_3:\exists \text{exists(Obj, P, i_3)} \wedge \text{starts(i_3, i_1)} \land \text{finishes(i_2, i_3)}\)

In the MDS an event that is external to the system is not recognised as having occurred unless it is observed by, or has a recorded effect on, the system. External events are recognised in the MDS architecture by modelling the creation of an instance of the Action Object meta-class. Events can also occur within the system and are modelled as production rules, typically involving Action Objects, that are formalised to the structure of RR-1. These rules represent system behaviour that is initiated by object interactions or the internal behaviour knowledge of instantiated objects. System events in the MDS architecture correspond to major changes of the state of the system that occur as a result of an object being created (instantiated) or destroyed within the system.

6. CAUSALITY IN THE MDS

The MDS architecture is made up of objects with temporal characteristics and supports generic causal association possibilities between its meta-class instantiations. The relations presented in this section will define the specific causal interactions between the MDS meta-classes that are inherent in its architecture. They will also be used for causal inference in production rules and in the definition of constructs for the MDS architecture, such as object creation and destruction. Allen’s ecause predicate is not used as it relates specifically to events, which are not explicitly modelled in the MDS but are instead represented by mechanisms such as Action Object instantiations.

The MDS further refines permitted casual associations between particular sub-classes by applying the instantiations as arguments to class-specific variations of the related predicate, REL-1 below, to specify a necessary condition for causality between any instantiations of the sub-classes. REL-1 reflects pragmatic associations between object classes within the medical domain, such as those indicated by a clinical expert.

REL-1: \(\forall C_1, \forall C_2: \text{related}(C_1, C_2) = true \lor false\)

Production Rules are one of the mechanisms used in the MDS architecture to represent causality. These have the form: IF antecedents THEN consequent
The consequent can be an action or a logical statement. The antecedents may be a set of conditions, joined by logical conjunction operators that, if all are true will imply that the consequent is true. Putting it in other terms, if all of the antecedent conditions are satisfied, then the production rule will fire, meaning that the consequent action will be carried out (or will be held true).

In the general sense the rule related predicate, RR-1 below, denotes that a relationship exists between the antecedent ant(R) and the consequent cons(R) of a production rule R. It provides a formal definition of a Production Rule. It asserts that the antecedent implies the consequent, that is, if the antecedent is true then the consequent will be true.

RR-1: Production Rule Definition
\[
\forall X, \forall Y: \text{rule-related}(X, Y) \Rightarrow \exists R: (Y = \text{cons(R)}) \land (X = \text{ant(R)}) \land (\text{ant(R)} \Rightarrow \text{cons(R)})
\]

\[X \Rightarrow Y\]

In answer to Zahedi’s (1993) objection that rule-based systems do not imply causality, a stronger temporal definition for causality that involves temporal considerations of the consequent cons(R) and antecedent(s) ant(R) of a production rule R over an interval is proposed The causal-rule-related predicate, CRR-1 below, specifically denotes that a causal relationship exists between the antecedent ant(R) and the consequent cons(R) of a production rule R, and that effects (consequent) cannot precede their causes (antecedent).

The definition CRR-1 covers the case when the consequent, cons(R), of a production Rule R, is fully determined by, or fully causally related to, its antecedent ant(R). CRR-1 considers X only as a single antecedent, but the result is extensible to the case when X is a conjunction of a set of antecedents, that is when \(X = \prod_{p=1}^{n} \text{ant}_p(R)\).

CRR-1 causally-rule-related
\[
\forall X, \forall Y, \forall i_1, \forall i_2: \text{causally-rule-related}(X, i_1, Y, i_2) \Leftrightarrow \text{rule-related}(X, Y) \land (\text{holds}(X, i_1) \Rightarrow \text{holds}(Y, i_2)) \land (\text{equals}(i_1, i_2) \lor \text{meets}(i_1, i_2) \lor \text{starts}(i_1, i_2) \lor \text{before}(i_1, i_2) \lor \text{overlaps}(i_1, i_2))
\]

CRR-1 uses the normal form of inference, so its converse need not necessarily be true, that is \((Y = \text{cons(R)}) \land (X = \text{ant(R)}) \land \text{holds}(Y, i)\) does not necessarily infer \(\text{holds}(X, i)\), as there may be other causes for Y over interval i. CRR-1 merely states that X is capable of causing Y, in the sense that X cannot be true without having Y true over any time interval.

The consequent of CRR-1 can be true either during the time the antecedent is true only (the case when \text{equals}(i_1, i_2) \lor \text{starts}(i_1, i_2) \lor \text{before}(i_1, i_2) \lor \text{overlaps}(i_1, i_2) \lor \text{meets}(i_1, i_2) \lor \text{meets}(i_1, i_2) \lor \text{starts}(i_1, i_2) \lor \text{before}(i_1, i_2) \lor \text{overlaps}(i_1, i_2) \lor \text{meets}(i_1, i_2))\) in CRR-1, or it can become true after the antecedent has become true (catered for by the conditions \text{meets}(i_1, i_2) \lor \text{starts}(i_1, i_2) \lor \text{before}(i_1, i_2) \lor \text{overlaps}(i_1, i_2) \lor \text{meets}(i_1, i_2))\) in CRR-1. An example of the former is “If the traffic light is red then cars must stop” and an example of the later is “If the pedestrian crossing button is pushed the Walk signal will appear”.

Sometimes the antecedent may be translated to an equivalent form that will be strictly true when the consequent is to be held true. For example, a necessary
condition for a gun shooting is that a trigger has been pulled. It is not the act of pulling the trigger or the duration of the trigger being pulled that is considered as the antecedent. It is the fact that the event occurred which must be held true over the period in which the truth of the consequent is being examined, that is the fact that the trigger has been pulled must be true over the interval being considered. In some cases there are other antecedents that must also occur before a consequent is true, and the times at which they become true may vary.

A variation of the rule-related predicate can also accept objects as arguments. When applied to objects the interpretation of causally-object-related(Objj, P, i, Objj, S, i) states that if object Objj exists with state P in interval i, then this causes object Objj to exist with state S during interval i:

\[ \forall \text{Objj}, \forall i, \forall \text{Objj}, \forall P, \forall S, \forall i_2: \text{causally-object-related}(\text{Objj}, P, i, \text{Objj}, S, i_2) \]
\[ \iff ( \exists \text{Objj}, P, i) \Rightarrow (\exists \text{Objj}, S, i_2) \]
\[ \land ( \text{equals}(i, i_2) \lor \text{meets}(i, i_2) \lor \text{starts}(i, i_2) \lor \text{before}(i, i_2) \lor \text{overlaps}(i, i_2) ) \]

As we have seen, CRR-1 applies when there is a single rule antecedent, or when there are multiple antecedents (either of objects, conditions or properties) and where all antecedents in the collection must be satisfied for the consequent to exist if it is an object, or to hold if it is a condition or a property.

CRR-1 and COR-1 formalise the relationships that are pragmatically formed by the clinical expert, and that are used to create production rules by the knowledge engineer during the MDS knowledge acquisition processes.

There is a complementary relation contributes, which considers the case of an item in the antecedent set that is required for the consequent to hold, but is not sufficient by itself to cause the consequent to hold. The formal definition of contributes is given by CCR-1 below. It follows from CRR-1 for the case where the Rule R has a single consequent Y and a set of antecedents X, formed by a conjunction of individual antecedents, one of which is the antecedent Z:

\[ \forall X, \forall Y: \text{rule-related}(X, Y) \]
\[ \Rightarrow \exists Z \in X: \text{contributes}(Z, Y) \]

7. CONCLUSIONS

Causality was introduced as a key modelling concept for medical domains. This paper has indicated the need for causality and has presented different approaches to causal modelling. Formalisations used in the MDS architecture that allow modelling causal approaches were also presented.

8. REFERENCES


BIOGRAPHY

Peter Summons is a Senior Lecturer in the Faculty of Science and IT at The University of Newcastle, Australia. He is a consultant and was a Systems Engineer in industry. He holds the degrees of Graduate Diploma in Education, Bachelor of Computer Engineering, Master of Computing and a PhD in Information Systems. His main research interests are health informatics, knowledge based systems, decision support systems and systems design.
A STEP OF MODELING FOR THE DESIGN OF SIMULATION MODELS
IN THE HOSPITAL SYSTEMS:
APPLICATION TO THE NEW HOSPITAL OF ESTAING

Julie Chauvet¹, Michel Gourmand¹
¹LIMOS UMR CNRS 6158
University of Blaise Pascal
Campus of Cézeaux
63000 Clermont-Ferrand, France
{chauvet, gourmand}@lism.fr

Sophie Rodier¹,²
²University Hospital of
Clermont-Ferrand
Bd Léon Malfreyt
63058 Clermont-Ferrand, France
rozier@isima.fr

INTRODUCTION

The hospital systems must have a business logic, and think of new methods. The public hospitals have to consider a more rigorous control of their processes to limit their expenditure and to optimize their organization. So, they need decision-making aid tools adapted to their specificities. To design such tools, it is necessary to think about modelling approaches and methods. It is an essential step for the knowledge formalization of any system. The methods, tools and levels of detail used for modelling depend primarily on the studied system and the appointed objectives.

The work presented in this article takes place within the framework of the New Hospital of Estaing (NHE) of Clermont-Ferrand. The opening of this new hospital is planned for December 2009. The managers of the Hospital University (HU) wishes to have decision-making aid tools for all the temporal horizons: strategic, tactical and operational. The main aim of these tools is to allow to act (i) upstream on the sizing of the structure, on the staffing requirement, on the choice of management rules in terms of planning and resource allocation, of flows managements (human flow, material flow, information flow); (ii) during the system working to change and adapt the management rules and the resource allocation; (iii) downstream, to evaluate the system performances. To formalize the knowledge and to design decision-making aid tools adapted, we chose to follow ASDI methodology (Analysis, specification, Design, Implantation) (Gourmand and Kellert, 1991). The latter is based on the design (i) of a knowledge model (formalization of the system knowledge); (ii) of one or several action models which can be designed starting from models of performance evaluation (analytical model, simulation), of optimization methods (mathematical model, heuristic, meta heuristic...) or of a coupling simulation-optimization; (iii) of a results model with the indicators necessary to act on the system.

In this article, we presente in a first section, a brief state of the art on the studies related to the hospital systems and on the methods and tools used. We propose a methodology for the hospital systems modelling and we present the tools chosen for the modelling of the various services of the NHE in the next sections. Then, we give two applications of this approach.

STATE OF THE ART

The patient, who is the central element of any hospital system, induced material, human, financial flows and generates value. (Dallery, 2004) proposes various breakdowns of physical, informational, financial flows (Fontant et al., 2003), (Boumejkarr et al., 2004) for a hospital system (Moedbeek et al., 2004; Leclercq et al., 2003). Many studied problems deal with modeling, simulation and optimization of patient flows. Table 1 gives some examples of treated problems these last years on the hospital systems.

<table>
<thead>
<tr>
<th>Problems treated</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modeling, simulation and optimization</td>
<td>Hadges et al., 2003</td>
</tr>
<tr>
<td></td>
<td>Rohleder et al., 2007</td>
</tr>
<tr>
<td></td>
<td>Vissers, 1998</td>
</tr>
<tr>
<td></td>
<td>Faure et al., 2003</td>
</tr>
<tr>
<td></td>
<td>Moreno et al., 1996</td>
</tr>
<tr>
<td>Information system</td>
<td>Fabbe-Costes and Romeyer, 2003</td>
</tr>
<tr>
<td></td>
<td>Blanc, 2003</td>
</tr>
<tr>
<td>Measure satisfaction of patients</td>
<td>Hollingsworth et al., 1999</td>
</tr>
<tr>
<td>Beds planning</td>
<td>Akcali et al., 2006</td>
</tr>
</tbody>
</table>

Table 1: Problems treated for the hospital systems

The concepts to model are composed of methods, formalisms and various languages. They are presented according to order of their appearance in the literature by taking into account classification suggested in (Vernadat, 1999). In a non exhaustive way, we can quote:
- The cartesian approach which is based on the function analysis that must fill the system: CIMOSA (CIM Open System Structures) (Cimosa, 1993), GRAI (Graph of Interrelated Results and Activities) (Petit and Doumeignets, 2002), PERA (Purdue Enterprise Reference Architecture) (Williams, 1992), GERAM (Generalized Enterprise Reference Architecture and Methodology), (Geram, 1999), SADT (Structured Analysis and Design Technical) (Lissandre, 1990)...
- The systemic approach, developed in the Eighties, which is inspired as for it by the organization systemic theory, combined with a conceptual approach. Here modeling is approached according to two complementary points of view, data and treatments: Merise (Dionosi, 1993), IDEF1x, the language Express train (ISO, 1992).
- The object approach, developed in the Nineties, initially in the field of programming, then in the field of design and analysis of systems: OMT (Technical Object Modeling) (Rumbaugh et al., 1991), UML (Unified Language Modeling) (Kim et al., 2003)...

333

In a recent work (Hernandez-Matias et al., 2008) enumerates tools and methods for analysis and modeling of industrial systems by showing that if each one provides solutions to several problems in the manufacturing sector, none is sufficient to analyze and to model complex systems. We find these same problems with hospital systems.

We present, in the following part, the methodology and the tools used to model the various departments of the NHE in order to design decision-making aid tools.

**PROPOSAL FOR A METHODOLOGY**

**Proposal for a modeling process for hospital systems**

From ASDI methodological approach (Analysis, Specification, Design, Implementation) (Gourgand and Kellert, 1991), the modeling process of a system for the field (figure 1) recommends the consecutive construction of a knowledge model (description model for working system) and one or more action models (model data processing) which are obtained from knowledge model. The design and the use of the knowledge models and the action models to design, configure and pilote a complex system constitute the modeling process of this last one. This iterative process is composed of four stages which are successively connected in order to be able to evaluate and interpret results of study. The resulyis allow the managers to act on the system.

**Figure 1: Modeling process of a hospital system**

**Breakdown of the knowledge model in 5 levels**

The methodology used to design the knowledge model of a hospital system (figure 2) recommends a system breakdown in 5 descriptive levels: general structure, pool structure, relation services, patient pathways and actor functions:
- The general structure proposes an overview for the hospital system.
- The pool structure presents the organization of the hospital system according to a pool logic.

- The services relations give all the services composing pool and their relations.
- The patient pathways represent all the activities related to patients pathology using the event-driven process chains.
- The actor functions give all the activities linked to hospital actors activities thanks to functions trees.

The ASDI methodology recommends a systemic breakdown of system studied in three communicating subsystems: (i) the Physical Subsystem (PSS) (for production, storage, handling and transport), (ii) the Logical Subsystem (LSS) (for flows), (iii) the Decision-making Subsystem (DSS) (for management rules). The system description is carried out according to various smoothness levels.

**Figure 2: Organization of the 5 descriptive levels**

**Design methodology of the knowledge model**

After having collected the knowledge on the system using the group interviews and the semi directive qualitative individual interviews, the knowledge model, with its three sub-systems, is designed using class diagrams with UML. This knowledge model represents the system studied before designing simulation models.

In order to complete the knowledge model, the patient pathways and health care workers activities are described using event-driven process chains and functions trees proposed by ARIS. The patient pathways present the sequence of the various activities related to patient pathology during his hospital stay. A patient pathway activity is characterized by one or more types of health care workers, a type of place, a range of time. The attributes values of the activity are defined in the data module such as activities duration or quantity of hospital actors for a given activity. The workers functions are composed of the activities of the hospital workers carried out during 24 hours. For a type of hospital workers, all the one day activities are described. An workers function activity includes a type of place, a range of time. As for the patient pathways activities, the attribute values of the workers function activity are given in the data module like the number of that kind of place for a definite activity. These elaborate data starting from the class diagrams, the event-driven process chains and the function trees constitute the knowledge base carried out using ARIS which propose several reference models (Schéer, 2001). The various health care workers activities are modeled using the function trees. These function trees uses elementary functions which are functions that it is not necessary to break up front to meet the needs for the management of firm. Various pathologies representing the service activity and described by the medical workers are modeled using the event-driven process chains. The pathology activities
compose the patient pathways. The event-driven process chains allow to describe the various processes of one system using functions and events. The events begin the functions and they are the result of these last. The knowledge model is translated into simulation model. The data necessary to the simulation model are provided by the data module of the knowledge base. The results model arises in the decisional instrument panels form and in the histograms form. After an analysis of these results, the managers can make decisions and act on the system.

Knowledge collection for the ancillary medical workers
To collect knowledge, group interviews were organized so that the data collection is structured (Chauvet et al., 2005). The work groups taking part in the study are composed, for example, of Hospital Services Agent, of the Nurse's aides, of the nurses and of the executives. The work carried out at the time of the meetings concerns the operational people. They carry out a whole task defining their function. This amount of work represents the research foundation for work. It is analyzed starting from the structure existing and transposed on the future structure by holding account of changes. To note the various tasks carried out for each function per day, a time screen is made up. It takes into account activities every 15 minutes; it starts at 5 a.m. and lasts 24 hours.

To build the knowledge model of care units, we place ourselves in a microscopic approach on the worker/actor function level. Inside a service, workers with their competences act around a patient using material resources coming from the system. These workers, to act around a patient, carry out tasks. These gathered tasks characterize the station of the various workers. They are defined for one day of a worker for each service.

Knowledge collection for medical workers
After having collected knowledge from the ancillary medical actors in order to design the actor function models, the data model on the patient courses were collected near the medical workers. Before carrying out individual talks near the medical workers, a list of the pathology groups was drawn up so that the medical actors can describe the various pathology groups composing the activity of their service. The IMD (Information Medical Department) managing the medical information collected within the various services of a hospital, for each service, transmitted a pathology group list treated during the year. To sort this information so as to have the most faithful possible service activity representation, two criteria were used: the number of Summary of Medical Unit (SMU) and the Duration Average of Stay (DAS). The selection was carried out using a coefficient resulting from the multiplication of these two parameters. The fact of multiplying the number of SMU of a pathology group by its DAS, amounted defining, for example, the quantity of rooms occupied in the year for a precise type of pathology group (according to the principle that only one patient occupies only one room). In order to identify the major pathology groups, a mathematical law was used: the law of Pareto. The law of Pareto (Thiombiano, 1999), called law of the 80/20, is an empirical law inspired by the observations of Pareto, economist and Italian sociologist. Applied to our study, it shows that 20% of the pathology group account for 80% of the number of SMU to knowing that 20% of pathologies account for 80% of the patients. Thus, the result obtained is an image of the service activity. After having drawn up a pathology list treated in the service and individual discussions list with the medical actors were organized. They describe the linked activities with various pathologies of the service. This knowledge collection was carried out according to a knowledge collection grid in order to design the patient pathways.

Detailed specification of the pathways with LAESH
The knowledge collection allowed us to identify the significant quantity and the complexity of the patient pathways in some system based on a transaction approach (operating room suite, emergency). We define the transaction approach as the description, in the formalism chosen, of the system operations while specifying, for each type of entities flow, the movement of these entities and their successive treatments.

The complexity of these pathways is mainly related to the use of probabilities and/or management rules in the pathways and the "elementary operations" which compose them (Boolean expressions on the resources, preferences...). Taking into account this complexity, we have proposed the use of an intermediate language: the language of Analysis and Evaluation of the Hospital Systems (LAESH).

LAESH allows to design a knowledge model by using a transaction approach; to write the model obtained in the form of a file says "running file" by using a simple description language and to exploit the running file directly with the evaluation tools QNAP2. It makes the development with other simulation tools easier too.

One of the advantages of LAESH is its simplicity but also its flexibility which makes it possible to consider extensions to adapt it, as well as possible, to the system studied.

The knowledge model is built on two levels: the first level, known as overall, uses the stage and path concepts, and the second level, known as detailed, gives the contents of each stage. The overall graphic representation is a tree structure revealing the various paths which can take a patient and the stages constituting the paths. The root is the system input stage. The sheets of the tree structure are obligatory the system output stages. A node other than the root cannot be an input stage. A path is the sequence of stages connecting the input stage and an output stage.

Each stage of the overall graphic representation is the subject of a detailed chart. This one revealed the sequence of the elementary operations and the symbols of structuring: stage beginning and end; elementary service execution; waiting delay; passive resource seizing and release; loops; parallel processing; stage and sub stage call. This language was more largely presented in (Chabrol et al., 2008).

Table 2 compares ARIS and LAESH tools for the processes modelling.

<table>
<thead>
<tr>
<th>Approaches</th>
<th>ARIS</th>
<th>LAESH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modeled systems</td>
<td>Station approach</td>
<td>Station approach</td>
</tr>
<tr>
<td>Industrial system, hospital system</td>
<td>Transaction approach</td>
<td></td>
</tr>
<tr>
<td>Level of granularity</td>
<td>Macroscopic, mesoscopic, microscopic</td>
<td>Macroscopic, mesoscopic, microscopic</td>
</tr>
<tr>
<td>Choice models</td>
<td>Event-driven process chains</td>
<td>Overall chart</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td></td>
<td>Function trees</td>
<td>Detailed chart</td>
</tr>
<tr>
<td>Complexity models</td>
<td>Linear pathways</td>
<td>Complex pathways</td>
</tr>
<tr>
<td>Specificities</td>
<td>Process approach</td>
<td>Process approach</td>
</tr>
<tr>
<td>Support of communication</td>
<td>Complex</td>
<td>Simple</td>
</tr>
<tr>
<td>Level of training</td>
<td>Complex</td>
<td>Simple</td>
</tr>
</tbody>
</table>

Table 2: Comparison of ARIS and LAESH

APPLICATION OF OUR METHODOLOGY

In this section we present the application of our methodology to the modeling of a care unit and to that of an obstetrical unit. The characteristics of each one of these systems led us to use different numerical models to formalize the patient courses. In the first example we have simple and linear pathways, we thus use language ARIS. In the second example, the pathways are much more complex and are not linear. We thus use LAESH. These two applications were already presented in preceding work, we give only here their objectives, the data which are in entry of each one and the results obtained by the actions models based on simulation models.

The Generic Care Unit of (GCU)

We propose, by analogy with Supply Chains industrial (Fenious et al., 2004) to define the contemporary hospital as Supply Chain hospital. Each SCH is composed of whole care units (US) which, taking into account pathologies treated in each one, appear very different. Nevertheless, from the point of view of the data processing specialist and the modeling specialist, these US uses the same type of human resources, have the same objective (to generate value for the patient, to cure it, look after it...), and function according to very close rules. This is why our research tasks were directed towards project GCU (Generic Care Unit). We define the generic care unit like a whole conceptual and software bricks allowing modeling any care unit belonging to the SCH in order to bring a decision-making aid for the decision maker. This GCU is a virtual unit including/understanding all the elements common to each US, and instanciable on any US with the objective to generate little specific development. Following the presentation of the three subsystems of the generic model of the GCU using class diagrams, we proposed the knowledge generic model design using the numerical model ARIS. This generic knowledge model allows to design the knowledge models of the NHE. After having designed the generic knowledge model, we carry out a software component library.

To generate the data input of the simulation model, the models under ARIS are exported thanks to language Perl.

According to the objective of our study which is to evaluate the manpower impact looking after on the US one, the entry variables are thus: the quantity of the medical staff, the arrival of the medical staff, with the service represented by the rate of filling of the beds.

In this study, the results model contains: utilization ratio of various human resources (min, avg, max), utilization ratio of the various material resources (min, avg, max), time lag of the activities compared to their hour of end envisaged (min, avg, max), in the case of activities described as urgent: latency before the realization of the activity (min, avg, max), time lag of the activities during the shift (min, avg, max), the variation enters the interruption of an activity and its recovery (maximum, average).

This model was more largely presented in (Chauvet et al., 2008).

Obstetrical unit

The future obstetrical unit of the NHE will gather in a single place two obstetrics units now distinct. The goal of our work is to provide to the medical teams a decision-making aid tool for: (i) the sizing: to validate the size of the physical structure and to help them to specify the staffing requirements (functions, allocation by area); (ii) the management: to test and to compare management rules (means assignment), to study the response of the system to random events, to test different scenarios (plannings, load,…), to improve of service working; (iii) the performance evaluation: evaluation of indicators, latencies calculation, occupancy rates, identification of possible bottlenecks (full waiting room, …).

The complexity of the system is primarily structural and functional with numerous patient pathways and a complexity of these pathways (Boolean expressions on the resources, preferences). We used, for the system modelling, UML for the systemic breakdown (classes diagrams) and the actors functions (activity charts) and LAESH for the patient pathways. The decision-making aid tool, based on a discrete events simulation model was designed with the Witness software.

The input variables of the decision-making aid tool are: human resources with the various time slots of presence (planning) and the quantity of persons by planning/time slot; the quantity of patients expected by week; the patients distribution by “pathway”; the patients arrival terms by “pathway” (programmed, emergency arrivals); the several probabilities (complications, multiple births, etc.); the times of elementary operating (constants, variables).

The interface can generate a patient arrival schedule based on the data and parameters captured (the quantity of patients and the distribution’s arrival law). This schedule can be reviewed by the user before running the simulation. The simulation’s duration has been set to one week.

The results obtained are: the passive resources (rooms) and active resources occupancy (recorded every fifteen minutes); the time passed through the system for each patient and the quantity of births (natural childbirth and cesarean section); the occupancy rate and the quantity of taking into in charge for each room; the times and occupancy rate by place, elementary operation for each type of human resources and each human resources ;the processing times and the total latencies, by place or elementary operation for each patient...

This model was more largely presented in (Chabrol et al., 2007)

CONCLUSION

We showed the interest to set up a methodology for the modelling and the design of decision-making aid tools which are used by the UH of Clermont-Ferrand. Several simulation
models were designed starting from this methodology. A simulation tool, in progress, was designed for the generic unit of care in order to be instantiated on the various units of care. A simulation tool for the obstetrical unit is installed in the maternity of the UH, so that the medical teams can test different scenarios from organization (resource allocation, slot time for the programmed interventions...). This tool was presented to International Conference ORAHS in July 2007 and, at the request of the doctors and midwives, at the National Days of the French company of Prenatal Medicine (October 2007), at the Day of Research in Obstetrics and Gynecology (December 2007).

REFERENCES


Blanc J.X. 2003. « Le soin apporté aux soins : de la modélisation des processus de soins à la mise en place d’un système d’information de gestion et de management de la qualité. » In 2ème conférence francophone en Gestion et Ingénierie de Systèmes Hospitaliers (GISEH), Mons.


Flette-Costes N., Romeyer C. 2003. « La tracabilité des activités de soins par les systèmes d’information hospitaliers : un constat d’échec ou un phénomène émergent ». In 1ère conférence francophone en Gestion et Ingénierie de Systèmes Hospitaliers (GISEH), Lyon.

Faure S., Dr Vermuulun B., Dr Wieser P. 2003. « Modélisation et réévaluation des systèmes hospitaliers ». In 1ère conférence francophone en Gestion et Ingénierie de Systèmes Hospitaliers (GISEH), Lyon.


COMPLEX SYSTEMS SIMULATION
MEDICAL SIMULATIONS
Case-based reasoning: a perspective for pharmaceutical decision making

Claude Dussart\textsuperscript{1,2}
François-Xavier Magaud\textsuperscript{1,3}
Joel Colloc\textsuperscript{1,4}
Claude Petit\textsuperscript{1,3}

\textsuperscript{1} European Center for Humanitary Health University Claude Bernard - Lyon 1 - France
\textsuperscript{2} Military Hospital Desgenettes - Lyon - France
\textsuperscript{3} Laboratory LIRIS “modelling health decision” CNRS UMR 5205 University Claude Bernard - Lyon – France
\textsuperscript{4} Laboratory CRIS EA 647, University Claude Bernard Lyon 1 – Lyon - France

e-mail : claude.dussart@cesh.org

ABSTRACT

The role of “Case-Based Reasoning in health care” is to aid decision-making. Although the number of evaluations carried out is constantly growing, little is known about their real application in decision-making. Decision making is a crucial element in the field of pharmacy. The pharmacist has to determine what is wrong with the patient and recommend treatment, while the patient has to decide whether or not to seek medical care, and go along with the treatment recommended by the physician. Following a short presentation of some case-based reasoning principles, we present a new approach: the model puzzle.

KEY-WORDS

Case-based reasoning, multi-agent decision systems, decision making, hospital, pharmacy

1. INTRODUCTION

The market for health care systems supporting hospital pharmacists and improving their daily routine is steadily growing. The development of these systems requires handling both pharmaceutical knowledge and process knowledge.

Knowledge Management (KM) in healthcare can be regarded as the confluence of formal methodologies and techniques to facilitate the creation, identification, acquisition, development, preservation, dissemination and finally the utilisation of the various facets of a healthcare enterprise’s knowledge assets.

KM encompasses a variety of statistical analysis, pattern recognition and machine learning techniques. One promising knowledge management technique is case-based reasoning (CBR). In many domains CBR has become a successful technique for knowledge-based systems. In medical domains, attempts to apply the complete CBR cycle are rather exceptional. The pharmaceutical domain still provides new problems and challenges, where a CBR-approach can be profitably applied.

2. CASE BASED REASONING

The case-based reasoning (CBR) method solves new problems by remembering previous similar situations and reusing information and knowledge of those situations (Aamodt, 1994). Thus, CBR is an appropriate and effective method to solve problems when previous cases are sufficiently similar to a target case and also consistently support the same results.

Case-based reasoning provides the decision maker with an ability to utilize the specific knowledge of previously experienced, concrete problem situations, or specific patient cases (Kolodner, 1993). Cases may be kept as concrete experiences, or a set of similar cases may form a generalized case. According to Aamodt and Plaza (Aamodt, 1994) central tasks of CBR methods are to identify the current problem situation, find a past case similar to the new one, use that case to suggest a solution to the current problem, evaluate the proposed solution, and update the system by learning from this experience. A general CBR cycle may be described by four processes:

- retrieve the most similar case or cases;
- reuse the information and knowledge in that case;
- revise the proposed solution; and
- retain the experience for future problem solving.

The retrieval process involves the tasks of situation assessment, initial match and final selection. Instead of relying on general knowledge of a problem domain, or making associations along generalized relationships between problem descriptors and conclusions, CBR is able to utilize the specific knowledge of previously experienced, concrete problem situations (cases). In medical domains, CBR has mainly been applied to diagnostic and partly to therapeutic tasks (Schmidt, 2001).

The CBR Paradigm is based on the model of Aamodt (Aamodt, 1994) represented by figure 1. Research in the field of the CBR has been earlier developed by Kolodner (Kolodner, 1988), Richter (Richter et al., 1993) and Veloso (Veloso, 1995), and then by Watson (Watson, 1997). CBR is characterized by an adaptation step of an existing solution. In CBR the intervention of the user is necessary and the research came up against the difficulty in formalizing this step in a generic way. Many authors follow the works of
Carbonell (Carbonell, 1986) by using analogies. The process of adaptation was developed by Voss (Voss, 1997) and (Wilke et al., 1998). The organization of knowledge in a hierarchy (Kayser, 1997) opened new approaches (Lieber, 2002). We will examine the best approaches for re-use and adaptation in case-based reasoning systems.

From the first research work on CBR, many proposals focussed on the problem of adaptability (Hinrichs, 1989), (Turner, 1989). Janet Kolodner has proposed a classification of the types of adaptation (Kolodner, 1993). She distinguishes the differences of methods and the differences of search for adaptation rules.

In 1996, Voss [Voss, 1996] added case identification criterion on the characteristics of the case type, with an estimation of the adaptation. (Purvis, 1995) suggested exploiting the constraints to guide the adaptation. This approach attempts to ensure obtaining consistent solutions. This implies that all constraints are both complete and correct.

A team at Trinity College in Dublin has studied the adaptation for complex systems [Smyth and al 93]. They are proposing in particular to learn rules of adaptation from a base of cases (Hanney, 1997). Veloso (Veloso, 1997) describes a mechanism which adapts parts of one case with a part of other cases. An interesting contribution has presented by Hanks (Hanks, 1995) with an algorithm that is based on cases and independent of the domain.

Leake (Leake, 1997) imagined a process of adaptation with three types of learning: case learning by creation of responses plans; learning by indexation of cases according to their use; and case learning by searching for similarity about their adaptation cost. However, we note that the improvement of abilities of adaptation and re-use go through new organizations of case-based reasoning systems (Kayser, 1997) (Lieber, 2002).

CBR systems have been used in the medical domain for the purposes of, classification, tutoring, and planning (such as in therapy support). But, in the medical field, attempts to apply the complete CBR cycle are rather exceptional (Schmidt, 2001). The most challenging task for the CBR method is that of adaptation. In medical applications it is almost impossible to generate adaptation rules to consider all possible important differences between current and former similar cases. Therefore, some adaptation solutions have been developed that are rather typical for medical domains. One of these solutions is to focus only on the retrieval of similar cases and present them as information to the user. The motivation for abandoning the adaptation task is two-fold: in health-related application domains it is too complicated, or even impossible, to acquire sufficient adaptation knowledge; also the physicians tend to be interested in getting information about former similar cases, but prefer to reason about current situations themselves.

Case-based reasoning (CBR) methods are applied to various target problems on the supposition that previous cases are sufficiently similar to current target problems, and the results of previous similar cases support the same result consistently. However, these assumptions are not applicable for all target cases. Some target cases have no sufficiently similar cases, or if they have, the results of these previous cases are inconsistent. That is, the appropriateness of CBR is different for each target case, even though they are about problems in the same pharmaceutical domain.

3. MODEL P: CONCEPTS

For this reason Model P (figure 1) is proposed, showing a new organization of knowledge.

In many pharmaceutical applications, the basis of knowledge can be divided into several subsets, which are common in several cases. It is interesting to store these fragments not in several distinct cases, but in the form of pieces capable of being assembled together. Similar to a jigsaw puzzle, a case is reconstituted from its pieces. The idea is not to store cases any more but instead to store pieces, which will be assembled and will form a new case: a “puzzle”.

A puzzle is a progressive assemblage of pieces. (figure 2) For each problem, the user provides information to reconstitute a case (puzzle) which will bring a solution to the problem. Pieces, and their possibility of being brought together, will have been defined beforehand by an expert. The process, which leads to the formation of the puzzle-solution, is therefore directed by the choices of this expert. This work establishes implicit bonds between the pieces, which guarantees to the user that nothing is omitted in its search for solution.

Figure 1: Principle of CBR technique

Figure 2: Principle of Model P

Model P framework
**phase 1**: identification of a possible piece for a puzzle (experts: Delphi method...)

**phase 2**: Construction mechanism of the puzzle and resolution of the problem

---

**Figure 3**: process of puzzle construction mechanism

---

4. **BASIC ENTITY: THE PIECE**

A piece of a puzzle is a cellular object, composed of two information types:
- Descriptive information, which is more or less complex, on a precise subject. This can, for example, be materialized in the simplest cases by an image or a HTML page that corresponds to a piece of puzzle.
- Information of pairing, guiding the choice of the pieces which will be connected to it. This information allows an automatic aggregation of the pieces to form the puzzle. This corresponds to the contours of a piece of the puzzle. The pieces do not behave like passive objects. They have their own techniques for the selection of their successors and, as we will see later, they are able to adapt their search according to pieces that have already been selected. The pieces can thus be compared with intelligent agents, holding knowledge of their environment and being semi-autonomous (because the user is sometimes solicited).

5. **CONSTITUTION OF A PUZZLE**

The aim of the puzzle model is, starting from the pieces of the puzzle base, to arrive at a solution to the user’s problem. The algorithm used is recursive and builds a puzzle. Starting from the choice of the first piece, the relations of this piece with the not-yet-visited pieces are explored. The mechanism builds a tree with an in-depth exploration initially. The target piece represents the set of information provided by the user to initialise the search. It does not have descriptive information, and thus does not appear in the final puzzle. The initialisation of the search for the first piece (called the target piece) of the puzzle is made by the user. He/she also has the choice of the technique for pairing the successors, as well as filtering. According to the technique chosen, the user can be asked to give more precise criteria.

The search is done in a logical order: either from most general to the more detailed, or in an order imposed by the nature of the problem. The choice of the possible successors is based on two processes:
- Filtering: the choice is reduced at pieces having a common criterion.
- Filtering by fields: The pieces belong to a field, that is represented by a name. They also have a field of call, in which the search for the possible successors is carried out.
- Filtering by dates: the base evolves with time and new information; therefore certain pieces can become null and void, or less adapted to the evolving situation. Each piece has a creation date and the selection is restricted by pieces former/posterior to a selected date.
- Techniques of pairing: in a field, a piece’s successors are chosen either by asking the user for the resolution, or by seeking the closest piece, according to criteria defined by the expert:
  - Techniques by interrogation: selection on photo/description/HTML page, expert system, etc ...
  - Techniques of the nearest neighbour: indexation, multicriterion distance from outclassing, etc ...

Each piece indicates, according to the technique of pairing used, a set of successors which will seek another set of successors. This tree structure establishes relations which bring together the information represented by the puzzle. We note that it is impossible to select a piece that has already been selected as this would create a circularity.

The combination of these two pairing techniques makes the work of the expert easier: he/she can divide the general problem and classify the pieces in the field, according to the search technique used.

This mixed method also has interest: the user is guided/directed automatically to certain tracks by the expert, and the user’s initiative can be used to redirect the search aided by the interrogation techniques.

Termination of the algorithm: There are two possibilities:
- The stock of pieces is exhausted
- All of the started branches are finished or terminated, ie for each branch a termination, or “sheet”, has been reached. Indeed, certain pieces can be defined as termination sheets by the expert. The pieces-sheets are those which do not search for a successor.

Associative bonds of the pieces: It is interesting to bind pieces whose association is necessary to enable understanding of the problem. These bonds are not commutative. Two associated pieces will thus be called in different contexts, but the selection of a piece will automatically involve the selection of its bonded piece.
This method makes it possible to choose at each stage the most adapted information, and provides in the end a "made to measure" answer to the question. The adaptation to the problem was made progressively and in a generic way. Thus, this approach, which builds a search for a solution as a puzzle, is a great improvement on the stages of adaptation.

5. GENERIC COGNITIVE AGENT TYPE ARCHITECTURE

The Multi-Agent Decision Support System (MADSS) architecture allows knowledge bases interoperability. The generic cognitive agent presented figure 4, is composed of six modules and three knowledge and databases, CBR databases.

The communication module is responsible of the message exchanges. The negotiation module implements the negotiation protocol that allows the agent to commit or not to complete the tasks specified by the supervisor. The decision module builds the decision. The task scheduler module is in charge to control the execution when the agent accepts the task. The environment evaluation module supplies an external system representation to the agent.

[Diagram of MADSS General Cognitive Agent]

The communication module wraps and unwraps the messages, containing queries and answers, to and from the supervisor and other agents. The interpreter module decodes the semantics of the messages. The interface provides mailbox in/out functionalities. The implementation of the MADSS is available in (Colloc 2003).

6. CONCLUSION

This original mechanism proposed is based on a postulate: that the cases of a field are decomposable (entities, sub problems, processes, diagrams, ...) and that a component can be divided into one or more other cases. Only the components of a field are preserved. The case solution is built automatically as is done for a puzzle. Each part of the puzzle brings an element of the solution and associates it with another part corresponding in width and in depth. A part has an information part and an associated behaviour. The mechanism is recursive so the depth of the puzzle is not limited. Several models of reasoning were implemented: an engine with binary rules of production; indexing; and a multicriterion search of a case. In the same puzzle, several types of reasoning can cohabit.

The user does not seek a particular case among a multitude of cases but instead reconstitutes the nearest case with several possible reasoning methods. The updates relate to the stored parts, the subsets of the puzzle, and the parameter setting of the models of reasoning.

In November 2004, the Joint Commission of Pharmacy Practitioners developed a consensus position on the future vision of pharmacy and described how pharmacy practice will benefit society. Part of this document stated that pharmacists should “achieve public recognition; that they are essential to the provision of effective health care by ensuring that: (1) medication therapy management is readily available to all patients, (2) desired patient outcomes are more frequently achieved, (3) overuse, underuse, and misuse of medications are minimized, (4) medication-related public health goals are more effectively achieved, and (5) cost effectiveness of medication therapy is optimized”.

In this perspective, the rationale for integrating data mining and CBR methodologies is to discover knowledge from past data using data mining, and to retrieve and enable the use of this knowledge through CBR for the purposes of pharmaceutical decision support. The MADSS (multi-agent decision support systems) is able to share many different knowledge models (CBR, Rules, Objects...) and make them cooperate together, during decision processes, that match the usual job practices.

7. REFERENCES


BIOGRAPHY

Claude DUSSART holds a Ph. D. (2003) in medical decision support “meta-learning of distributed experiences” and a Pharm.D. (1998) by the pharmaceutical faculty of Lyon (France). He is a researcher, free teacher at the faculty of pharmacy (Lyon) and works at the military hospital Desgenettes.

François-Xavier MAGAUD (Ph.D.) is a researcher at University Claude Bernard – Lyon 1 in France. His research topic is about meta-learning. He works on meta-learning strategies.

Joel COLLOC holds a Ph.D. (1990) in computer sciences by the National Institute of Applied Sciences (INSA) of Lyon and a MD (1985) by the medical faculty of the Claude Bernard University (Lyon 1) France. Joël Colloc is currently professor in computer sciences in Le Havre University and was previously associate professor in the Jean Moulin Lyon 3 University. His main research topics are DSS and knowledge bases in medicine, CBR and multi-agent systems, nervous system modelling and cognitive sciences. He is member of the CRIS laboratory and of CESH University Lyon 1.

Claude PETIT: Habilitation for research director « approach for meta-model », Doctor of philosophy « Generativ modelisation », Doctor of philosophy « Global model for entreprise ». He is a research director in the European Center for Health Humanitary (CESH Lyon) and researcher in the laboratory LIRIS “modelling health decision” (Lyon-France).
PREGNANCY-BIRTH COURSE IN VENETO REGION:  
DEVELOPMENT OF A FLEXIBLE SIMULATION MODEL  
FOR DESCRIPTION AND FORECASTING

Alessia Barella, Paola Facchin  
Epidemiology and Community Medicine Unit  
Department of Paediatrics - University of Padova  
I 35129, Padova  
Italy  
E-mail: epi@pediatria.unipd.it

Giorgio Romanin-Jacur  
Department of Management and Engineering  
University of Padova  
I 36100, Vicenza  
Italy  
E-mail: romjac@dei.unipd.it

KEYWORDS
Model, simulation, pregnancy, birth, newborn.

ABSTRACT

In the paper a simulation model of pregnancy and birth in Italian Veneto Region has been studied and built up to forecast future requirement of assistance. Main factors conditioning fertility and raise of pathological pregnancy have been statistically determined, in order to represent the dynamical pregnancy and birth evolution of the whole women population. The resulting model, implemented in MicroSaint, is able to foresee the amount of physiological and pathological pregnancies, of normal and pathological newborns and of medical inspections and hospital admissions; that may be done both in current situation and in modified ones, consequent to increase of age of Italian pregnant women and to increase of foreign population.

INTRODUCTION

Pregnancy and birth are of high interest both for the various elements that affect them and for the amount of health care they require from the conception to the after birth phase. During recent years, many different factors have conditioned pregnancy and birth in the Italian Veneto Region, therefore it is interesting to describe in detail all interactions which influence pregnancy and birth course and related consequences for what concerns assistance requests. Even if many authors consider single phoenomena, from our knowledge no complete simulation model dynamically describing the total course has been built up. In the paper such a model is studied, built up and implemented; applications on actual situation of Veneto Region, with the objective of describing current conditions and of forecasting future ones as consequences of currently revealed trends are reported. In the first section main phoenomena ruling pregnancy and birth course in Veneto Region are analyzed; in the second section data sources and data elaboration methods are examined; in the third section the model is discussed in detail; finally model applications are reported.

MAIN PHENOMENA RULING PREGNANCY AND BIRTH COURSE IN VENETO REGION

As was pointed out in the Introduction, during recent years many different factors have conditioned pregnancy and birth in Veneto Region. First of all, the increase of mean woman age of first birth, from 25 years in 1981 to 32 years in 2004; such a phenomenon implies not only fertility reduction, but also an increase of pathological pregnancies and of pathological newborns, which requires a larger health assistance and therefore the necessity of improving existing services to face a higher request, by suitable health care plans. A second change factor concerns the dramatic recent immigration flow, which increased the amount of foreign citizens in the territory, in particular those in prolific age. The importance of such a factor concerns both a higher birth rate because of higher fertility of foreign women with respect to Italian ones, and a different behaviour of foreign women for what concerns access to services, as they prefer public consulting rooms and hospitals instead of general practitioners, and then they effect a lower number of inspections and diagnostic examinations. The study of the effects of these demographic changes on the women pregnancy course and on births is interesting for the future consequences on service requests. As is well known, pregnancy is the state of a woman who carries the product of conception; during pregnancy woman's organism undergoes many variations which are useful to assure the best condition for embryo installation, nutrition and growing. Normal pregnancy course may be interrupted by miscarriage or abortion; miscarriage risk factors are a past mother miscarriage history and a high mother age; abortion may be executed within the first three months from conception, if required by the mother. For what concerns going on pregnancy, we may distinguish two types of pregnancy: physiological, which follows a natural course, without complications, and pathological, with complications both for the mother and for the embryo. During physiological pregnancy the only instrumental examination is echography (ultrasound scan) which is normally made three times during the whole pregnancy course; a higher number of echographies is possible, generally due to excessive anxiety and permitted by money to be spent.
During pathological pregnancy one or more hospital admissions are required, while echographical examinations are made anyway.

At birth newborns may be normal, i.e., called "of first level", or alternatively "of second level" when they require neonatal pathologic care and "of third level" when they require neonatal intensive care and they present life risk. Because of some of the causes seen above, in particular mother high age, the state of newborns' health is worsening; in Veneto Region only some of the Hospitals with obstetric wards are equipped for giving high level neonatal cares; in the case the obstetric ward is not equipped for neonatal special cares, it is necessary the newborn to be immediately transferred to one of more specialized centres, which have often to satisfy large demand in comparison to their capacity. For this reason it is important to know the amount and the trend of these pathological newborns, so to analyse whether their amount is going to increase, and by which rate, related to the variation of factors such as the age of the mother or the presence in Veneto of foreign women.

DATA SOURCES AND DATA ELABORATION METHODS

The Epidemiology and Community Medicine Unit, coordinated by one of the authors (Paola Facchin), manages a large amount of national and regional data concerning population health and health assistance.

The most interesting data for our work are contained in two papers, the CEDAP (birth assistance paper) and the SDO (hospital discharge paper). The CEDAP reports information about the parents, the pregnancy, the birth and the newborn. The SDO reports information about the patient, the hospital, the admission, the admission department (or departments), the disease and the cares supplied; it is emitted also for newborns.

Data were recorded in four data sets by employing software SAS; two data sets come out from CEDAP and SDO respectively and the other two are mixtures, one on the basis of the mothers and the other on the basis of the newborns.

The main model parameters, ruling all actions during pregnancy, are the mother ethnic group and the mother age. We considered five ethnic groups, Italy and West Europe, East Europe, Africa, America and Asia: related women age distributions are respectively specular lognormal, normal, lognormal, normal, lognormal. Other interesting distribution are weibull for the number of echographies and for the number of admissions, gamma for the length of stay after miscarriage and after abortion, weibull for the pregnancy age. Dichotomical variables concern previous miscarriages or abortions and previous living and dead newborns. Links with dichotomic or polotomic variables have been expressed by means of logistic regression.

THE MODEL

The many phenomena and aspects that condition the whole pregnancy course make this reality very complex to be represented and analyzed. It is therefore necessary to use an instrument that allows to synthetically describe the whole real system, even if less detailed, but with the advantage of being more easily manageable. Dynamic discrete simulation give a suitable answer to this aim. Therefore a simulation model was built up and implemented; for model implementation we used the software package Micro Saint 3.2, a powerful and flexible instrument, which can be used both by simulation’s experts and programmers and by less qualified users. Model implementation uses a graphical model representation where essential elements are: ellipses (tasks) representing activities (time running and/or variable computing), directed arcs representing activity sequences, rhombs representing (multiple, probabilistic or deterministic) decisions and striped rectangles representing queues.

The model graphical representation is reported in Fig. 1: it includes 38 tasks. The model’s structure includes three main components: the first component (tasks 1 to 11) concerns the starting phase of pregnancy, the second component (tasks 12 to 25) represents the health assistance course during pregnancy and the last one (tasks 26 to 38) describes the births and the different birth’s results.

In the first model component pregnant women are generated in a random independent way; to every pregnant woman a set of attributes, reporting age and ethnic group (Italy and Western Europe, East Europe, Africa, America and Asia), which are essential for pregnancy course behaviour and possible birth pathologies, is stated (see Fig. 2); pregnancy may be interrupted by miscarriage or abortion or may go on (see Fig.3); miscarriage requires echography and either hospital admission or day hospital, as reported in Fig. 4; abortion requires the same assistance (Fig. 5); all probabilities, probability distributions and time distributions have been obtained and inferred from data contained in hospitals’ and region’s information systems.

In the second model component the pregnancy course for non interrupted pregnancies is described. Pregnant women are divided among the ones characterised by a “physiological pregnancy” and those characterised by a “pathological pregnancy” (Fig. 6); physiological pregnancies require only three regularly scheduled (and sometimes more irregularly added) echographies before birth, as shown in Fig. 7; pathological pregnancies are characterised by random (possibly repeated more times) hospital admissions, and in any case by echographies, as reported by Fig. 8, where pregnant women are temporarily "doubled" to perform both actions; all functions and statistics are obtained from region information system data bases.

The third model component describes the birth (single, two or three twins) and possible newborn pathologies; we distinguish the newborns according to the three different levels of health care they need; as seen above: the newborns with low risk are at the first level and require an orderly obstetric ward, the mildly premature or moderately abnormal newborns are at the second level and require an obstetric ward with incubators and more qualified assistance, the strongly premature newborns or those characterized by severe pathologies are at the third level and require a specialized ward with ventilated incubators and dedicated personnel.

The model parameters are the amounts of resident population and the age distribution of the different ethnic groups; the model outputs are the amount of miscarriages and abortions, the amount of physiological and pathological pregnancies, the amount of newborns at the three levels, the
amount of hospital admissions, the amount of echographies, in other words they give the amount of hospital services required.

MODEL APPLICATIONS

The model ran with current parameters (related to years between 2002 and 2004) and it was possible to prove that the model’s results were in good accordance with actual ones; once validated the model we carried out three experiments in order to forecast the effects of different factors on the amount of newborns with health at risk in year 2009, in order to foresee the impact on admission requests to the present scarce wards; the considered factors were:

- the increasing age of pregnant women;
- an increase of the proportion of pregnant women of different origins than Italy and West Europe;
- both above factors considered at the same time.

In all three cases we considered pregnancies to increase by about 17.5% between 2004 and 2009, by adopting the same trend of the period 2002 - 2004.

In the first experiment, the results were the following: miscarriages reported an increase of 26.5%, followed by abortion (17.6%), pregnancies with pathological course (17.5%) and pregnancies with physiological course (15.7%). The number of newborns increased by 16.25% between 2004 and 2009: the third level newborns reported the highest increase (19.06%), followed by the second level newborns (16.17%) and the first level (16.03%). Echographies expected in 2009, caused by the increase of woman’s age, have been 247,497, with an increase of 16.15% on the total of 213,067 echographies resulted from the initial model in the current 2004 situation. The total number of admissions at hospitals have been 13,048 instead of 11,228 on the initial model (17.5% larger).

In the second experiment abortions registered the highest percentage increase between 2004 and 2009, that is 33.25%, followed by miscarriages (16.7%), pathological pregnancy’s course (16.48%) and physiological pregnancy’s course (15.86%); newborns increased by 16%; the third level newborns increased by 21.97% between 2004 and 2009, the second level ones increased by 19.9%, and the first level newborns increased by 14.98%. Echographies in the second experiment were 240,008, showing a 12.6% increase with respect to the initial model result. The increase of women from East Europe, Asia, Africa and America caused the comparative increase of the women that effected three echographies during pregnancy and the decrease of the women that effected four or more. The number of admissions into hospitals increased by 18.3%, up to 13,278 (second experiment) from 11,228 (initial model result).

In the third and last experiment, the abortions had the highest increase of 33.16%, followed respectively by miscarriages (24.14%), pregnancies with pathological course (15.14%) and pregnancies with physiological course (13.84%); the newborns increased by 14.2% in the total; the third level newborns increased by 19.64% between 2004 and 2009, the second level newborns increased by 18.17% and the first level ones by 13.11%.

The model showed to be highly flexible and therefore extendable to other forecasts for the same Veneto Region, or alternatively to different regions, after changing the essential parameters.

CONCLUSION

A dynamical simulation model of pregnancy-birth course in Italian Veneto Region has been built up, implemented and validated. Beside descriptive results, it was used to forecast future situations and consequent assistance requests in correspondence to changes in the resident population and in some population habits. The model revealed to be very flexible and may be used for other forecasting or may be easily applied to other Italian or foreign regions.

REFERENCES

- http://jhi.sagepub.com/ site related to journal “Health Informatics Journal”
- http://www.schattauer.de/ site related to journal “Methods of Information in Medicine”
- http://www.iee.org/Publish/Journals/ProfJourn/Proc/SYB/ site related to journal “Systems Biology”
- www.regione.veneto.it

BIOGRAPHIES

Alessia Barella was graduated in Statistics at the University of Padova and obtained a PhD in Epidemiology after three years of research at the Epidemiology and Community Medicine Unit.

Paola Facchin obtained a medical degree at the University of Padova where she is now Associate Professor of Paediatrics; she is also coordinator of the Epidemiology and Community Medicine Unit.

Giorgio Romanin-Jacur was graduated in Electrical Engineering at the University of Padova where he is now Full Professor of Operations Research.

350
Fig. 1: Pregnancy-birth model implementation in MicroSaint

Fig. 2: Statement of main parameters: 2 ethnic group, 3 age

Fig. 3: 4 miscarriage, 8 abortion, 12 pregnancy course

Fig. 4: 4 miscarriage, 5 echography, 6 hospital admission, 7 day hospital

Fig. 5: 8 abortion, 9 echography, 10 hospital admission, 11 day hospital

Fig. 6: 12 pregnancy course, 13 physiological pregnancy, 17 pathological pregnancy

Fig. 7: 13 physiological pregnancy, 14 echography, 15 wait echography, 16 wait birth.
Fig. 8: 17 pathological pregnancy, 18 doubling, 19 echography, 20-21 wait and exit echography, 22 admission, 23-24 wait and exit admission, 25 wait birth

Fig. 9: 26 birth, 27-28-29 triplet-twin-single birth, 30-31-32-33-34 twins, 35 difference, 36-37-38 first-second-third level
KINEMATIC MODELISATION OF JOINT DISPLACEMENT: VALIDATION IN HUMAN POINTING TASKS

Clément Villars, Jérôme Bastien, Karine Monteil and Pierre Legreneur
Centre de Recherche et d’Innovation sur le Sport, U.F.R.S.T.A.P.S.,
Université de Lyon
27-29, Bd du 11 Novembre 1918, 69622 Villeurbanne Cedex,
France
E-mail : pierre.legreneur@univ-lyon1.fr

KEYWORDS
Sigmoid, model, biomechanics, pointing task

ABSTRACT

The aim of this study was to characterize a human pointing task and to validate a sigmoid model for the joint displacement in this movement. Nine subjects pointed three times ten targets located in their upper limb workspace. These targets were distributed in ipsilatéral, contralateral, proximal and distal spaces. Kinematic data showed that ipsilateral task is shorter. Finger peak velocity is higher and earlier than in contralateral task. Moreover, its trajectory curvature is greater in ipsilateral movements. Concerning joint displacement sigmoid model, predictive displacements are similar with experimental results. The model reliability is correlated with joint amplitude.

INTRODUCTION

Procedural animation in robotic and video games is actually an industrial challenge aimed to reduce the pre-production charges. For this purpose, heuristic models of joint displacements have to be performed aimed to deduce behavioral general laws while various movement generations, like pointing, grasping or dynamic tasks.

Joint displacement results from activation of synergistic actuators, i.e. agonist and antagonist muscles. Motor control of a mono-articular movement has to take into account two constraints (Van Ingen Schenau 1989), so-called anatomical and geometrical ones. During dynamic motion of an articulation, it is necessary to attain a zero velocity of the joint when its maximal amplitude is reached, in order to preserve its physical integrity (anatomical constraint). Moreover, in most cases, the purpose is to transfer the angular velocity of the joints into linear velocity of the distal extremity of the articular chain (geometrical constraint).

Therefore, joint displacement could be modeled under a sigmoid shape controlled by 7 adjusting variables, i.e. 3 temporal parameters ($t_i, t_o, t_f$), 3 spatial variables ($\theta_i, \theta_o, \theta_f$) and the sigmoid slope corresponding to velocity peak ($K$) (Figure 1).

Thus, the main purpose of this study was to validate such a joint displacement model in human pointing tasks and to characterize these movements.

![Figure 1: Joint kinematic's model and adjusting variables.](image)

MATERIALS AND METHODS

Nine men right-handed pointed three times ten targets located in their upper limb workspace. Referring to the shoulder, targets were located along five axis ($30^\circ$, $60^\circ$, $90^\circ$, $120^\circ$, $150^\circ$) and distributed between proximal and distal spaces (Figures 2). The position of the finger at the start of the movement was fixed at 40cm of the shoulder along the $90^\circ$ axis.
The kinematic parameters of the movement were calculated in order to characterize the pointing task.

Figure 2: Schematic representation of targets’ localizations and upper limb initial position.

**Mathematical expression of sigmoid**

We try to determine a function $\theta \in [0, T]$ to $\mathbb{R}$ of class $C^2$. Let $t_0, t_0' \leq T$ be 3 times such that

$$0 \leq t_0 \leq t_0' \leq T.$$

We assume that function $\theta$ satisfies:

- $\theta = \theta_i$ on $[0, t_0]$.
- $\theta = \theta_f$ on $[t_f, T]$.
- It exists a number $\varepsilon \in (-1, 1)$, such that $\varepsilon \theta$ is strictly increasing on $[t_i, t_f]$ with
  $$\varepsilon = \text{Sign}(\theta_f - \theta_i)$$
- $\varepsilon \theta$ is strictly convex on $[t_i, t_0]$;
- $\varepsilon \theta$ is strictly concave on $[t_f, T]$.

Let $K$ be a number defined by

$$K = \begin{cases} \frac{\max_{x \in [t_i, t_f]} \theta'(x)}{\varepsilon} & \text{if } \varepsilon = 1, \\ \frac{\min_{x \in [t_i, t_f]} \theta'(x)}{\varepsilon} & \text{if } \varepsilon = -1. \end{cases}$$

Some works are relative to expression of sigmoid (Drakopoulos et al. 1995; Menon et al. 1996; Plamondon et al. 1995a, 1995b, 1998; Plamondon et al. 2003; Singh and Chandra 2003; Yun and Kim 2003) but these results can not be used here.

We consider $a$, $b$ et $k$ defined by

$$a = \frac{t_0 - t_i}{t_f - t_i},$$
$$b = \frac{\theta_0 - \theta_i}{\theta_f - \theta_i},$$
$$k = K \frac{t_f - t_i}{\theta_f - \theta_i}.$$  

For all $a, b > 0$ et $k > 2$, we define the function $H(a, b, \kappa)$ over $[0, \alpha]$ by

$$H(a, b, \kappa)(u) = a \left(1 - e^{-bu}\right).$$

We set

$$\gamma_0 = \frac{1}{e^{\frac{1}{2}} - 1} \approx 1.54.$$  

The numbers $a$, $b$ and $k$ are given by the following method: we set

$$r = \frac{\beta}{k\alpha} \in \left[0, \frac{1}{2} - 1\right].$$

There exists a unique $x \in \left[0, \frac{1}{2}\right]$ such that

$$\left(e^x - 1\right)\left(1 - \frac{1}{x}\right) = y,$$

and $a$, $b$ and $k$ are given by

$$k = \frac{1}{1 - x},$$
$$a = \frac{1}{e^{-x}},$$
$$b = \frac{x}{e^x}.$$  

For what follows, we set

$$(a, b, r) = \mathcal{G}(a, b, r).$$

We consider the vector $\mathbf{A} = (t_0, t_0', \theta_i, \theta_0, \theta_f, K) \in \mathbb{R}^7$; we can define $\sigma_a$ by the following method:

We assume

$$k \in \left[\gamma_0 \max \left(\frac{b}{a}, \frac{1 - \beta}{1 - a}\right), +\infty\right].$$

And we set for all $u \in [0, 1]$

$$g(u) = \begin{cases} H(g(u), 0, \kappa)(u), & \text{if } u \leq a, \\ 1 - H(g(u), 1 - a - \kappa)(1 - u), & \text{if } u > a. \end{cases}$$

Finally, we consider the function $\sigma_a$ defined by for all $t \in [0, T]$

$$\sigma_a(t) = \begin{cases} \theta_i, & \text{if } t \leq t_i, \\ \theta_f - \theta_i \cdot \frac{t - t_i}{t_f - t_i} + \theta_i, & \text{if } t_i < t < t_f, \\ \theta_f, & \text{if } t \geq t_f. \end{cases}$$

**RESULTS**

Experimental results showed that coordinations in pointing tasks depended on target localization, i.e. contralateral vs ipsilateral spaces. So, elbow and wrist contributions were predominant for ipsilateral targets whereas shoulder’s displacement was greater for contralateral ones. Angular joints’ velocities presented higher mean and maximal values when targets were ipsilateral and distal. Concerning hand-path curvature, it depends on target distance and decreased with the distance between finger initial position and target localization.
When comparing the sigmoid model predictions and experimental upper limb joints' trajectories, the mean deviations along the 270 trials were $7.1 \pm 4 \text{ mm}$, $9 \pm 5 \text{ mm}$ and $10.8 \pm 7 \text{ mm}$ for the shoulder, the elbow and the wrist respectively. These differences are slight in comparison with the upper limb length ($80 \pm 1.7\text{cm}$). Moreover, as shown in figure 3, wired schematic representations of the pointing task for experimental and predicted movements are similar.

Figures 3: Wired schematic representation of upper limb displacement for experimental movement (on left) and model movement (on right).

CONCLUSION

Experimental data are in agreement with those of the literature (Tseng and Scholz 2003) and allowed to consider the general sigmoid model of joint displacement as valid for slow movements like pointing tasks. The interest of such a model is to formalize any articular movement and deducing, without degradation of the signal induced by its own noise, the joint velocity and acceleration curves. Moreover, regarding to video games context, the sigmoid model seemed to be a valid tool in regard to procedural animation constraints.

REFERENCES


Using a GA to Determine Genotype and Phenotype Relationships

Clyde Meli
Assistant Lecturer
CIS Department
University of Malta
Tel: 2340-2509
cmeli @ cis.um.edu.mt

ABSTRACT

A method using genetic algorithms is being proposed to help advance the analysis of genotype and phenotype relationships so as to help in the determination of which genes contribute to a particular disease or condition. The genetic algorithm has potential in mapping out the genotype-phenotype relationship in a computationally efficient way. The GAGENES package (Meli, 2006) was used for this purpose.

General Terms
Algorithms, Bioinformatics, SNP, Genotype, Phenotype.

Keywords
ga, gagenes, bioinformatics, SNP, genotype, phenotype, c++, object-oriented

1.INTRODUCTION

1.1 GENETIC GENOTYPE AND PHENOTYPE

Every organism has characteristics which are determined by the genes in its chromosomes. Each gene can take different forms, known as alleles, which produce different characteristics associated with that gene. The interactions of these genes build up the biological system's complexity.

The Genotype is the genetic identity of a biological entity, including allele or gene forms, that are not externally visible. It is akin to a biological specification.

On the other hand, the Phenotype is the outwardly physically expressed manifestation of the organism. It is made up of the physical parts of the organism, including cells, tissues, organs, molecules and proteins.

QTL, Quantitative Trait Loci, are a statistical creation that identifies a particular region of the genome as containing one or more genes that are associated with a trait being measured. This theoretical principle was first proposed by (Sax, 1923) who was the first to associate pattern and pigment markers with seed size in beans. Statistical studies attempt to identify which DNA allele markers are associated with the complex trait by showing that their being in a genotype is consistently accompanied by the phenotype of interest. A significant association indicates genetic linkage between the marker gene and one or more QTLs affecting the trait. Pleiotropy refers to the case where a single gene affects several traits simultaneously. A strategy named functional mapping has been proposed by (Ma et al., 2002) to measure complex function-valued traits, i.e. Those that change as a function of an independent and continuous variable e.g. body size and shape which change with age and metabolic rate. (Carlberg et al., 2000) utilised a Genetic Algorithm (Holland, 1975) to simultaneously map multiple interacting QTLs.

A candidate gene for a complex trait is one for which some a priori basis exists for suspecting that the gene may affect the trait (Lesch et al., 1996).

There is a many-to-many relationship between the genotype and phenotype. Any particular genotype corresponds to a number of different phenotypes,
and different genotypes correspond to a particular phenotype. In "At Home in the Universe" (Kauffman, 1995), Stuart Kauffman writes that "one of the most important presuppositions of Darwin's entire thesis - is gradualism, the idea that mutations to the genome, or genotype, can cause minor variations in the organism's properties - that is, the phenotype." Kauffman tried to demonstrate that it is not the case that gradualism always holds.

Aspects of the phenotype are characteristics which can be measured by observation. Some of these are connected to certain conditions and diseases for example blood pressure percentage, hemoglobin levels. Genetic disorders may play havoc with these characteristics. For example decrease in hemoglobin leads to symptoms of anemia. The phenotype itself as a detectable characteristic can be used to determine an existing condition. For example, (Strojek et al, 2005) determined that the salt sensitive blood pressure may be a phenotype in individuals with a future risk of getting type 2 diabetes. Thus known existing phenotypes may be measured and used the process of determining the genotype-phenotype relationships. (Pezzini et al, 2006) concluded that by analysing specific subcategories of patients, defined by the distribution of established risk factors, one can consistently analyse both phenotype-disease and genotype-disease.

*Single-Nucleotide Polymorphisms (SNPs)* can be found throughout the human genome and some are typically positioned next to an important disease-related gene e.g. a polymorphism might be located next to a defective tumour suppressor gene. SNPs are a primary means for determining the possibilities that one may develop a particular disease. SNP data is available online at the International Hapmap Project, dbSNP and HGVbase databases but typically not with accompanying phenotype data. There are some diseases and conditions arising most commonly from deleitional forms of SNP, for example alpha thalassemia (Baysal and Huisman, 1994).

A *Phenotype-Genotype Grid* (Levy et al, 2006) has been used to provide a visual display of the phenotype-genotype relationship. (Dave-Sharma et al, 1998) found it difficult to correlate phenotype with genotype due to the small number of patients with identical mutations. It is being proposed that this method should work well with patients with different mutations, but the higher number of identical mutations should increase the detection of the phenotype-genotype mapping.

**2. METHOD**

The method that is being proposed utilises a Genetic Algorithm. Genetic Algorithms are search algorithms, based upon principles of natural selection and genetics. They were first proposed by John Holland (Holland, 1975). GA's have been used as computational models of creativity, as well as a basis for machine learning, evolutionary systems and evolutionary programming. The GA does not guarantee to find the best solution, however most likely it will come up with a partially optimal solution. An open source genetic algorithm package named GAGENES (Meli, 2006) has been implemented for general purpose usage and to proceed with this research. It has been applied to this problem. It can be called from C++, C, or other languages (via a C api) and is freely available on http://www.gagenes.com (hosting by Sourceforge). It works on several computer platforms. To avoid confusion of terminology the GA genes and GA chromosomes will be referred to as ga-genes and ga-chromosomes.

Rather than taking a 'mapping population' such as an F2 generation as is done in QTL studies, one takes the genetic DNA sequence for a number of individuals, most of them suffering from one particular condition or disease. Healthy individuals should be chosen for this population
too. Also a known measured phenotype characteristic e.g. blood pressure %, etc must be measured for all the individuals. If the condition is usually diagnosed using a set of measured characteristics, then all of them should be measured. It is essential that the individuals have different mutations (SNPs) or the GA may have difficulty to do its work. A special C++ class representing the population is implemented which represents the possible alterations in the SNP. If one may restrict the possible range of loci for possible SNP’s, this can reduce the storage space needed for representation. A locus in genetics is the position on a chromosome of a gene or other chromosome marker. It can also refer to the DNA at that position. In the current implementation, gene position has been simply “numbered” (i.e. 1\textsuperscript{st} gene, 2\textsuperscript{nd} gene, 3\textsuperscript{rd} gene along a chromosome).

The purpose of this GA will be to determine which pairs (or more) of genes are responsible for affecting the expression of the phenotypes (e.g. protein levels) and result in triggering the disease or condition. It will search the space of possible genes. A Fitness function needs to be devised to be able to determine whether a particular pair (or more) of genes fit the phenotype picture. This requires a known set of ranges of characteristic phenotype values which map to different forms of the condition or disease. One range will specify a healthy individual not suffering from the condition/disease, while another range will indicate a particular form of the condition. Other forms of the condition may exist and will be indicated by other ranges. If these ranges are not known, the determination of the responsible genes may not be possible using this method.

For simplicity, at present an assumption has been taken that there are two genes affecting the disease which are respectively on two different chromosomes, and one range of values as phenotype. The principle should hold for more genes and more phenotypes, however.

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{diagram.png}
\caption{Outline for the Genetic Algorithm Implementation.}
\end{figure}

The genetic algorithm (Figure 1) is initiated by randomly generating a ga-population of ga-chromosomes (candidate solutions to the problem) using the Mersenne Twister random number generator (Matsumoto and Nishimura, 1998). Usage of different random number generators in GAGENES was tested (Meli, 2006). Each member of the population, a ga-chromosome constitutes a vector of coded values, each one of which is called a ga-gene. A ga-gene can take the values 0 or 1, as a binary representation was used. Every ga-chromosome (representing a point in the genetic search space) in this method can store, at present, the position (location) values for two genes on each of the two chromosomes respectively. In the implementation one locus value is stored and used on both maternal and paternal chromosomes. Its fitness is calculated as specified above. A ga-chromosome whose gene position values happen to be higher in number within the patient population will tend to have a higher fitness value. Thus the genetic algorithm will tend to converge towards the gene positions which are influencing the phenotype value.
A suitable fitness function for this GA has been devised as follows:

- Define a function `rangeidentcount` to count the number of individuals within a particular phenotype range of values (inclusive).

- Define a function `rangecountidentalpha` which given the two (biological) genes on the first chromosome, takes those individuals from the population with the phenotype values within a particular given range (specified as percentages), and counts how many of them have the same two pair of alleles within that chromosome.

- Implemented the corresponding function `rangecountidentbeta` for two genes on the second chromosome.

- Implemented the fitness function which takes specific different fixed phenotype ranges and adds the counts returned by calling the above functions `rangecountidentalpha` and `rangecountidentbeta` – for every phenotype range and for both chromosomes being investigated. It should also do some sanity checks to prevent nonsense like an integer-valued GA parameter out of range.

The standard Fitness Proportional selection method is normally used. Another method has been devised and implemented, the Binary Fitness Selection Method, which will be tested in future. The Selection operator selects ga-chromosomes in the ga-population for reproduction and is implemented as the standard “roulette” wheel. A fitter ga-chromosome is more likely to be selected to reproduce. The number of pairs of ga-chromosomes selected as “parents” is equal to the ga-population size divided by 2. Selection is done with replacement, meaning the same ga-chromosome may be selected more than once.

The standard single-point crossover operator roughly mimics biological recombination between single-chromosome (haploid) organisms. It randomly chooses a ga-locus and exchanges the parts split at that single point. With probability $p_c$ (the crossover probability), a pair of “parents” are crossed over at a randomly chosen point, to form two offspring ga-chromosomes. If no crossover takes place, the two offspring formed are copies of their parents. The two offspring are mutated at each locus with probability $p_m$ (the mutation probability). Mutation randomly flips bits in a ga-chromosome.

The current ga-population is replaced by the new ga-population, every generation (iteration), until the termination condition is reached for this run.

3. RESULTS AND EVALUATION

The current implementation only handles deletional forms of a condition or disease (referred to as “-” whereas the existing gene is referred to as “a” which stands for allele), but it should prove possible to handle non-deletional forms by future extension.

To count the number of individuals within a phenotype range, two methods were implemented to compare identical strings as part of this process and both had the same results. The length of a ga-chromosome in this implementation is of 56 bits.

Currently random data has been used to evaluate this system successfully for a fictional deletional-only condition. This patient population size is of 100. The phenotype value in this implementation is a floating point number ranging from 0 to 100.0, representing a percentage. The genetic algorithm loop converges until the last three highest fitness values are identical or 29000 generations have been processed.

For this synthetic test, the correct positions for the genes in the genotype were taken to be 5 and
10 on the both paternal and maternal chromosomes. For every person, phenotype values are calculated. This is done as follows: a check is done of the alleles as were generated randomly earlier, and if both are equal to “aa”, then the person’s phenotype is equal to 0 indicating the absence of the condition. If the 2nd allele pair is “--” and the 1st pair is “-a” or “-a”, the phenotype value is generated randomly within the range 5 to 15. If the two pairs are identical and equal to “-a” or “a”, the value is generated within the range 15 to 25. If the 2nd pair is “aa” and the 1st is either “-a” or “a-”, then the value is generated within the range 32.5 to 37.5. If the 1st pair is “--” and the 2nd pair is “aa”, then the value is generated within the range 37.5 to 42.5. These ranges are the same used in the fitness function.

Within the first generation of a sample run, for example the current fittest ga-chromosome was representing the positions 3789 and 7150 on the first and second chromosomes.

Using a ga-population size of 100, a crossover probability of 0.7, the genetic algorithm was run for a number of times, every time with a different mutation probability - ranging from 0.001 to 0.5. In the first run, it was found that the highest fitness value of 133, was found when the mutation probability was equal to 0.181. Over a number of runs the highest fitness value of 137 was found when the mutation probability was equal to 0.145. The genetic algorithm outputs the fittest value and the positions found for the corresponding four genes which correspond to this high fitness value according to the fixed phenotype ranges specified in the fitness function.

4.CONCLUSION
This study has shown that the genetic algorithm has potential in mapping out the genotype-phenotype relationship in a computationally efficient way. Using simulated data the results from this study show that the method is a robust one. The genetic algorithm turns out to be not too susceptible to changes to the mutation rate. Within a typical run, the lowest fitness value was of 123. The best solution cannot be guaranteed by a genetic algorithm however using repeated runs one may find better solutions and lessen the risk of ending up in a lower optima.

Future development of this promising method will explore enhancements to the genetic algorithm itself and the possibilities of different numbers of genes effecting the phenotype as well as the influence of environmental agents involved in certain multifactorial disorders which need to be accounted for, by suitable representation in the GA.

5.REFERENCES


COMPLEX ENVIRONMENTS SIMULATION
Finite element modelling of underground water flow in low permeability lands

Abhijeet Kulkarni, Bahareh Kaveh-Baghbaderani, Vahid Nassehi*

Chemical Engineering Department
Loughborough University
Loughborough LE11 3TU
Email: v.nassehi@lboro.ac.uk

KEYWORDS
Underground flow, Hydrodynamics, Darcy flow, Finite element modelling, Three dimensional porous flow.

ABSTRACT
The present work deals with three dimensional (3-D) finite element analysis of water flow under the ground. The model applied here is used to simulate the porous flow occurring in the saturated subsurface porous regions. The analysis of this type of flow is a common and important feature of contaminants mobility under the ground and hence development of efficient computer simulation tools for such regimes is of significant environmental importance. In general, two types of governing equations can be used to represent water flow through porous lands. The choice of a particular equation does depend on the type of porous medium under consideration and flow Reynolds number. For low permeability cases the Darcy’s equation provides the best choice whilst in higher permeability cases the Brinkman equation must be used (Darcy, 1856; Brinkman, 1947). In the present work the use of Darcy equation provides a more realistic representation for underground flow of water. Therefore first, we have considered the validation of the Darcy’s law for isotropic, homogenous, incompressible, saturated and isothermal porous media. We have subsequently extended the analysis the case of an underground domain in which the flow is partly obstructed by the presence of an impermeable barrier. It is shown that the model is capable of yielding theoretically expected results for the obstructed domain as well as a simple domain and is hence suitable for the prediction of appropriate locations for the construction of reactive barriers for the treatment of contaminated underground flow.

INTRODUCTION
Underground flow regimes constitute one of the most vital features in the natural hydro-environmental systems. These regimes often have complicated characteristics due to the heterogeneity of soils through which water seepage occurs. Heterogeneity of soil media arises for various reasons such as staggered layers of soil with different porosity or fractures. Therefore mathematical description and modelling of underground flow requires robust techniques both for the formulation of realistic governing equations and boundary conditions for given situations and use of robust and efficient numerical solution techniques.

Underground flow models have been discussed by many researchers (Pinder and Celia, 2006; Fitts, 2002; Nassehi and Das, 2007). Basic transport equations through porous media have been shown (Bear, et al. 1991) to depend on the properties of the fluid as well as factors such as temperature, and permeability and porosity of media. However, in the present paper which is mainly concerned with the simulation of hydro-environmental cases the porous flow regime is assumed to be saturated and isothermal. On the other hand the emphasis has been given to the problem of soil heterogeneity. Delleur (1999) has classified different types of porous media such as soils as:

- Isotropic and homogenous
- Isotropic and heterogeneous
- Anisotropic and homogenous
- Anisotropic and heterogeneous

While considering the realistic models all the above mentioned types may coexists together or in combination. In this work we have considered the first and fourth case. Model developed for these conditions can be readily extended to the other situations.

It has been proven that saturated flow in a porous medium should be modelled using either Darcy or Brinkman equations depending on the porosity of the medium and flow Reynolds number (Wakeman and Tarleton 2005). For very low Reynolds number (creeping flow) and porosity less than 0.6, Darcy’s equation is most suitable. This is the situation in most types of underground flow of water in natural environment. However this equation inherently implies perfect slip conditions at domain boundary walls and cannot include wall effects (Ishizawa and Hori, 1966). Therefore an accurate solution scheme for this equation should be capable of yielding a slip velocity on the porous domains. The finite element scheme used in this work can very effectively cope with such boundary conditions. This technique can also very effectively cope with irregular geometries of solution domain. However, considering the large scale of lands where environmental phenomena needs to be studied, any irregularity of the domain walls can be ignored. In this work we have therefore used a rectangular block domain. Despite the geometrical simplicity of the overall domain, it has also been used to study the impact of an impermeable barrier within the flow field the flexibility of the present finite element scheme in coping with flow...
divisions due to abrupt changes domain geometry has been tested.

GOVERNING MODEL EQUATIONS

The governing model equations for the incompressible creeping, isothermal flow of a Newtonian fluid such as water can be described by following equations

A. EQUATION OF CONTINUITY

The continuity equation representing conservation of mass for an incompressible fluid can be represented using vector notations as

$$\nabla \cdot \vec{u} = 0$$  \hspace{1cm} (1)

Where \( \vec{u} \) is the velocity vector.

B. EQUATION OF CONSERVATION OF MOMENTUM

As mentioned earlier in this work we have selected the Darcy equation to represent the conservation of momentum as it is better suited for hydro-environmental problems. Again using vector notation this equation can be written as (Nield and Bejan, 1992):

$$\rho \frac{\partial \vec{u}}{\partial t} + \nabla p + \frac{\mu}{K} \cdot \vec{u} = 0$$  \hspace{1cm} (2)

Where \( K \) is the permeability of the porous medium, \( \mu \) is the viscosity of the fluid and \( \vec{u} \) is the velocity vector. In its most general description \( K \) should be regarded as a second order tensor which is represented, in the matrix form, as

$$K = \begin{bmatrix}
K_{xx} & 0 & 0 \\
0 & K_{yy} & 0 \\
0 & 0 & K_{zz}
\end{bmatrix}$$  \hspace{1cm} (3)

Where, \( K_{xx}, K_{yy} \) and \( K_{zz} \) are the principle components of the permeability tensor along the \( x, y \) and \( z \) directions of a Cartesian coordinate system. Any anisotropy in a porous medium can hence be modelled via assigning appropriate values to the components of the permeability tensor.

Conjunctive solution of equations (1) and (2) poses a mathematical problem as the first equation does not include a pressure term. Full mathematical analysis of the problem originated because of this situation is somewhat abstract and requires relatively lengthy descriptions. Such discussions can be found in text books dedicated to function analysis. However, it is shown that an stable and accurate solution can be obtained provided that the scheme used satisfies a condition known as the LBB condition (Reddy, 1986). A more convenient way of resolving this problem is to replace equation (1) which is the expression of incompressibility constraint with a modified form as

$$\frac{1}{\rho c^2} \frac{\partial P}{\partial t} + \nabla \cdot \vec{u} = 0$$  \hspace{1cm} (4)

Where \( \rho \) is the fluid density, \( c \) is the speed of sound in the fluid and \( \dot{p} \) is the temporal derivative of the pressure. Equation (4) represents conservation of mass for a slightly incompressible fluid and as it contains a pressure term its conjunctive solution with equation (2) automatically satisfies the LBB condition (Zienkiewicz and Wu 1991).

BOUNDARY CONDITIONS

In the present work the underground flow of water is represented by the conjunctive finite element solution of equations (2) and (4) subject to the following boundary conditions:

*Inlet:* a plug flow velocity boundary condition is imposed

*Walls:* Perfect slip conditions along porous walls and no-slip conditions along the impervious walls are imposed.

*Exit:* Zero pressure (arbitrary datum) is imposed

THE FINITE ELEMENT SOLUTION SCHEME

The finite element method used here is based on the UVWP scheme. The unknown variables \( u, v, w \) and \( p \) are approximated over an element as

$$u \approx \tilde{u} = \sum_{j=1}^{k} N_j u_j \quad v \approx \tilde{v} = \sum_{j=1}^{k} N_j v_j$$  \hspace{1cm} (5)

$$w \approx \tilde{w} = \sum_{j=1}^{k} N_j w_j \quad p \approx \tilde{p} = \sum_{j=1}^{k} N_j p_j$$

Where, \( k \) is the number of nodes per element and \( N_j \) is the shape function associated with the nodes \( j = 1, \ldots, k \).

After substituting these approximations into the governing equations weighting of the generated residuals via the standard Galerkin method, the weak variational formulation for the porous flow model is obtained. To reduce the inter element continuity requirement of the field variables to \( C^0 \) the first order differential equations of pressure, and to maintain the consistency of the formulation with the velocity in the statement, are treated by Divergence theorem. The surface integrals appearing are taken to the right-hand side of the variational statement (Zienkiewicz and Taylor, 1991). These surface integrals cancel out during the final assembly of the stiffness matrices or are replaced...
by using the boundary conditions. The final working equations for the porous flow model are represented as:

$$
\begin{align*}
[ A_{y} ]^{11} & = \int_{\Omega} \left( N_i^j N_j + \alpha \Delta t \frac{\mu}{K_\rho} N_i^j N_j \right) dxdydz \\
A_{y}^{12} & = 0 \\
A_{y}^{13} & = 0 \\
A_{y}^{14} & = -\int_{\Omega} \alpha \Delta t N_i^j \frac{\partial N_j}{\partial x} dxdydz \\
A_{y}^{21} & = 0 \\
A_{y}^{22} & = \int_{\Omega} \left( N_i^j N_j + \alpha \Delta t \frac{\mu}{K_\rho} N_i^j N_j \right) dxdydz \\
A_{y}^{23} & = 0 \\
A_{y}^{24} & = -\int_{\Omega} \alpha \Delta t N_i^j \frac{\partial N_j}{\partial y} dxdydz \\
A_{y}^{31} & = 0 \\
A_{y}^{32} & = 0 \\
A_{y}^{33} & = \int_{\Omega} \alpha \Delta t N_i^j \frac{\partial N_j}{\partial z} dxdydz \\
A_{y}^{34} & = \int_{\Omega} \alpha \Delta t N_i^j \frac{\partial N_j}{\partial z} dxdydz \\
A_{y}^{41} & = \int_{\Omega} \alpha \Delta t N_i^j \frac{\partial N_j}{\partial z} dxdydz \\
A_{y}^{42} & = \int_{\Omega} \alpha \Delta t N_i^j \frac{\partial N_j}{\partial z} dxdydz \\
A_{y}^{43} & = \int_{\Omega} \alpha \Delta t N_i^j \frac{\partial N_j}{\partial z} dxdydz \\
A_{y}^{44} & = \int_{\Omega} \alpha \Delta t N_i^j \frac{\partial N_j}{\partial z} dxdydz \\
C_{i}^{2} & = -\int_{\Omega} \alpha \Delta t N_i^j P_i^r n_j dA_e \\
C_{i}^{3} & = -\int_{\Omega} \alpha \Delta t N_i^j P_i^r n_j dA_e \\
C_{i}^{4} & = -\left( \frac{1}{2} \alpha \Delta t c^2 \right) N_i^j \frac{\partial P_e}{\partial x} n_j dA_e - \left( \frac{1}{2} \alpha \Delta t c^2 \right) N_i^j \frac{\partial P_e}{\partial y} n_j dA_e \\
& - \left( \frac{1}{2} \alpha \Delta t c^2 \right) N_i^j \frac{\partial P_e}{\partial z} n_j dA_e \\
\end{align*}
$$

Where

The discretized working equations are then transformed into the local coordinate system using iso-parametric mapping. The elements used are 8 nodded hexahedral elements. The integrals in the elemental stiffness equations are calculated using Gauss-Legendre quadrature. After evaluation of the members of the elemental stiffness matrix, they are assembled over the common nodes in the computational grid to obtain a system of algebraic equations. After substituting the boundary condition, the system becomes determinate and the algebraic equations can be solved using a frontal solution method (Irons 1970). The described simulations are carried out using an in-house developed Intel Visual FORTRAN code in Pentium IV processor. The results are plotted using Techplot 360 ® (evaluation version). The time stepping technique used here is second order Taylor-Galerkin method. (Nassehi, 2002)

**COMPUTATIONAL RESULTS AND DISCUSSIONS**

Results are presented for different geometrical flow domain and discussed with increasing order of the intricacies involved, in terms of the geometrical features and the boundary condition imposed. For all the simulation done, the fluid under consideration is water and the properties of water, at 20° C, are considered. Consistency coefficient is taken to be 0.001 Kg m^-1 s^-1 and the density is assumed to be 1000 Kg m^-3. Depending upon the property of the permeable medium, isotropic or anisotropic, the permeability of the medium was selected to cover the range applicable to the realistic situations.

- **Case I: Rectangular block domain with isotropic permeable medium**

For the first case, a simple rectangular block (test) domain is considered where the dimension of the domain is 100 (L) x 75 (B) x 75 (W) m along x, y and z axis respectively. The computational grid comprises of 2000 elements and 2541 nodes which was used for simulation, as shown in Figure 1. The porous domain is considered to be isotropic (K_x = K_y = K_z = 10^{-12} m^2) and homogeneous. The schematic representation of the boundary condition imposed over the different faces is shown in Figure 2 and Figure 3, along xy and xz plane respectively.
flow. The corresponding nodal pressures are shown in Figure 6. This is as theoretically expected, the fluid seeps the low permeability porous matrix and the viscous stress associated with the bulk of the fluid gets transferred to the solid porous matrix.

Figure 1: Computational finite element mesh for the test domain.

Figure 5: Velocity vector plot for the test domain

Figure 2: Schematic representation of the boundary condition in the xy plane.

Figure 6: Pressure plot for the test domain

The discrepancy between the inlet and the outlet mass flow rate was found to be about 0.08 percent which ensures the continuity of the solution and thus validating the model.

- Case II: Rectangular block domain with anisotropic permeable media with an impermeable solid barrier

The dimensions of the porous domain are the same as that of the previous one considered, however with an impermeable solid barrier obstructing the flow of fluid. As shown in Figure 7, the impermeable barrier forms the inner domain with dimension 25 x 37.5 x 25 m along x, y and z axis respectively. The computational mesh grid consists of 3000 hexahedral elements and 3751 nodes. For the outer domain the velocity boundary condition applied are same as mentioned above. As a necessary boundary condition the velocity boundary condition imposed on the surfaces of the barrier are zero (\(V_x = V_y = V_z = 0\)), Figure 8. However for the present case the medium is considered to be saturated, anisotropic and homogeneous. Thus more realistic value of the permeability in the subsurface flow as mentioned by Nassihi and Das (Nassiehi, Das 2007).

\[
K_{xx} = 3.36 \times 10^{-9} \text{ m}^2, \quad K_{yy} = 2.37 \times 10^{-9} \text{ m}^2, \quad K_{zz} = 1.48 \times 10^{-9} \text{ m}^2
\]

Figure 4: Three dimensional schematic representation of the boundary points imposed different faces of the domain.
Figure 7: Computational domain for the simulation of the flow across a non permeable barrier in the porous media.

Figure 8: Three dimensional representation of the boundary points on which the conditions are imposed for flow past an impermeable solid barrier.

The velocity contour and vector plot are shown in different plane sections in Figure 9 to Figure 14. The pressure plot is shown in Figure 15, which is plotted on a plane parallel to the xy plane.

Figure 9: Velocity contour in the section parallel to the xy plane at z=37.5m.

Figure 11: Velocity contour in the plane parallel to yz plane at x = 50m.

Figure 12: 2D Velocity vector plot in the surface parallel to the xy plane at z=37.5m.

Figure 13: 2D Velocity vector plot in the surface parallel to the xz plane at y=18.75 m.

Figure 10: Velocity contour in the section parallel to the xz plane at y = 18.7 m.

Figure 14: Velocity vector plot in the surface parallel to yz plane at
\( x = 50 \text{m} \).

**Figure 15:** Predicted pressure plot in a sectional plane for the flow past an impermeable barrier.

As expected the presence of an impermeable barrier does affect the flow of the fluid in the porous domain, which can be seen in the vector plot in Figure 12 and Figure 13. However very important feature for the present case study is the pressure plot, as shown in Figure 15, which shows that the pressure varies through out the domain, unlike as observed in the first case. The pressure is observed to be very low, along the length, close to the barrier however, as the go away from the barrier the pressure increase in an abrupt manner. The maximum pressure observed at the upper tip of the barrier.

**CONCLUSIONS**

The described simulation is based on the finite element solution of 3 dimensional Darcy equation, which provides a powerful means to the investigation of seepage flow of water in the subsurface regions. The accuracy of the model is verified by the ability of the model to preserve the mass balance and the agreement between the predicted velocity and pressure results with the theoretical expectations. The model is shown to yield very useful results for saturated, homogenous, isothermal and anisotropic permeable medium. It is also shown that the present model can be extended to obtain the realistic results for domains with heterogeneous features.

**REFERENCES**


MODELING AND SIMULATION OF FEEDBACK SEISMIC STRUCTURAL CONTROL

Svetla Radeva
University of Architecture,
Civil Engineering and Geodesy
1 Hristo Smirnensky Blvd.
1046 Sofia, Bulgaria
E-mail: svetla_fec@abv.bg

Izabella Lokshina
SUNY Oneonta
324B Netzer Administration Bldg.
Oneonta, NY 13820, USA
E-mail: lokshiv@oneonta.edu

KEYWORDS
Modeling and Simulation, Structural Engineering, Structural Control, Seismic Processes.

ABSTRACT

Most of the current research on active structural control for seismic protection has focused on either full state feedback strategies or velocity feedback strategies. However, accurate measurement of the necessary displacements and velocities of the structure is difficult to achieve directly, particularly during seismic activity. As accelerometers are inexpensive and can readily provide reliable measurements of the structural acceleration at the strategic points on the structure, development of control methods based on acceleration feedback control is an ideal solution to this problem. An approach to modeling of feedback seismic structural control with active mass drivers is shown. This paper experimentally demonstrates efficiency of the acceleration feedback mass driver systems to reduce response of the seismically excited structures.

INTRODUCTION

Structural control provides possibility to recognize the measures to reduce seismic vulnerability of the high risk structures, like nuclear power plants, bridges, lifelines, dams, and high buildings (Radeva et all, 2005). There is active, semi-active and passive structural control (Nasu et all, 2001). Most of the current research on active structural control for seismic protection has been focused on either full state feedback strategies (i.e., the structural displacements and velocities) or velocity feedback strategies (Gluck et al. 2000). Displacements and velocities are not absolute, but dependent upon inertial reference frame in which they are taken. Their direct measurements or arbitrary locations on large-scale structure are difficult to achieve. During seismic activity this difficulty is exacerbated, because the foundation to which the structure is attached, is moving along with the ground, and does not provide an inertial reference frame. Thus, the control algorithms that depend on direct measurements of the displacements and velocities may be impracticable for full scale implementations (He et al. 2003). Alternatively, the accelerometers can provide inexpensive and reliable measurements of the accelerations at strategic points on the structure, making use of the structural absolute acceleration measurements for control force determination that could be an ideal way to avoid the problem. This paper considers acceleration feedback strategies for active structural control with active mass driver systems to reduce structural response during seismic activity.

EXPERIMENTAL SETUP

The simulator used for this investigation consists of a hydraulic actuator servo/valve assembly that drives a 122cm × 122 cm aluminum slip table mounted on high-precision, low-friction linear bearings. The capabilities of simulator are: maximum displacement ±5 cm, maximum velocity ±90 cm/sec, and maximum acceleration ±4 g/s with a 450 kg test load. The operational frequency range of the simulator is nominally 0-50 Hz. The test structure, shown in Fig.1, was a three-storey single-bay scale model building controlled by an active mass driver. The building frame was constructed on steel with a height of 160 cm. The floor masses of the model weighted a total of 230 kg, distributed evenly between the three floors. The time scale factor was 0.2 making the natural frequencies of the model approximately five times those of the prototype.

Figure 1: Experimental Setup

A simple implementation of the active mass driver was placed on the third floor of the structure for control purposes. The active mass driver consisted of a high pressure hydraulic actuator with steel masses attached to each end of the piston rod. A Duvall servo valve was employed that has an operational frequency range of 0-45 Hz. This hydraulic actuator was fitted with low friction Teflon seals to reduce nonlinear frictional effects. The total
mass of the structure including the frame and the active mass driver was 300 kg. The moving mass of the active mass driver was 1.7% of the total mass of the structure. As the hydraulic actuators are inherently open loop unstable, position feedback was employed to stabilize the control actuator. The position of the actuator was obtained with a linear variable differential transformer rigidly mounted between the piston rod and the third floor.

As shown in Fig. 1, the accelerometers positioned at each floor of the structure and at the active mass driver, measured the absolute accelerations of the model, and an accelerometer located on the base measured the ground excitation. The displacement of the active mass driver was measured using the linear variable differential transformer. Only the three floor acceleration measurements and the absolute acceleration of the active mass driver were employed for purposes of control force determination.

To develop a high quality, control-oriented model, an eight channel data acquisition system consisted of eight Syminex XFM82 3 decade programmable anti-aliasing filters were employed. Additionally, the data acquisition system consists of an Analogical CTRTM-05 counter/timer board and the Snap-Master software package. The XFM82 offer programmable pre-filter gains to amplify the signal into the filter, programmable post-filter gains to adjust the signal so that it falls in the correct range for the A/D converter, and analog anti-aliasing filters which are programmable up to 25 kHz.

Implementation of the digital controller was performed using the Spectrum Signal Processing Real-Time Signal Processor (DSP) System. The on-board A/D system has two channels with 16 bit precision and a maximum sampling rate of 200 kHz. The two D/A channels, also with 16 bit precision, allow for even greater output rates so as not to be limiting.

**SYSTEM IDENTIFICATION**

Developing of an accurate mathematical model of structural system is very important in the control design (Scruds & Iwan 2003). For linear structures system identification techniques fall into two categories: time domain and frequency domain (Yuen & Beck 2003). Time domain techniques such as recursive least squares system identification method (Zhang & Iwan 2003) are superior when limited measurement time is available (Nishitani et al. 2002).

Frequency domain techniques are generally preferred when significant noise is present in the measurements and the system is assumed to be linear and time invariant. In the frequency domain approach the first step is to experimentally determine the transfer functions (also termed frequency response functions) from each of the system inputs to each of the outputs. This approach was used to develop a mathematical model of the structural system. A block diagram of structural system identification is shown in Fig. 2, where the two inputs are the ground excitation \( \ddot{x}_g \) and the command signal of the actuator \( u \).

The four measured system outputs include the absolute acceleration of the actuator \( \ddot{x}_w \) and the absolute accelerations \( \ddot{x}_{a1}, \ddot{x}_{a2}, \ddot{x}_{a3} \), of the three floors of the test structure.

![Figure 2: System Identification](image)

Then, a 4×2 transfer function matrix (i.e., eight input/output relations) has to be defined to describe the characteristics of the system.

**EXPERIMENTAL DETERMINATION OF TRANSFER FUNCTIONS**

Methods for experimental determination of transfer function are divided in two main types: the swept-sine and broadband approaches using fast Fourier transforms. Both methods can produce accurate transfer function estimates. The swept-sine approach is rather time-consuming, because it analyzes the system one frequency at a time. The broadband approach estimates the transfer function simultaneously over a band of frequencies. The first step is to independently excite each of the system inputs over the frequency range of interest. Exciting the system at frequencies outside this range is typically counter productive; thus the excitation should be bound limited (e.g., pseudo-random). Assuming the two continuous signals (input \( u(t) \) and output \( y(t) \)) are stationary, the transfer function is determined by dividing the cross spectral density of the two signals \( S_{uy} \) by the auto-spectral density of the of the input signal \( S_{uu} \) as (1).

\[
H_{uy}(j\omega) = \frac{S_{uy}(\omega)}{S_{uu}(\omega)}
\]

(1)

However, experimental transfer functions are usually defined using discrete-time data. The continuous time records of the specified system input and the resulting responses are sampled at \( N \) discrete-time intervals with an A/D converter yielding a finite duration, discrete-time representation of each signal \( u(nT) \) and \( y(nT) \), where \( T \) is the sampling period and \( n=1,2,\ldots,N \). For the discrete case (1) can be presented as (2).

\[
H_{uy}(jk\Omega) = \frac{S_{uy}(k\Omega)}{S_{uu}(k\Omega)}
\]

(2)

where \( \Omega = \omega_0/N, \omega_0 \) is the sampling frequency, \( k=0,1,\ldots,N-1 \). The discrete spectral density functions are obtained with standard digital signal processing methods. The frequency transfer function can be considered as frequency sampled version of the continuous transfer function in (1).

In practice, one collection of samples of length \( N \) does not produce very accurate results. Better results are obtained
by averaging the spectral densities of a number of collections for samples of the same length. Given that \( M \) collection of samples are taken the equations for averaged functions are (3),

\[
\begin{align*}
\bar{S}_{uu}(k\Omega) &= \frac{1}{M} \sum_{i=1}^{M} S_{uu}^{i}(k\Omega) \\
\bar{S}_{yy}(j\Omega) &= \frac{1}{M} \sum_{i=1}^{M} S_{yy}^{i}(k\Omega) \\
\bar{H}_{yu}(j\Omega) &= \frac{\bar{S}_{uu}(k\Omega)}{\bar{S}_{yy}(k\Omega)}
\end{align*}
\]  

(3)

where \( S \) denotes the spectral density of the \( i \)-th collections samples and the over bar represents the ensemble average. Let’s note that increasing number of samples \( N \) increases the frequency resolution, but does not increase the accuracy of the transfer functions. Only increasing number of averages \( M \) will reduce the effects of noise and nonlinearity in the results. To define the discrete spectral density functions in (3) a finite number of samples is acquired and a fast Fourier transform is performed.

The transfer functions from the ground accelerations to each of the four measured responses were obtained by exciting the structure with band-limited white noise ground acceleration (0-50 Hz) to the actuator command while the ground was held fixed.

Next step in the system identification procedure is to model the transfer functions as a ratio of two polynomials in the Laplace variable \( s \). This task was accomplished via a least squares fit to the ratio of numerator and denominator polynomials, evaluated on the \( j\omega \) axis, to the experimentally obtained transfer functions. The algorithm requires the user to input the number of poles and zeros to use in estimating the transfer function, and then determines the location for the poles/zeros and the gain of the transfer function for a best fit. This algorithm was used to fit each of the eight transfer functions.

**CONTROL ALGORITHM**

Let’s consider the general block diagram description of the control problem given in Fig. 3. Here \( y \) is the measured output vector of structural responses, \( z \) is the vector of structural responses which are desired to control, \( u \) is the control input vector, and \( d \) is the input excitation vector. For this experiment the measured output vector \( y \) includes the actuator displacement and the accelerations of the three floors of the test structure.

![Figure 3: Principle Structural Control Block Diagram](image)

The regulated output vector \( z \) may consists of any linear combination of the states of the system and components of the control input vector \( u \), thus allowing a broad range of control design objectives to be formulated through appropriate choice of elements of \( z \). Weighting functions can be added to elements of \( z \) to specify the frequency range over which each element of \( z \) is minimized.

The structural control block diagram in Fig. 4 contains the test structure, the active mass driver, in addition to filters and weighting functions in the frequency domain.

![Figure 4: Structural Control Block Diagram](image)

The task here is to design a controller \( K \) that stabilizes the system and, within the class of all controllers which do so, minimizes the \( H_2 \) norm of the transfer function matrix \( H_{zd} \) from \( d \) to \( z \), where the \( H_2 \) norm is given according to (4).

\[
\|H\|_{H_2} = \sqrt{\text{trace}\left(\frac{1}{2\pi} \int_{-\infty}^{\infty} |H(j\omega)H^*(j\omega)| d\omega\right)}
\]  

(4)

To obtain the transfer function \( H_{zd} \), we refer to Fig. 3 and partition the system transfer function matrix \( P \) into its components, as shown in (5).

\[
P = \begin{bmatrix} P_{zd} & P_{zu} \\ P_{yd} & P_{yz} \end{bmatrix}
\]  

(5)

The matrix \( P \) includes the weighting functions employed in the control design and is assumed to be strictly proper. The overall transfer function from \( d \) to \( z \) is written as (6).

\[
H_{zd} = P_{zd} + P_{zu} K (I - P_{yu} K)^{-1} P_{yd}
\]  

(6)

Let’s consider a structure experiencing a one-dimensional earthquake excitation \( \ddot{x}_g \) and active control input \( u \). The structural system, which includes the structure and the active mass driver, can be represented in state space form as (7),

\[
\dot{x} = Ax + Bu + Ex_g
\]  

(7)

\[
y = Cy + D_j u + v
\]  

(8)

where \( x \) is the state vector of the system, \( y \) is the vector of measured responses, and \( v \) represents the noise in the measurements. A detailed block diagram of the system...
given in (7) and (8) could be demonstrated in Fig. 4, since the transfer function \( G \) is given by (9).

\[
G = (sI - A)^{-1}B = (sI - A)^{-1}[B E] = [G_1, G_2] \quad (9)
\]

The filter \( F \) shapes the spectral content of the disturbance modeling the excitation, \( C_1 \) and \( C_2 \) are constant matrices that dictate the components of structural response comprising the measured output vector \( y \) and the regulated response vector \( z \), respectively. The matrix weighting functions \( \alpha_1 W_1 \) and \( \alpha_2 W_2 \) are generally frequency dependent, with \( \alpha_1 W_1 \) weighting the components of regulated response and \( \alpha_2 W_2 \) weighting the control force vector \( u \). The input excitation vector \( d \) consists of a white noise excitation vector \( w \) and a measurement noise vector \( v \). The scalar parameter \( k \) is used to express a preference in minimizing the norm of the transfer function from \( w \) to \( z \) versus minimizing the norm of the transfer function from \( v \) to \( z \). For this block diagram representation, the partitioned elements of the system transfer function matrix \( P \) in (5) are given by (10) – (13).

\[
P_{zd} = \begin{bmatrix} P_{z_1w} & P_{z_1y} \\ P_{z_2w} & P_{z_2y} \end{bmatrix} = \begin{bmatrix} k\alpha_1 W_1 C_1 G_1 F & 0 \\ 0 & 0 \end{bmatrix} \quad (10)
\]

\[
P_{zw} = \begin{bmatrix} P_{z_1w} & P_{z_1y} \\ P_{z_2w} & P_{z_2y} \end{bmatrix} = \begin{bmatrix} \alpha_1 W_1 C_1 G_1 \\ \alpha_2 W_2 \end{bmatrix} \quad (11)
\]

\[
P_{wy} = C_2 G_2 + D_y \quad (12)
\]

\[
P_{uw} = \begin{bmatrix} P_{w_1w} & P_{w_1y} \\ P_{w_2w} & P_{w_2y} \end{bmatrix} = \begin{bmatrix} k\alpha_2 W_1 G_2 F & 1 \end{bmatrix} \quad (13)
\]

Equations (10) – (13) can then be substituted into (6) to yield an explicit expression for \( H_{zd} \).

Designs that minimize various linear combinations of the three structures and active mass driver absolute accelerations were considered. In all of the controller designs considered, the weighting function on the regulated output, \( \alpha_1 W_1 \), and the weighting function on the control force, \( \alpha_2 W_2 \), were constant matrices (i.e., independent of frequency). When included in the design process, the earthquake filter \( F \) was modeled based on the Kanai-Tajimi spectrum.

The loop gain transfer function was examined with various control designs. Here, the loop gain transfer function is defined as transfer function of the system formed by breaking the control loop in the input to the system, as shown in Fig. 5.

Using the plant transfer function given in (12), the loop gain transfer function is given as (14).

\[
H_{loop} = KP_{yw} = K(C_y G_1 + D_y) \quad (14)
\]

By “connecting” the measured outputs of the analytical system model to the inputs of the mathematical representation of the controller, the loop gain transfer function from the actuator command input to the controller command output was calculated.

The loop gain transfer function was used to provide an indication of the closed-loop stability when the controller is implemented on the physical system. For this purpose, the loop gain should be less than one at the higher frequencies where the model poorly represents the structural system (i.e., above 35 Hz). Thus, the magnitude of the loop gain transfer function, at higher frequencies, should roll-off steadily and be well below unity. Herein, a control design was considered to be acceptable for implementation if the magnitude of the loop gain at high frequencies was less than -5 dB at frequencies greater than 35 Hz.

The method of “elimination” was used for the design of the discrete-time controller. Using this technique a continuous-time controller was approximated or “emulated” with an equivalent digital filter using the bilinear transformation. The digital filter was implemented on the in the state space form (15) and (16),

\[
x(kT + T) = Ax(kT) + By(kT) \quad (15)
\]

\[
u(kT) = Cx(kT) + Dy(kT) \quad (16)
\]

where \( y \) represents the vector of measurement sampled inputs to the controller and \( u \) represents the vector of outputs of the digital filter. In the method of emulation, the controller samples the measured outputs of the plant and passes the samples through the digital filter. The output of the digital filter is then passes through a hold device to create a continuous-time signal which becomes the control input to the plant. The series combination of sampler, digital filter and hold device emulates the operation of the continuous-time controller upon which it is based. Usually, within the period of closed-loop system bandwidth, as in this experiment, the discrete equivalent system would adequately represent the behavior of the emulated continuous-time system over the frequency range of interest.

**EXPERIMENTAL RESULTS**

Two series of experimental tests were provided to evaluate the performance of the controllers that were designed. First a broadband signal (0-50 Hz) was used to excite the structure and root mean square responses were calculated. In the second series of the tests an earthquake-type excitation was applied to the structure and peak responses were determined.

The results of two representative control designs are given, where the Controller A was designed by placing an equal weighting on the absolute accelerations of the top two
floors of the structure. The second controller (Controller B) was designed using the same weighting matrix as Controller A, but in addition, with loop shaping techniques to roll-off the control effort at higher frequencies. The analytical transfer functions for these two controllers are compared to the experimentally obtained transfer functions in Fig. 6 and Fig. 7.

![Figure 6: Transfer Function for Controller A](image)

![Figure 7: Transfer Function for Controller B](image)

The performance of each controller was tested by exciting the structure with broadband ground acceleration (0-100Hz) and the response values for the controlled and uncontrolled configurations of the structural system are shown in Table 1.

<table>
<thead>
<tr>
<th>$x_r$</th>
<th>$\ddot{x}_{o_1}$</th>
<th>$\ddot{x}_{o_2}$</th>
<th>$\ddot{x}_{o_3}$</th>
<th>$f_N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>cm</td>
<td>cm/s$^2$</td>
<td>cm/s$^2$</td>
<td>cm/s$^2$</td>
<td></td>
</tr>
<tr>
<td>Uncontrolled</td>
<td>0.693c$^2$</td>
<td>159.4</td>
<td>179.5</td>
<td>232.0</td>
</tr>
<tr>
<td>Zerod-</td>
<td>0.197c$^2$</td>
<td>121.9</td>
<td>137.8</td>
<td>168.5</td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controlled</td>
<td>0.310c$^2$</td>
<td>99.8</td>
<td>78.23</td>
<td>90.4</td>
</tr>
</tbody>
</table>

The results include responses for the relative displacement of the actuator $x_r$, which is also a measure of the first floor relative displacement, the absolute accelerations of the three floors, $\ddot{x}_{o_1}, \ddot{x}_{o_2}, \ddot{x}_{o_3}$, and the applied control force $f$. The zeroed-control case corresponds to the case in which the actuator is attached, but the command signal is set equal to zero (i.e., $u=0$). From the zeroed configuration response it is shown that the “stiffness” of the actuator has a significant effect on the displacement (97.4%) and a moderate effect on the accelerations. Let’s notice that with control, the absolute accelerations of the three floors are reduced by 37.8%, 56.4% and 61.0%, respectively, over the uncontrolled responses, and the first floor displacement is reduced by 95.6%. The controlled responses are achieved by using less force than the zeroed-control case.

Comparison of the uncontrolled, zeroed and controlled transfer functions for the ground acceleration to the first floor absolute acceleration is shown in Fig. 8. Let’s notice that the peaks of the controlled transfer functions from the ground acceleration to the structural responses are significantly smaller those of the zeroed transfer functions. Only the controlled transfer function from the ground acceleration to the actuator displacement is larger in magnitude than the zeroed response, because in the zeroed configuration the actuator attempts to remain in the locked position.

![Figure 8: Comparison of Uncontrolled, Zerod and Controlled Transfer Functions: Ground Acceleration to the First Floor Absolute Acceleration](image)

Comparison of the uncontrolled, zeroed and controlled transfer functions for the ground acceleration to the second floor absolute acceleration is shown in Fig. 9.

![Figure 9: Comparison of Uncontrolled, Zerod and Controlled Transfer Functions: Ground Acceleration to the Second Floor Absolute Acceleration](image)
For the broadband disturbance tests, Table 2 compares the responses for the two controllers, to the uncontrolled responses. The percent reductions are indicated in parentheses. For this experiment, uncontrolled refers to the case in which the active mass driver was attached to the structure and the command signal was set equal to zero.

<table>
<thead>
<tr>
<th>Controller</th>
<th>$\ddot{x}_{n1}$</th>
<th>$\ddot{x}_{n2}$</th>
<th>$\ddot{x}_{n3}$</th>
<th>$\ddot{x}_a$</th>
<th>$\ddot{x}_{cal}$</th>
<th>$u$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncontrolled</td>
<td>159.4</td>
<td>179.5</td>
<td>232.0</td>
<td>-</td>
<td>226.8</td>
<td>-</td>
</tr>
<tr>
<td>Controller A</td>
<td>34.8</td>
<td>36.9</td>
<td>39.6</td>
<td>0.219</td>
<td>606</td>
<td>0.057</td>
</tr>
<tr>
<td>(77.4)</td>
<td>(80.5)</td>
<td>(82.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controller B</td>
<td>42.6</td>
<td>43.5</td>
<td>45.3</td>
<td>0.2424</td>
<td>303</td>
<td>0.055</td>
</tr>
<tr>
<td>(72.7)</td>
<td>(80.5)</td>
<td>(80.1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Both controllers were able to achieve at least an 80% reduction in the third floor absolute acceleration. Let’s notice that Controller A was able to achieve moderately better results than those of Controller B, but the acceleration of the actuator was almost twice that of Controller B while displacements remained approximately the same, as shown in Fig.10. This difference is due to the loop shaping used in design of Controller B, which resulted in less control effort being applied at the higher frequencies.

![Figure 10: Uncontrolled and Controlled Relative Displacements](image)

**CONCLUSIONS**

An approach to modeling of feedback seismic structural control with active mass drivers was developed. Recommended acceleration feedback control strategies were implemented and verified on a three-storey single-bay test structure controlled with an active mass driver. The effects on actuator dynamics and control structure interaction were incorporated into the system identification procedure. Under the broadband excitation the active mass driver controller was able to achieve approximately 80% reduction of acceleration responses and a significant response reduction was achieved in all three modes of the system. When excited with an earthquake disturbance, the peak response reduction of the top floor acceleration was 68%. The obtained results show that acceleration feedback control strategies should be considered as practical and effective for mitigation of structural response due to seismic excitations.

**ACKNOWLEDGEMENTS**

This research is a part of the project TH 1511/05 sponsored by Bulgarian National Science Fund of Ministry of Science, Education and Technologies.

**REFERENCES**


**BIOGRAPHY**

**SVETLA RADEVA**, PhD is Associate Professor at Department of Computer Aided Engineering at University of Architecture Civil Engineering and Geodesy, Sofia, Bulgaria. Her main research interests are Structural Control, Simulation and Modeling of Earthquake Processes, and Seismic Response Reduction; svetla_fce@abv.bg

**IZABELLALOKSHINA**, PhD is Associate Professor of MIS and chair of Management, Marketing and Information Systems Department at SUNY Oneonta, USA. Her main research interests are in Complex System Modeling, Industrial Simulation and Artificial Intelligence; lokshiv@oneonta.edu
FORWAVER – A WAVELET-BASED PREDICTOR FOR NON STATIONARY SIGNALS

Dan Stefanoiu \(^1\), Janetta Culita \(^1\), Florin Ionescu \(^2\)

\(^1\) „Politehnica” University of Bucharest, ROMANIA
Dept. of Automatic Control and Computer Science (www.acs.pub.ro)
\(^2\) University of Applied Sciences in Konstanz, GERMANY
Dept. of Mechatronics (www.htwg-konstanz.de)

E-mails: danny@indinf.pub.ro, jculita@yahoo.com, ionescu@htwg-konstanz.de

KEYWORDS
Prediction, orthogonal wavelet packets, prediction quality.

ABSTRACT
FORWAVER is a simulator that provides high accuracy time series forecasting with the help of wavelet-based models. Within this paper, the models rely on signal bases consisting of Daubechies orthogonal wavelet packets and ARMA modeling. Such a basis is adaptively selected, depending on data, while the ARMA model is obtained by maximizing a cost function, namely the prediction quality. Simulation on real time series have proven that, when comparing to the classical time series model (built by adding the deterministic trend, the seasonal component and the stochastic auto-regressive component) the wavelet-based model is more accurate and less subjective.

1. INTRODUCTION

Prediction of natural phenomena (seismic, climatic, celestial, etc.) is nowadays of great importance, because of unexpected and rapid changes they might exhibit. Many engineering processes also require quality forecasting in order to be modeled, simulated, automatically controlled, etc. Some of such phenomena or processes reveal their behavior through finite length data strings \(\{y(n)\}_{n=1}^{N}\) referred to as time series (ts). Although the governing dynamics of such a ts is unknown, the entity providing the data can be associated to a non stationary model (i.e. with variable frequency contents).

Any ts is subsuming two types of behaviors: deterministic and stochastic. Classically, the deterministic component consists of a trend \(y_t\) and a seasonal (periodic) signal \(y_s\), whereas the stochastic component is an auto-regressive (AR) process \(y_{ar}\) [Söderström, 1989], [Stefanou, 2005]. The classical model is then:

\[
\hat{y} = y_t + y_s + y_{ar}.
\]  

(1)

The trend is a low degree polynomial, estimated by using the Least Squares Method (LSM). The periodic behavior is perhaps the most difficult to decode, although the LSM can be employed as well. Finally, a Levinson-Durbin Algorithm (LDA) returns the stochastic component. The accuracy of ts prediction is sensitive to the accuracy of deterministic component. Expressing the deterministic behavior like in (1), inhibits the modeling of non stationary behavior. Thus, the classical model is quite accurate only in case of quasi-stationary ts. Moreover, the user has to select the optimal period (if any) by considering three sources: the closest average periodic waveform to the ts (in the least squares sense), the maximum point of spectral representation (periodogram) and the preliminary information about the process that generated the data. Therefore, when detecting and selecting the seasonal component, the user’s subjectivism is crucial and can sensibly increase the prediction error.

The non stationary behavior of a ts is more suitably decoded through a time-frequency-scale analysis [Cohen, 1995]. Wavelets are typical instruments of this analysis in context of multi-resolution theory. This theory has mostly been devised within the works of S. Mallat and I. Daubechies, in the late ’80s [Daubechies, 1988], [Mallat, 1989]. (A generalization has been introduced in [Coifman, 1992].) Orthogonal wavelets are employed to build bases within the multi-resolution structure endowing the space of stable, finite energy signals. Such a wavelet basis is uniquely associated to a filter bank that produces a specific sub-band segmentation of ts spectrum. Moreover, the filter bank can be configured according to the energy distribution of ts, both in time and frequency. Two modifications are proposed in this paper: the classical deterministic model of a ts is replaced by a wavelet-based model and the AR modeling is replaced by ARMA modeling (auto-regressive, with moving average).

2. ON THE WAVELET-BASED MODEL

The deterministic modeling with wavelets is actually a problem of data denoising. Thus, the original data are projected on a subspace of signal space, \(\mathcal{U}\), generated by a finite collection of wavelet packets, like in Figure 1.

![Figure 1: Data Denoising Principle](image)

The projection, \(y_w\), is associated not only to the deterministic component of ts, but also to its useful
The overall model is then:

\[ \hat{y} = y_r + y_{\text{ARMA}}. \]  

(2)

The colored noise corrupting the data, \( v = y - y_r \), is orthogonal on the wavelet subspace and, thus, on the useful component \( y_r \), as illustrated in Figure 1. The problem is to build and use the overall model (2). This goal can be reached after completing the following steps: (a) estimate the useful component \( y_r \) according to the data \( y \) and to a given multi-resolution structure; (b) estimate the ARMA (stochastic) component \( y_{\text{ARMA}} \) starting from the colored noise \( v \); (c) predict the ts by means of best predictor. All together constitute the core of FORWAVER simulator. The details of its construction can be found in [Stefanoiu, 2007]. Hereafter, only a short description of the three steps above is provided.

The \( \mathbf{H} \) subspace is generated by using wavelet packets and an entropy based criterion. More specifically, the deterministic (useful) component is:

\[ y_r(t) = \sum_{(m,p) \in \mathbf{H} \times \mathbf{P}} \sum_{n \in \mathbf{N}_{m,p}} c_{m,p,r}(t) \quad \forall t \in \mathbb{R}, \]  

(3)

where \( t \) is the continuous time, \( c_{m,p,r}(t) = \{y_0, \psi_{m,p,r}(t)\} \quad \forall (m,p) \in \mathbf{H} \times \mathbf{P}, \forall n \in \mathbf{N}_{m,p} \) are wavelet coefficients, \( \psi_{m,p,r} \) is a generic wavelet from the current packet and \( y_0 \) is the projection of ts \( y \) on the subspace \( \mathbf{H} \equiv \mathbf{N}_0 \) from the multi-resolution structure [Mallat, 1989]. Coefficients are computed by means of the filter bank corresponding to the multi-resolution structure. The finite sets \( \mathbf{H}, \mathbf{P} \) and \( \mathbf{N}_{m,p} \) determine the structure of such filter bank. Figure 2 displays an example of filter bank (binary tree) together with the frequency segmentation it realizes.

![Wavelet Tree and Frequency Effect](image)

**Figure 2:** A Wavelet Tree and its Frequency Effect

In this example, the set of direct product depth-azimuth is: \( \mathbf{H} \times \mathbf{P} = \{(2,3), (3,1), (3,2), (3,5), (4,0), \ldots, (4,9)\} \) (corresponding to tree leaves), while the time offsets \( \mathbf{N}_{m,p} \) are determined by accounting the supports of ts and basic couple of father-mother wavelets \( \{\phi, \psi\} \) [Daubechies, 1988]. On the tree branches lie mirrored versions of QMF filters \( h \) (low-pass) and \( g \) (high-pass), followed by decimators. Wavelet coefficients are provided by the tree leaves, after entering the first projection coefficients (on time-shifted father wavelets):

\[ c_{m,p}(t) = \{y_0, \phi_{m,p}(t)\} = \sum_{k=1}^{M} y[n]\phi(k-n). \]  

(5)

Interestingly, the filter bank can be employed to compute wavelet values as well, according to the recursive equation below:

\[ \psi_{m,p,r}(t) = \sqrt{2} \sum_{n \in \mathbb{Z}} \gamma_{m,p} \psi_{m-1,p/2,j}(2t-n), \quad \forall t \in \mathbb{R}, \]  

(6)

where \( \gamma = h, \quad \text{if } p \text{ is even}, \quad \text{or } \gamma = g, \quad \text{otherwise.} \) The recursive equation (6) starts from the father wavelet \( \psi_{0,0} \equiv \phi \). The other wavelets of packet \( \{\psi_{m,p,r}\}_{m \in \mathbb{Z}} \) are just time-shifted versions of \( \psi_{m,p,0} \).

The structure of filter bank can adaptively be set, by minimization of wavelet coefficients entropy. By definition, the entropy of a discrete time signal \( x \) is:

\[ \mathbf{H}(x) = -\sum_{x \in \mathbb{X}} [x[n]] \log \frac{x[n]}{\mathbf{E}(x)} \]  

(7)

where \( \mathbf{E}(x) = \sum_{x \in \mathbb{X}} [x[n]]^2 \) is the signal energy. According to Shannon’s interpretation (given in 1948, see [Shannon, 1948]), the entropy (7) quantifies the average number of bits that are necessary to encode a signal sample. Also, the entropy expresses both the signal redundancy and its intrinsic order. The smaller the entropy, the less redundant and the more ordered the signal. Small redundancy allows the signal information to be concentrated in a small number of samples, due to their high intrinsic order. Note that the entropy is not an additive measure, i.e. the entropy of a signal built by concatenation of two or more signals is different, in general, from the sum of each concatenating signal entropy. Finding the tree leaves that yield minimum entropy is a problem that can be solved by using IDA* from Artificial Intelligence [Russel, 1995].

Hence, the problem of minimum entropy wavelet basis selection is equivalent to searching for the optimum within a tree endowed with costs on the arcs. IDA* strategy relies on a tree pruning mechanism that allows the optimum path to be found very efficiently. In case of QMF banks and the signal entropy, the search of minimum entropy configuration of associated binary tree must be adapted to the context of IDA*. Thus, the search is realized within a meta-tree defined as follows (see Figure 3): (a) meta-nodes are associated to a binary tree shape (or boundary); (b) a meta-node is generated by expanding a binary tree node up to its two children; (c) the meta-node label is the string of indexes associated to binary tree leaves; (d) the cost of a transition from a meta-node to one of its children is the entropy of the wavelet coefficients lying on the boundary of the QMF bank associated to the meta-child; (e) the estimated cost is null. The optimal filter bank corresponds to a meta-leaf for which the total entropy on the path from root is minimal.

The final version of \( y_r \) is obtained only after removing the weak wavelet coefficients from representation (3). Thus, for any \( (m,p) \in \mathbf{H} \times \mathbf{P} \), all wavelet coefficients whose amplitudes are inferior to \( \mu \mathbf{E}(c_{m,p}) \) (where \( \mu \geq 0 \) is the masking threshold) have to be removed (or masked).
then the accuracy of each predicted value is inversely proportional to its variance, defined as below:

\[ \hat{\lambda}_k = \lambda^2 (\hat{b}_k^2 + \hat{b}_k^2 + \ldots + \hat{b}_k^2), \quad \forall k \in 1, K. \]  

(11)

By assembling equations (3) and (9), the overall prediction model is then:

\[
jy[N+k] = \sum_{(m,p) \in \Phi} \sum_{\psi \in \psi_{m,p}} c_{n,p} \psi_{m,p} (N+k) + \hat{\bar{y}}_y[N+k], \quad \forall k \in 1, K. 
\]  

(12)

(The accuracy of predicted values (12) is given by (11).)

Usually, the accuracy of predicted values is represented by means of the confidence intervals:

\[ I_k = [jy[N+k] - \beta \lambda_k, jy[N+k] + \beta \lambda_k], \forall k \in 1, K. \]  

(13)

In case of normally distributed noises, in definition (13), \( \beta = 3 \) and the true value \( y[N+k] \) belongs to \( I_k \) with a probability superior to 95%.

Several predictors are available, by varying the structural indices \( na \) and \( nb \). In order to select the best one, the prediction quality criterion has to be employed. This cost function has been introduced in [Stefanou, 2007] with the following definition (for a \( K \)-length prediction horizon):

\[
PQ[K] = \frac{100}{\sum_{k=1}^{K} \hat{\lambda}_k} \left[ \frac{1}{N+1} \sum_{k=1}^{N+1} [y[N+k] - jy[N+k]] \right] \% , \]  

(14)

where \( \sigma_y \) is the standard deviation of \( y \). According to criterion (14), the predictor improves its performance when the confidence tube tightens and/or the predicted values are closer to the true ones. The only problem is to implement definition (14), because the true values are unknown on the prediction horizon. Clearly, the only way is to preserve the last \( K \) data of \( y \), as if they were unknown. The wavelet-based ARMA models are then built by using \( N-K \) data only. The other \( K \) data are used to select the best predictor, by maximizing PQ surface, as function of \( na \) and \( nb \).

3. SIMULATION RESULTS AND DISCUSSION

The FORWAVER simulator has been implemented within MATLAB 7.04 environment. Some parameters have to be set in advance, in order to tune the simulation, such as: the support length of father wavelet, \( 2 \text{N}_\phi \) (which leads to the analysis wavelets \{0, \psi\}, according to Daubechies’ algorithm [Daubechies, 1988]), the stop test for IDA* procedure \( \varepsilon > \) 0) and the masking threshold for awavelet coefficients \( \mu \in (0,1) \). Usually, \( N_\phi \in [2, 75] \) and \( \varepsilon = 10^{-7} \), whereas the masking threshold can be increased up to 10%. A technical parameter is also employed: the resolution index \( L \in \mathbb{N} \) of father wavelet. Thus the projection (5) is computed by using values of wavelet father, which are sampled with period \( 10^{-L} \). Usually, \( L \in [0, 6] \). If \( L > 0 \), the wavelet family of expansion (3) becomes a tight frame of signals space; its
constant is determined by means of LSM (see [Stefanoiu, 2007] for details).

The predictor has been tested on a collection of 15 ts extracted from various natural phenomena. Note that all the signals are genuine and not artificially generated, as artifacts. The following pictures exhibit the performance of classical and wavelet-based predictors for one of them, representing the evolution of USD-EURO exchange rate, since the formal introduction of European currency. In order to compare the two predictors, the last \( K = 5 \) data has completely been removed from the models construction. From the \( N - K \) remaining data, another \( K \) (the most recent ones) have been removed, such that the optimal predictors be found. Thus, the models are actually built by using the oldest \( N - 2K \) data.

Figure 4 displays the ts variation and the corresponding trend, whereas Figure 5 reveals the performance of classical best predictor.

Figure 4: USD-EURO Exchange Rate

The PQ criterion has been employed to find the best polynomial degree of trend and AR order. One can see that the value of evaluated PQ on the last \( K \) data (which did not partake to the model construction or selection) is quite modest: 73.49\%. Thereby, Figure 5 shows that the measured data hardly fall inside the confidence tube.

Figure 5: Performance of Classical Predictor

Figure 6 exhibits the performance of wavelet based predictor.

Figure 6: Performance of Wavelet-Based Predictor

The improvement is quite obvious, since PQ jumped to 93.86\% and the predicted values are close enough to the measured ones, even for far away prediction instants.

The father and mother wavelets selected by the simulator after varying \( N_{s} \) from 2 to 25 are illustrated in Figure 7. Wavelets correspond to the maximum of \( N_{s} \) in this case (i.e. 25). Also, the resolution index \( L \) has been varied in order to reach for the optimal predictor, in PQ sense. One has been found that the optimal value is 6 (which generated a very dense tight frame).

Figure 7: Optimal Wavelets Selected by FORWAVER

Selection has been made by using the maxima of PQ surface, as shown in Figure 8. One can see that the overall maximum of PQ is 90.3\%, for \( na = 11 \) and \( nc = 2 \). Interestingly, in this case, the effective PQ is higher (93.86\%). (Usually, the PQ decreases after considering the last \( K \) data.)

In Figure 9, the structure of filter bank corresponding to minimum entropy of wavelet coefficients is displayed. Beside the fact the structure is quite simple and effective, it reveals an interesting phenomenon: the information about prediction is seemingly concentrated into the middle frequency sub-bands for this ts (recall the time-frequency...
correspondence of Figure 2). This characteristic is confirmed by the scalogram represented in Figure 10.

Figure 8: Prediction Quality Surface and Optimal Indices

Figure 9: Optimal Filter Bank Selected by IDA* procedure

Figure 10: Time Series Scalogram

It would be dishonest to claim that the performance of wavelet-based predictor is by far superior to the one exhibited by the classical predictor for all 15 tested ts. For some ts, PQ gained only few percents. This is the case of ts for which the seasonal component is quite easy to determine by classical methods. It seems that wavelets tremendously improve the classical predictor performance when the seasonal component is difficult to detect and/or derive or the ts does not exhibit periodic variations (like the one taken as example in this paper). While the wavelet-based model can naturally decode the repeatable patterns carried by the ts, the seasonal component is critical for classical prediction (and requires user’s intervention). In general, the wavelet-based predictor is very likely superior to the classical one, at the expense of complexity increase.

4. CONCLUSION

This paper integrated two types of signal modeling in order to perform data compression and especially accurate prediction. The deterministic model is concerned with wavelets and adaptive frequency sub-band configurations, while the stochastic model is of ARMA class. The combination led to an overall prediction model with higher prediction accuracy, comparing to the classical model, on all tested ts. Future improvements are concerned with replacing the orthogonal wavelets by a dictionary of waveforms, from which the most fitted ones have to be selected. Removing the orthogonal nature of wavelets might produce superior performance in terms of prediction.

REFERENCES


ACKNOWLEDGMENTS

This research has been supported by Alexander von Humbold Foundation (Germany) and, subsequently, by the National Center for Programs Management (Romania). The authors are extremely grateful to both of them.
TEXTILE SIMULATION
MODELLING AND SIMULATION OF TEXTILE PROCESSES
Modelling and simulation of an active anti-ulcer mattress system

Jalloul Ellehri
François Boussu
Vladan Koncar

Univ. Lille North of France, F-59100, ROUBAIX, ENSAIT, GEMTEX
E-mail : jallou.elfehri@ensait.fr

INTRODUCTION

The involuntary behaviour of a human being is amazing. Even while unconscious, the human brain continues to handle various operations, such as a heart rate control, breathing, etc. However, failures are possible, due to accidental or natural causes. The ulcer is a direct example of consequences of these defects.

Having regarded the aging population, the ulcer prevention has become a major concern of hospitals and other long-stay facilities.

The goal of our study is to design a comfortable bed which will prevent the ulcer forming in an effective and autonomous way.

Ulcer Definition

An ulcer is a cutaneous lesion of ischemic origin related to a compression of soft tissue between a hard plan and the osseous projections (National Presses Ulcer Advisory Panel 1989).

STATE OF THE ART

It is possible to find in the market different kinds of anti ulcer bed classified under two categories (www.escarre.fr or 13th Talks of the Garches Institute).

Static support, where:

- bed supports are from foams or gel and hollow silicone fibres,
- a constant pressure or water mattresses.

Dynamic supports with:

- alternate pressure supports,
- air loss supports,
- flotation bed supports.

Figure 1 Foam support

Figure 2 Air loss support

All these solutions do not take into account the differences among the various parts of the body regarding the possibility of ulcer formation. For instance, the most exposed areas are those where the zones exposed to a high pressure, as buttocks, heels and head, are located. Aforementioned beds distribute the pressure in uniform or
alternate way, whereas the human body is not uniform and cannot withstand the same pressure on different parts. The solution presented in this article is adapted to the complexity of the human body by using the non uniform distribution of the pressure.

ANTI-ULCER BED WITH_THRESHOLDS VALUES

Presentation

The system is composed of a pressure sensors matrix and a device with motorized parts making possible the pressure modification in various zones (Figures 3 and 4).

The top layer is made of comfortable textiles based on non-woven or 3D woven structures, in order to minimize the effect of moving parts.

The matrix of sensors may be assimilated to a CCD matrix; with analogy a pressure sensor output is treated as a pixel with different levels of gray. This approach enables the utilization of the classification and decision theory dedicated to the image processing (POSTAIRE 1987 or JOLION 2001 or BUREL 2001).

![Figure 3 Model Diagram](image)

Initially, a matrix of 10 rows and 5 columns (20 cm per 20 cm by cell) has been defined. Each cell is controlled by a motorized bed part. However, after considering the proportions of the human body, this number of sensors was not sufficient to give relevant information about the body position. For this reason this number has been increased to 20 rows and 10 columns, but without modifying the number of actuators. Therefore, each motorized part coincides with a zone of 4 sensors.

The matrix of sensors enables to make a classification and to carry out a form reconstruction in order to find the occupied zones (by the patient) and to define her or his position. In the second time, a threshold value of pressure is assigned to each part of the body. This value depends on the morphology of the patient and his health condition (dermatological issues...) (www.escarre.fr)

In order to generate a control strategy able to maintain the pressures on the body under those threshold boundary values, it is necessary to have tools to predict the evolution of the pressure in real time, if one acts on the system.

For this purpose, the finite element model has been stressed in the first time. The complexity of this method and the necessary computational load reoriented our researches toward a simpler approach.

Human model

The human body can be compared to a whole of weights which are bounded by mechanical joints. Generally, they are ball pivot joint or pivot joint. When the torsors of forces transmission for these joints are observed, the possible force transfer may be envisaged. Thus, a human body may be decomposed to independent parts. This theory is valid only for small changes of posture where the skin does not support a force transfer (COLOMBARI & GIRAUD 1998).

To implement this model, data are required, as the distribution of the weights in the body: weight of the head, chest, hands, legs...

If this distribution is not known for a patient, the system must be able to find an acceptable distribution.
Masses distribution

Initially a Principal Components Analysis (PCA) is realised with the maximum of data. Inputs are defined as follows: the height, the weight, the age and the sex. Then, the distribution of body masses is computed. The PCA enables to eliminate the redundancies and the elements which do not influence the global system behaviour. Finally, the weight and the height are selected as prevalent inputs. The fuzzy logic tools are then used to compute the masses of different parts of the body that are considered as outputs. It is important to notice that each part of the body has its own learning function, excepting symmetrical elements (hands and legs).

The problem which arises is that it is necessary to carry out boundaries between different learning functions, because these functions give independent result whereas the sum of the body parts masses gives the total mass (weight). Hence a forced function which readjusts the results has been added.

Example for the hand

Fuzzification DATA MATLAB RESULT

![Figure 5 Weight Input](image)

Thin (sure between 0 and 50, fuzzy between 50 and 65)
Normal (sure between 65 and 90, fuzzy between 50 and 65 or 90 and 100)
Fat (sure between 100 and 150, fuzzy between 90 and 100)

![Figure 6 Length Input](image)

Short (sure between 0.8 and 1.5, fuzzy between 1.5 and 1.7)
Normal (sure between 1.7 and 1.8, fuzzy between 1.5 and 1.7 or 1.8 and 1.9)

Long (sure between 1.9 and 2, fuzzy between 1.8 and 1.9)

![Figure 7 Hand-Weight Output](image)

Small (sure between 0 and 1, fuzzy between 1 and 2)
Normal (sure between 2 and 3, fuzzy between 1 and 2 or 3 and 4)
Big (sure between 4 and 6, fuzzy between 3 and 4)

Results

For total weight of 95.6 Kg and the height 1.82 m the hand weight will be 3.5 Kg.

Effort transfer

The human body is a complex structure so that it is difficult to determine with accuracy the stress and energy transfers. For the anti ulcer application the exact data are not required, but rather a tendency of evolution and if possible an order of magnitude.

![Figure 8 Mechanical simplified model of a human body](image)

In a ball pivot joint there is no transmissible torque. For the pivot joint, there are 2 transmissible torques. On a laying
position as there are no transmissible stresses, it is possible to separate the body part between 2 joints.

After taking each part alone, the simple law of physics may be applied taking a body form and a weight in order to compute the distribution of pressures generated by the system body - mattress.

On the human body the support points (between the body and the mattress) are in general those with the greatest pressure which implies a non linear pressure distribution.

Because of this positive weighting factors have been assigned to the support points. These positive weighting factors are determined from the cartography of the pressure obtained when the patient lies on the bed, by sensor matrix. Then the system re computes the positive weighting factors for different support points.

Example

Let consider a matrix of pressures \( P(i, j, k) \), for an area of “L” rows and “C” columns where \( i \) the row number, \( j \) the column number and \( k \) the iteration index.

\[ S = \text{surface of the cell } (i, j); \]

\[ Pd = \text{weight generated by the body part, on the occupied area}; \]

\( P(i, j, k) \) is a three-dimensional matrix which size is defined by \([L, C, K]\) where \( L \) and \( C \) denote the dimensions of the area and \( K \) the total number of iterations necessary to compute a final \( P \) matrix.

\( k = 1 \) is the initial state.

\[ \sum_{i=1}^{L} \sum_{j=1}^{C} P(i, j, k) = \frac{Pd}{S} \]

Then the positive weighting factor \( \alpha(i, j, k) \) is defined as:

\[ P(i, j, k) = \alpha(i, j, k) \times \frac{Pd}{S} \]

\[ \sum_{i=1}^{L} \sum_{j=1}^{C} \alpha(i, j, k) = 1 \]

At the initial state \( (k = 1) \), \( P(i, j, k), S \) and \( Pd \) are known, then:

\[ \alpha(i, j, k) = \frac{P(i, j, k) \times S}{Pd} \]

Simulation and iteration

The control strategy is described in this section. The necessary adjustment of actuator positions (motorized bed cells) will be obtained by computing the pressure evolutions from \( \alpha(i, j, k) \), \( P(i, j, k) \) and the \( PS(i, j) \) where \( PS \) is the thresholds matrix defined in function of the patient, health condition.

The control algorithm is given below:

From \( k = 1 \),

The first \( P(i, j, k) > PS(i, j) \) is detected and its \( i \) and \( j \) values are recorded as \( id \) and \( jd \),

Then we set \( P(id, jd, k + 1) = PS(id, jd) \) and other \( P(i, j, k + 1) \) with \( (i \neq id \text{ and } j \neq jd) \) are computed.

Then the pressure differences and transitional positive weighting factors are determined

\[ \Delta P = P(id, jd, k) - P(id, jd, k + 1) \]

\[ \alpha(i, j, k + 1) = \alpha(i, j, k) + \frac{\alpha(id, jd, k)}{(L \times C) - 1} \]

\[ P(i, j, k + 1) = P(i, j, k) + (\alpha(i, j, k + 1) \times \Delta P) \]

Where \( \alpha(i,j,k+1) \) is a transitional positive weighting factor allowing the computation of \( P(i, j, k + 1) \).

The previous procedure is repeated until obtaining the condition of pressures regulation for \( k = K \).

\[ P(i, j, K) \leq 1.05 \times PS(i, j), \forall i,j \]

5% of overshoot pressure is allowed to improve the efficiency of the algorithm.

Control

When the final sequence of matrix \( P(i, j, k) \) for \( k \in [1, ..., K] \) and \( i \in [1, ..., L], j \in [1, ..., C] \) is computed (the condition \( P(i, j, K) \leq PS(i, j), \forall i,j \) should be satisfied), this sequence is used to determine the new positions of actuators that are motorized bed parts.

To calculate these positions, the effect of a foam memory has also to be taken into account. In fact, the presence of a memory effect is always accompanied by a hysteresis.

The procedure of the position determining for the \( i,j \) motorized part is given below.

Figure 9 shows the hysteresis where positions are on x axis and pressures on y axis. Therefore, if the motorized part is moving from the position Pos1 to Pos2 and the back to Pos1 the pressure corresponding to the position Pos1 will be different from those recorded at the beginning.
Figure 9  Foam Hysteresis

To calculate the references positions \((i, j)\), the procedure is defined as follows:

- The previous positions will be considered as initial ones.

- The hysteresis diagram is used to compute the new position from the necessary pressure modification \(K\) times, using the sequence of matrix \(P(i, j, k)\) for \(k \in [1, \ldots, K]\) and \(i \in [1, \ldots, L]\), \(j \in [1, \ldots, C]\) computed before. Directions of all the actuator movements from the previous regulation are taken into account during the computation of their new positions because of the foam “memory” effect.

- Final positions of motorized parts are then applied and the pressures measured by the sensor matrix are compared with the threshold matrix \(PS\) in order to validate the control sequence obtained by simulation.

- If the result is not correct the new simulation in order to recompute the actuators positions is started.

**Regulation Example**

From an initial state with overshoot pressure, we apply rules to obtain a stabilised state.

For simulation we use Simulink, to build our model. (At this time we use just an area of 2 rows and 4 columns for simulation).

For \(k = 1\)

\[
P(:, :, k) = \begin{bmatrix} 3 & 3 & 2 & 3 \\ 3 & 6 & 3 & 3 \end{bmatrix}
\]

\[
PS = \begin{bmatrix} 3 & 6 & 3 & 3 \\ 3 & 3 & 3 & 2 \end{bmatrix}
\]

We apply the algorithm, until \(P(i, j) \leq 1.05 \times PS(i, j)\).

\[
P(:, :, 1) = \begin{bmatrix} 3 & 3 & 2 & 3 \\ 3 & 6 & 3 & 3 \end{bmatrix}
\]

\[
P(:, :, 2) = \begin{bmatrix} 3.4451 & 3.4451 & 2.3297 & 3.4451 \\ 3.4451 & 3 & 3.4451 & 3.4451 \end{bmatrix}
\]

\[
P(:, :, 3) = \begin{bmatrix} 3.0000 & 3.5125 & 2.3780 & 3.5125 \\ 3.5125 & 3.0598 & 33.5125 & 33.5125 \end{bmatrix}
\]

At the end \(k = 34\), and then \(K = 34\).

\[
P(:, :, 34) = \begin{bmatrix} 3.0224 & 5.5345 & 3.082 & 3.1385 \\ 3.1146 & 3.0632 & 3.0426 & 2.000 \end{bmatrix}
\]

Now the motorized parts positions are computed.

\[
Pos = \begin{bmatrix} 3.7724 & 5.5173 & 3.8342 & 3.8885 \\ 3.8646 & 3.8132 & 3.7926 & 0 \end{bmatrix}
\]

These positions are as motorized parts setpoint.

**CONCLUSION**

Ulcer prevention requires intelligent and active system. To be efficient this system has to adapt to the patient in real time. The human model made of articulated parts has been used to fulfil this objective.

At this stage, the entire human body has not yet be simulated. Therefore, our future work will be focused on the generalization of the model.

Finally, a prototype will be realized in order to validate the control algorithm.

Similar approach may be used in other applications as automotive interior (car seats), planes, trains ...

**REFERENCES**

National Presses Ulcer Advisory Panel, 1989

www.escarre.fr, Specialized website in France with regard to Ulcer

The Ulcer Acts of the 13th Talks of the Garches Institute

Jack-Gérard POSTAIRE. « From Image to decision « analyse des images numériques et théorie de la décision ». Dunod Informatique, 1987

Jean Michel JOLION, The Vision Systems. HERMES, 2001
Gilles BUREL, “Introduction to the image treatment «simulation under Matlab». HERMES, 2001

Gerard COLOMBARI, and Jacques GIRAUD, “Industrial sciences (Classes preparing students for university entry 2nd year). FOUCHER, 1998


BIOGRAPHIES

Jalloul ELFEHRI is preparing his PhD at ENSAIT Roubaix in France since 2006. He is doing her research in the field of Modelling and simulation of an active anti-ulcer mattress system. He participates actively to collaborations with industrial partners specialized in textile devices for medical applications.

François BOUSSU is Associate Professor at ENSAIT, Roubaix in France. He is teaching in the field of weaving and composite structures. He is doing his research in the field of multifunctional textiles, ballistic applications and special 3D woven structures. He is also Director of external affairs of ENSAIT and is managing industrial partnerships.

Vladan KONCAR is Professor at ENSAIT. He is teaching in the field of virtual reality, computer networks, and communicative & intelligent textile structures. He is doing his research in the field intelligent textiles. V. KONCAR is Scientific Director in charge of International Development of ENSAIT and since 2007 President of AUTEX, Association of Universities for Textiles.
MODELING THE FIBER-YARN INTERFACE FOR CLOTHING DESIGN

Ionut Neagu  
Pascal Bruniaux  
Xavier Legrand  
ENSAIT " L'ecole Nationale Supérieure des Arts et Industries Textiles "  
2, allée Louise et Victor Champier, Roubaix, 59056, France.  
ionut.neagu@ensait.fr

Dorin Avram  
Technical university "Gh. Asachi"  
Iasi, Roumania  
dorin.avram@tuiasi.ro

KEYWORDS  
Fibre-yarn, simulation interfaces, model design, virtual reality, Computer Aided Design.

ABSTRACT

The goal of this study is to present the methodology, in order to develop a three-dimensional environment for the virtual design and the simulation of clothing, the emphasis being on the need to model a yarn structure before modeling the fabric. To integrate the model of the yarn in the fabric model, we must define a geometry adapted to the yarn and describe the dynamic behavior of the yarn model. For that, an algorithm of modeling the fibers in 3D must be developed under various mechanical constraints. These requests produce very complex states of deformation, on a fabric scale, an infinite number of varieties of structure combination for fiber-yarn-fabric. Those generate disturbances on the nature of fiber, the structure of the yarn, its three-dimensional organization, the torsion of the yarn and the fabric. The geometrical complexity of the various structures and the presence of a hierarchical levels of the structure lead to a high complexity of the predictive models, of which degrees of approximation from the lowest level at the highest level, this can lead to a hierarchical management of the forecast uncertainty, not to forget the accumulated errors from one hierarchical level to another. The fundamental need is to be able to accurately predict the behavior of the final structure based on the available physical behavior and the mechanical properties under various loads applied to the range of structures. Beyond the scientific character of the forecast, one can use the same computerized base to virtually simulate the tests and laboratory experiments or to generate realistic animations in artistic matter and (design of textile clothing).

ORIGINALITY, POSITION

The two great problems which contribute to the difficulty in modeling textiles are the geometrical complexity of the majority of the textile structures and the anisotropy, also the non-linear behavior of much of the fibers which one finds in the yarn (Ngo Ngoc 2003). The response of fibers solicited to a mechanical load is guided by nature of fibers and with their interactions. Beyond this observation, there was and there will always be a great difficulty to characterize the physical behavior of small fibers, often of irregular form. In general, the characterization of textile materials follow a hierarchical strategy around a multidimensional and structural concept leading to successive representations of geometrical models (Bruniaux et al 2006a; Legrand et al 2000b; Legrand 2000; Ngo Ngoc et al 2002b), associated to the mechanical models for the description of the behavior (Ngo Ngoc 2003; Legrand et al 2001). The geometrical complexity of the various structures and the presence of a hierarchical levels in structure and scaling of observation (fibers : 10-5 m, yarn: 10-3 m , fabrics: 10-1 m, composite materials: multi-layer) (Bruniaux et al 2002) lead to a high complexity of the predictive models, from here the degree of different approximation from the lowest level at the highest level, leading to a hierarchical management of the uncertainty in the forecast, not to forget the errors that are accumulated from one hierarchical level to another. The fundamental need is to be able to accurately model the final behavior of the structure based on the physical behavior and the mechanical properties under various loads applied of the range of structure available. Beyond the scientific character of the forecast, one can use the same computerized base to virtually simulate the tests and experiments of laboratory (Legrand et al 2002; Bruniaux and Castelain 2003; Ngo Ngoc et al 2002a). Consequently, to model and predict the various behaviors in the textile chain of creation by the data-processing tool provide a saving of time and considerable labor. The simulations of the laboratory tests of textile products for the virtualization in the clothing sector can lead to improvements in products, lead to higher rates of success or reduce the quantity of unsold products, and raise the sales turnover. With the computer graphics, it is possible to implement a virtual chain of comparative product tests using a specific quality standard based from the obtained simulation. The presence of effective virtual technologies in the textile and clothing industry (Figure 1) can provide real benefits because the speed of the conception-production chain can be accelerated.

The advantages from which the companies could profit from the deployment on an industrial scale of virtual technologies are as follows: By using the tools 3D, the companies can reduce the hour consumed by the process of product development. This alternatively has like consequence the reduction of the costs for raw material. The progresses in computers technology, especially in power and speed, make it possible at this time for the development phase and the virtual tools to improve the technical quality of the simulation. Integrating the virtual technologies will make it possible for the companies to produce products more quickly, by eliminating the need for building the prototypes, the samples of the market and perhaps even the products themselves before they are sold. This would have as a consequence the reduction of the time of arrival on the market and would allow a process to be practically integrated in to the development and to be adaptable to the geographical distances.
The mechanical model of the yarn describe the integration of the yarn structure in a three-dimensional model is a field of investigation where many research were undertaken. Why such an effort? The industrial sector of development for physical products requires many tests and repetitive tests, expensive in time and money but essential to validate the quality and the certification of the product. However none of this work is really usable in a textile industrial context. Indeed, the majority of the manufacturers would like to visualize and test their products by using computer graphics techniques. But none of the existing physical models for image synthesis doesn’t takes account of the fibers parameters that one employs to design the yarn and fabric into final products. The most interesting approach was made by Adabala in which the concept of micro-scale geometry represents the subject of a study to model of yarn made up from fiber assembly. In order to visualize the close-up aspect of the yarn, Adabala and Thalmann have presented a technique to create in procedural form the texture and torsions of the yarn. The yarn torsional degree and the fiber roughness are the principal parameters that were presented. The geometrical model presenting the torsion is simulated with the help of a goniometrical function. The textures are then combined with parameters of color; employees can create models of fabric with objective to actually simulating the yarn virtually. Adabala and the others widened their work to produce textured textiles starting from the parameters of weaving machines.

**METHODOLOGY FOR SIMULATION**

The methods and technologies for virtually 3D simulating the textile products must create the footbridge and link the abstract and physical barrier data exchange, between the various sectors of creation, the design, the communication and manufacture. They are essential nowadays, following the phenomenon of delocalization, in a logistic chain of provisioning for textiles. This study is in upstream of the widened concept at the unit of the textile and will aim to develop methods and techniques to define, characterize and evaluate the structure of the yarn in a 3D virtual environment. The yarn for technical use must be adapted to its purpose, where the efforts generated in its structure modify its dynamic behavior by the deformation. For this, it is necessary to anticipate and model the nonlinear behavior of the yarn at the time of its elongation starting from the deformation of the 3D structure and the characterization of fibers. The characterization of the product is carried out in privileged directions and the transfer of the mechanical properties is directly integrated in the components of the substructures. For several years, many scientists (university and industrial) have worked on the modeling of the textile structures by regarding this material as a deformable surface (Bruniaux et al 2006a; LeGrand 2000; LeGrand, et al 2000b). That led for certain developed of CAD software for clothing; most known in the industrial sector are Optitex (Israel), Modaris 3D (Lectra, France), Browzwear-VStitcher (America), TPC (Honk-Kong), and in the university laboratories (INRIA, MIRALAB). But it proves that this work represents only one link of the new chain of creation and development of clothing. Upstream models of the manufacturing process have a direct interest to simulate the very expensive mechanical tests which provide the entry parameters of the simulated yarn or fabric (Ngo Ngoc 2003).
This project thus positions upstream European projects on the modeling of clothing by the use of an interactive data base parameters-fabric with the manufacturing process. But its other advantage is to go further in the knowledge of material and its dynamic behavior. Today the clothing industry is not completely satisfied with the current simulators (good overall, but lack of realism in detail). The current strategies for modeling without taking into account the structural phenomena and some nonlinearity characteristics unfortunately damage the studies undertaken on the fabric for 3d clothing simulation.

**APPROACH**

We intend to build the textile structures starting from the setting in cascade of the different interfaces under behavior models while starting with the study of the interface fiber-yarn with input data as fiber properties, after that the two interface yarn-fabric (in progress with the Lectra company, confidential) and fabric-product as in figure 1 by defining structural and geometrical models (fabric, woven). This approach is new and original because it has as starting point the mechanical properties of the fibers to represent a fabric or a textile product, whereas preceding work started only with the properties the yarn or fabric. This vision implies the creation of a database containing all information necessary to characterize the state of fibers /filaments used to design the yarn, and the installation of geometrical models defining the structural state of woven or knitted materials. The simulation leads us to incorporate the mechanical properties of the yarn starting from fiber in order to have a field of validity and very broad application. The interest of this multi interface concept starting from knowledge of the raw material makes it possible to observe, control and adapt the product according to the scheduled conditions and this to any stage in the manufacturing process. Thus, by this principle, it is possible to shorten the design and the time of production, to act on the implementation costs significantly and controlled. One can imagine that we can test new virtual products beyond some limit that it would not be possible to really make with the textile industry nowadays. For example, it is less expensive to simulate a virtual machine of tests. In the same way, one can also imagine a very complex structure, whose realization would be impossible in its entirety in a laboratory, obtained via the structural simulator. In other words, the principal profits for textile industry will be:

- To widen the possibilities of development of 3D products;
- To design more innovating products for a greater competitiveness;
- To bring value added and expertise to the customers;
- To increase the product quality and the delivery period with cost control optimization.

Many work of investigation were completed in this field by using modern methods of modeling which are the discrete elements. The interest of these techniques in comparison with the finite element method more traditional lies in particular in their capacities to integrate into nonlinear structures. Indeed, the textiles are strongly nonlinear as well by their laws of behavior as by their geometries (Bruniaux et al 2006b). The work of the thesis carried out at the ENSAIT made it possible to set up one of the links of the new modeling textile perception (Ngo Ngoc 2003; Legrand et al 2000b). The originality of the model developed by the team of the GEMTEX is based on a fabric/scaling yarn (Legrand et al 2001) which requires setup adequate measures of measurement. Inter yarn friction is for example an important parameter to take into account for modeling (Ngo Ngoc et al 2002a). It thus was the subject of a very detailed study made in 2002. This made it possible to define the intermediate results of the total concept from figure 1. Moreover, the model being very greedy in computing times, the modeled sample is of reduced size. Also, it required work of homogenization to allow the use of model in industry. It is in this spirit of improvement that we wish to supplement the upstream of this research while working on the modeling the interface fiber/yarn. This model must be imperatively connected with other work on the yarn/fabric interface. This interface represents the logical continuity of the studies on the dynamic modeling of the textile structures carried out with the GEMTEX for several years. An industrial validation of this study is considered with the department of research from company Lectra which works lies downstream from our aims. A scientific collaboration is already in considered. The goal of this study is to sensitize itself in a first approach with the influence of each parameter on the behavior of the yarn model. For that, starting from a first vector of parameters, each component of this vector is separately changed in order to observe and to quantify the relative variation of the outputted model of yarn (dynamic behavior of the yarn under specific requests) according to each variation of the fiber entry (length, title, torsion, techniques of spinning ...). Then, the method most adapted to simulate the distribution of fibers in the yarn is based on the observation of the transversal sections of a yarn associated with a 3D rebuilding of the yarn structure. The distribution function used in the 3D geometrical model must be sufficient to represents structure and integrate with realism the mechanical properties of fibers. The objective is to position with regular intervals the yarns cross section over its length. Each section is turned of a certain times relatively compared to the preceding one, to take account of the yarn torsion and the fibers ideal migration parameters (Avram 2004). To respect the fibers alignment in order to allow the reconstitution of fiber curves, defining the median directions of fibers, it is necessary to optimize the difference between two successive cross sections. Moreover, the generation of these direct curves also controlled by the torsion of each fiber in the yarn is determined by a configuration of central line to generate the fibers model. The simulated yarn structure employing the previously describes algorithm can model a very large variety of yarn structured surfaces containing fibers and bring an improved visual representation. Lastly, it is important to analyze the structures diversity which a yarn according to the fibers organization can have. For that, two base forms can describe the circular arrangement of fibers. If this arrangement is open, we can consider that the radial dispersion of fibers in the yarn is carried out in the shape of various successive layers.
This can be symbolized by concentric circles in which the fibers are organized circularly. The first layer represents the heart of the yarn. The second is pressed on this core and surrounds it by six fibers in contact. In this same state of mind, the third layer has twelve arranged fibers so that they touch initially the circle which surrounds the second layer. Additional layers are then added until the respect the yarn count (Hearle et al 1969; Hickie et al 1974; Keeffe et al 1994; Langenhoove 1997). Another idealized arrangement of fibers is to be taken into account, i.e. small arrangement (Jiang et al 2005; Keeffe et al 1992). In this case, the fibers organization is around a fiber and takes a hexagonal configuration leading to the tightening of peripheral fibers on this core. The interest of this geometry results in a permanent contact of fibers between them. Since the number of fibers in the cuts increase quickly according to the number of layers, contours tend to become more complicated and deviate from the desired hexagonal form (Lomov et al 2001; Lawrence 2003). Also, a model of these two can then be planned to improve dispersion of the fibers (Djaja et al 1992; Grishanov et al 1997).

CONCLUSIONS

Today, in spite of the significant number of work in the textiles modeling field, there is not any study showing feasibility to simulate the dynamic behavior of a yarn starting from the fiber characteristics. More precisely, the modeling of a yarn and the 3D fibers representation, the mechanical and physical characteristics and the integration into a yarn model, is still at the stage of fiction. The principal objective of this thesis is to supplement the research completed in the textile field, by widening the studied sector significantly and by including new mythologies of research. This can be carried out by the widened vision described by the conceptual diagram of figure 1. It shows that the setting in coherent cascade of the various models of interface fiber-yarn, yarn-fabric and fabric-clothing leads to a complete analysis of the textiles. This concept avoids the technological fractures which one at the present time finds at the various stages of the product: spinning mill, weaving, clothes industry. The team of the GEMTEX working on the modeling of the textile structures has knowledge to recognize the problems of interface yarn-fabric and fabric-clothing, but it remains to supplement the upstream of this multi-model structure, i.e. the interface fiber-yarn, to control the whole modeling domain. This can be articulated around three large axes which can be arranged around the conceptual diagram of figure 2.

The first axis treats geometrical modeling of yarn by taking account of and characteristic of the fibers structural parameters to carry out a virtual 3D (Ghith et al 1997; Ghith et al 1998). Methodology for the installation of this geometrical model is described previously. The principal characteristics used are: variation of diameter, angle of torsion determined by the inclination of fibers, pilosity, position and the shape of the yarn in the textile structure. The second axis is characterized by the use of the artificial neural networks in order to establish the transfer of the fibers characteristics towards the yarn (Brudaru and Copaceanu 1999; Neagu 2006). This model can be widened and must adapt to the transfer of the characteristics of the yarn towards weaving and clothing. It requires the installation of a data base taking into account these considerations. The dynamic model will consider behaviors in inflection, traction of fibers and frictions inter-fiber in order to limit the already consequent study. We will limit also the study to certain fibers which we will define according to the complexity of the problem. The interest of the artificial neural networks is to integrate automatically the nonlinear character of the study (Grishanov et al 1999). The third axis consists in analyzing the deformations problems mathematically (Harwood et al 1997) caused by the integration of the yarn in a textile structure such as fabric. That analyzes the effects of negative feedback on woven structures to the various geometries of the yarn.

REFERENCES


Neagu I, D.Avram, P.Bruniaux, 2006 “The state of using the artificial neuronal networks in forecasting the characteristics of textile fibers”, Master en Roumanie, research in GEMTEX


APPAREL INDUSTRY AND CLOTHING MODELLING AND SIMULATION
TOWARDS A COUPLED SIMULATION OF CLOTH AND SOFT TISSUE

Bastian Linneweber
Arnulph Fuhrmann
Fraunhofer Institute for Computer Graphics Research IGD
Fraunhoferstr. 5
D-64238 Darmstadt, Germany
email: {bastian.linneweber|arnulph.fuhrmann}@igd.fraunhofer.de

KEYWORDS
cloth simulation, soft tissue simulation, collision detection, collision response

ABSTRACT
In this paper, we present important steps towards a coupled simulation of textiles and human soft tissues. We have developed a new simulator for soft tissue which is able to simulate the skin of a non-moving virtual human in a realistic manner. This simulator is coupled with a cloth simulator so that both human skin and textiles can be simulated in parallel with mutual interaction and deformation. The described system allows the adjustment of several material parameters of the skin and many aspects of the skin’s behavior especially shape and volume preservation are considered. The simulation of the functionality of certain cloth types such as forming and supporting the human skin as well as an improved control of the correct fit is permitted by the coupled simulation of cloth and skin. Additionally, we present an efficient method for detecting collisions between cloth and the deformable virtual human.

INTRODUCTION
Physically-based simulation is one of the major topics in today’s computer graphics research. Amongst others, the simulation of cloth has received much attention as it is of interest for various branches such as the film industry, computer game development or the clothing industry. A large number of works on that topic have been published until today, showing plausible or even realistic simulations of garments. Many of these publications are designed to simulate static or animated virtual humans wearing these clothes. Most of these works, however, do not address the fact that the human skin is deformable and that there are interactions between the garments and the human body in reality. In this paper, we present a simulator which overcomes this lack of realism. There is a growing interest in the simulation of soft tissue, in particular in the fields of medical simulation and human animation. Medical environments such as virtual realities for surgeons require a high level of accuracy and the possibility to attach haptic interfaces. Contrary, human or facial animation approaches often do not employ physically realistic simulation but prefer rigging and skinning approaches to facilitate the animation process. The simulation of the human skin in the context of cloth simulation has to be at least physically plausible but also fast to calculate and furthermore easy to combine with a cloth simulator. Therefore, one needs to exert methods of collision detection and collision response between several objects.

The treatment of collisions and contact is a very important area of research, too. Collisions are usually one of the major bottlenecks in a simulation environment. A large variety of different collision detection approaches have been proposed so far, each of them having its own advantages and disadvantages being also more or less specialized for certain simulation environments. Therefore, it is necessary to find a suitable collision detection algorithm in combination with an adequate collision response method regarding the coupled simulation of cloth and human skin.

Our contribution
In this paper, we present a simulator which is able to simulate both cloth and human skin in parallel at interactive frame rates. Interactions between cloth and skin are possible and the simulator can mimic deformations on the human body caused by cloth while worn. These are important steps towards what we call a coupled simulation of cloth and soft tissue.

We also present a refined model for simulating human skin as well as a novel geometric constraint for volume preservation. We introduce fast algorithms for collision detection and response which allow the interaction between cloth and skin. As a result we may simulate the functionality of certain cloth types like brassieres or waist belts and provide a realistic simulation as well as an improved control of the correct fit of tight cloth that squeezes or constricts the wearer.

RELATED WORK
Physically-based simulation has received much attention during the last two decades and is still of major interest for the community. Publications on cloth and soft tissue simulation as well as collision detection and response are the most important ones for this paper. As there is such a great variety of publications, we are not able to give an extensive list but a quick overview of the most important works concerning our topic.
Figure 1: (a) The schematic structure of our soft tissue model. The morph targets and their corresponding triangles in the final simulation model can be seen as the separating layers between the skeleton, the muscles and subcutis as well as the dermis. The tetrahedrons on the other hand symbolize the volumetric layers of the human skin, especially the fat cells.
(b) Morph targets (right) are used to generate the three layers of the simulation mesh (left). The first morph target is always used without change while the proportions \( f_{\text{layer}} \) of the second and third are adjustable. The simulation mesh is sliced and colored to visualize the skin layers (gray: skeleton, red: muscle + subcutis, pink: subcutis + dermis).

A quite recent state of the art report on today’s simulation methods can be found in (Nealen et al. 2005). Cloth simulation was pioneered twenty years ago by Terzopoulos and Witkin (Terzopoulos and Fleischer 1988). There have been several important publications since then, one of the most prominent might be the work of Baraff and Witkin (Baraff and Witkin 1998), introducing a semi-implicit integration method enabling larger time steps than previous approaches. This method is also of importance for this paper, as we use it for the soft tissue simulation. Cordier et al. (Cordier et al. 2002) presented a real-time cloth simulation allowing animated figurines including a simulation of the skin. Their approach however did not provide mutual interaction and deformations between cloth and skin. Volino and Magnenat-Thalman (Volino and Magnenat-Thalmann 2005) focused on the prototyping of garments which is a main feature of our cloth simulator, too.

Provot (Provot 1995) was the first to propose the use of geometric constraints on the mesh edges in combination with the integration of forces circumventing the stiffness problem while still providing a fast simulation. Recently, Müller et al. (Müller et al. 2007) generalized the idea of geometric constraints. The geometric constraints we use for the simulation of skin is based on their approach. Goldenthal et al. (Goldenthal et al. 2007) also presented a Constrained Lagrangian Mechanics formulation based on the constraint’s gradient and proposed a non-iterative, fast projection method for enforcing the constraints. According to the authors’ results, the approach is much faster than previous ones, especially when dealing with a large number of constraints.

In order to reproduce the deformations induced on the surface of a virtual human by worn cloth, it is necessary to simulate this surface. Human skin is a special type of soft tissue and its simulation has gained a certain amount of interest in several fields, e.g. in medical simulation. Notable works in this area comprise for example (Bro-Nielsen 1998), (Hauth et al. 2003) and (Mosegaard and Sorensen 2005). In the field of human animation, there are some approaches to simulate the skin upon the moving skeleton to increase the physical accuracy of the simulation (for example (Capell et al. 2005)).

An overview of current collision detection techniques is provided by Teschner et al. (Teschner et al. 2005). The work of Sigg and Peikers (Sigg et al. 2003) for distance field generation is significant for this paper, as their ideas provide the basis for our collision detection algorithm. Collision response is another important issue in physically-based simulation. An impulse-based solution as presented by Bridson et al. (Bridson et al. 2002) will be used in this paper.

**SOFT TISSUE SIMULATION**

To simulate the human skin in a realistic but also interactive manner, we need to develop an adequate model of this organ. Cloth simulation, notably in the clothing industry, requires an interactive environment. The accuracy of the simulation regarding a cloth-skin simulation is therefore not as important as the calculation speed. One of our main goals is to achieve frame rates sufficient for an interactive simulation. Hence, our first step is to find a suitable underlying model. The most prominent choices are either mass-spring systems
or finite element methods. While finite elements methods usually provide a much higher accuracy, this advantage comes with a certain increase of computational costs. The simulation mesh we use for the human skin consists of thousands of elements so it should prove difficult to implement a simulator based on a finite element approach with interactive simulation rates on today’s standard computer equipment. The recent paper by Irving et al. (Irving et al. 2007) presenting a finite element approach including volume preservation showed a rather high computational complexity for such objects of high resolution. Therefore, we have opted for a mass-spring model which has been extended by some more features to increase the realism of this simulation as a first step. Other authors have already shown convincing results using mass-spring systems for soft tissue simulation, for example (Mosegaard and Sørensen 2005) and (Hong et al. 2006).

**Modeling**

We use a multi-layered model of triangles and tetrahedrons for this simulation which is similar to the one devised by Lee et al. (Lee et al. 1995) for facial animation. We have developed a novel approach to initialize the simulation mesh of the skin which is based on morph targets of the human body. Using three morph targets, we generate three layers of triangles which separate another two layers of tetrahedrons during our initialization process (see Figure 1a). The proportion of the morph targets two and three regarding the final thickness of their corresponding layer is variable as depicted in Figure 1b. Our simulator is nevertheless completely independent from the described initialization method. The coupled simulation can be used with any simulation mesh which comprises tetrahedron layers.

**Forces and Constraints**

The calculation of internal forces which model the natural behavior of a simulated object is one of the most important aspects of a simulation. The human skin is a highly elastic organ although it is rather incompressible and preserves its volume. In order to reproduce these qualities we use different types of constraints in conjunction.

On the one hand, we compute force based constraints as described in (Baraff and Witkin 2003) to calculate forces that cause acceleration or deceleration of the particles during the integration step. On the other hand, we use geometric constraints which directly affect the positions of the particles and are applied after the integration. We also include a global damping force into our system.

The forces are described by the function

\[ \mathbf{f}(\mathbf{x}, \mathbf{x}) = (-k_s \cdot C(\mathbf{x}, \mathbf{x})) \cdot \frac{dC(\mathbf{x}, \mathbf{x})}{dx} \]  

(1)

where \( k_s \) is the spring rate and thus a material parameter to describe the skin’s behavior under deformation. \( C(\mathbf{x}, \mathbf{x}) \) is the constraint function subject to positions and velocities. This constraint yields the desired behavior like distance or volume preservation for \( C(\mathbf{x}, \mathbf{x}) = 0 \).

**Distance preservation**

The simplest and most commonly used force based constraint is the distance preservation constraint applied to the edges of the simulation mesh. It models the elasticity of the skin and thus preserves the shape of this organ. The distance preservation constraint can be described as

\[ C_{distance} = |\mathbf{x}_{ij}| - l_{ij} \]  

(2)

where \( |\mathbf{x}_{ij}| \) means \( |\mathbf{x}_i - \mathbf{x}_j| \) and \( l_{ij} \) is the rest length of the edge between the particles \( p_i \) and \( p_j \). The stiffness \( k_s \) in equation (1) can be computed using a measured modulus of elasticity for human skin as shown in (Gelder 1998) for triangle meshes.

**Volume preservation in tetrahedrons**

Concerning volume preservation we pursue two different approaches. The first one preserves the volume of each tetrahedron element by applying forces to restore the original or rest volume. This idea is mainly based on the work of Teschner et al. (Teschner et al. 2004) on the simulation of deformable solids. The formula to calculate the volume of a tetrahedron is required for deriving the forces to apply on every vertex of a tetrahedron. It is given by

\[ V_{tetrahedron} = \frac{1}{6} (\mathbf{x}_j - \mathbf{x}_i) \cdot (\mathbf{x}_k - \mathbf{x}_i) \times (\mathbf{x}_l - \mathbf{x}_i) \]  

(3)

where \( \mathbf{x}_{i,j} \) are the positions of the four vertices of a tetrahedron. Equation 4 yields the signed volume of a tetrahedron, so we may detect inverted tetrahedrons as described in (Teschner et al. 2004). The general constraint for volume preservation constraints is defined as

\[ C_{volume} = \frac{(V_{current} - V_{rest \, state})}{V_{rest \, state}} \]  

(4)

When dealing with tetrahedrons \( V_{current} \) is calculated using equation (3). \( V_{rest \, state} \) is the volume of a tetrahedron at its rest state. The material parameter \( k_s \) to calculate the volume preservation forces (1) has to be determined by experiments.

**Volume preservation of the whole object**

The volume preservation force which acts on the object as a whole is based on the work of Hong et al. (Hong et al. 2006). They use the divergence theorem to derive a formula to calculate the volume of a closed object considering the triangles and normals of its surface. This formula is given by

\[ V = \frac{1}{3} \sum_{i=1}^{nt} A_i \cdot n_i \cdot (\mathbf{x}_i^1 + \mathbf{x}_i^2 + \mathbf{x}_i^3) \]  

(5)

where \( nt \) stands for the number of surface triangles, \( A_i \) is the area, \( n_i \) the normal and \( \mathbf{x}_i^1 \) the positions of the vertices of the \( i \)-th triangle. Using equation (4) we can devise a constraint that preserves the volume of the object. \( V_{current} \) is determined by equation (5) and \( V_{rest \, state} \) is the global volume of the object in its rest state. We also multiply the force acting on every particle \( \mathbf{p}_i \) around a deformation zone with the weight factor \( w_i \) as described in (Hong et al. 2006). Thus, we create a bulging effect around the regions of the skin which are deformed by the cloth as depicted in Figure 2.
Integration

We have opted for an implicit Euler integrator as a solver for the differential equation of our simulation for several reasons. One aspect is the stability of this integrator even under large time steps. As the calculation of the above mentioned forces and also the following geometric constraints is computationally very costly, it is necessary to reduce these computations by using larger time steps. The use of an implicit integrator slightly increases the complexity and the computational cost compared to an explicit solver. However, this increase is negligible in comparison to the reduction of constraint calculations due to the larger time steps. We are therefore able to achieve interactive frame rates.

GEOMETRIC VOLUME CONSTRAINT

In order to simulate the skin’s behavior realistically when compressed, we propose a novel geometric constraint. The human soft tissue always preserves its volume when deformed as it mostly comprises incompressible water. As we have already mentioned in the previous section, we apply forces in our simulation to preserve volume. These methods however only produce a linear force that restores the volume of the simulated skin. Furthermore, the forces do not guarantee that the layers of the simulation mesh do not penetrate each other.

We have therefore developed a new volumetric constraint based on the ideas of Provot (Provot 1995) for cloth simulation and the recently published work of Müller et al. (Müller et al. 2007) on Point Based Dynamics. This constraint acts on the tetrahedrons of the mesh. It ensures that the volume of one tetrahedron remains in a predefined range. If the volume of a tetrahedron falls below or exceeds this range, its vertices will be moved so that the volume constrained is fulfilled as shown in Figure 3. In combination with the force constraints, we obtain a non-linear volume preservation rate for the simulated skin. Additionally, the constraint ensures that the layers will not interpenetrate as it detects and avoids tetrahedron inversion likewise the above mentioned volume force for tetrahedrons. The equation describing the displacement of each vertex by the constraint

\[ \Delta x_i = \lambda \cdot w_i \cdot d_i \]

is similar to the one devised by (Müller et al. 2007). In equation (6) \( w_i = \frac{1}{m_i} \) is a weight factor to cope with individual particle masses and \( d_i \) is the direction of the constraint projection. As this direction should cause the maximal change in volume it is equal to the derivative of the tetrahedron volume equation (3) with respect to the vertex position \( x_i \). \( \lambda \) is the scalar we need solve the equation for. This can either be done by the linearization as described in Point Based Dynamics. In this case, it is also possible to solve the equation analytically as the constraint for the four vertices of a tetrahedron results in a polynomial of third order.

COUPLED SIMULATION

The coupled simulation of cloth and soft tissue can be split up into two major tasks. Firstly, one needs to find the regions where cloth and skin may possibly interact. Thus, a collision detection between cloth and soft tissue objects is required. Secondly, it is necessary to react on a collision in a physically accurate manner. We therefore need a physically correct collision response.

There are requirements which both the collision detection and the collision response need to fulfill. It is obvious that both algorithms shall not be expensive in calculation time and in their data structure updates as we are dealing with an interactive simulation environment. Furthermore, the collision response requires various information from the collision detection to accomplish its task, i.e. the collision triangle on one object for a collision particle on the other.

The integration of the system of linear equations as well as
the calculation of the constraints are independent of the interaction between cloth and soft tissue. The necessary calculations can therefore be executed separately and in parallel. The two simulations need to be synchronized only at the collision detection and response step.

Collision detection

A large number of different collision detection algorithms have been proposed and investigated in the context of animation and simulation. For our purpose we have developed a new collision detection algorithm based on two existing ideas which provides the necessary information for the collision response.

We call our method a Distance Prism Hierarchy (DPH) and it can be seen as an extension of the well known Bounding Volume Hierarchies (BVHs). While the construction of the tree and the inner nodes is very similar, its leaves do not only comprise the triangle but also a three-sided pyramid frustum enclosing the triangle. The sides of this prism are tilted to the half angle plane between two neighboring triangles. The caps of the prism are planes parallel to the triangle and shifted to given positive and negative distance thresholds (see Figure 4).

Each prism encloses all points in \( \mathbb{R}^3 \) up to the distance thresholds for which the closest point lies on its corresponding triangle. Additionally, it also encloses a part of the points in \( \mathbb{R}^3 \) for which the closest point lies on the vertices or edges of the triangle. The missing points in \( \mathbb{R}^3 \) for edges and vertices are enclosed by the neighboring prisms. Those prisms have been introduced by Sigg et al. (Sigg et al. 2003) for the GPU-based generation of distance fields.

In our hierarchy the bounding volume of the parent node of a leaf node will encapsulate its prism. As bounding volume any standard volume like k-DOPs can be used. In our implementation we use axis-aligned bounding boxes. The hierarchy is built top down by splitting the triangles recursively along the longest side of their enclosing bounding box once in a pre-processing step.

For the collision detection of vertices against a DPH, the tree is traversed exactly like a BVH. However, the collision test for the leaves is different. A candidate vertex is tested against all the prisms found during traversal. If the vertex lies within the prism, the signed distance to the triangle can be computed easily. The triangle with the smallest absolute distance is closest to the vertex. The corresponding signed distance and triangle are the result of our collision query.

Our approach combines the advantages of distance fields, which provide the necessary information for collision response, with the flexibility of BVHs, which allow to determine quickly the corresponding triangle.

In our coupled simulation we are testing the particles of the cloth mesh against the surface triangles of the soft tissue mesh. We create a DPH for the skin only and test every cloth particle for a collision with this structure. A particle and the reported closest triangle will exchange energy during collision response. During simulation the skin deforms and the DPH is updated after each simulation step. This refitting of bounding volumes is efficient for deformable objects with fixed topology (Teschner et al. 2005).

Collision response

Our collision response is based on the ideas of Bridson et al. (Bridson et al. 2002) using impulses and a friction approximation. After a collision between a cloth particle and a soft tissue triangle is detected, we use the barycentric coordinates of the collision point to interpolate the momenta of the triangles vertices. Thus, we create a virtual collision particle on the triangle. We consider a collision between cloth and skin as a completely inelastic one. Therefore, the resulting velocities \( \mathbf{v}^{+}_{\text{cloth}} \) and \( \mathbf{v}^{+}_{\text{virtual}} \) in the direction of the triangles normal \( \mathbf{n} \) of two colliding particles (\( \mathbf{p} \text{cloth} \) and \( \mathbf{p} \text{virtual} \)) are equal. They can be described as

\[
\mathbf{v}^{+}_{\text{cloth}} = \mathbf{v}^{+}_{\text{virtual}} = \frac{m_{\text{cloth}} \cdot \mathbf{v}^{+}_{\text{cloth}} + m_{\text{virtual}} \cdot \mathbf{v}^{+}_{\text{virtual}}}{m_{\text{cloth}} + m_{\text{virtual}}} \tag{7}
\]

where \( m_{\text{cloth}} \) and \( m_{\text{virtual}} \) are the masses of the respective particles. If \( \mathbf{v}^{+}_{\text{cloth}} \cdot \mathbf{n} < 0 \) an impulse between cloth and skin particles will be exchanged. Otherwise, this is only a contact region. In case of a change of momentum the resulting velocity of the cloth particle is simply \( \mathbf{v}^{+}_{\text{cloth}} \), while the velocities for the real particles of the skin triangles have to be calculated according to the barycentric coordinates \( w_{1,3} \) and the impulse \( \mathbf{I}_{\text{virtual}} \) on the virtual particle

\[
\mathbf{I}_{\text{virtual}} = (\mathbf{v}^{+}_{\text{virtual}} - \mathbf{v}^{+}_{\text{cloth}}) \cdot \frac{1}{w_1 + w_2 + w_3} \tag{8}
\]

\[
\mathbf{v}^{+}_i = \mathbf{v}_i + w_i \cdot (\mathbf{I}_{\text{virtual}} / m_i) \tag{9}
\]

where \( \mathbf{v}_i \) is the velocity and \( m_i \) is the mass of the \( i \)-th triangle particle.
Figure 5: Female figurine wearing a bra. (a) Soft tissue simulation is inactive. The bra does not deform or support the figurines breast; (b) Soft tissue simulation active, the breast are pushed up and formed; (c) The differences between the two in the breasts region.

<table>
<thead>
<tr>
<th>Simulation</th>
<th>Tetrahedrons</th>
<th>FPS</th>
<th>forces + integrator</th>
<th>geometric constraints</th>
<th>collision</th>
</tr>
</thead>
<tbody>
<tr>
<td>bracelet</td>
<td>15360</td>
<td>10.9</td>
<td>74 %</td>
<td>14 %</td>
<td>12 %</td>
</tr>
<tr>
<td>bra</td>
<td>11148</td>
<td>4.8</td>
<td>66 %</td>
<td>8 %</td>
<td>26 %</td>
</tr>
<tr>
<td>skirt (some simulation regions active)</td>
<td>15951</td>
<td>4.1</td>
<td>67 %</td>
<td>8 %</td>
<td>25 %</td>
</tr>
<tr>
<td>skirt (all simulation regions active)</td>
<td>40722</td>
<td>2.2</td>
<td>70 %</td>
<td>20 %</td>
<td>10 %</td>
</tr>
</tbody>
</table>

Table 1: Performance measurements of different test scenes, including the percentage of CPU time required by the most important calculation steps (performed on a Athlon X2 3800+).

We also use a Coulomb approximation of friction

\[ v_i^+ = \max \left( 1 - \mu \cdot \frac{|v_i|}{|v_{\text{normal}}|}, 0 \right) \cdot v_i \]  \hspace{1cm} (10)

in order to simulate this energy dissipation, where \( v_i \) is the velocity of a particle tangential to the surface. For the sake of simplicity we use the same friction coefficient \( \mu \) for both static and kinetic friction.

As the changes in velocity do not have an immediate effect on the particles’ positions, we separate the cloth particles from the soft tissue mesh by moving them back to the surface if necessary. The correction of the positions does not have an impact on the particles’ velocities, though. We subdivide the cloth triangles and test the subdivisions against the soft tissue surface, too, in order to avoid inner triangle collisions.

RESULTS AND DISCUSSION

In this work, we have presented a method for simulating both cloth and skin in parallel with mutual interaction and deformation. The coupled simulation enables us to simulate not only cloth itself, but also to reproduce the effect that cloth has on the body. Whenever a virtual garment incises or squeezes the wearer for example, the simulated skin will be deformed by the forces the garment applies on it. The skin even bulges around the contact area as one would expect it to do in reality. The clothing may also compress and push a whole region of skin in the direction determined by the design of the garment. We can therefore simulate cloth with a particular functionality, like brassieres (see Figure 5).

Our simulator also provides an improved control of the correct fit, in particular concerning tight garments like trousers (Figure 6). Cloth of different sizes deforms the wearer differently as shown in Figure 7. The tighter the cloth, the stronger the force applied on the skin by the cloth and the more pronounced the deformation of the skin. The differences in the visual result might not be drastic at the first sight, but they are very important for the fit and hence for the virtual prototyping of garments. Furthermore, the deformations induced on the virtual human give information about whether the clothing is comfortable for a person in reality or not.

The soft tissue simulation can be controlled by several material parameters like elasticity, thickness and compressibility. Many aspects of the skin’s behavior in reality, like shape and volume preservation have been considered. The novel geometric constraint presented in this paper turned out to be a great asset regarding the reproduction of volume preservation and incompressibility of the skin as well as for an increase.
in stability of the simulation. All forces and constraints
described in our work can be parametrized, so we may simulate
different types of skin as shown in Figure 8 which depicts
that soft skin deforms more than hard skin.
The simulator fulfills our requirement of an interactive sys-
tem in all our examples. Depending on the complexity of
the scene, especially the number of triangles and tetrahedrons, as
well as the number of activated simulation regions, our simu-
lator can produce up to 10 frames per second on a today’s
standard PC (see Table 1).

FUTURE WORK

One of our main goals for the future is the simulation of fully-
animatable and deformable avatars wearing clothings. We
therefore need to develop a system simulating a non-static
avatar with moving extremities which is still deformable by
the worn textiles. We would then be able to simulate for
example sports wear in action, for example.
Furthermore, we would do more research in a soft tissue sim-
ulation system based on a finite element approach to increase
the physical accuracy. Improvements in physical accuracy
could also be achieved by including creeping and relaxation
in the skin’s behavior which is neglected right now.
Another important issue is the increase of simulation speed.
This could be achieved by improving the collision detection
and response algorithm, further parallelizing the simulation,
especially in the light of the new generation of multi-core
CPUs, and maybe also by some kind of adaptive meshing.

Another interesting field of future research is the assignment
of material parameters measured in real-life to the simula-
tion. We are also considering the possibility of simulating
the inner organs. Currently, everything below the skin is re-
garded as stiff which is not true in reality especially in the
abdominal region.

REFERENCES

Baraff D. and Witkin A., 1998. Large steps in cloth simula-
conference on Computer graphics and interactive tech-

In SIGGRAPH ’03: ACM SIGGRAPH 2003 Course Notes.
ACM Press, 276.

Bridson R.; Fedkiw R.; and Anderson J., 2002. Robust treat-
ment of collisions, contact and friction for cloth animation.
In SIGGRAPH ’02: Proceedings of the 29th annual
conference on Computer graphics and interactive tech-


Capell S.; Burkhart M.; Curless B.; Duchamp T.; and
Popovic Z., 2005. Physically based rigging for deformable


BIOGRAPHY

BASTIAN LINNEWEBER studied computer science at the University of Technology in Darmstadt from 2001 to 2007, with Anglistics as minor field of study. After finishing his studies with the diploma thesis in 2007, he continued to work as an assistant at Fraunhofer IGD. In 2008 he became a member of the department Realtime Solutions for Simulation and Visual Analytics where he pursues his work on cloth and soft-tissue simulation and his PhD thesis.

Dr.-Ing. ARNULPH FUHRMANN studied computer science with focus on mathematics at the University of Technology in Darmstadt. After receiving his diploma in 2001 he became a member of the research group Realtime Solutions for Simulation and Visual Analytics at the Fraunhofer Institute for Computer Graphics Research. Since 2005 he is deputy head of this group. He received his doctoral degree from the faculty of computer science at the Technical University of Darmstadt in 2006 for his research on “Interactive animation of textile materials”. His main research interests are 3D simulation, physically based modelling, garment simulation and collision detection for deformable objects. Arnulph Fuhrmann is a member of the European Association for Computer Graphics (Eurographics) and the European Technology Platform for the Future of Textiles and Clothing.
Figure 7: Comparison of skirts of two different sizes of the same garment. The figurine in the left image wearing a smaller skirt is more squeezed in the waist than the one right image with a larger skirt.

Figure 8: Female figurine wearing a body. (a) Hard skin type avatar deformed by the body; (b) Difference between the non-deformed avatar and the deformed, hard skin type avatar: the deformation of the skin is visible; (c) Soft skin type avatar deformed by the body; (d) Difference between the non-deformed avatar and the deformed, soft skin type avatar: the deformation of the skin is more pronounced compared to the hard skin type avatar.
STRASTRATEGIC ANALYSIS OF SPEED AND FLEXIBILITY IN SOURCING TEXTILE PRODUCTS

Lisa Hartman, Jeffrey A. Joines, K.A. Thoney, and Russell E. King
North Carolina State University
College of Textiles, Campus Box 8301
Raleigh, NC 27606 USA
E-mail:JeffJoines@ncsu.edu

KEYWORDS
Supply Chain, Sourcing Decisions, Inventory, Lead-Time

ABSTRACT
This research investigates sourcing basic and seasonal garments as a function of lead time and ability to provide replenishment orders. The Sourcing Simulator™ is used to determine the effect of forecasting error, drift, and seasonality on retail performance (i.e., service levels, gross margins, and inventory levels) for a variety of case studies, which were based upon data obtained from apparel manufacturers and retailers. Results show that choosing suppliers with shorter lead times provides significant advantages for basic goods as well as seasonal goods, which in some instances justify using more expensive suppliers.

INTRODUCTION
When making decisions about sourcing merchandise, many buyers base their decisions on what supplier can provide them with the merchandise at the lowest first cost in an attempt to improve their gross margin (i.e., revenue minus cost of goods sold). Many buyers and decision makers believe that the choice of supplier only affects the cost of goods sold and will not affect the revenue in the gross margin calculation. However, this is not always the case. A manufacturer that is faster can provide a mid-season replenishment of items that have sold well. Also, having large amounts of merchandise does not necessarily ensure high service levels. If the high inventory is carried in the wrong SKUs, some popular styles of garments may sell quickly while others will not. Some studies have shown that if the SKU mix is based on current, mid-season sales, then the amount of unsold merchandise is reduced. Many long lead-time suppliers cannot provide a replenishment order before the end of the selling season. Buyers need to know how their sourcing decisions will impact their company’s finances in relation to customer service levels, inventory levels, etc.

In this research, the differences between suppliers with different lead times are investigated with respect to profitability and customer service levels using a quantitative model. The analysis determines when using fast, more expensive suppliers, over low cost, slow suppliers is beneficial to a company (i.e., under what conditions or characteristics). The focus will be on basic and seasonal items. The original goal was to study sourcing to the U.S. from manufacturers in the following four regions of the world: U.S., Central America / South America / Caribbean, Pakistan / India, and other Asian countries. A survey of U.S. retailers and apparel manufacturers was conducted that asked about their perceptions of sourcing from those four regions. The survey revealed that there was not much agreement about the lead time’s experienced from these four regions. Proximity to the U.S. was expected to be directly related to lead times, but this was not always evident in the survey responses. Some retailers and apparel manufacturers interviewed also provided data that was used to create case studies. Due to the various perceptions of lead times from different regions of the world, analysis of the cases studies will be performed for the overall range of lead times indicated in the survey. This will allow the reader to draw his/her own conclusions about sourcing to the U.S. based on his/her own beliefs about lead times in sourcing from different regions of the world.

LITERATURE REVIEW

Sourcing Decisions

Several researchers have looked at various sourcing decisions. Zeng (2000) considers the type of outsourcing a company should use: multiple sourcing, single sourcing, single/dual hybrid or network sourcing, and global sourcing, according to what the company considers the most important aspects of purchasing. More flexibility and better protection against stockouts is often associated with a multiple supplier system. Single sourcing, on the other hand, can result in cost reduction, stability, and improved communication.

Lowson (2003) determines how lead-time, inventory, supplier performance, and customer service levels affect the sourcing decision. Flexibility and speed in a supply chain can compensate for the uncertainty in demand. Using a low cost manufacturer is not always the best choice.

According to Smith and Magnussen (2005) and Velazquez (2004), companies in Central and South America, CBI, and United States are not capable of competing with other Asian countries on price alone. Countries in the western hemisphere must capitalize on their proximity to the U.S. market. Instead of focusing on creating basic garments that can be produced cheaper elsewhere, companies in these regions need to start focusing on manufacturing fashion garments that can then be quickly shipped to or within the U.S.

Sourcing Simulator

The Sourcing Simulator is a stochastic simulation model for the consumer product retailing process developed by Nuttle et al. (1991) and Hunter et al. (1992). The model allows investigation into the effects of alternative retailing
procedures on financial and other performance measures for a retail store. The value of the model lies in the fact that it captures the random nature of consumer behavior at the retail store within a robust framework that allows investigation of buyer strategies.

Consumer arrivals at the retail store are modeled as a time-dependent Poisson process. The rate each week is based on a specified season arrival pattern. Customers arrive at the store and attempt to purchase garments according to a defined sales and choice distribution. For a particular customer, if the desired stock keeping unit (SKU) is in stock, a sale is recorded and inventory decremented. If the SKU is out of stock, a stockout is recorded. After a stockout a customer may look for another item with certain probabilities. A forecast of consumer demand is expressed in terms of customer volume, SKU mix, and presumed seasonality. The model assumes that the forecast is in error, which is specified as a volume error, SKU mix error, and actual seasonality.

The simulator models alternative mechanisms for supplying product to a retail store. The store defines an initial inventory to start the selling season according to the store buyer’s plan. The store may issue replenishment orders on the vendor. Replenishment may be based upon the original buyer’s plan or may reflect the use of actual Point-of-Sale (POS) data. In this way the selling season is played out and performance statistics are computed. Detailed information about development of the Sourcing Simulator and a number of case studies on sourcing, forecasting and replenishment analysis using the Sourcing Simulator is included in Lawson et al. (1999) and Hunter et al. (2002).

SIMULATION ANALYSIS FOR BASIC GOODS

For this study, basic items are considered to be merchandise that is carried in a store all year with little change in color, style, size, or price. Examples of basic items include bras and men’s white button down shirts. Although demand for basic goods is easier to forecast then for seasonal or fashion goods, when retailers place orders for basic merchandise, it is still difficult to exactly forecast the types of garments needed and the amount of each garment to order according to SKU (color, style, and size of the garment). Any errors in the ordering of merchandise can be very costly to the retailer in the form of lost sales or markdowns on excess merchandise, which will cut into the profitability of the retailer.

The Sourcing Simulator™ was used to simulate a men’s basic pant over a five year period. The focus was on determining how lead times affect the average inventory levels retailers must carry to meet desired customer service levels. Experiments were performed for target service levels of 80%, 85%, 90%, 95%, and 97%. Since there was a lack of agreement on lead times that are typical for each region, lead times for the simulation experiments were varied from 1 to 25 weeks, which is the range from the industry surveys. Specifically, supplier lead times of 1, 5, 10, 15, 20, and 25 weeks were investigated. The simulations were run using two different customer demand patterns. The first demand pattern was flat demand. While the other demand pattern was a seasonal pattern representing the back to school and Christmas selling seasons (see Figure 1).

To determine the average inventory levels, experiments were undertaken to figure out the best value for the amount of merchandise that should be initially ordered and the number of weeks of supply the retailer should carry in the store in order to meet desired service levels for a basic product. The target weeks of supply is the “optimal number” of weeks of inventory that the retailer needs to carry between reorders in order to meet service levels, but not have an excess or shortage of merchandise. For example, if one wants a ten week target, the average inventory needed is the sum of the next ten weeks of forecasted demand. To determine the best inventory targets, two different service levels were used: the service level before the first reorder and the service level before the first markdown. The “optimal” initial weeks of inventory ordered as well as the target weeks of supply were determined by systematically changing the specified number of initial weeks of inventory and weeks of supply and then simulating the scenario until the desired service levels were met.

![Figure 1: Seasonality of Demand for Five Year Selling Period](image)

**Initial Experiments**

Initially, the forecasted trend was assumed equal to the actual trend, and there was no SKU drift. Therefore, the first experiments assumed perfect forecasted demand, which, of course, rarely happens but will produce the most optimistic view for longer lead times.

**Average Inventory Levels**

In the initial experiments to find average inventory level, two different ordering strategies were analyzed. In the first strategy, reorders were placed every four weeks, regardless of supplier lead time. Demand was assumed to be flat over the five year period. In the second ordering strategy, simulations were run with the weeks between reorders equaling the supplier lead time. For example, if the supplier has a ten week lead time, then reorders are made every ten weeks. This strategy was used since it is more representative of industry, where typically an order is not made until the previous one has been received (i.e. no multiple outstanding orders).

Figures 2 and 3 should the results of the experimentation using the flat demand and seasonal demand, respectively. The inventory levels of demand in weeks was found by dividing average inventory by the average weekly demand. Figures 2 and 3 demonstrates that the longer the lead time of the supplier, the more inventory the retailer must carry to meet the desired service level for a given type of order strategy. This pattern was observed for all target service
levels tested. For example, if a retailer has a supplier with a thirteen week lead time and is using order strategy one with flat demand, then the retailer must carry about 6.94 weeks of inventory in order to meet a 95% service level. In contrast, if the supplier had a three week lead time and is using order strategy one with flat demand, then the retailer has to carry only 3.76 weeks of inventory to meet that same service level.

Although the pattern of results for order strategy two is similar to that seen for the results of strategy one, the slope of each of the target service level lines is steeper, indicating even more inventory is needed when supplier lead times are long (see Figure 2). This is the result of ordering less frequently under the second ordering strategy than in the first (i.e. the more frequent ordering effectively reduces the lead times). By reordering frequently, the effect of longer lead times smooths out, effectively reducing the lead time impact. The takeaway is that suppliers with longer lead times should encourage their customers to order more often. Since reorder strategy 2 is more likely to be used in industry, it was used for the rest of the simulations on basic goods.

![Image](image1)

**Figure 2:** Inventory Levels Needed to Achieve Desired Service Levels for Flat Demand

![Image](image2)

**Figure 3:** Inventory Levels Needed to Achieve Desired Service Levels for Seasonal Demand

**Breakeven Analysis Calculation**

A cost analysis was performed for retailers with suppliers of varying lead times. Specifically, how much more a retailer could pay a supplier with shorter lead times in landed costs and still achieve the same margins if fixed ordering costs are negligible? The cost increase is possible because the retailer is paying less to carry inventory with a shorter lead time supplier. Notation required for this analysis is as follows: $i = \text{inventory carrying costs in }$/\$ of inventory/year; $C_j = \text{landed cost (procurement costs) for choice } j$; $D = \text{yearly demand}$; and $I_j = \text{average inventory levels for choice } j$. Thus, the total annual inventory carrying costs plus the total annual procurement cost is given as follows:

$$ic_i I_j + C_j D.$$

Let Choice 1 be a higher cost, shorter lead time option and Choice 2 be a lesser cost, longer lead time option. The breakeven point between Choice 1 and 2 can be expressed as follows:

$$ic_i I_j + C_j D = ic_i I_j + C_j D.$$

Therefore, the percent increase in landed costs for Choice 1 is given as follows:

$$(C_i/C_j - 1) = (IC_i + D)/(IC_j + D) - 1.$$

Table 1 shows the results of the breakeven analysis for a 90% service level with flat demand with $i = 20$. If a retailer had a supplier with a 25 week lead time, they could pay 10% more per garment for a supplier with a one week lead time or 8.9% more in landed costs per garment for the same garment for a supplier with a five week lead time and still achieve the same margins. The same trend is also true if a retailer had a supplier with a 15 week lead time and wanted to switch to a supplier with a five week lead time. The retailer could pay 3.9% more in landed costs per garment for the same garment and still achieve the same margins. All the results show conservative numbers for inventory levels for various lead times in idealized conditions which means that the actual percentage increases represent best case scenarios for longer lead times. The relatively small difference can be explained by the lack of error in forecasting demand in these initial experiments.

**Table 1: Breakeven Analysis of the Increase in Costs Per Garment Retailers Could Pay for Shorter Lead Times for a 90% Service Level with Flat Demand with No Drift**

<table>
<thead>
<tr>
<th>New Lead Time</th>
<th>Current Lead Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 Weeks</td>
<td>10 Weeks</td>
</tr>
<tr>
<td>1</td>
<td>1.0%</td>
</tr>
<tr>
<td>5</td>
<td>0.0%</td>
</tr>
<tr>
<td>10</td>
<td>0.0%</td>
</tr>
<tr>
<td>15</td>
<td>0.0%</td>
</tr>
<tr>
<td>20</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

**Drift**

The previous simulations assumed idealized conditions (i.e. no forecasting error, no demand error, or SKU errors). Drift is the variation in SKU mix demanded by the customer from week to week. The next set of experiments looked at the effect of drift on the inventory levels of demand. Simulations were run looking at a drift of 1%, 2.5%, 5%, and 10%. The 1% and 2.5% drift were applied to seasonal and flat demand, and the 5% and 10% were only applied to flat demand. For the 1% drift a 0% range of deviation was applied. For the 2.5%, 5%, and 10% drift, a 2.5% range of drift deviation was applied to the simulations. For example, with a 5% drift and a 2.5% range of deviation, simulations would allow the actual drift to be between 2.5% and 7.5%.
For the 1%, the average inventory levels were not much different than the cases with no drift. To achieve a 97% service level for high lead times, the average inventory levels were somewhat higher for the 1% drift than no drift. For 2.5%, inventory levels were similar to inventory levels for no drift at an 80% service level, but were higher for a 90% service levels. The differences at a 90% service level were greater as lead time increased. For high lead times (20, 25), a 97% service level could not be achieved without carrying more than 1.5 years of inventory. Similar patterns occurred for 5% and 10% drift. But for 5% and 10% drift, inventory to satisfy more than 1.5 years of demand would need to be carried for high lead times to achieve service levels of at least 90% and 80%, respectively, and much more would be needed to achieve even higher service levels.

Table 2 shows the landed cost increase a retailer could pay for suppliers with shorter lead times for the 5% Drift with Flat Demand for a 90% service level. Note these numbers are higher than the base case in Table 1.

Table 2: Break-even Analysis of the Increase in Costs Per Garment Retailers Could Pay for Shorter Lead Times for a 90% Service Level with Flat Demand with a 5% Drift

<table>
<thead>
<tr>
<th>Current Lead Time</th>
<th>New LT</th>
<th>5 Weeks</th>
<th>10 Weeks</th>
<th>15 Weeks</th>
<th>20 Weeks</th>
<th>25 Weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>3.4%</td>
<td>8.7%</td>
<td>14.1%</td>
<td>22.8%</td>
<td>30.7%</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0%</td>
<td>5.2%</td>
<td>10.3%</td>
<td>18.7%</td>
<td>26.4%</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.0%</td>
<td>0.9%</td>
<td>12.9%</td>
<td>20.2%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.0%</td>
<td>0.0%</td>
<td>7.6%</td>
<td>14.5%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
<td>0.0%</td>
<td>6.4%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Results of Simulations-Forecasting Error

Forecasting error is defined in this section to specifically be the difference between the expected demand and the actual demand for a garment each week, (the weekly volume error). Figure 4 shows both the forecasted and actual seasonal demand pattern investigated. The average inventory levels that would need to be carried to achieve each service level were higher (as much as 12 weeks) than those necessary if the forecasting error was not present. Table 3 shows the increase in costs that could be paid for shorter lead times for this scenario.

![Figure 4: Comparison of Forecasting Error with Seasonal Demand Base Case for 5 Year Selling Period](image)

Table 3: Break-even Analysis of the Increase in Costs Per Garment Retailers Could Pay for Shorter Lead Times for a 90% Service Level with Seasonal Demand with Forecasting Error

<table>
<thead>
<tr>
<th>Current Lead Time</th>
<th>New LT</th>
<th>5 Weeks</th>
<th>10 Weeks</th>
<th>15 Weeks</th>
<th>20 Weeks</th>
<th>25 Weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>1.6%</td>
<td>4.0%</td>
<td>6.4%</td>
<td>8.6%</td>
<td>13.8%</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0%</td>
<td>2.3%</td>
<td>4.7%</td>
<td>6.9%</td>
<td>12.0%</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.0%</td>
<td>0.0%</td>
<td>2.3%</td>
<td>4.5%</td>
<td>9.5%</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.0%</td>
<td>0.0%</td>
<td>2.1%</td>
<td>7.0%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.0%</td>
<td>0.0%</td>
<td>4.8%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SIMULATION RESULTS FOR SEASONAL ITEMS

Seasonal merchandise is merchandise that has a limited selling season as compared to basic items. Usually a seasonal item is based on a holiday or a time of year such as winter. Examples of seasonal items are bathing suits or parkas. Seasonal garments have different characteristics than basic garments. Therefore, it was important to study the effect of the special characteristics of seasonal garments. Since seasonal garments have a shorter selling season, it is imperative that retailers have a good forecasting plan. If the forecasting plan is inaccurate, then the retailer will lose money in the form of lost sales or markdowns on surplus merchandise. One way to counterbalance a poor forecasting plan is to have a supplier that can provide replenishments during the selling season based on point of sale information.

Ordering Strategy Methodology

A men’s seasonal pant was studied with a selling season of 20 weeks. Four ordering scenarios were investigated. Ordering strategy one looked at the effect of no reorders during the selling season, which represents the traditional method of ordering merchandise. All inventory is purchased and carried in the stores prior to the start of the selling season. The results from this ordering strategy are used as a base case when comparing the other strategies. Without reorders, lead times were not taken into consideration since there is no replenishment. The second ordering strategy looked at the effect of one reorder, while the third and fourth ordering strategies studied the effect of having two and three reorders, respectively.

For each ordering strategy except the first, four different initial amounts of inventory were considered: 20%, 40%, 60%, and 80%. For a 60% initial inventory, 60% of the planned amount of merchandise was carried in the store prior to the start of the selling season, and the rest of the needed merchandise was ordered during the selling season. For the second, third, and fourth ordering strategies, three different supplier lead times were also studied: 1, 5, and 10 weeks. For each initial inventory amount, the optimal week of first reorder was found for strategy one, as well as the optimal time between orders for strategy two and three. The optimal week of first reorder and optimal time between reorders were found by systematically investigating all possible combinations to find the one that produced the highest overall service level. The metrics that were used to evaluate each ordering strategy were the overall service level, average inventory, and gross margin.

Analysis of Service Levels
Table 4 shows that the overall service level is considerably smaller for the traditional method than the service levels that are able to be obtained with at least one reorder. The numbers in parenthesis represent the time of the first reorder and the time between reorders (e.g., (1, 4) represents reorder starting week one with four weeks between reorders). In the traditional method, there is no chance for the retailer to adjust to any trends or SKU mix error. Within any initial drop level, higher lead times lead to significantly lower service levels. This is because the retailer would not be able to receive replenishments soon enough to prevent lost sales from not having the needed merchandise unless they carried a lot of initial inventory. Note that the first reorder generally happens after the first week. This indicates that enough information is obtained in the first week about consumer preferences for this garment to provide a good estimate of what should be reordered, and the value of getting this reorder as soon as possible outweighs the benefits of waiting to obtain more consumer information from subsequent weeks of sales. Table 4 also indicates that there is really no difference between two and three reorders.

### Table 4: Service Levels for Seasonal Product

<table>
<thead>
<tr>
<th>Traditional</th>
<th>0 Reorders</th>
<th>1 Reorder</th>
<th>2 Reorders</th>
<th>3 Reorders</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>58.3%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80% Initial</td>
<td></td>
<td>97.7% (1,0)</td>
<td>99.1% (1,4)</td>
<td>99.1% (1,2)</td>
</tr>
<tr>
<td>LT1</td>
<td>97.7% (1,0)</td>
<td>99.1% (1,4)</td>
<td>99.1% (1,2)</td>
<td></td>
</tr>
<tr>
<td>LT5</td>
<td>98.9% (2,0)</td>
<td>96.0% (1,4)</td>
<td>96.0% (1,2)</td>
<td></td>
</tr>
<tr>
<td>LT10</td>
<td>91.0% (1,0)</td>
<td>91.0% (1,4)</td>
<td>91.0% (1,2)</td>
<td></td>
</tr>
<tr>
<td>60% Initial</td>
<td></td>
<td>91.0% (1,0)</td>
<td>91.0% (1,4)</td>
<td>91.0% (1,2)</td>
</tr>
<tr>
<td>LT1</td>
<td>89.5% (1,0)</td>
<td>91.0% (1,4)</td>
<td>91.0% (1,2)</td>
<td></td>
</tr>
<tr>
<td>LT5</td>
<td>92.4% (1,0)</td>
<td>96.0% (1,4)</td>
<td>96.0% (1,2)</td>
<td></td>
</tr>
<tr>
<td>LT10</td>
<td>72.6% (1,0)</td>
<td>78.6% (1,4)</td>
<td>78.6% (1,2)</td>
<td></td>
</tr>
<tr>
<td>40% Initial</td>
<td></td>
<td>90.0% (1,0)</td>
<td>90.0% (1,4)</td>
<td>90.0% (1,2)</td>
</tr>
<tr>
<td>LT1</td>
<td>89.5% (1,0)</td>
<td>91.0% (1,4)</td>
<td>91.0% (1,2)</td>
<td></td>
</tr>
<tr>
<td>LT5</td>
<td>92.4% (1,0)</td>
<td>96.0% (1,4)</td>
<td>96.0% (1,2)</td>
<td></td>
</tr>
<tr>
<td>LT10</td>
<td>77.8% (1,0)</td>
<td>83.8% (1,4)</td>
<td>83.8% (1,2)</td>
<td></td>
</tr>
<tr>
<td>20% Initial</td>
<td></td>
<td>90.0% (1,0)</td>
<td>90.0% (1,4)</td>
<td>90.0% (1,2)</td>
</tr>
<tr>
<td>LT1</td>
<td>89.5% (1,0)</td>
<td>91.0% (1,4)</td>
<td>91.0% (1,2)</td>
<td></td>
</tr>
<tr>
<td>LT5</td>
<td>92.4% (1,0)</td>
<td>96.0% (1,4)</td>
<td>96.0% (1,2)</td>
<td></td>
</tr>
<tr>
<td>LT10</td>
<td>77.8% (1,0)</td>
<td>83.8% (1,4)</td>
<td>83.8% (1,2)</td>
<td></td>
</tr>
</tbody>
</table>

### Analysis of Gross Margins

Tables 3 and 4 showed that the smaller lead times resulted in higher service levels but also required carrying more inventory. Now, the gross margin will be analyzed to see if the smaller lead times leads to better margins. The gross margins were normalized by dividing the gross margin obtained for each problem by the gross margin for the traditional method. As seen in Table 5, The gross margin obtained for the traditional ordering strategy is much lower than those obtained with at least one reorder for most cases. In terms of gross margin, the number of reorders does not seem to matter as much since the numbers are very similar. However, they are generally much better than the traditional method. The smaller initial drops (20% and 40%) with lead times of ten produces gross margins that are worse than the traditional methods. The reason is that these strategies basically plan to have inventory shortages during the first ten weeks before the first reorder arrives. Overall, there were not any significant differences between two and three reorders.

### Table 6: Gross Margins for Seasonal Products

<table>
<thead>
<tr>
<th>Traditional</th>
<th>0 Reorders</th>
<th>1 Reorder</th>
<th>2 Reorders</th>
<th>3 Reorders</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>905,642</td>
<td>1.66</td>
<td>1.67</td>
<td>1.30</td>
</tr>
<tr>
<td>80% Initial</td>
<td></td>
<td>1.70</td>
<td>1.66</td>
<td>1.27</td>
</tr>
<tr>
<td>LT1</td>
<td>1.66</td>
<td>1.70</td>
<td>1.66</td>
<td>1.27</td>
</tr>
<tr>
<td>LT5</td>
<td>1.67</td>
<td>1.66</td>
<td>1.66</td>
<td>1.27</td>
</tr>
<tr>
<td>LT10</td>
<td>1.30</td>
<td>1.27</td>
<td>1.27</td>
<td>1.27</td>
</tr>
<tr>
<td>60% Initial</td>
<td></td>
<td>1.69</td>
<td>1.69</td>
<td>1.69</td>
</tr>
<tr>
<td>LT1</td>
<td>1.69</td>
<td>1.69</td>
<td>1.69</td>
<td>1.69</td>
</tr>
<tr>
<td>LT5</td>
<td>1.70</td>
<td>1.69</td>
<td>1.69</td>
<td>1.69</td>
</tr>
<tr>
<td>LT10</td>
<td>1.13</td>
<td>1.08</td>
<td>1.08</td>
<td>1.08</td>
</tr>
<tr>
<td>40% Initial</td>
<td></td>
<td>1.81</td>
<td>1.81</td>
<td>1.81</td>
</tr>
<tr>
<td>LT1</td>
<td>1.81</td>
<td>1.81</td>
<td>1.81</td>
<td>1.81</td>
</tr>
<tr>
<td>LT5</td>
<td>1.64</td>
<td>1.63</td>
<td>1.63</td>
<td>1.63</td>
</tr>
<tr>
<td>LT10</td>
<td>0.84</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
</tr>
<tr>
<td>20% Initial</td>
<td></td>
<td>1.82</td>
<td>1.82</td>
<td>1.82</td>
</tr>
<tr>
<td>LT1</td>
<td>1.82</td>
<td>1.82</td>
<td>1.82</td>
<td>1.82</td>
</tr>
<tr>
<td>LT5</td>
<td>1.42</td>
<td>1.40</td>
<td>1.40</td>
<td>1.40</td>
</tr>
<tr>
<td>LT10</td>
<td>0.46</td>
<td>0.45</td>
<td>0.45</td>
<td>0.45</td>
</tr>
</tbody>
</table>

### Analysis of Inventory Levels

Table 5 shows the average inventory levels specified in weeks of supply for each order strategy. This is obtained by dividing the average inventory by the planned number of units to be sold. One trend seen is that smaller initial inventory drops lead to smaller average weeks of supply. When looking at this table in isolation, it also appears that longer lead times would enable the retailer to carry smaller inventory amounts which is counter intuitive. However, recall that the longer lead times had very poor service levels. The shorter lead times on average resulted in carrying more inventory, but they did so to satisfy their customers (i.e., higher service). Also, it was apparent that two and three reorders were basically identical. At the lower initial drops, two and three reorders carried smaller amounts of inventory each week.

### Table 5: Inventory Levels Based on Demand in Weeks

<table>
<thead>
<tr>
<th>Traditional</th>
<th>0 Reorders</th>
<th>1 Reorder</th>
<th>2 Reorders</th>
<th>3 Reorders</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>58.3%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80% Initial</td>
<td></td>
<td>97.7% (1,0)</td>
<td>99.1% (1,4)</td>
<td>99.1% (1,2)</td>
</tr>
<tr>
<td>LT1</td>
<td>97.7% (1,0)</td>
<td>99.1% (1,4)</td>
<td>99.1% (1,2)</td>
<td></td>
</tr>
<tr>
<td>LT5</td>
<td>98.9% (2,0)</td>
<td>96.0% (1,4)</td>
<td>96.0% (1,2)</td>
<td></td>
</tr>
<tr>
<td>LT10</td>
<td>91.0% (1,0)</td>
<td>91.0% (1,4)</td>
<td>91.0% (1,2)</td>
<td></td>
</tr>
<tr>
<td>60% Initial</td>
<td></td>
<td>91.0% (1,0)</td>
<td>91.0% (1,4)</td>
<td>91.0% (1,2)</td>
</tr>
<tr>
<td>LT1</td>
<td>89.5% (1,0)</td>
<td>91.0% (1,4)</td>
<td>91.0% (1,2)</td>
<td></td>
</tr>
<tr>
<td>LT5</td>
<td>92.4% (1,0)</td>
<td>96.0% (1,4)</td>
<td>96.0% (1,2)</td>
<td></td>
</tr>
<tr>
<td>LT10</td>
<td>72.6% (1,0)</td>
<td>78.6% (1,4)</td>
<td>78.6% (1,2)</td>
<td></td>
</tr>
<tr>
<td>40% Initial</td>
<td></td>
<td>90.0% (1,0)</td>
<td>90.0% (1,4)</td>
<td>90.0% (1,2)</td>
</tr>
<tr>
<td>LT1</td>
<td>89.5% (1,0)</td>
<td>91.0% (1,4)</td>
<td>91.0% (1,2)</td>
<td></td>
</tr>
<tr>
<td>LT5</td>
<td>92.4% (1,0)</td>
<td>96.0% (1,4)</td>
<td>96.0% (1,2)</td>
<td></td>
</tr>
<tr>
<td>LT10</td>
<td>77.8% (1,0)</td>
<td>83.8% (1,4)</td>
<td>83.8% (1,2)</td>
<td></td>
</tr>
<tr>
<td>20% Initial</td>
<td></td>
<td>90.0% (1,0)</td>
<td>90.0% (1,4)</td>
<td>90.0% (1,2)</td>
</tr>
<tr>
<td>LT1</td>
<td>89.5% (1,0)</td>
<td>91.0% (1,4)</td>
<td>91.0% (1,2)</td>
<td></td>
</tr>
<tr>
<td>LT5</td>
<td>92.4% (1,0)</td>
<td>96.0% (1,4)</td>
<td>96.0% (1,2)</td>
<td></td>
</tr>
<tr>
<td>LT10</td>
<td>77.8% (1,0)</td>
<td>83.8% (1,4)</td>
<td>83.8% (1,2)</td>
<td></td>
</tr>
</tbody>
</table>

### CONCLUSIONS
The results of this study show that there is significant value in having a supplier with shorter lead times for basic as well as seasonal goods. For basic items, results showed that shorter lead times result in having to carry less average inventory to achieve service level targets and that these savings would allow the retailer to pay more for shorter lead times. Furthermore, decreases in inventory for shorter lead times were greater as the demand uncertainty increased in terms of larger SKU drift and forecast errors. In these cases, the retailer could afford to pay even more. For seasonal items, results showed that the ability to place at least one reorder within the selling season greatly increased the gross margin when compared to not reordering. In addition, lower lead times for receiving reorders tended to also increase the gross margin. Retailers that do not take the value of lead time and replenishment capabilities into account when making sourcing decisions may be missing significant opportunities for increased profits, including possibly sourcing products from the wrong parts of the world.

REFERENCES


BIOGRAPHY

LISA HARTMAN was an Institute of Textile Technology fellow in the College of Textiles at NC State where she completed her Masters of Science in Textile and Apparel, Technology and Management in 2006. She received a BS in Fashion Merchandising and Apparel Design from Georgia Southern.

JEFFREY A. JOINES is an Associate Professor of Textile Engineering at NC State University and is the current TE Program Director. He received his B.S.I.E., B.S.E.E., M.S.I.E, and Ph.D. from NC State University. He received the 1997 Pritsker Doctoral Dissertation Award from IIE. His research interests include evolutionary optimization, object-oriented simulation, simulation-based scheduling and supply chain optimization. His email address is <JeffJoines@ncsu.edu>.

KRISTIN A. THONEY, an Associate Professor in the Textile and Apparel, Technology and Management Dept. at NC State, joined the faculty in 2000. She earned her Ph.D. in Industrial Engineering and Operations Research and her MS in Operations Research also from NC State. Kristin’s research interests include production scheduling, logistics, and supply chain modeling. She is a member of INFORMS and IIE. Her email is <Kristin_Thoney@ncsu.edu>.

RUSSELL E. KING is a Professor of Industrial and Systems Engineering at NC State University. He received a Bachelors degree in Systems Engineering and a Masters and Ph.D. in IE from the University of Florida. His research interests include control and scheduling of production systems and demand activated manufacturing/supply systems. His email address is <king@eos.ncsu.edu>.
MODELLING THE FABRIC - CLOTHING COMPATIBILITY USING A SPECIALISED PROGRAMME

Antonela Curteza and Adela Florea
"Gherge Asachi" Technical University of Iasi
B-dul. D. Mangeron, Nr. 53bis,
70050, Iasi, Romania
e-mail: acurteza@yahoo.com

Viorica Scobioala
Technical University of Moldavia
Chisinau, Republic of Moldavia
e-mail: viorique@yahoo.fr

INTRODUCTION
An important issue to solve for the textile industry consists in establishing the compatibility between material and clothing ever since the stage of designing the fabrics. The lack of unitary data bases related to fibers, yarns, fabrics and clothing makes difficult the intersectoral circulation of information with a negative impact on the transparency of carrying on the commands. As the technical language of the yarn, fabric and clothing manufacturers is is different depending on the specific field of activity, it is necessary to provide information which might build up a database that can be also consulted by the specialists involved in such research areas.

The insurance of a quality match between the fabric and the clothing requires the study of the fabric properties, establishing the variation span of their values and, for each particular fabric, a global quality index, based on which the compatibility levels between the clothing and its destination are set. The applications of this assessment process are difficult to use in enterprises because the effect-effect report is in most cases correlated with the consumption, sometimes exaggerated, of time and human resources. In order to increase the degree of applicability within profile enterprises, for the selection stage of the best type of fabric, a computerized program was developed in order to obtain an optimization of the time required by the statistic date processing and by the precision in processing data related to the clothing compatibility.

The purpose of this study is to describe the facilities of this specialized soft and the results obtained for a group of wool type fabrics.

PRESENTATION OF THE SPECIALISED SOFT
The programme developed can be accessed through Windows 9x, ME, 200 or XP operation system. The programme can be installed in any computer file. The database is created in a current manual. The programme interface presents: the menu for navigation and package selection including the results of the experimental tests, the case specifying the type of fabric, the general buttons and a set of cases (the initial data cases, the cases including the introduced and calculated data and the cases including the graph representation of the initial data results).

The programme is structured on the following sections:
- Processing of the experimental results on the assessment of fabric characteristics, physical dimensions, graph representations;
- Quality characteristics of the fabrics, relative dimensions;
- The variation field of the quality features on quality degrees;
- Change of the dimensional indices into adimensional quality ones;
- Calculation of optional indices of the quality features and of the global index;
- Graph creation of the 3D charts;
- Calculation of the complex quality indices;
- Fabrics hierarchy based on their importance.

When starting the programme, the working window is activated, and this is made up of three cases: the working package, the “Constants” constant case and the setting case. The “Introducted and calculated data” case includes the input data (original and calculated), which refer to the following characteristics: rigidity (K<sub>EL</sub>), draping (R<sub>D</sub>%), abrasion hardness by friction (m, g), painting resistance to dry friction (points), air permeability (P<sub>A</sub>, p/minm<sup>2</sup>), vapour permeability (μ, g/m<sup>2</sup>h), hydrophilicity (γ, m/s). As an example, figure 1 shows the introductory and calculation means of the data related to vapour permeability.

Figure 1 Data related to the fabrics air permeability

FACILITIES PROVIDED BY THE PROGRAMME
The programme allows:

- To store the identity features of the fabric in an initial database (figures 2 and 3).

![Figure 2 Presentation of the general picture of the experimental results; physical dimensions](image2)

![Figure 3 Fabric characteristics; relative dimensions](image3)

- To divide the features of the fabrics into variation spans, using six quality degrees: exceptional, very good, good, satisfactory, bad and very bad;

In this respect, the identification of \(a_0\) and \(a_1\) constants (for the mathematical modelling) requires some knowledge about the reference values of the optional indicator \(d\). The determination of the reference values of the optional indicator \(d\), on four quality degrees according to the specialty literature, is replaced within this work by the reference values of the optional indicators \(d\), according to six quality degrees, establishing new sensitivity intervals based on the original system (figure 4).

In relation with the variation intervals shown in figure 4, the variation fields for each fabric feature according to the quality degrees are mentioned (figure 5).

- To determine the mathematical modelling, which helps to establish the canonical form of the equation which describes different quality features within the study (figure 6).

![Figure 4 The variation intervals of the adimensional indicators (d) and dimensional indicators (y)](image4)

![Figure 5 Delimitation of the variation field of the fabric quality features](image5)

![Figure 6 Mathematical modelling of the fabric quality features](image6)

- To calculate the complex fabric indices.

In this respect, first of all it is necessary to calculate the material quality optional index (figure 7) and then the quality global index (figure 8).

![Figure 7 Calculation of the fabric quality optional indicator and the quality global indicator](image7)
Figure 8 Calculation of the quality complex indicators

- To establish the hierarchy of the fabrics, in relation with their present destination within the clothing (figure 9).

Figure 9 The hierarchy of the fabrics according to their significance

CASE STUDY

The group of fabrics on which the programme developed was applied is represented by wool and wool-like fabrics. In order to establish the compatibility with certain kind of garments, the following multi-criteria evaluation methods were used (table 1).

<table>
<thead>
<tr>
<th>Table 1 The quality complex indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality indicators</td>
</tr>
<tr>
<td>----------------------------------------</td>
</tr>
</tbody>
</table>
| Arithmetical mean | K
| G
| H
| P
| The quality number |
| Quality global index according to the optional indicators |

Where:

- \( Q_{ij} \) – is the mean of the relative feature values;
- \( J_i \) – is the characteristic weight;
- \( n \) – is the number of the defining weight for the analysed product;
- \( Q_{ij} \) – is the mean of the \( j \) characteristic values;
- \( c_i \) – are the individual values of the dimensional indicators;
- \( m \) – is the number of characteristics included in the analysis.

Considering the wide range of the fabrics present on the market, as well as the lack of normative restrictions on the criteria of establishing their destination in the clothing product, the selection criteria of the analyzed group were based mainly on the fibre composition and the thickness. The main characteristics of the wool type fabrics are shown in tables 2 and 3.

<table>
<thead>
<tr>
<th>Table 2 Identity features of the wool type fabrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labelling</td>
</tr>
<tr>
<td>1715/329</td>
</tr>
<tr>
<td>1680/434</td>
</tr>
<tr>
<td>1167/5612</td>
</tr>
<tr>
<td>11328/723</td>
</tr>
<tr>
<td>11338/725</td>
</tr>
<tr>
<td>1191/6341</td>
</tr>
<tr>
<td>1522/331</td>
</tr>
<tr>
<td>2300</td>
</tr>
<tr>
<td>1461/329</td>
</tr>
<tr>
<td>1512/419</td>
</tr>
<tr>
<td>1511/212</td>
</tr>
<tr>
<td>1522/429</td>
</tr>
<tr>
<td>1501/210</td>
</tr>
<tr>
<td>2145/7911</td>
</tr>
<tr>
<td>1501/216</td>
</tr>
<tr>
<td>1501/220</td>
</tr>
<tr>
<td>1505/431</td>
</tr>
<tr>
<td>8581/210</td>
</tr>
<tr>
<td>1567/419</td>
</tr>
</tbody>
</table>

The complex assessment of the fabric quality involved the assessment through analysis of a global index integrating several quality features. Because the quality feature values
419


Table 8 Other relative dimensions of the fabrics

<table>
<thead>
<tr>
<th>Fabric code</th>
<th>Air permeability ( P_a, \text{P/mm}^2 )</th>
<th>Vapour permeability, ( \mu, \text{g/m}^2 \text{h} )</th>
<th>Hydropophyly, ( y, \text{m/s} )</th>
<th>( Q_0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>L 1.1</td>
<td>0.2916</td>
<td>0.1443</td>
<td>0.1818</td>
<td>0.205</td>
</tr>
<tr>
<td>L 1.2</td>
<td>0.4333</td>
<td>0.1855</td>
<td>0.0181</td>
<td>0.212</td>
</tr>
<tr>
<td>L 2.1</td>
<td>0.346</td>
<td>0.1573</td>
<td>0.1363</td>
<td>0.213</td>
</tr>
<tr>
<td>L 2.2</td>
<td>0.3566</td>
<td>0.1290</td>
<td>0.0181</td>
<td>0.161</td>
</tr>
<tr>
<td>L 2.3</td>
<td>0.1</td>
<td>0.1542</td>
<td>0.0272</td>
<td>0.393</td>
</tr>
<tr>
<td>L 3.1</td>
<td>0.214</td>
<td>0.3352</td>
<td>0.0909</td>
<td>0.213</td>
</tr>
<tr>
<td>L 4.1</td>
<td>0.73</td>
<td>0.1791</td>
<td>0.4454</td>
<td>0.451</td>
</tr>
<tr>
<td>L 5.1</td>
<td>0.0833</td>
<td>0.1840</td>
<td>0.4272</td>
<td>0.231</td>
</tr>
<tr>
<td>L 6.1</td>
<td>0.89</td>
<td>0.1569</td>
<td>0.0090</td>
<td>0.352</td>
</tr>
<tr>
<td>L 7.1</td>
<td>0.51933</td>
<td>0.1750</td>
<td>0.4454</td>
<td>0.379</td>
</tr>
<tr>
<td>L 7.2</td>
<td>0.166</td>
<td>0.1629</td>
<td>0.2454</td>
<td>0.191</td>
</tr>
<tr>
<td>L 8.1</td>
<td>0.42066</td>
<td>0.3703</td>
<td>0.5363</td>
<td>0.442</td>
</tr>
<tr>
<td>L 9.1</td>
<td>0.45833</td>
<td>1</td>
<td>1</td>
<td>0.819</td>
</tr>
<tr>
<td>L 9.2</td>
<td>0.156</td>
<td>0.1588</td>
<td>0.0181</td>
<td>0.111</td>
</tr>
<tr>
<td>L 9.3</td>
<td>0.5666</td>
<td>0.1503</td>
<td>1</td>
<td>0.572</td>
</tr>
<tr>
<td>L 9.4</td>
<td>0.61</td>
<td>0.19765</td>
<td>1</td>
<td>0.602</td>
</tr>
<tr>
<td>L 9.5</td>
<td>0.092</td>
<td>0.1616</td>
<td>1</td>
<td>0.417</td>
</tr>
<tr>
<td>L 10.1</td>
<td>0.14833</td>
<td>0.15043</td>
<td>0.51818</td>
<td>0.272</td>
</tr>
<tr>
<td>L 11.1</td>
<td>0.15866</td>
<td>0.29240</td>
<td>0.34545</td>
<td>0.265</td>
</tr>
</tbody>
</table>

Table 9 The variation limits of the \( a_n, a_1, a_2 \) coefficients

<table>
<thead>
<tr>
<th>Quality degrees</th>
<th>Type I</th>
<th>Type II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( d )</td>
<td>( y ) values</td>
</tr>
<tr>
<td>Very good</td>
<td>( \geq 0.80 )</td>
<td>( \geq 4.50 )</td>
</tr>
<tr>
<td>Good</td>
<td>( \geq 0.65 )</td>
<td>( \geq 2.18 )</td>
</tr>
<tr>
<td>Satisfactory</td>
<td>( \geq 0.37 )</td>
<td>( \geq 1.00 )</td>
</tr>
<tr>
<td>Bad</td>
<td>&lt;0.37</td>
<td>&lt;1.00</td>
</tr>
</tbody>
</table>

Table 10 Variation intervals of the adimensional (d) and dimensional (z) indicators

<table>
<thead>
<tr>
<th>Indicators</th>
<th>Quality degrees</th>
<th>Very Bad</th>
<th>Bad</th>
<th>Satisfactory</th>
<th>Good</th>
<th>Very Good</th>
<th>Exceptional</th>
</tr>
</thead>
<tbody>
<tr>
<td>d</td>
<td>&lt;0.2</td>
<td>0.2-0.37</td>
<td>0.38-0.62</td>
<td>0.63-0.8</td>
<td>0.81-0.9</td>
<td>&gt;1.00</td>
<td></td>
</tr>
<tr>
<td>y</td>
<td>&lt;0</td>
<td>0.0-0.38</td>
<td>0.39-0.77</td>
<td>0.7-1.13</td>
<td>1.1-1.49</td>
<td>&gt;1.5</td>
<td></td>
</tr>
</tbody>
</table>

The application of the specialized soft described before conducted to the results presented in tables 10-14. The calculation of the variation intervals of the adimensional indicators (d) and dimensional indicators (z) was made in the reference data correspondence (presented in table 10).

The mathematical modelling of the quality features of the wool materials is shown in table 12.

For a full description of the selected materials several assessment methods were used, the results being written in table 13.

Next, the quality complex indicators were set on types of materials in numerical order from 1 to 19, number 1 being attributed to the features with the least values depending on the established purpose. After that, the compatibility between the results of the described assessments was tested, according to the \( \chi^2 \) test. The results are shown in table 14.

CONCLUSIONS

The soft can be used within enterprises for:
- The management of designing the quality features of the fabrics during their making-up, corresponding to their present destination;
- The correct setting of the fabric – clothing compatibility within the clothing design step;
- Setting the optimal manufacture technological parameters for the processing of the fabrics.

The results after the case study allow for the following conclusions:
- The selection of the best type of material for a known destination depends on the precision of the experimental tests, on the observance of the measurement taking over conditions, as well as on the number of calculation possibilities of the applied quality index. Thus, the greater the number of methods, the better the compatibility between material-clothing product.
- The soft is easier to implement in the profile enterprises, without requiring major investments in the field of IT.

REFERENCES

Florea, A., "Research regarding the complex evaluation of the fabrics quality used in garment manufacture", PhD theses, Iasi, Romania, 1998
Scheiber, E., Lixandroiu, D., "Math-CAD", Ed. Tehnica, Bucharest, Romania, 1994, pag. 95-100
Soloviev, O., "Otena prognozirovania kacestva tekstilnih materialov, Legkaia I Piscevaia promaslennosti", Moscova, 1985

BIBLIOGRAPHIE

ANTONELA CURTEZA is the design course leader in her Department, an authorised trainer for Sympatex Technologies GmbH Germany and director of research projects. She has a number of trainings in research and design/product development management.
ADELA FLOREA is an authorised trainer for REFA – Germany, an expert in Total Quality Management and garment technology.
VIORICA SCOBIAALA has graduated the Technical University of Moldavia, Chisinau, and is an expert in CAD
### Table 11 The variation field of the quality features for the wool type fabrics

<table>
<thead>
<tr>
<th>The fabric group</th>
<th>The quality features</th>
<th>Weight</th>
<th>Quality degrees</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Very bad</td>
</tr>
<tr>
<td>wool + PES</td>
<td>Abrasion hardness by friction, $P_{ab}$, g</td>
<td>0.2116</td>
<td>&lt;0.11</td>
</tr>
<tr>
<td></td>
<td>Painting resistance to dry friction, points</td>
<td>0.2116</td>
<td>&lt;4.00</td>
</tr>
<tr>
<td></td>
<td>Rigidility, $R_{bl}$</td>
<td>0.1893</td>
<td>&lt;0.29</td>
</tr>
<tr>
<td></td>
<td>Draping, $K_D$</td>
<td>0.1198</td>
<td>&lt;8.03</td>
</tr>
<tr>
<td></td>
<td>Air permeability, $P_{a}$, m/min m$^2$</td>
<td>0.0945</td>
<td>&lt;2.60</td>
</tr>
<tr>
<td></td>
<td>Vapour permeability, $P_V$, $\mu$, g/m$^2$h</td>
<td>0.0886</td>
<td>&lt;12.45</td>
</tr>
<tr>
<td></td>
<td>Hydrophily, $\gamma$, m/s</td>
<td>0.08452</td>
<td>&lt;0.02</td>
</tr>
</tbody>
</table>

| PES              | Abrasion hardness by friction, $P_{ab}$, g | 0.2116 | <0.12 | 0.13-0.41 | 0.42-0.83 | 0.84-1.14 | 1.15-1.48 | >1.49 |
|                  | Painting resistance to dry friction, points | 0.2116 | <3.00 | 3.01-3.42 | 3.43-4.05 | 4.06-4.50 | 4.51-5.00 | >5.01 |
|                  | Rigidility, $R_{bl}$ | 0.1893 | <0.54 | 0.55-0.64 | 0.65-0.80 | 0.81-0.91 | 0.92-1.03 | >1.04 |
|                  | Draping, $K_D$ | 0.1198 | <8.38 | 8.39-18.2 | 18.21-31.18 | 31.19-40.5 | 40.53-50.90 | >50.91 |
|                  | Air permeability, $P_{a}$, m/min m$^2$ | 0.0945 | <2.87 | 2.88-6.30 | 6.31-11.34 | 11.35-4.98 | 14.99-19.01 | >19.02 |
|                  | Vapour permeability, $P_V$, $\mu$, g/m$^2$h | 0.0886 | <14.50 | 14.51-31.92 | 31.93-57.55 | 57.56-75.99 | 76.0-96.49 | >96.5 |
|                  | Hydrophily, $\gamma$, m/s | 0.0845 | <0.04 | 0.041-0.5 | 0.51-1.17 | 1.18-1.66 | 1.67-2.20 | >2.21 |

### Table 12 The mathematical modelling of the main features of fabrics used for making up caps

<table>
<thead>
<tr>
<th>Fabrics</th>
<th>Quality features</th>
<th>$Y_1=1.5$</th>
<th>$Y_2=0$</th>
<th>$x_1$ max</th>
<th>$x_3$ min</th>
<th>$a_0$</th>
<th>$a_1$</th>
<th>$Y=a_0+a_1x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>wool + PES</td>
<td>Abrasion hardness by friction, $P_{ab}$, g</td>
<td>1.5</td>
<td>0</td>
<td>0.98</td>
<td>0.11</td>
<td>-0.1928</td>
<td>1.72390</td>
<td>Y0=-0.1928+1.72390X</td>
</tr>
<tr>
<td></td>
<td>Painting resistance to dry friction, points</td>
<td>1.5</td>
<td>0</td>
<td>5.00</td>
<td>4.00</td>
<td>-6</td>
<td>1.5</td>
<td>Y1=-6+1.5X</td>
</tr>
<tr>
<td></td>
<td>Rigidility, $R_{bl}$</td>
<td>1.5</td>
<td>0</td>
<td>1.04</td>
<td>0.29</td>
<td>-0.5810</td>
<td>2.0003</td>
<td>Y2=-0.5810+2.0003X</td>
</tr>
<tr>
<td></td>
<td>Draping, $K_D$</td>
<td>1.5</td>
<td>0</td>
<td>46.82</td>
<td>8.03</td>
<td>-0.3102</td>
<td>0.0386</td>
<td>Y3=-0.3102+0.0386X</td>
</tr>
<tr>
<td></td>
<td>Air permeability, $P_{a}$, m/min m$^2$</td>
<td>1.5</td>
<td>0</td>
<td>31.17</td>
<td>2.60</td>
<td>-0.1363</td>
<td>0.03249</td>
<td>Y4=-0.1363+0.03249X</td>
</tr>
<tr>
<td></td>
<td>Vapour permeability, $P_V$, $\mu$, g/m$^2$h</td>
<td>1.5</td>
<td>0</td>
<td>35.73</td>
<td>12.45</td>
<td>-0.8018</td>
<td>0.06441</td>
<td>Y5=-0.8018+0.06441X</td>
</tr>
<tr>
<td></td>
<td>Hydrophily, $\gamma$, m/s</td>
<td>1.5</td>
<td>0</td>
<td>1.18</td>
<td>0.02</td>
<td>-0.0258</td>
<td>1.29310</td>
<td>Y6=-0.0258+1.29310X</td>
</tr>
<tr>
<td>PES</td>
<td>Abrasion hardness by friction, $P_{ab}$, g</td>
<td>1.5</td>
<td>0</td>
<td>1.48</td>
<td>0.12</td>
<td>-0.1263</td>
<td>1.0974</td>
<td>Y0=-0.1263+1.0974X</td>
</tr>
<tr>
<td></td>
<td>Painting resistance to dry friction, points</td>
<td>1.5</td>
<td>0</td>
<td>5.00</td>
<td>3.00</td>
<td>-2.25</td>
<td>0.75</td>
<td>Y1=-2.25+0.75X</td>
</tr>
<tr>
<td></td>
<td>Rigidility, $R_{bl}$</td>
<td>1.5</td>
<td>0</td>
<td>1.03</td>
<td>0.54</td>
<td>-1.6513</td>
<td>3.0569</td>
<td>Y2=-1.6513+3.0569X</td>
</tr>
<tr>
<td></td>
<td>Draping, $K_D$</td>
<td>1.5</td>
<td>0</td>
<td>50.90</td>
<td>9.38</td>
<td>-0.3387</td>
<td>0.0361</td>
<td>Y3=-0.3387+0.0361X</td>
</tr>
<tr>
<td></td>
<td>Air permeability, $P_{a}$, m/min m$^2$</td>
<td>1.5</td>
<td>0</td>
<td>19.01</td>
<td>2.87</td>
<td>-0.2664</td>
<td>0.0929</td>
<td>Y4=-0.2664+0.0929X</td>
</tr>
<tr>
<td></td>
<td>Vapour permeability, $P_V$, $\mu$, g/m$^2$h</td>
<td>1.5</td>
<td>0</td>
<td>96.49</td>
<td>14.50</td>
<td>-0.2653</td>
<td>0.0182</td>
<td>Y5=-0.2653+0.0182X</td>
</tr>
<tr>
<td></td>
<td>Hydrophily, $\gamma$, m/s</td>
<td>1.5</td>
<td>0</td>
<td>2.20</td>
<td>0.04</td>
<td>-0.0277</td>
<td>0.6944</td>
<td>Y6=-0.0277+0.6944X</td>
</tr>
</tbody>
</table>
### Table 13 Calculation of the complex quality index

<table>
<thead>
<tr>
<th>The fabric code</th>
<th>The quality complex indicator</th>
<th>Arithmetical mean, $R_j$</th>
<th>Geometrical mean, $G$</th>
<th>Harmonic mean, $H_j$</th>
<th>Polygonal area, $P$</th>
<th>Quality number, $N$</th>
<th>Quality global index, $G_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>L 1.1</td>
<td></td>
<td>0.18034</td>
<td>0.38382</td>
<td>0.12716</td>
<td>0.21370</td>
<td>0.62351</td>
<td>0.760117</td>
</tr>
<tr>
<td>L 1.2</td>
<td></td>
<td>0.21520</td>
<td>0.47333</td>
<td>0.14485</td>
<td>0.34635</td>
<td>1.04854</td>
<td>0.751092</td>
</tr>
<tr>
<td>L 2.1</td>
<td></td>
<td>0.17082</td>
<td>0.40298</td>
<td>0.13277</td>
<td>0.23543</td>
<td>0.66473</td>
<td>0.775964</td>
</tr>
<tr>
<td>L 2.2</td>
<td></td>
<td>0.18002</td>
<td>0.39934</td>
<td>0.11615</td>
<td>0.25884</td>
<td>0.81925</td>
<td>0.75983</td>
</tr>
<tr>
<td>L 2.3</td>
<td></td>
<td>0.18149</td>
<td>0.46561</td>
<td>0.17208</td>
<td>0.28499</td>
<td>0.68358</td>
<td>0.742614</td>
</tr>
<tr>
<td>L 3.1</td>
<td></td>
<td>0.14989</td>
<td>0.30043</td>
<td>0.10431</td>
<td>0.1289</td>
<td>0.43173</td>
<td>0.713837</td>
</tr>
<tr>
<td>L 4.1</td>
<td></td>
<td>0.22038</td>
<td>0.53448</td>
<td>0.19305</td>
<td>0.37640</td>
<td>0.91111</td>
<td>0.801340</td>
</tr>
<tr>
<td>L 5.1</td>
<td></td>
<td>0.22988</td>
<td>0.46775</td>
<td>0.14688</td>
<td>0.32771</td>
<td>1.01832</td>
<td>0.774040</td>
</tr>
<tr>
<td>L 6.1</td>
<td></td>
<td>0.25044</td>
<td>0.60982</td>
<td>0.20550</td>
<td>0.52423</td>
<td>1.42467</td>
<td>0.768525</td>
</tr>
<tr>
<td>L 7.1</td>
<td></td>
<td>0.26061</td>
<td>0.59881</td>
<td>0.20480</td>
<td>0.49588</td>
<td>1.34488</td>
<td>0.817095</td>
</tr>
<tr>
<td>L 7.2</td>
<td></td>
<td>0.22179</td>
<td>0.44043</td>
<td>0.13309</td>
<td>0.30399</td>
<td>0.92935</td>
<td>0.767191</td>
</tr>
<tr>
<td>L 8.1</td>
<td></td>
<td>0.248722</td>
<td>0.57947</td>
<td>0.21053</td>
<td>0.451062</td>
<td>1.16674</td>
<td>0.833158</td>
</tr>
<tr>
<td>L 9.1</td>
<td></td>
<td>0.306879</td>
<td>0.74540</td>
<td>0.27828</td>
<td>0.724432</td>
<td>1.69083</td>
<td>0.854699</td>
</tr>
<tr>
<td>L 9.2</td>
<td></td>
<td>0.114807</td>
<td>0.25277</td>
<td>0.07772</td>
<td>0.104296</td>
<td>0.32625</td>
<td>0.723145</td>
</tr>
<tr>
<td>L 9.3</td>
<td></td>
<td>0.256329</td>
<td>0.56683</td>
<td>0.20232</td>
<td>0.426257</td>
<td>1.04597</td>
<td>0.812553</td>
</tr>
<tr>
<td>L 9.4</td>
<td></td>
<td>0.255911</td>
<td>0.56480</td>
<td>0.20493</td>
<td>0.419129</td>
<td>0.99746</td>
<td>0.808786</td>
</tr>
<tr>
<td>L 9.5</td>
<td></td>
<td>0.301747</td>
<td>0.70683</td>
<td>0.24088</td>
<td>0.698707</td>
<td>1.86638</td>
<td>0.895852</td>
</tr>
<tr>
<td>L 10.1</td>
<td></td>
<td>0.262609</td>
<td>0.55339</td>
<td>0.17590</td>
<td>0.456056</td>
<td>1.33297</td>
<td>0.827924</td>
</tr>
<tr>
<td>L 11.1</td>
<td></td>
<td>0.240287</td>
<td>0.52143</td>
<td>0.16855</td>
<td>0.399330</td>
<td>1.14050</td>
<td>0.802724</td>
</tr>
</tbody>
</table>

### Table 14 Wool material hierarchy according to the coefficients of significance

<table>
<thead>
<tr>
<th>The fabric code</th>
<th>The quality complex indices</th>
<th>$\sum R_j$</th>
<th>level</th>
<th>$R_{ij}$</th>
<th>$\sum R_{ij}$</th>
<th>$\gamma_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>R_j</td>
<td>G</td>
<td>H_j</td>
<td>P</td>
<td>N</td>
<td>G_i</td>
<td>R_j</td>
</tr>
<tr>
<td>L 1.1</td>
<td>15</td>
<td>17</td>
<td>16</td>
<td>17</td>
<td>14</td>
<td>96</td>
</tr>
<tr>
<td>L 1.2</td>
<td>13</td>
<td>11</td>
<td>13</td>
<td>11</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>L 2.1</td>
<td>17</td>
<td>15</td>
<td>15</td>
<td>16</td>
<td>10</td>
<td>89</td>
</tr>
<tr>
<td>L 2.2</td>
<td>16</td>
<td>16</td>
<td>17</td>
<td>15</td>
<td>14</td>
<td>93</td>
</tr>
<tr>
<td>L 2.3</td>
<td>14</td>
<td>13</td>
<td>10</td>
<td>14</td>
<td>15</td>
<td>83</td>
</tr>
<tr>
<td>L 3.1</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>19</td>
<td>109</td>
</tr>
<tr>
<td>L 4.1</td>
<td>12</td>
<td>9</td>
<td>8</td>
<td>10</td>
<td>13</td>
<td>61</td>
</tr>
<tr>
<td>L 5.1</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>L 6.1</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>L 7.1</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>L 7.2</td>
<td>11</td>
<td>14</td>
<td>14</td>
<td>13</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>L 8.1</td>
<td>8</td>
<td>5</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>L 9.1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>L 9.2</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>18</td>
<td>113</td>
</tr>
<tr>
<td>L 9.3</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>L 9.4</td>
<td>6</td>
<td>7</td>
<td>5</td>
<td>8</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>L 9.5</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>L 10.1</td>
<td>3</td>
<td>8</td>
<td>9</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>L 11.1</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>9</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>

Sum | 16563 |
Optimization of the liquor distribution in yarn package dyeing by CFD

Fabiano Scalora, Carlo Beltramo, Marco Vanni, Giorgio Rovero and Silvio Sicardi

Department of Materials Science and Chemical Engineering, Politecnico di Torino
Corso Duca degli Abruzzi 24
10100 Torino, Italy
fabiano.scalora@polito.it

KEYWORDS
Yarn dyeing, hydrodynamics simulation, dyeing diagnostics, CFD.

ABSTRACT
The aim of this work was to simulate the fluid flow in the spindle-bobbin system using computational fluid dynamics (CFD). In the frame of a wider project finalized at dyeing machinery diagnostics, the CFD simulations in the distribution apparatus offer an interesting tool to investigate various spindle geometries and choose the best structure afore of designing industrial scale dyeing equipment. The flow behaviour has been predicted using the k-ε model in the commercial CFD package of Fluent 6.2.

INTRODUCTION
An even distribution of the dye-liquor through the bobbins packed in a dyeing machine is essential to attain dyeing uniformity: this equipment feature can be achieved by optimizing the perforated hollow spindle design. In spite of this fundamental consideration the hydraulics of these components has never been studied in detail at the machinery construction stage, thus causing maldistribution also in the textile material. Traditionally, rather than applying a thorough fluid dynamics analysis, semi-empirical criteria based on the manufacturer experience have been preferred.

The hydrodynamics of an equipment can be predicted in detail by modelling the fluid flow making use of Computational Fluid Dynamics (CFD), which is capable to solve mass, momentum and energy conservation equations, also in turbulent conditions and multiphase systems. It is widely known that the application of CFD is very common in many engineering fields (aeronautics, environmental, combustion and chemical processing, etc.), while the textile science has neglected this potential.

Depending on the initialization conditions given to the simulation program, novel geometrical arrangements can be consequentially devised. These boundary conditions include fluid properties, flow rate as well as the textile substrate permeability. As a result, innovative dyeing machinery can be foreseen, based on detailed design conditions.

SYSTEM GEOMETRY AND FUNDAMENTAL CONSIDERATIONS
A horizontal single-tube dyeing machine type API/01/273/SP by OBEM S.p.A. was considered. It is suitable to hold up twelve piled 2 kg yarn packages. A centrifugal pump induces a radial circulation of the liquor through the bobbins. In this work the dyeing kier and the distribution spindle system was considered either empty or loaded by the yarn packages. A schematic representation of the system at its closed end side is shown in Figure 1.

![Figure 1: Section of spindle-bobbin-kier system](image)

The openings in the spindle are circumferentially distributed according to a four-hole radial symmetry, while the side repartitions are 45° rotated. The controlling factor to axially balance the liquid exit velocity from the openings into each yarn package is given by the hole density as well as the distance from the inlet.

Actually, when inertial effects associated with velocity changes are dominant, the discharge flow from the holes increases toward the closed end, where the pressure is higher. On the contrary, when the frictional pressure drop dominates, the flow rate is higher at the feed end of the pipe (R.H. Perry et al., 7th edition).

Fig. 2 schematically depicts the spindle geometry.
Two types of distribution spindles (A and B configurations) have been examined up to now:

A - configuration presently used in a commercial dyeing equipment (manufactured by OBEM), where the 59 pitch-diameters are more frequent nearby the spindle inlet.

B - a basic pipe geometry where an equal number of pitch-diameters are uniformly distributed.

**CFD MODELLING**

In the frame of a wider project finalized at dyeing machinery diagnostics, CFD simulations of the distribution apparatus were carried out using the Fluent™ 6.2 commercial code, which makes use of a finite volume method. A geometrical model of a dyeing system was created using the modelling package Gambit™. The simulations were performed in a range of conditions matching industrial operation.

**Mathematical formulation**

An incompressible fluid and a steady state system were assumed as basic conditions. Yarn bobbins were described as an isotropic porous system with uniform properties along the three directions. The mean value of the viscous resistance measured along the bobbin radial direction (1.25×10¹¹ m⁻²) and the average porosity among yarns (0.65) were obtained from a previous experimental work (Beltramo et al., 2005). A porous model according to Darcy’s law was defined. The lack of isotropy of the textile substrate was considered irrelevant since both the measurements and the actual dyeing liquor flow follow identical directions.

Outside of the bobbins the flow is highly turbulent. Therefore in these regions the standard \( k-\varepsilon \) model (Launder and Spalding, 1972) for turbulence closure was coupled to mass and momentum equations to generate the predictions and compare them against the experimental measurements.

\[
\frac{\partial (\rho k)}{\partial t} + \frac{\partial (\rho U_i k)}{\partial x_i} = G_k + G_b - \rho e + \frac{\partial}{\partial x_j} \left[ \frac{\mu_t}{\sigma_k} \frac{\partial k}{\partial x_j} \right] \tag{1}
\]

\[
\frac{\partial (\rho e)}{\partial t} + \frac{\partial (\rho U_i e)}{\partial x_i} = C_1 \varepsilon \frac{e}{\kappa} [G_k + (1 - C_3 \varepsilon) G_b] - \rho C_2 \varepsilon^2 \frac{e^2}{k} + \frac{\partial}{\partial x_j} \left[ \frac{\mu_t}{\sigma_e} \frac{\partial e}{\partial x_j} \right] \tag{2}
\]

\[
\rho \frac{\partial U_i}{\partial x_i} = 0 \tag{3}
\]

\[
\frac{\partial (\rho U_i)}{\partial t} + U_j \frac{\partial (\rho U_i)}{\partial x_j} = - \frac{\partial p}{\partial x_i} + (\mu_t + \mu) \frac{\partial}{\partial x_j} \left( \frac{\partial U_i}{\partial x_j} + \frac{\partial U_j}{\partial x_i} \right) + \rho g_i + F_i \tag{4}
\]

\[
\mu_t = \rho C_{\mu} \frac{k^2}{\varepsilon} \tag{5}
\]

\( G_k \) is the rate of production of turbulence kinetic energy:

\[
G_k = \mu_t \left( \frac{\partial U_i}{\partial x_j} + \frac{\partial U_j}{\partial x_i} \right) \frac{\partial U_i}{\partial x_j} \tag{6}
\]
It is well known that the $k-\varepsilon$ model is robust and accurate for highly turbulent flows with low or no swirl components, as in this case (S.B. Pope, 2000). In this model turbulent stresses are calculated by the eddy viscosity assumption, and the eddy viscosity is evaluated by combining turbulent kinetic energy $k$ and turbulence dissipation rate $\varepsilon$ (Eqn. 1).

The model is based on the transport equations for the turbulence kinetic energy (Eqn. 3), and the dissipation rate (Eqn. 4), combined with the mass conservation equation (Eqn. 1) and the momentum equation (Eqn. 2).

The values of the closure coefficients are:

\[ C_{e1} = 1.44 \quad C_{e2} = 1.92 \quad C_\mu = 0.09 \]
\[ \sigma_k = 1.0 \quad \sigma_\varepsilon = 1.3 \]

Standard wall functions were used in all simulations to model fluid-wall interaction.

System modelling

Both 2D and 3D simulations of the system were performed. The 2D simulations consider the longitudinal section of the system and use a mesh with a total number of about $2.3 \cdot 10^3$ cells. The results obtained permitted to achieve adequate predictions on the fluid flow pattern at the holes outlet and the cell size to be adopted in the subsequent 3D configuration.

Figure 3 is a visual output of the simulations: it illustrates the velocity vector flow patterns of the liquid along the perforated spindle in a 2D view, for a flow rate of 300 L/min.

Vectors indicate the liquid streamlines and colours the magnitude in m/s. The three sections shown in the Figure 3 depict the results as a function of spindle length. Thanks to system symmetry, 3D simulations could be restricted to one quarter of the geometry, thus minimising computing requirements. Three different series of simulations were performed. The first and second series considered respectively the type A geometry distribution spindle by OBEM and the type B geometry. These two configurations were represented with about $6 \cdot 10^5$ hexahedral cells. The third simulation series was based on the same geometry as the first one, with a refined grid across the distribution spindle wall. As a convergence criterion, the invariance of residuals of all the considered equations in the $k-\varepsilon$ model was adopted, which were monitored against the iteration number.

RESULTS

Figure 4 shows the dimensionless flow rate distribution (the local flow rate to the mean flow rate ratio) for the spindle alone (no bobbin present): a flat profile around the unity corresponds to a perfect liquid distribution. Both configurations provided similar profiles with a very uneven distribution (i.e. a ratio value of 4.5 between the dead end and the inlet sections). A better distribution could have been obtained by increasing the resistance to flow and by reducing the opening diameter, which would conversely cause an unacceptable liquor exit velocity. The actual flow rate fed to each package would alternatively been obtained by adjusting the opening density.

![Figure 3: Magnitude and direction of the velocity field at the hole outlet along the spindle axial coordinate according the 2D analysis.](image-url)
Figure 4: Dimensionless flow rate vs spindle axial length for the spindle alone.

Figure 5 provides the liquor distribution by introducing the bobbin fluid dynamical resistance into the model. This additional resistance becomes the flow rate controlling parameter, thus reducing the misdistribution in the ratio range between 1.2 to 1.5.

Figure 5: Dimensionless flow rate vs spindle axial length for the spindle alone.

The pressure profiles along the type A spindle at two different configurations (alone and with bobbins) are shown in Figures 6 and 7.

Figure 6: Static pressure vs spindle axial length for the spindle alone at a flow rate of 300 L/min.

Figure 7: Static pressure vs spindle axial length for the spindle-and-bobbin system at a flow rate of 300 L/min.

Both pressure profiles are in a good qualitative agreement with the data measured by the machinery instrumentation. This represents a first validation of the data obtained by the simulation. The static pressure profile appears to be more regular in the spindle-and-bobbin system than in the unloaded spindle; moreover, the static pressure in the loaded configuration exceeds more than a magnitude order with respect to the spindle alone. The additional resistance is given by the yarn package which locally helps to redistribute the liquid flow rate.

Figures 8a and 8b show the axial velocity as a function of the axial position for a flow rate of 300 L/min considering a A-type spindle.

Figures 8a and 8b: Axial velocity vs spindle axial length for the spindle alone at a flow rate of 300 L/min.
A visual representation of the results in terms of velocity magnitude is given in Figure 9. It is a contour illustration along the perforated spindle in a 2D view, for a 3D simulation of the spindle-and-bobbin system with a flow rate of 300 L/min.

![Figure 9: Magnitude of the velocity field at the opening outlet along the spindle axial coordinate in 2D view, according the 3D analysis.](image)

Schematically, the top part of each figure (the blue zone) represents a portion of the yarn bobbin, with an almost zero-velocity field. This confirms that the yarn package operates in a laminar flow. The picture in the right hand side shows the closed end of the distribution spindle, where the velocity field dissolves against the wall. The distribution of the liquor along the spindle can be obtained by locally integrating the velocity field in a chosen cross section. The integral dimensionless flow rate across each bobbin is given in Fig. 10.

![Figure 10: Integral dimensionless flow rate vs spindle axial length for the spindle-and-bobbin system at 300 L/min.](image)

The B-type geometrical configuration appears to be more favourable to maintain an almost homogeneous liquid distribution along the spindle length, independently of the total flow rate in the system.

**CONCLUSIONS**

The results of the present study indicate that the CFD gives both good qualitative and quantitative representations of the hydrodynamics acting in the spindle distributor, as well as in a complete dyeing machinery. The CFD model allows a visualisation of the flow patterns at the outlet of the opening in the spindle and inside of the spindle itself; moreover the computation indicates that the inertial effects associated with velocity changes are dominant. The prove of this effect is given by the fact that the discharge flow from the holes increases towards the closed end, where the pressure is higher. After a suitable experimental validation, now in progress, the CFD analysis will offer an interesting and sound tool to investigate various spindle geometries and choose the best structure afore of application in actual dyeing equipment.

**ACKNOWLEDGEMENTS**

The Authors have operated within the HITEX Research Project of Regione Piemonte and OBEM S.p.A. for its technical support and a specific know-how made available to benefit this work.

**REFERENCES**


AUTHOR LISTING
<table>
<thead>
<tr>
<th>Author</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aleisa E.E.</td>
<td>249</td>
</tr>
<tr>
<td>Al-Gattoufi S.</td>
<td>17</td>
</tr>
<tr>
<td>Andreescu C.</td>
<td>133</td>
</tr>
<tr>
<td>Angeli C.</td>
<td>143</td>
</tr>
<tr>
<td>Antonescu F.</td>
<td>194</td>
</tr>
<tr>
<td>Armesto L.</td>
<td>156</td>
</tr>
<tr>
<td>Atrianfar H.</td>
<td>189</td>
</tr>
<tr>
<td>Avram D.</td>
<td>393</td>
</tr>
<tr>
<td>Ayatollahi Z.</td>
<td>189</td>
</tr>
<tr>
<td>Barella A.</td>
<td>348</td>
</tr>
<tr>
<td>Bastien J.</td>
<td>353</td>
</tr>
<tr>
<td>Bayrak A.G.</td>
<td>148</td>
</tr>
<tr>
<td>Beckelbans R.</td>
<td>181</td>
</tr>
<tr>
<td>Beltramo C.</td>
<td>423</td>
</tr>
<tr>
<td>Ben Said L.</td>
<td>235</td>
</tr>
<tr>
<td>Ben Salah K.</td>
<td>52</td>
</tr>
<tr>
<td>Bernabé J.A.</td>
<td>161</td>
</tr>
<tr>
<td>Boissière J.</td>
<td>273</td>
</tr>
<tr>
<td>Bouani F.</td>
<td>40</td>
</tr>
<tr>
<td>Boussu F.</td>
<td>387</td>
</tr>
<tr>
<td>Bruniaux P.</td>
<td>393</td>
</tr>
<tr>
<td>Byrne P.J.</td>
<td>265</td>
</tr>
<tr>
<td>Campagne J.P.</td>
<td>241</td>
</tr>
<tr>
<td>Caramihai S.</td>
<td>194</td>
</tr>
<tr>
<td>Cauvin A.</td>
<td>101</td>
</tr>
<tr>
<td>Champiromati E.</td>
<td>104</td>
</tr>
<tr>
<td>Chatzinikolaou A.</td>
<td>143</td>
</tr>
<tr>
<td>Chauvet J.</td>
<td>333</td>
</tr>
<tr>
<td>Colloc J.</td>
<td>343</td>
</tr>
<tr>
<td>Crotorescu V.</td>
<td>133</td>
</tr>
<tr>
<td>Culita J.</td>
<td>377</td>
</tr>
<tr>
<td>Curtea Z.</td>
<td>416</td>
</tr>
<tr>
<td>Dangelmaier W.</td>
<td>306</td>
</tr>
<tr>
<td>Darmoul S.</td>
<td>31</td>
</tr>
<tr>
<td>De La Cruz A.M.L.</td>
<td>313/317</td>
</tr>
<tr>
<td>Dewanns S.</td>
<td>283</td>
</tr>
<tr>
<td>Dimon C.</td>
<td>71</td>
</tr>
<tr>
<td>Djeridi R.</td>
<td>101</td>
</tr>
<tr>
<td>Domenech L.</td>
<td>156</td>
</tr>
<tr>
<td>Dussart C.</td>
<td>343</td>
</tr>
<tr>
<td>Eikel B.</td>
<td>306</td>
</tr>
<tr>
<td>Elfahri J.</td>
<td>387</td>
</tr>
<tr>
<td>Facchin P.</td>
<td>348</td>
</tr>
<tr>
<td>Fernández M.</td>
<td>173</td>
</tr>
<tr>
<td>Fischer M.</td>
<td>306</td>
</tr>
<tr>
<td>Florea A.</td>
<td>416</td>
</tr>
<tr>
<td>Fuhrmann A.</td>
<td>401</td>
</tr>
<tr>
<td>Ghédira K.</td>
<td>52/241/235</td>
</tr>
<tr>
<td>Ghelam S.</td>
<td>101</td>
</tr>
<tr>
<td>Gimeno J.</td>
<td>173</td>
</tr>
<tr>
<td>Glade M.</td>
<td>101</td>
</tr>
<tr>
<td>Gourgrand M.</td>
<td>333</td>
</tr>
<tr>
<td>Grieco A.</td>
<td>81</td>
</tr>
<tr>
<td>Grossschmidt G.</td>
<td>224</td>
</tr>
<tr>
<td>Habchi G.</td>
<td>273</td>
</tr>
<tr>
<td>Hajri-Gabouj S.</td>
<td>31</td>
</tr>
<tr>
<td>Harf M.</td>
<td>224</td>
</tr>
<tr>
<td>Harşescu F.</td>
<td>219</td>
</tr>
<tr>
<td>Hartman L.</td>
<td>410</td>
</tr>
<tr>
<td>Heavey C.</td>
<td>265</td>
</tr>
<tr>
<td>Hemapala U.M.</td>
<td>5</td>
</tr>
<tr>
<td>Hmiden M.</td>
<td>235</td>
</tr>
<tr>
<td>Hong J.</td>
<td>215</td>
</tr>
<tr>
<td>Hwang D.</td>
<td>215</td>
</tr>
<tr>
<td>Ionescu F.</td>
<td>377</td>
</tr>
<tr>
<td>Irinia B.</td>
<td>71</td>
</tr>
<tr>
<td>Jánosy J.S.</td>
<td>209</td>
</tr>
<tr>
<td>Joines J.A.</td>
<td>410</td>
</tr>
<tr>
<td>Kasper K.</td>
<td>166</td>
</tr>
<tr>
<td>Kaveh-Baghbaderani B.</td>
<td>365</td>
</tr>
<tr>
<td>Kim S.-J.</td>
<td>122</td>
</tr>
<tr>
<td>King R.E.</td>
<td>410</td>
</tr>
<tr>
<td>Koch M.</td>
<td>181</td>
</tr>
<tr>
<td>Kolar D.</td>
<td>34</td>
</tr>
<tr>
<td>Koncar V.</td>
<td>387</td>
</tr>
<tr>
<td>Ksouri M.</td>
<td>40</td>
</tr>
<tr>
<td>Kulkarni A.</td>
<td>365</td>
</tr>
<tr>
<td>Kvetonova S.</td>
<td>34</td>
</tr>
<tr>
<td>Ladbrook J.</td>
<td>126</td>
</tr>
<tr>
<td>Landalouze J.</td>
<td>173</td>
</tr>
<tr>
<td>Laroque C.</td>
<td>306</td>
</tr>
<tr>
<td>Lee M.</td>
<td>215</td>
</tr>
<tr>
<td>Lee S.</td>
<td>215</td>
</tr>
<tr>
<td>Lee W.S.</td>
<td>293</td>
</tr>
<tr>
<td>Legrand X.</td>
<td>393</td>
</tr>
<tr>
<td>Legreneur P.</td>
<td>353</td>
</tr>
<tr>
<td>Leopoulos V.</td>
<td>104</td>
</tr>
<tr>
<td>Linneweber B.</td>
<td>401</td>
</tr>
<tr>
<td>Lodewijks G.</td>
<td>65/288/313/317</td>
</tr>
<tr>
<td>Author Name</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Lokshina I.</td>
<td>199/371</td>
</tr>
<tr>
<td>Lupu C.</td>
<td>71</td>
</tr>
<tr>
<td>Maciac A.</td>
<td>133</td>
</tr>
<tr>
<td>Magaud F.-X.</td>
<td>343</td>
</tr>
<tr>
<td>Marin A.</td>
<td>133</td>
</tr>
<tr>
<td>Marzbanrad J.</td>
<td>111</td>
</tr>
<tr>
<td>Masoumi A.</td>
<td>117</td>
</tr>
<tr>
<td>Matheron F.</td>
<td>101</td>
</tr>
<tr>
<td>McLean C.R.</td>
<td>323</td>
</tr>
<tr>
<td>Mebrahtu H.</td>
<td>126</td>
</tr>
<tr>
<td>Meli C.</td>
<td>356</td>
</tr>
<tr>
<td>Michelini R.C.</td>
<td>5/256</td>
</tr>
<tr>
<td>Miciu I.</td>
<td>219</td>
</tr>
<tr>
<td>Monteil K.</td>
<td>353</td>
</tr>
<tr>
<td>Mora M.C.</td>
<td>161</td>
</tr>
<tr>
<td>Munteanu C.</td>
<td>47/194</td>
</tr>
<tr>
<td>Najibi A.</td>
<td>117</td>
</tr>
<tr>
<td>Nassehi V.</td>
<td>365</td>
</tr>
<tr>
<td>Neagu I.</td>
<td>393</td>
</tr>
<tr>
<td>Neumann L.</td>
<td>283</td>
</tr>
<tr>
<td>Nucci F.</td>
<td>81</td>
</tr>
<tr>
<td>Odubasteanu C.</td>
<td>47/194</td>
</tr>
<tr>
<td>Oprean M.</td>
<td>133</td>
</tr>
<tr>
<td>Ottjes J.A.</td>
<td>65/288/293</td>
</tr>
<tr>
<td>Petit C.</td>
<td>343</td>
</tr>
<tr>
<td>Petrescu C.</td>
<td>71</td>
</tr>
<tr>
<td>Peyman M.J.</td>
<td>111</td>
</tr>
<tr>
<td>Pierreval H.</td>
<td>31</td>
</tr>
<tr>
<td>Pipard N.</td>
<td>89</td>
</tr>
<tr>
<td>Polat F.</td>
<td>148</td>
</tr>
<tr>
<td>Popela P.</td>
<td>301</td>
</tr>
<tr>
<td>Popescu D.</td>
<td>71</td>
</tr>
<tr>
<td>Pujana-Arrrese A.</td>
<td>173</td>
</tr>
<tr>
<td>Radev D.</td>
<td>199</td>
</tr>
<tr>
<td>Radeva S.</td>
<td>371</td>
</tr>
<tr>
<td>Razzoli R.P.</td>
<td>5/256</td>
</tr>
<tr>
<td>Reuter M.</td>
<td>25</td>
</tr>
<tr>
<td>Revello E.</td>
<td>93</td>
</tr>
<tr>
<td>Revetria R.</td>
<td>93</td>
</tr>
<tr>
<td>Rijstenbrij J.C.</td>
<td>293</td>
</tr>
<tr>
<td>Röben S.A.</td>
<td>288</td>
</tr>
<tr>
<td>Rodier S.</td>
<td>333</td>
</tr>
<tr>
<td>Romanin-Jacur G.</td>
<td>348</td>
</tr>
<tr>
<td>Rovero G.</td>
<td>423</td>
</tr>
<tr>
<td>Ruescas A.</td>
<td>161</td>
</tr>
<tr>
<td>Ryan P.</td>
<td>265</td>
</tr>
<tr>
<td>Scalora F.</td>
<td>423</td>
</tr>
<tr>
<td>Schrage J.</td>
<td>181</td>
</tr>
<tr>
<td>Scobiola V.</td>
<td>416</td>
</tr>
<tr>
<td>Sepasi M.</td>
<td>189</td>
</tr>
<tr>
<td>Shao G.</td>
<td>323</td>
</tr>
<tr>
<td>Siala M.</td>
<td>241</td>
</tr>
<tr>
<td>Sicardi S.</td>
<td>423</td>
</tr>
<tr>
<td>Sklenar J.</td>
<td>301</td>
</tr>
<tr>
<td>Stefanou D.</td>
<td>377</td>
</tr>
<tr>
<td>Suh J.</td>
<td>215</td>
</tr>
<tr>
<td>Summons P.F.</td>
<td>328</td>
</tr>
<tr>
<td>Thoney K.A.</td>
<td>410</td>
</tr>
<tr>
<td>Thornton S.</td>
<td>89</td>
</tr>
<tr>
<td>Ticlea A.</td>
<td>71</td>
</tr>
<tr>
<td>Tili B.</td>
<td>40</td>
</tr>
<tr>
<td>Tonelli F.</td>
<td>93</td>
</tr>
<tr>
<td>Tornero J.</td>
<td>156/161</td>
</tr>
<tr>
<td>Tounsi J.</td>
<td>273</td>
</tr>
<tr>
<td>Touš M.</td>
<td>301</td>
</tr>
<tr>
<td>Tsikonis L.</td>
<td>104</td>
</tr>
<tr>
<td>Tsogkas C.</td>
<td>104</td>
</tr>
<tr>
<td>Udrea A.</td>
<td>71</td>
</tr>
<tr>
<td>van den Dool A.</td>
<td>288</td>
</tr>
<tr>
<td>Vanni M.</td>
<td>423</td>
</tr>
<tr>
<td>Veeke H.P.M.</td>
<td>65/293/313/317</td>
</tr>
<tr>
<td>Villars C.</td>
<td>353</td>
</tr>
<tr>
<td>Vlasenko D.</td>
<td>166</td>
</tr>
</tbody>
</table>