



# **ICCGI 2013**

The Eighth International Multi-Conference on Computing in the Global  
Information Technology

ISBN: 978-1-61208-283-7

July 21 - 26, 2013

Nice, France

## **ICCGI 2013 Editors**

John Terzakis, Intel, USA

Constantin Paleologu, University Politehnica of Bucharest, Romania

Tibor Gyires, Illinois State University, USA

# ICCGI 2013

## Foreword

The Eighth International Multi-Conference on Computing in the Global Information Technology (ICCGI 2013), held between July 21 and July 26, 2013 in Nice, France, continued a series of international events covering a large spectrum of topics related to global knowledge concerning computation, technologies, mechanisms, cognitive patterns, thinking, communications, user-centric approaches, nanotechnologies, and advanced networking and systems. The conference topics focused on challenging aspects in the next generation of information technology and communications related to the computing paradigms (mobile computing, database computing, GRID computing, multi-agent computing, autonomic computing, evolutionary computation) and communication and networking and telecommunications technologies (mobility, networking, bio-technologies, autonomous systems, image processing, Internet and web technologies), towards secure, self-defendable, autonomous, privacy-safe, and context-aware scalable systems.

We take here the opportunity to warmly thank all the members of the ICCGI 2013 Technical Program Committee, as well as the numerous reviewers. The creation of such a broad and high quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and efforts to contribute to ICCGI 2013. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations, and sponsors. We are grateful to the members of the ICCGI 2013 organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that ICCGI 2013 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the field of computing in the global information technology.

We are convinced that the participants found the event useful and communications very open. We hope that Nice, France provided a pleasant environment during the conference and everyone saved some time to enjoy the charm of this city.

### **ICCGI 2013 Chairs:**

#### **ICCGI Advisory Committee**

Constantin Paleologu, University Politehnica of Bucharest, Romania

Tibor Gyires, Illinois State University, USA

Luc Vouligny, Institut de Recherche d'Hydro-Québec - Varennes, Canada

John Terzakis, Intel, USA

Yasushi Kambayashi, Nippon Institute of Technology, Japan

Mirela Danubianu, "Stefan cel Mare" University of Suceava, Romania

#### **ICCGI Special Area Chairs**

##### **Knowledge/Cognition**

Constandinos Mavromoustakis, University of Nicosia, Cyprus

**e-Learning/Mobility**

José Rouillard, Université Lille Nord, France

**Industrial Systems**

Beata Czarnacka-Chrobot, Warsaw School of Economics, Poland

Bernhard Freudenthaler, Software Competence Center Hagenberg GmbH, Austria

## **ICCGI 2013**

### **Committee**

#### **ICCGI Advisory Committee**

Constantin Paleologu, University Politehnica of Bucharest, Romania  
Tibor Gyires, Illinois State University, USA  
Luc Vouligny, Institut de Recherche d'Hydro-Québec - Varennes, Canada  
John Terzakis, Intel, USA  
Yasushi Kambayashi, Nippon Institute of Technology, Japan  
Mirela Danubianu, "Stefan cel Mare" University of Suceava, Romania

#### **ICCGI Special Area Chairs**

##### **Knowledge/Cognition**

Constandinos Mavromoustakis, University of Nicosia, Cyprus

##### **e-Learning/Mobility**

José Rouillard, Université Lille Nord, France

##### **Industrial Systems**

Beata Czarnacka-Chrobot, Warsaw School of Economics, Poland  
Bernhard Freudenthaler, Software Competence Center Hagenberg GmbH, Austria

#### **ICCGI 2013 Technical Program Committee**

Pablo Adasme, Universidad de Santiago de Chile, Chile  
El-Houssaine Aghezzaf, Gent University, Belgium  
Werner Aigner, FAW, Austria  
Johan Akerberg, ABB Corporate Research, Sweden  
Nadine Akkari, King Abdulaziz University, Kingdom of Saudi Arabia  
Areej Al-Wabil, King Saud University - Riyadh, Saudi Arabia  
Cesar Alberto Collazos, Universidad del Cauca, Colombia  
Cristina Alcaraz, University of Malaga, Spain  
Jose M. Alcaraz Calero, University of Valencia, Spain  
Panos Alexopoulos, iSOCO S.A. - Madrid, Spain  
Ali Alharbi, The University of Newcastle, Australia  
Fernando Almeida, University of Porto, Portugal  
Hala Mohammed Alshamlan, King Saud University, Saudi Arabia  
José Enrique Armendáriz-Iñigo, Universidad Pública de Navarra, Spain  
Stanislaw Ambroszkiewicz, Institute of Computer Science Polish Academy of Sciences, Poland  
Christos Anagnostopoulos, Ionian University, Greece  
Plamen Angelov, Lancaster University, UK  
Josep Arnal Garcia, Universidad de Alicante, Spain

Ezendu Ariwa, London Metropolitan University, UK  
Mustafa Atay, Winston-Salem State University, USA  
Ali Barati, Azad University - Dezfoul Branch, Iran  
Reza Barkhi, Virginia Tech - Blacksburg, USA  
Tristan Barnett, University of South Australia, Australia  
Carlos Becker Westphall, Federal University of Santa Catarina, Brazil  
Hatem Ben Sta, University of Tunis, Tunisia  
Jorge Bernardino, Institute Polytechnic of Coimbra - ISEC, Portugal  
Ateet Bhala, Oriental Institute of Science and Technology, Bhopal, India  
Fernando Bobillo, University of Zaragoza, Spain  
Mihai Boicu, George Mason University - Fairfax, USA  
Eugen Borcoci, University 'Politehnica' of Bucharest, Romania  
Djamila Boukredera, University of Bejaia, Algeria  
Daniela Briola, University of Genova, Italy  
Xiaoqiang Cai, The Chinese University of Hong Kong, Hong Kong  
Ani Calinescu, Oxford University, UK  
George Caridakis, University of the Aegean and National Technical University of Greece, Greece  
Laura Carnevali, University of Florence, Italy  
Cheng-Yuan Chang, National United University, Taiwan  
Maiga Chang, Athabasca University, Canada  
Chantana Chantrapornchai, Silpakorn University, Thailand  
Emmanuel Chaput, IRIT-CNRS, France  
Savvas A. Chatzichristofis, Democritus University of Thrace, Greece  
Chi-Hua Chen, National Chiao Tung University - Taiwan, R.O.C.  
David Chen, University of Bordeaux, France  
Shu-Ching Chen, Florida International University - Miami, USA  
Wen-Shiung Chen (陳文雄), National Chi Nan University, Taiwan  
Zhixiong Chen, School of Liberal Arts, Mercy College - Dobbs Ferry, USA  
Albert M. K. Cheng, University of Houston, USA  
Dickson Chiu, Dickson Computer Systems, Hong Kong  
Michal Choras, University of Technology and Life Sciences Bydgoszcz (UTP), Poland  
Jose M. Claver, University of Valencia, Spain  
Francesco Colace, DIEII - Università degli Studi di Salerno, Italy  
Rebeca Cortázar, University of Deusto, Spain  
Beata Czarnacka-Chrobot, Warsaw School of Economics, Poland  
Kevin Daimi, University of Detroit Mercy, USA  
Mirela Danubianu, "Stefan cel Mare" University of Suceava, Romania  
Bernard De Baets, Gent University, Belgium  
José de Oliveira Guimarães, Federal University of São Carlos, Brazil  
Vincenzo Deufemia, Università di Salerno - Fisciano, Italy  
Kamil Dimililer, Near East University - Nicosia, Cyprus  
Alexandre Dolgui, Ecole des Mines de Saint-Etienne, France  
Ludek Dolihal, Masaryk University - Brno, Czech Republic  
Juan Carlos Dueñas, ETSI Telecomunicación, Universidad Politécnica de Madrid, Spain  
Chanaka Edirisinghe, The University of Tennessee - Knoxville, USA  
Hans-Dieter Ehrich, Technische Universität Braunschweig, Germany  
Nabil El Kadhi, Ahlia University, Kingdom of Bahrain  
Moez Essehghir, Technology University of Troyes, France

Fausto Fasano, University of Molise, Italy  
Maria Fazio, University of Messina, Italy  
Javier Dario Fernandez Ledesma, Universidad Pontificia Bolivariana - Medellín, Colombia  
Kurt B. Ferreira, Sandia National Laboratories, USA  
Joerg Fliege, The University of Southampton, UK  
Rana Forsati, Shahid Beheshti University - Tehran, Iran  
Panagiotis Fotaris, University of Macedonia - Thessaloniki, Greece  
Rita Francese, Università di Salerno - Fisciano, Italy  
Bernhard Freudenthaler, Software Competence Center Hagenberg GmbH, Austria  
Matjaz Gams, Jozef Stefan Institute - Ljubljana, Slovenia  
Raghu Ganti, IBM Thomas J. Watson Research Center, U.S.A.  
Félix J. García, University of Murcia, Spain  
Vanessa Gardellin, Institute for Informatics and Telematics CNR, Pisa, Italy  
David Garcia Rosado, University of Castilla-La Mancha, Spain  
Joseph Andrew Giampapa, Carnegie Mellon University, USA  
Debasis Giri, Haldia Institute of Technology, India  
Apostolos Gkamas, University Ecclesiastical Academy of Vella of Ioannina, Greece  
Gustavo González, Mediapro Research - Barcelona, Spain  
Feliz Gouveia, University Fernando Pessoa, Portugal  
William Grosky, University of Michigan, USA  
Francesco Guerra, University of Modena and Reggio Emilia, Italy  
Carlos Guerrero, Universitat de les Illes Balears Palma de Mallorca, Spain  
Nalan Gulpinar, University of Warwick - Coventry, UK  
Chris Guy, University of Reading, UK  
Tibor Gyires, Technology Illinois State University, USA  
Maki K. Habib, The American University in Cairo, Egypt  
Jana Hájková, University of West Bohemia, Pilsen, Czech Republic  
Hani Hamdan, École Supérieure d'Électricité (SUPÉLEC), France  
Petr Hanáček, Brno University of Technology, Czech Republic  
Sven Hartmann, TU-Clausthal, Germany  
Mohammad Mehedi Hassan, King Saud University, Kingdom of Saudi Arabia  
Iem Heng, New York City College of Technology, USA  
Wladyslaw Homenda, Warsaw University of Technology, Poland  
Wei-Chiang Hong, Oriental Institute of Technology, Taiwan  
Jun Hu, Eindhoven University of Technology, The Netherlands  
Tee Sim Hui, Multimedia University, Malaysia  
Rosziati Ibrahim, Universiti Tun Hussein Onn Malaysia (UTHM), Malaysia  
Larisa Ismailova, National Research Nuclear University "MEPhI" - Moscow, Russia  
Kyoko Iwasawa, Takushoku University - Tokyo Japan  
Helge Janicke, De Montfort University, UK  
Mehrshid Javanbakht, Azad University - Tehran, Iran  
Guorui Jiang, Beijing University of Technology, China  
Maria João Ferreira, Universidade Portucalense - Porto, Portugal  
Paul Johannesson, Royal Institute of Technology - Stockholm, Sweden  
Matjaz B. Juric, University of Ljubljana, Slovenia  
Imed Kacem, Université de Lorraine, France  
Hermann Kaindl, TU-Wien, Austria  
Yasushi Kambayashi, Nippon Institute of Technology, Japan

Georgios Kambourakis, University of the Aegean - Samos, Greece  
Byoung Uk Kim, US Air Force Research Laboratory, USA  
DongKyun Kim, Korea Institute of Science and Technology Information, Daejeon, South Korea  
Georgios Kioumourtzis, Center for Security Studies, Greece  
Wojciech Kmiecik, Wroclaw University of Technology, Poland  
Mehmet Koc, Bilecik Seyh Edebali University, Turkey  
Leszek Koszalka, Wroclaw University of Technology, Poland  
Janet L. Kourik, Webster University - St Louis, USA  
Piotr A. Kowalski, SRI Polish Academy of Sciences and Cracow University of Technology, Poland  
Bartosz Krawczyk, Wroclaw University of Technology, Poland  
Panos Kudumakis, Queen Mary University of London, UK  
Gyu Myoung Lee, Telecom SudParis, France  
Tracey K.M. Lee, School of EEE Singapore, Singapore  
Arno Leist, Massey University, New Zealand  
Daniel Lemire, Université du Québec à Montréal (UQAM), Canada  
Ricardo Lent, Imperial College - London, UK  
Isaac Lera, Universitat de les Illes Balears Palma de Mallorca, Spain  
Tiberiu S. Letia, Technical University of Cluj-Napoca, Romania  
Chendong Li, University of Connecticut - Storrs, USA  
Abdel Lisser, University of Paris Sud, LRI - Orsay, France  
Arne Løkketangen, Molde University College, Norway  
Angela Locoro, University of Genova, Italy  
Francesca Lonetti, CNR-ISTI, Italy  
Pericles Loucopoulos, Harokopio University of Athens, Greece / Loughborough University, UK  
Alen Lovrencic, University of Zagreb, Croatia  
Szymon Lukasik, Cracow University of Technology, Poland  
Stephane Maag, Telecom SudParis, France  
Olaf Maennel, Loughborough University, UK  
José María Luna, University of Córdoba, Spain  
Francesco Maiorana, University of Catania, Italy  
Viliam Makis, University of Toronto, Canada  
Giuseppe Mangioni, University of Catania, Italy  
Yannis Manolopoulos, Aristotle University - Thessaloniki, Greece  
Gregorio Martinez, University of Murcia, Spain  
Maristella Matera, Politecnico di Milano, Italy  
Jean-Denis Mathias, National Research Institute of Science and Technology for Environment and Agriculture (IRSTEA), France  
Constandinos Mavromoustakis, University of Nicosia, Cyprus  
Natarajan Meghanathan, Jackson State University, USA  
Angelos Michalas, Technological Educational Institute of Western Macedonia, Greece  
Jerzy Michnik, University of Economics in Katowice, Poland  
Marius Minea, University Politehnica of Bucharest, Romania  
Vladimir Modrak, Technical University of Kosice - Presov, Slovakia  
Lars Moench, FernUni-Hagen, Germany  
Ghodrat Moghadampour, Vaasa University of Applied Sciences Technology and Communication, Finland  
Valérie Monfort, Université Paris 1 Panthéon Sorbonne, France  
Paula Morais, Universidade Portucalense - Porto, Portugal  
Mary Luz Mouronte López, Ericsson S.A., Spain

Isabel Muench, German Federal Office for Information Security), Germany  
Antonio Muñoz, University of Malaga, Spain  
Phivos Mylonas, Ionian University, Greece  
Pablo Najera, University of Malaga, Spain  
Tomoharu Nakashima, Osaka Prefecture University, Japan  
Joan Navarro, Ramon Llull University, Spain  
Antonio Navarro Martín, Universidad Complutense de Madrid, Spain  
Leila Nemmiche Alachaher, ISIMA, France  
Marcio Katsumi Oikawa, Universidade Federal do ABC, Brazil  
Hichem Omrani, CEPS/INSTEAD, Luxembourg  
Kok-Leong Ong, Deakin University, Australia  
Constantin Paleologu, University Politehnica of Bucharest, Romania  
Thanasis G. Papaioannou, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland  
Marcin Paprzycki, Systems Research Institute / Polish Academy of Sciences - Warsaw, Poland  
Mukaddim Pathan, Telstra Corporation Limited, Australia  
Al-Sakib Khan Pathan, International Islamic University Malaysia, Malaysia  
Kunal Patel, Ingenuity Systems, USA  
Jose J. Pazos Arias, Universidad Vigo, Spain  
Marek Penhaker, VSB - Technical University of Ostrava, Czech Republic  
Yoseba K. Peña, University of Deusto, Basque Country, Spain  
Andrea Perego, European Commission - Joint Research Centre, Italy  
Fernando Pereñíguez García, University of Murcia, Spain  
Dana Petcu, Western University of Timisoara, Romania  
Willy Picard, Poznan University of Economics, Poland  
Selwyn Piramuthu, University of Florida, USA  
Jan Platoš, VSB-Technical University of Ostrava, Czech Republic  
Kornelije Rabuzin, University of Zagreb - Varazdin, Croatia  
Thurasamy Ramayah, Universiti Sains Malaysia, Malaysia  
Stefan Rass, Universitaet Klagenfurt, Austria  
Danda B. Rawat, Eastern Kentucky University, USA  
Marek Reformat, University of Alberta - Edmonton, Canada  
Luis Paulo Reis, University of Minho/LIACC, Portugal  
Paolo Romano, INESC-ID/IST, Portugal  
Agos Rosa, LaSEEB, Portugal  
José Rouillard, Université Lille Nord de France  
Pawel Rózycki, University of Information Technology and Management - Rzeszow, Poland  
Serap Sahin, Izmir Institute of Technology, Turkey  
Ozgur Koray Sahingoz, Turkish Air Force Academy, Turkey  
Manuel Filipe Santos, Universidade do Minho - Guimarães, Portugal  
Ana Šaša, University of Ljubljana, Slovenija  
Peter Schartner, Klagenfurt University, Austria  
Isabel Seruca, Universidade Portucalense - Porto, Portugal  
Marc Sevaux, Université de Bretagne-Sud, France  
Qiao Shaojie, Southwest Jiaotong University, China  
Ashok Sharma, Mahindra Satyam, India  
Mei-Ling Shyu, University of Miami - Coral Gables, USA  
Patrick Siarry, Université Paris 12 (LiSSi) - Creteil, France  
Spiros Sirmakessis, Technological Educational Institution of Messolongi, Greece



Tomas Skersys, Kaunas University of Technology, Lithuania  
Martin Stanton, Manchester Metropolitan University, UK  
Kathryn E. Stecke, University of Texas at Dallas - Richardson, USA  
Anca-Juliana Stoica, Uppsala University, Sweden  
Renate Strazdina, Ernst&Young Baltic SIA, Latvia  
Vadim Strijov, Computing Centre of the Russian Academy of Sciences - Moscow, Russia  
Weifeng Sun, Dalian University of Technology, China  
Ryszard Tadeusiewicz, AGH University of Science and Technology, Poland  
John Terzakis, Intel, USA  
Zenonas Theodosiou, Cyprus University of Technology, Cyprus  
Ousmane Thiare, Gaston Berger University of Saint-Louis, Senegal  
Maria Tortorella, University of Sannio, Italy  
Guglielmo Trentin, National Research Council - Genoa & University of Turin, Italy  
Chrisa Tsinaraki, Technical University of Crete, Greece  
Meltem Sonmez Turan, NIST, USA  
Ion Tutanescu, University of Pitesti, Romania  
Theodoros Tzouramanis, University of the Aegean, Greece  
Eleni I. Vlahogianni, National Technical University of Athens, Greece  
José Vasconcelos, Instituto Superior de Tecnologias Avançadas de Lisboa, Portugal  
Maria-Esther Vidal, Universidad Simon Bolivar - Caracas, Venezuela  
Ante Vilenica, University of Hamburg, Germany  
Patravadee Vongsumedh, Bangkok University, Thailand  
Luc Vouligny, Institut de Recherche d'Hydro-Québec - Varennes, Canada  
Mihaela Vranic, University of Zagreb, Croatia  
Chieh-Yih Wan, Intel Corporation, USA  
Gerhard-Wilhelm Weber, METU - Ankara, Turkey  
Viacheslav Wolfengagen, Institute "JurInfoR-MSU", Russia  
Ouri Wolfson, University of Illinois, USA  
Min Wu, Oracle Inc., USA  
Mudasser F. Wyne, National University - San Diego, USA  
Farouk Yalaoui, University of Technology of Troyes, France  
Xia Yan, Hunan University, China  
Yugang Yu, The University of Science and Technology of China, China  
Fernando Zacarias Flores, Benemerita Universidad Autonoma de Puebla, Mexico  
Marcelo Zanchetta do Nascimento, Federal University of ABC, Brazil  
Xuying Zhao, University of Notre Dame, USA  
Zuqing Zhu, University of Science and Technology, China  
Iveta Zolotová, Technical University of Kosice, Slovakia

## Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission to reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

## Table of Contents

Restoring Information Needed for Social Internetworking Analysis from Anonymized Data <i>Francesco Buccafurri, Daniele Caridi, Gianluca Lax, Antonino Nocera, and Domenico Ursino</i>	1
The Impact of the Internet and the World Wide Web On Distance and Collaborative Learning <i>Franklyn Chukwunonso, Roliana Binti Ibrahim, Ali Bin Selamat, Adamu Idama, and Wadzani A. Gadzama</i>	7
A New Approach Based on Computer Vision and Collaborative Social Networking for Environmental Preservation: Theory, Tools and Results of Italian ACI Project <i>Maria Grazia Albanesi and Roberto Albanesi</i>	16
Impact of the Network Structure on the SIR Model Spreading Phenomena in Online Networks <i>Marek Opuszko and Johannes Ruhland</i>	22
Using Quick Response Codes For Student Interaction During Lectures <i>Robert Law</i>	29
Administration of Knowledge Assessment at Riga Technical University <i>Natalija Prokofjeva, Alla Anohina-Naumeca, and Olga Lebedeva</i>	34
ICT Related Tasks and Challenges in the New Model of Technical Teacher Training <i>Gyorgy Molnar and Andras Benedek</i>	40
The Impact of Requirements on Software Quality Across Three Product Generations <i>John Terzakis</i>	45
A Review of Domain-Specific Modelling and Software Testing <i>Teemu Kanstren</i>	51
Lack of Software Engineering Practices in the Development of Bioinformatics Software <i>Dhawal Verma, Jon Gesell, Harvey Siy, and Mansour Zand</i>	57
Pinpoint Analysis of Software Usability <i>Divya K. V. Dasari, Dan E. Tamir,, Oleg V. Komogortsev,, Gregory R. LaKomski, and Carl J. Mueller</i>	63
Gender Classification of Face with Moment Descriptors <i>Wen-Shiung Chen, Po-Yi Lee, and Lili Hsieh</i>	72
Artificial Immune Memory Formed by Dynamics of Antibody Networks <i>Chung-Ming Ou and Chung-Jen Ou</i>	76

High Performance Grid Environment for Parallel Multiple Biological Sequence Alignment <i>Plamenka Borovska, Veska Gancheva, and Nikolay Landzhev</i>	82
A New Method of Vehicle Initiative Safety: Heart Sound Acquisition and Identification Technology <i>Cheng Yu-han and Ma Yong</i>	88
Global Illumination-Invariant Fast Sub-Pixel Image Registration <i>Andrew Gilman and Arno Leist</i>	95
Filtering of Large Signal Sets: An Almost Blind Case <i>Anatholi Torokhti, Phil Howlett, and Hamid Laga</i>	101
A P300-Based Word Typing Brain Computer Interface System Using a Smart Dictionary and Random Forest Classifier <i>Faraz Akram, Hee-Sok Han, and Tae-Seong Kim</i>	106
Research and Application of the Radar Intelligent Fault Diagnosis System Based on Dual-Mode Fusion <i>Cheng Xie-feng and Cheng Hui-zhong</i>	110
Automated Analysis of CT Slices for Detection of Ideal Midline from Brain CT Scans <i>Xuguang Qi, Sharad Shandilya, Ashwin Belle, Rosalyn Hargraves, Charles Cockrell, Yang Tang, Kevin Ward, and Kayvan Najarian</i>	117
Actual Brain Midline Detection using Level Set Segmentation and Window Selection <i>Xuguang Qi, Ashwin Belle, Sharad Shandilya, Rosalyn Hargraves, Charles Cockrell, Yang Tang, Kevin Ward, and Kayvan Najarian</i>	122
Detection of Brain Tumor Using Zernike Moments on Magnetic Resonance Images <i>Kiran Thapaliya and Goo-Rak Kwon</i>	127
A Meta Model-Based Web Framework for Domain Independent Data Acquisition <i>Dominic Girardi, Johannes Dirnberger, and Johannes Trenkler</i>	133
Cancer and Deadly Infection in Institutions: Developing Use Cases for an MBE Application to Prevent another Enron or Barings <i>Thang Nguyen</i>	139
Temporal Data Management <i>Michal Kvet and Anton Lieskovsky</i>	145
Evaluation of Visual structure for Industrial size Software Product Line Architecture <i>Abeer Khalid and Salma Imtiaz</i>	152

A Novel Multiple Attributes Decision Making Approach For Multimedia Session Selection <i>Tein-Yaw Chung, Ibrahim Mashal, Fong-Ching Yuan, Yuan-Hao Chiang, and Osam Alsaryrah</i>	158
A Novel Algorithm for Selecting Multimedia Network Connections in Next Generation Networks <i>Tein-Yaw Chung, Ibrahim Mashal, Fong-Ching Yuan, Yuan-Hao Chiang, and Osama Alsaryrah</i>	164
Reducing Power Consumption using Improved Wakelock on Android Platform <i>Joonkyo Kim and Jaehyun Park</i>	171
Using an Individual-Based Model of Uneven-Aged Forests for Studying Trade-off Between Timber Production and Deadwood Preservation <i>Bruno Bonte, Valentine Lafond, Thomas Cordonnier, and Jean-Denis Mathias</i>	175
Stability analysis of global FCFS and presorting service discipline <i>Willem Melange, Joris Walraevens, Dieter Claeys, Bart Steyaert, and Herwig Bruneel</i>	181
Development of Logistical Model Based on Integration of Ontology, Multi-Agent Approach and Simulation <i>Konstantin Aksyonov, Eugene Bykov, Olga Aksyonova, and Alena Nevolina</i>	188
An Eclipse Plug-in for Aspect-Oriented Bidirectional Engineering <i>Oscar Pulido-Prieto and Ulises Juarez-Martinez</i>	195
Strategic Goal Oriented Supplier Selection <i>Chang Joo Yun and Chung-Hsing Yeh</i>	202
Enacting a Requirement Engineering Process with Meta-Tools: an Exploratory Project <i>Sana Damak Mallouli and Said Assar</i>	208
Finding an Optimal Model for Prediction of Shock Outcomes through Machine Learning <i>Sharad Shandilya, Xuguang Qi, Kayvan Najarian, Kevin Ward, Michael Kurz, and Rosalyn Hargraves</i>	214
A Design of Hybrid Automatic Repeat Request Scheme based on FlexRay used for Smart Hybrid Powerpack <i>Leilei Shi, Jekwang Choi, and Hunmo Kim</i>	219
Sensor Web Deployment Using Informed Virtual Geographic Environments <i>Mehdi Mekni</i>	224
Block Algorithm and Its Implementation for Cholesky Factorization <i>Jianping Chen, Zhe Jin, Quan Shi, Jianlin Qiu, and Weifu Liu</i>	232
Multiagent Genetic Optimization to Solve the Project Scheduling Problem <i>Konstantin Aksyonov and Anna Antonova</i>	237

Robust Optimization for Stochastic Wireless CDMA/TDMA Networks <i>Belarmino Nunez, Pablo Adasme, Ismael Soto, and Abdel Lisser</i>	244
Towards an Efficient Handling of the Maximum Triangle Packing Problem <i>Youcef Abdelsadek, Francine Herrmann, Imed Kacem, and Benoit Otjacques</i>	249
Trial Testing Efficiency of Algorithms for Task Execution in Multi-Processor Systems <i>Magdalena Respondek, Leszek Koszalka, Iwona Pozniak-Koszalka, and Andrzej Kasprzak</i>	253
Preprocessing of Binary Executable Files Towards Retargetable Decompilation <i>Jakub Kroustek and Dusan Kolar</i>	259
EKOCA: Energy Aware Overlapping Multihop Clustering for Wireless Sensor Networks <i>Eman Ramadan, Moustafa A. Youssef, Magdy Abd-ElAzim Ahmed, and Mohamed Nazih El-Derini</i>	265
Quantum Key Distribution over Collective Amplitude Damping Quantum Channels <i>Elloa B. Guedes and Francisco M. de Assis</i>	271
Live Replication of Virtualized VoIP Servers <i>Jiri Hlavacek and Robert Bestak</i>	277
Reasoning About Consistency of Relational Knowledge Bases <i>Tadeusz Pankowski</i>	283
Predicting Early Students with High Risk to Drop Out of University using a Neural Network-Based Approach <i>Miguel Angel Gil Rios, Norma Griselda Reyes Avila, Maria Dolores Juarez Ramirez, Emmanuel Espitia Rea, Julio Cesar Mosqueda Gomez, and Myriam Soria Garcia</i>	289
A Kinetic Light Shelf Unit as an Integrated Intelligent Control Device for Optimizing Interior Illumination <i>Ok-Kyun Im, Kyoung-Hee Kim, and Seung-Hoon Han</i>	295

# Restoring Information Needed for Social Internetworking Analysis from Anonymized Data

Francesco Buccafurri, Daniele Caridi, Gianluca Lax, Antonino Nocera and Domenico Ursino  
 Dept. of Information, Infrastructure and Sustainable Energy Engineering  
 University "Mediterranea" of Reggio Calabria  
 Reggio Calabria, Italy  
 Email: {bucca,daniele.caridi,lax,a.nocera,ursino}@unirc.it

**Abstract**—The interaction among distinct social networks is the basis of a new emergent internetworking scenario (called *Social Internetworking Scenario* or, simply, SIS), enabling a lot of strategic applications whose main strength will be just the integration of possibly different communities yet preserving their diversity and autonomy. As a consequence, studying this new scenario from a Social-Network-Analysis perspective is certainly an important and topical issue, also for the possibility of discovering a lot of relevant knowledge about multiple aspects of people life. However, not always the analyst is able to deal with the hard problem of collecting data through the execution of a crawler. In this case, she could exploit graph-based social data, collected by another party, and usually anonymized for privacy reasons. Unfortunately, even the most frequent and trivial anonymization (i.e., the elimination of URLs associated to nodes), handicaps a lot of SIS-oriented investigations, due to the lack of some relevant explicit information. In this paper, we deal with this problem, by proposing and by experimentally validating a clustering-based technique able to restore part of the missing explicit information, thus allowing the profitable analysis of anonymized multi-social-network data.

**Keywords**-*Social Network; Social Network Analysis; Social Internetworking System; Anonymized Data; Clustering*

## I. INTRODUCTION

In the last years, (on line) social networks have become one of the main actors not only of the Cyberspace but also of real life. Indeed, an always increasing number of persons joins one or more social networks, and large areas of economy, politics, communication and, more in general, of all the aspects of real life, make a large use of this innovative tool. The extraordinary development of this phenomenon has led to the presence of several different social networks, whose number is expected to increase, each incorporating peculiar features making it more or less attractive for the market. Thus, a user joins Facebook to talk to her friends, YouTube to share her videos, Flickr to store her photos, LinkedIn to look for a job, and so on. The resulting scenario is not the one of single, isolated, independent social networks, but a universe composed of a constellation of several social networks, each forming a community with specific connotations, but strongly interconnected with each other.

It is a matter of fact that, despite the inherent underlying heterogeneity, the interaction among distinct social networks is the basis of a new emergent internetworking scenario enabling a lot of strategic applications, whose main strength will be just the integration of possibly different communities yet

preserving their diversity and autonomy. Clearly, social mining and analysis approaches should not miss this huge multi-network source of information, which also reflects multiple aspects of people personal life [19], thus enabling a lot of powerful discovering activities. As a matter of fact, this scenario represents the natural substrate for the development of multi-social-network applications, following a process that is already started (for instance, think of Google Open Social [4], Power.com [5], Gathera [3] and Friendfeed [2]), and will lead to increasingly powerful and innovative systems, thus increasing the importance of the so called field of *social internetworking systems* [28], [14], and determining the raising attraction of both research and industry.

Classical social networks have been studied since several years, initially by sociologists and, then, with the advent of on line social networks, by computer science researchers [20]. Studying *Social Internetworking Scenarios* (SIS's, for short) [9], [11] means analyzing specific aspects (mainly concerning the interconnection among different social networks [10]) and adopting investigation methodologies which cannot be trivially derived from single-social-network analyses. However, the starting point is always the same.

In order to carry out meaningful analyses, we must have the availability of sufficiently large social dataset, extracted from real-life social networks, including as much information as possible. However, not always the analyst is able to deal with the hard problem of collecting data through the execution of a crawler. In fact, this is a very time-consuming activity also requiring highly performing hardware. In this (frequent) case, she could exploit data made available on the Internet by researchers who collected them in the past. For instance, Mislove et al. [26] make available a set of 11 millions users and 328 millions links among them referred to four different social networks. As it typically happens also in the context of data mining, the party that conducts the analysis on data is different from the party that collects them.

Hence, the party that exports data has also to deal with privacy issues, by purifying data from information which can be related to individuals. In the context of social network analysis, this means that collected graph-data, which represent friendships among different users, do not allow a node of the graph to be related with a real user of the social network [27]. Thus, URLs associated to nodes cannot be given.

Observe that an anonymization based on an alphabetic substitution encryption done on URLs is not secure, because

it is well known that alphabetic substitution is vulnerable w.r.t. frequency-analysis attacks. Therefore, in privacy-aware contexts, anonymized social data simply consist in a graph, with no information about the URLs associated to nodes. However, in a SIS context, working with such anonymized data results in the lost of relevant explicit information. As a matter of fact, the evidence of the existence of different social networks and their interconnections is lost. Clearly, this problem does not occur whenever a single social network is investigated.

In this paper, we deal with the issue of restoring part of the missing explicit information, crucial for SIS-oriented analysis, by partitioning the whole graph in subgraphs, each corresponding as much as possible to an original social network, and, as difference, by discovering the interconnections among social networks. Once this task has been done, the analyst will be able to conduct most of the SIS-oriented investigations in a direct way, starting from the basic information concerning the membership of two or more users either to the same or to distinct social networks.

In order to carry our investigation, it appears reasonable to apply clustering techniques. Clustering represents one of the most important sectors of Data Mining. It aims at grouping a set of objects into homogeneous groups called clusters. In the Data Mining context, clustering has been exploited in a very large number of application contexts, and often extremely important results have been obtained. Despite the seemingly high execution time, clustering has been extensively exploited to analyze single social networks [34], [15], [12], [30], [22], [18], [31], [35]. Indeed, there exists a great number of clustering techniques specifically conceived to operate on large datasets. However, to the best of our knowledge, it was never adopted to investigate SIS's. Nevertheless, the intrinsic nature of a SIS makes it well suited to be investigated by means of clustering. As a matter of fact, a SIS can be seen as a set of social networks which cooperate each other. Each social network can be seen as a cluster. Therefore, the analysis of the features of a cluster can contribute to define the features of the corresponding social network.

In this paper, we verify the effectiveness of this idea. For this purpose, first we applied a crawling technique (namely, BFS [23]) on the SIS to extract a sample of data. The considered SIS consists of 5 social networks, namely Twitter, YouTube, Flickr, MySpace, LiveJournal. We chose these social networks because they are compliant with FOAF [8] and XFN [7] standards. These allow a crawler to access some not private information stored in the corresponding social networks. Then, we anonymized these data and we applied different clustering techniques on them. Finally, we compared obtained clusters with the original social networks to test the accuracy of clustering techniques in finding clusters coincident with the original social networks. In this activity, we did not exploit anonymized data available on the Web since they did not report the social network associated to each anonymized node and, hence, they were not able to allow us to measure the accuracy of the clustering techniques.

This paper is organized as follows: in Section II, we examine related literature. In Section III, we illustrate in detail the experimental analysis aimed at testing the capability of some clustering techniques to obtain clusters coincident

with the original social networks of a SIS when data are anonymized. Finally, in Section IV, we draw our conclusions.

## II. RELATED WORK

In the context of Social Network Analysis [12], many of the approaches follow techniques based on clustering, analogously to what we propose in this paper. As a matter of fact, despite the seemingly high execution time, clustering has been extensively exploited to analyze single social networks. In this section, we survey the main contributions that clustering techniques have given to the investigation of social networks.

In [34], social network users and their relationships are modeled by means of weighted graphs, and a suitable measure of the density of a sub-graph, which is an index of user correlation, is defined. The proposed density measure is used to decide whether adding a node to a cluster. The clustering task of this approach aims at detecting community structures in a social network.

The authors of [15] use a genetic algorithm as a heuristic search technique to cluster social networks. This algorithm allows the social network graph to be represented in a succinct way, thus increasing the efficiency of the clustering algorithm. In order to identify clusters, it adopts a distance measure based on random walks. The results of the experimental evaluation show that the adoption of random-walk-based distances, instead of the Euclidean ones, returns more accurate clusters. [12] proposes a methodology to discover possible aggregations of nodes covering specific positions in a graph (e.g., central nodes), as well as very relevant clusters.

In [30], the authors carry out a study of temporal relation co-clustering on directional social network and author-topic evolution. Interestingly enough, the corresponding approach includes the time dimension typically ignored by the other related approaches. The analysis performed by these authors obtains meaningful results about evolution patterns, the evolution of publication topics and the evolution of e-mail communication patterns over time.

The authors of [22] developed a model and a Bayesian technique to infer four features in a social network. These features are: (i) *transitivity* - if A relates to B and B to C, then A is more likely to relate to C; (ii) *homophily* - nodes with similar characteristics are more likely to be related; (iii) *clustering into groups* - ties are more dense within groups than between them; - (iv) *degree heterogeneity* - the tendency of some actors to send and/or receive links more than others.

The authors of [18] compare three approaches devoted to engineer a hierarchical ontology on the basis of user interests in a social network. The first approach uses Wikipedia to find interest definitions, the latent semantic analysis technique to measure the similarity between interests based on their definitions, and an agglomerative clustering algorithm to group similar interests into higher level concepts. The second one uses the Wikipedia Category Graph to extract relationships between interests. The third one uses Directory Mozilla to extract relationships between interests. The results of the authors' investigations show that, although the third approach is the simplest one, it is the most effective in building a hierarchy of user interests.



In [31], the authors investigate the use of data mining techniques to identify intra- and inter-organization clusters of people with similar profiles that could have relationships among them. The proposed approach exploits a clustering method, along with a link mining-based technique that uses the minimum spanning tree to construct group hierarchies. The authors performed their analyses on a scientific social network in Brazil; in this network two scientists are connected if they have co-authored a paper. As a result of their investigation, the authors show that it is possible to derive relationships between educational intuitions by analyzing the relationships among the scientists working in them.

The authors of [35] propose the use of automatic text analysis for clustering a social network. They applied their technique to the values contained in the Enron e-mail dataset [1] and obtained the following results: (i) individuals communicate more frequently with individuals sharing similar value patterns than with individuals having different value patterns; (ii) people who communicate more frequently with each other do not necessarily all fit into a particular value type.

All the above techniques discovering social communities, despite their relationship with our approach, are not comparable with it, because they need a number of information we do not assume available in anonymized data.

### III. DERIVING SOCIAL NETWORKS FROM ANONYMIZED DATA

In this section, we describe in detail our research efforts devoted to verify the effectiveness of applying clustering techniques to derive social networks from anonymized data. This section is organized as follows: first, we describe the testbed. Then, we discuss the issues concerning the construction of the Dissimilarity Matrix, which plays a key role in clustering techniques. Finally, we compare several clustering techniques against their capability of supporting our goal.

#### A. Experimental settings

As pointed out in the introduction, in order to perform our analyses, we had to extract some samples from a SIS. This last one contained social networks compliant with XFN and FOAF, which are two standards encoding human relationships in social networks. XFN [7] simply uses an attribute, called *rel*, to specify the kind of relationship between two users. Some possible values of *rel* are *me*, *friend*, *contact*, *co-worker*, *parent*, and so on. In particular, *rel* set to *me* denotes the presence of a *me* edge, which is an edge exploited to link two accounts of the same user in two different social networks. A (presumably) more complex alternative to XFN is FOAF (Friend-Of-A-Friend) [8]. A FOAF profile is essentially an XML file describing a person, her links with other people and the links to the objects created by her. It is worth pointing out that the technicalities concerning these two standards are not to be handled manually by the user. As a matter of fact, each social network has suitable mechanisms to automatically manage them in a way transparent to the user, who has just to specify her relationships in a user-friendly fashion.

The social networks of the SIS into consideration were five, namely Twitter, YouTube, Flickr, MySpace and LiveJournal. We choose these five social networks because they have been

largely analyzed in the past in Social Network Analysis papers devoted to study a single social network or to compare different social networks [21], [25], [13], [33].

The crawling technique we adopted was Breadth First Search (BFS, for short). It is the most popular and widely used strategy for performing topology measurements at several levels, including sampling large networks, and is preferred with respect to other graph traversal techniques, like Depth-First Search, Forest Fire and Snowball Sampling [23]. The crawled samples presented both connections between the accounts of different users of the same social network and connections between the accounts of the same users in different social networks. They can be represented by a direct graph whose nodes correspond to user accounts and whose edges correspond to the connections between these accounts.

Once samples were obtained, they were anonymized by replacing each URL with a numeric identifier unique for all the SIS. After this, several clustering techniques were applied on anonymized samples and the clusters obtained by these techniques were compared with the original social networks. As for the adopted clustering techniques, three of them (i.e., SimpleKMeans, EM and Hierarchical) are classical and well known in literature [17]. The fourth one, called Sequential Information Bottleneck (SIB, for short) integrates an agglomerative procedure in a sequential clustering algorithm [29]. In our experiments, we set the number of clusters to be found to 5, i.e., to the number of the social networks of the SIS. Observe that this fact does not represent a real limitation since, even when the data of a SIS are anonymized, it is always specified which are the social networks composing the SIS even if the membership of a node to a certain social network has been lost during the anonymization process. We adopted the implementations of the considered clustering techniques provided by WEKA [6].

For our experiments, we exploited a server equipped with a 2 Quad-Core E5440 processor and 16 GB of RAM with the CentOS 6.0 Server operating system. We performed the crawling tasks from February 5, 2012 to March 20, 2012. After this task, we performed the other activities described above.

We considered 10 samples for this analysis; each sample referred to 5000 visited nodes.

As a first quantitative index to evaluate the effectiveness of applying clustering techniques to derive social networks from anonymized data, we adopted the Jaccard coefficient, which is a statistical index used to measure the similarity and the diversity of two sets. Given two sets  $A$  and  $B$ , the Jaccard coefficient is defined as  $J(A, B) = \frac{|A \cap B|}{|A \cup B|}$  [24].

As a second quantitative index, we adopted an index called Average Cluster Partitioning (ACP) which operates at a higher abstraction level w.r.t. the Jaccard coefficient. Specifically, ACP is computed by averaging the maximum Jaccard Coefficient of each obtained cluster.

Both indexes range in the real interval  $[0, 1]$ . The higher their value, the better the performance of the corresponding clustering technique.

### B. Dissimilarity Matrix construction

In order to apply the clustering techniques, we had to construct the Dissimilarity Matrix. In fact, it is the structure which most of the clustering techniques operate on. This matrix is  $n \times n$ , where  $n$  is the number of the graph nodes. Its generic element  $(i, j)$  represents the dissimilarity degree between the nodes  $i$  and  $j$ . As for the dissimilarity measure to exploit for the construction of the Dissimilarity Matrix, we chose the minimum distance between the corresponding nodes of the graph. This distance was computed by fixing the weight of each edge equal to 1 and by exploiting the Dijkstra Algorithm for the computation of the minimum paths in the graph. If it did not exist a path between two nodes, the corresponding distance was set equal to  $\infty$ .

As for this matrix, two important decisions must be made, namely: (i) it must be symmetric or asymmetric? (ii) how to normalize it in such a way that its elements range in the real interval  $[0, 1]$  (i.e., in the form taken in input by clustering techniques)?

In order to answer the first question, we made the following reasoning: four of the five social networks belonging to our SIS have asymmetrical links; this makes it natural to adopt an asymmetric Dissimilarity Matrix.

In order to answer the second question, we considered three different normalization strategies which weight the absence of a path between two nodes in a different way.

In the first one, we computed the maximum among the lengths of the minimum paths between the pairs of the nodes of the graph and we divided each matrix element by this maximum incremented of 1 (this length ranged from 15 to 20 in the various samples). When the distance between two nodes was  $\infty$ , the corresponding matrix element was set to 1. This way, the normalized coefficients of the Dissimilarity Matrix for a pair of nodes not linked by a path is comparable with the normalized coefficients of the pair of nodes linked by the path with the maximum length. This choice aimed at not excessively penalizing the nodes not linked by any path. It was motivated by the idea that, actually, two nodes whose minimum path has a high length were not in a situation very different from the one of two nodes not linked by any path. This idea is also substantiated by sociological theories (see, for instance, the six-degree of separation and the small-world theories [32]) As for this normalization strategy, the ACP values obtained by applying SimpleKMeans, EM, Hierarchical and SIB were 0.762, 0.812, 0.622 and 0.760, respectively.

In the second normalization strategy, we divided the coefficients corresponding to nodes linked by paths by 100 (i.e., by a value much higher than the length of the maximum path, which ranged from 15 to 20), whereas we set to 1 the coefficients corresponding to nodes not linked by a path. In this way, we established a significant difference between the coefficients associated with the nodes linked by the maximum path and those ones corresponding to nodes not linked by a path. At the end of this experiment, the ACP values obtained by applying SimpleKMeans, EM, Hierarchical and SIB were 0.652, 0.812, 0.582 and 0.654, respectively. From the analysis of these values we obtained that this new normalization strategy negatively influenced the results of SimpleKMeans, Hierarchical and SIB, whereas the results produced by EM were identical to the

ones obtained with the previous normalization strategy. As a consequence, on the whole, this last normalization strategy appears worse than the previous one.

In the third normalization strategy, we repeated this experiment by dividing the coefficients corresponding to nodes linked by a path by 250, instead of by 100, in such a way as to obtain a more significant difference between the coefficients associated with nodes linked by the maximum path and the coefficients corresponding to nodes not linked by a path. Obtained ACP values for SimpleKMeans, EM, Hierarchical and SIB were 0.621, 0.812, 0.567 and 0.622, respectively. At the end of this experiment we observed that SimpleKMeans, Hierarchical and SIB produced worse results with respect to the corresponding ones returned by the first two strategies. EM, instead, obtained the same results as the ones of the previous cases; this implies that it is not influenced by the normalization strategy.

At the end of this experiment, we concluded that the best normalization strategy was the first one.

In order to verify if our conjecture about the decision (i) was correct, we applied the symmetric Dijkstra algorithm, instead of the classical one, in order to obtain a *symmetric* Dissimilarity Matrix, as it generally happens for the input of clustering techniques. In particular, for each pair of nodes  $n_a$  and  $n_b$ , we computed the corresponding value of the Dissimilarity Matrix by taking the minimum between the distance from  $n_a$  to  $n_b$  and the one from  $n_b$  to  $n_a$ . Clearly, in order to normalize the Dissimilarity Matrix, we chose the first strategy. Once the Dissimilarity Matrix was constructed, we applied the four clustering techniques. Finally, we computed the corresponding ACP values and we obtained 0.382 for SimpleKMeans, 0.700 for EM, 0.486 for Hierarchical and 0.498 for SIB. From the analysis of these values it is possible to conclude that the adoption of a symmetric Dissimilarity Matrix produces worse results than the ones returned by the adoption of an asymmetrical Dissimilarity Matrix. Interestingly enough, choosing an asymmetric Dissimilarity Matrix is not the standard choice to adopt in clustering. As a matter of fact, Dissimilarity Matrixes provided in input to clustering techniques are generally symmetric. Our reasoning about the characteristics of the involved social networks allowed us to make the right decision.

### C. Clustering technique comparison

The first clustering technique we applied was SimpleKMeans. For each sample, we computed the Jaccard coefficient between each cluster generated by SimpleKMeans and each social network of the SIS. Obtained results, averaged across all samples, are reported in Table I. From the analysis of this table, we can observe a correlation between clusters and social networks. In fact, clusters are capable of identifying the involved social networks. In particular, Cluster 3 perfectly corresponds to MySpace since the associated Jaccard coefficient is 1. In other three clusters, namely Clusters 1, 2 and 4, only nodes belonging to a single social network are contained. Only Cluster 0 contains all the nodes of YouTube but also nodes of Twitter, LiveJournal and Flickr. At the end of this analysis, we may conclude that SimpleKMeans finds quite a good correspondence between social networks and clusters.

TABLE I. JACCARD COEFFICIENTS REGARDING THE CLUSTERS OBTAINED BY SIMPLEKMEANS.

	Flickr	YouTube	MySpace	LiveJournal	Twitter
Cluster 0	0.01	0.55	0.00	0.21	0.10
Cluster 1	0.00	0.00	0.00	0.00	0.76
Cluster 2	0.00	0.00	0.00	0.54	0.00
Cluster 3	0.00	0.00	1.00	0.00	0.00
Cluster 4	0.96	0.00	0.00	0.00	0.00

TABLE II. JACCARD COEFFICIENTS REGARDING THE CLUSTERS OBTAINED BY EM.

	Flickr	YouTube	MySpace	LiveJournal	Twitter
Cluster 0	0.96	0.00	0.00	0.00	0.00
Cluster 1	0.01	0.80	0.00	0.01	0.05
Cluster 2	0.00	0.00	0.00	0.00	0.76
Cluster 3	0.00	0.00	1.00	0.00	0.00
Cluster 4	0.00	0.00	0.00	0.54	0.00

We have then applied EM to the same samples and we have computed the Jaccard coefficient between each returned cluster and each social network. Obtained results are reported in Table II. From the analysis of this table, it is possible to see that the results returned by EM are better than the ones returned by SimpleKMeans. Indeed, there is an optimal correspondence between clusters and social networks.

When we applied Hierarchical to the same samples we obtained worse results than the ones returned by SimpleKMeans and EM. The corresponding Jaccard coefficients are shown in Table III. This algorithm behaves as the previous ones as far as MySpace, Flickr and LiveJournal are concerned, whereas it shows a worse behavior for the other two social networks. In particular, in Cluster 4, only nodes of YouTube occur, but the Jaccard coefficient between YouTube and Cluster 4 is significantly lower than the one between YouTube and Cluster 0. Hierarchical tends to put the nodes of YouTube and Twitter in the same cluster. We may conclude that the clusters generated by this algorithm reflect the structure of the SIS in a less precise way than the clusters returned by SimpleKMeans and EM. It seems that Hierarchical is incapable of distinguishing clusters in presence of a certain number of  $m_e$  edges (see below). This can be explained by considering that Hierarchical is well suited when it is necessary to construct cluster hierarchies by proceeding in an agglomerative fashion. In our scenario, we would have a one-level hierarchy. Furthermore, the presence of  $m_e$  edges would make it very difficult to proceed in an agglomerative fashion because the corresponding aggregation process would be quite irregular with frequent hops from a social network to another.

The last clustering technique we applied is SIB. In Table IV, we report the Jaccard coefficients between the social networks of the SIS and the clusters returned by this algorithm. We may observe that obtained clusters are equivalent to the ones returned by SimpleKMeans.

At the end of this analysis, it emerged that the best

TABLE III. JACCARD COEFFICIENTS REGARDING THE CLUSTERS OBTAINED BY HIERARCHICAL.

	Flickr	YouTube	MySpace	LiveJournal	Twitter
Cluster 0	0.01	0.31	0.00	0.16	0.45
Cluster 1	0.96	0.00	0.00	0.00	0.00
Cluster 2	0.00	0.00	0.00	0.54	0.00
Cluster 3	0.00	0.00	1.00	0.00	0.00
Cluster 4	0.00	0.16	0.00	0.00	0.00

TABLE IV. JACCARD COEFFICIENTS REGARDING THE CLUSTERS OBTAINED BY SIB.

	Flickr	YouTube	MySpace	LiveJournal	Twitter
Cluster 0	0.01	0.54	0.00	0.21	0.10
Cluster 1	0.00	0.00	0.00	0.00	0.76
Cluster 2	0.00	0.00	0.00	0.54	0.00
Cluster 3	0.00	0.00	1.00	0.00	0.00
Cluster 4	0.96	0.00	0.00	0.00	0.00

clustering technique is EM. This is also confirmed by all the values of ACP obtained in the experiments shown in Section III-B. However, the analysis above was useful because, with respect to ACP values, Tables I - IV provide more detailed results.

#### IV. CONCLUSION AND FUTURE WORK

In this paper, we have studied the effectiveness of deriving social networks from anonymized data. We have explained that this problem is very important when data exploited for the analysis are taken from publicly available repositories. In this case, a lot of data are available but they have been anonymized in order to protect user privacy. The motivation of the work is that the derivation of social networks from anonymized data is needed for Social Internetworking Analysis.

As a future work, we plan to extend our research efforts in several directions. First of all, we observe that this paper is the first attempt of reconstructing information from anonymized SIS data. In order to facilitate our tasks we considered not very large samples, even though in line with those exploited by clustering techniques applied to Social Network scenarios in the past literature. We plan to analyze much larger samples in the future also considering clustering techniques specifically conceived for large datasets [16] as well as incremental clustering techniques [17]. Finally, we plan to define approaches for SIS analysis based on other Data Mining tasks (association rule extraction, classification and, above all, outlier analysis).

#### Acknowledgment

This work has been partially supported by the TENACE PRIN Project (n. 20103P34XC) funded by the Italian Ministry of Education, University and Research.

#### REFERENCES

- [1] Enron email dataset, <http://www.cs.cmu.edu/~enron> [retrieved: May, 2013].

- [2] FriendFeed, <http://friendfeed.com> [retrieved: May, 2013].
- [3] Gathera, <http://www.gathera.com> [retrieved: May, 2013].
- [4] Google Open Social, <http://code.google.com/intl/it-IT/apis/opensocial> [retrieved: May, 2013].
- [5] Power.com, <http://techcrunch.com/2011/04/21/power-com-shuts-down-domain-name-up-for-sale> [retrieved: May, 2013].
- [6] WEKA - Waikato Environment for Knowledge Analysis, <http://www.cs.waikato.ac.nz/ml/weka> [retrieved: May, 2013].
- [7] XFN - XHTML Friends Network, <http://gmpg.org/xfn> [retrieved: May, 2013].
- [8] D. Brickley and L. Miller, "The Friend of a Friend (FOAF) project," <http://www.foaf-project.org> [retrieved: May, 2013].
- [9] F. Buccafurri, V.D. Foti, G. Lax, A. Nocera, and D. Ursino, "Bridge Analysis in a Social Internetworking Scenario," *Information Sciences*, Elsevier, vol. 224, March 2013, pp. 1–18.
- [10] F. Buccafurri, G. Lax, A. Nocera, and D. Ursino, "Discovering Links among Social Networks," *Proc. European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2012)*, Springer, September 2012, pp. 467–482.
- [11] F. Buccafurri, G. Lax, A. Nocera, and D. Ursino, "Crawling Social Internetworking Systems," *Proc. International Conference on Advances in Social Analysis and Mining (ASONAM 2012)*, IEEE/ACM, August 2012, pp. 505–509.
- [12] P. Carrington, J. Scott, and S. Wasserman, *Models and Methods in Social Network Analysis*. Cambridge University Press, 2005.
- [13] X. Cheng, C. Dale, and J. Liu, "Statistics and Social Network of Youtube Videos," *Proc. International Workshop on Quality of Service (IWQoS 2008)*, IEEE press, June 2008, pp. 229–238.
- [14] P. De Meo, A. Nocera, G. Terracina, and D. Ursino, "Recommendation of similar users, resources and social networks in a Social Internetworking Scenario," *Information Sciences*, Elsevier, vol. 181(7), April 2011, pp. 1285–1305.
- [15] A. Firat, S. Chatterjee, and M. Yilmaz, "Genetic clustering of social networks using random walks," *Computational Statistics and Data Analysis*, Dec. 2007, vol. 51, pp. 6285–6294.
- [16] V. Ganti, R. Ramakrishnan, J. Gehrke, A.L. Powell, and J.C. French, "Clustering Large Datasets in Arbitrary Metric Spaces," *Proc. IEEE International Conference on Data Engineering (ICDE'99)*, IEEE Press, March 1999, pp. 502–511.
- [17] J. Han and M. Kamber, *Data Mining: Concepts and Techniques - Second Edition*. Morgan Kaufmann Publishers, 2006.
- [18] M. Haridas and D. Caragea, "Exploring Wikipedia and Dmoz as Knowledge Bases for Engineering a User Interests Hierarchy for Social Network Applications," *Proc. International Workshop On the Move to Meaningful Internet Systems (OTM 2009)*, Springer Press, Nov. 2009, pp. 1238–1245.
- [19] OFCOM The independent regulator and competition authority for the UK communications industries, "Social Networking: A quantitative and qualitative research report into attitudes, behaviours and use". <http://stakeholders.ofcom.org.uk/binaries/research/media-literacy/report1.pdf> [retrieved: May, 2013].
- [20] J. Kleinberg, "The convergence of social and technological networks," *Communications of the ACM*, vol. 51, Nov. 2008, pp. 66–72.
- [21] B. Krishnamurthy, P. Gill, and M. Arlitt, "A few chirps about Twitter," *Proc. First Workshop on Online Social Networks (WOSN 08)*, ACM Press, Aug. 2008, pp. 19–24.
- [22] P.N. Krivitsky, M.S. Handcock, A.E. Raftery, and P.D. Hoff, "Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models," *Social Networks*, vol. 31(3), 2009, pp. 204–213.
- [23] M. Kurant, A. Markopoulou, and P. Thiran, "On the bias of BFS (Breadth First Search)," *Proc. IEEE International Teletraffic Congress (ITC 22)*, IEEE Press, Sept. 2010, pp. 1–8.
- [24] C.D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge University Press, 2008.
- [25] A. Mislove, H.S. Koppula, K.P. Gummadi, F. Druschel, and B. Bhattacharjee, "Growth of the Flickr Social Network," *Proc. First Workshop on Online Social Networks (WOSN 08)*, ACM Press, Aug. 2008, pp. 25–30.
- [26] A. Mislove, M. Marcon, K.P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," *Proc. SIGCOMM International Conference on Internet Measurement (IMC'07)*, ACM Press, October 2007, pp. 29–42.
- [27] A. Narayanan and V. Shmatikov, "De-anonymizing social networks," *Proc. IEEE International Symposium on Security and Privacy (S&P 2009)*, IEEE Press, May 2009, pp. 173–187.
- [28] Y. Okada, K. Masui, and Y. Kadobayashi, "Proposal of Social Internetworking," *Proc. International Human.Society@Internet Conference (HSI 2005)*, Springer Press, July 2005, pp. 114–124.
- [29] J. Peltonen, J. Sinkkonen, and S. Kaski, "Sequential information bottleneck for finite data," *Proc. International Conference on Machine Learning (ICML'04)*, ACM Press, July 2004, pp. 82–88.
- [30] W. Peng and T. Li, "Temporal relation co-clustering on directional social network and author-topic evolution," *Knowledge and Information Systems*, Springer, 2011, vol. 26(3), pp. 467–486.
- [31] V. Ströele et al., "Mining and Analyzing Organizational Social Networks Using Minimum Spanning Tree," *Proc. International Workshop On the Move to Meaningful Internet Systems (OTM 2008)*, Springer Press, Nov. 2008, pp. 18–19.
- [32] J. Travers and S. Milgram, "An experimental study of the small world problem," *Sociometry*, 1969, pp. 425–443.
- [33] S. Ye, J. Lang, and F. Wu, "Crawling online social graphs," *Proc. IEEE International Asia-Pacific Web Conference (APWeb'10)*, IEEE Press, April. 2010, pp. 236–242.
- [34] P. Zhao and C. Zhang, "A new clustering method and its application in social networks," *Pattern Recognition Letters*, vol. 32(15), 2011, pp. 2109–2118.
- [35] Y. Zhou, K.R. Fleischmann, and W.A. Wallace, "Automatic text analysis of values in the enron email dataset: Clustering a social network using the value patterns of actors," *Proc. IEEE Hawaii International Conference on System Sciences (HICSS 2010)*, IEEE Press, January 2010, pp. 1–10.

# The Impact of the Internet and the World Wide Web On Distance and Collaborative Learning

Franklyn Chukwunonso  
Department of Information Systems  
Universiti Teknologi Malaysia  
Johor Bahru, Malaysia  
e-mail: franconicostelo@yahoo.com

Roliana Binti Ibrahim  
Department of Information Systems  
Universiti Teknologi Malaysia  
Johor Bahru, Malaysia  
e-mail: roliana@utm.my

Ali Bin Selamat  
Department of Information Systems  
Universiti Teknologi Malaysia  
Johor Bahru, Malaysia  
e-mail: aselamat@utm.my

Adamu Idama  
Department of Statistics and Operations Research  
Modibbo Adama University of Technology  
Yola, Nigeria  
e-mail: apostleidama@yahoo.com

Wadzani A. Gadzama  
Department of Information Security  
Universiti Teknologi Malaysia  
Johor Bahru, Malaysia  
e-mail: ask4gadzama@gmail.com

**Abstract**—The advent of the Internet and the World Wide Web has changed the face of technology especially in the way we communicate and interact. The increasing use of this technology especially in teaching and learning has attracted enormous research efforts particularly in the development of distance and collaborative learning. Even though much gain have been recorded in this direction, there still remains a gap as to how well these highly proliferating web technologies have impacted on the collaboration between students and teachers, students and students, and its overall contribution to distance learning. This paper considered how these web-based technologies contribute to collaborative learning by enabling people to interact with each other from different locations. It identified the factors necessary to optimise the impact of web-based technologies in distance and collaborative learning and concludes by proposing a best-practice guideline that will enhance the impact of the Internet and the World Wide Web for effective distance and collaborative learning.

**Keywords**—Internet; World Wide Web; distance learning; e-learning; collaborative learning.

## I. INTRODUCTION

The last decade has witnessed two major trends in the world of education: the transformation of the Web from a repository of hyper textual documents to a highly interactive communication media, and a shift in learning theory from the traditional educational theory of behaviourism to that of cognitivism and constructivism [1][2]. With this trend, web applications have become important candidates for educational activities, which often include collaborative

sessions. The traditional emphasis on print in the educational system is constantly threatened by technology enhanced learning (TEL) [3]. Over the years, the use of technology in education has witnessed a tremendous growth not only in size, but also in the innovative way it is used.

Distance learning is a way of delivering instructional and educational materials to learners in different places and at different or the same time, on an individual basis [4]. The use of technology in teaching brought about a transformational change in education, especially in distance learning. Distance education has come a long way, but it witnessed a rapid growth in popularity and use with the advent of more advanced technologies. From a humble beginning in 1728, when Caleb Philip taught his students new method of Short Hand through weekly mailed lessons [5], through 1840s, when Isaac Pitman used correspondence to teach shorthand in Great Britain [6], the introduction of technology to education introduced online learning programs at the K-12 level in 2008 [7]. Today, more impact of adopting Internet and web-based technologies have completely transformed the way we learn, play and work. Web applications are now considered very useful in enhancing educational activities, particularly for collaborative sessions [8][9]. A peoples wealth and competitiveness, human competency and social development, is highly dependent on the importance it attaches to education. With the current trend of globalization and rapid technological advancement, there is a need to adopt new methods of teaching and learning in training the population into a knowledge based society through the

adoption of more robust and collaborative web-based groupware technologies such as e-learning.

These new mode of learning (e-learning) has the potential of meeting present day challenges in the acquisition and dissemination of knowledge and skill. The problem however is not in the adoption of the technology, but in its use to achieve the student's learning outcomes. Unlike in the era of audiovisuals where the television played a lasting and remarkable role in distance learning (e.g., Telekolleg in Germany 1967, the Open University in Milton Keynes in Great Britain 1971, and the FernUniversität in Hagen Germany 1975), e-learning tools are yet to record a remarkable impact on learning as these tools are mostly underutilised for the benefit of distance and collaborative learning. This study is therefore convinced that if fully utilized, the Internet and web-based technologies can be used to address these challenges as e-learning is perceived to be a viable resource for teaching and collaborative learning either by synchronous or asynchronous methods.

Section 2 takes a look at the background of the study from two perspectives: historical and educational perspectives. In Section 3, we discuss the role of technology in distance and collaborative learning. In Section 4, we discuss the impact of these technologies on teaching and learning based from the findings of our study and propose a guideline of best practice that will enhance the impact of the Internet and the WWW for effective distance and collaborative learning. Section 5 concludes the paper.

## II. BACKGROUND OF THE STUDY

Recent times have seen a tremendous increase in the application of information and communication technologies (ICT) for educational purposes, resulting to a proliferation of networked technologies and evolution of e-learning methods [10]. As technology advanced, particularly Internet and web-based technologies, the mode and method of delivering educational and instructional materials, otherwise known as distance learning evolved from distance education to e-learning.

Historically, the use of technology was first noted when the University of Illinois (in 1960) employed computer terminals to enable students access recorded lectures using remote audio or television devices [11]. From 1963 to date, educational institutions have been taking advantage of these new opportunities as they witnessed these technologies evolve in different and innovative ways. Early e-learning systems began to emerge such as Computer-Based Learning/Training (CBT) in the 1970s and 1980s [12], and Computer Supported Collaborative Learning (CSCL) in the 1980s [13]. The introduction of CSCL coupled with the advance in technology from large low processing and expensive computers to today's handheld computing devices changed the face of learning and gave birth to the first Open University in Britain and University of British Columbia [13][14], which also saw the development of the first Web CT (Blackboard Inc.), thus facilitating the use of Internet to deliver education [15], giving birth to web-based training and online distance learning, and online discussions between students at different locations [16]. The birth of the WWW in

the early 1990s gave rise to the use of hypertext materials and online course websites – virtual learning [17]. Other advances in this regard are mentioned in [18][19][20].

From an educational perspective, the adoption of Internet and web-based technologies in education delivery has led to overthrow of more traditional methods of teaching and learning. From a simple and casual application of ICT to facilitate face-to-face classroom teaching and learning to the intensive use of ICT for educational purpose as in the case of virtual or online learning, e-learning, an end product of this evolution process has brought to light a completely different method of learning [21][22][23][24].

In a simple form, E-learning is the application of information and communication technologies (ICT) and electronic media in education, inclusive of multimedia learning, technology-enhanced learning (TEL), computer-based instruction (CBI), computer-based training (CBT), computer-assisted instruction or computer-aided instruction (CAI), internet-based training (IBT), web-based training (WBT), online education, virtual education, virtual learning environments (VLE) (which are also called learning platforms), m-learning, and digital educational collaboration, and can take place within or outside a conventional classroom, in synchronous or asynchronous ways [25]. E-learning is best tailored for distance and collaborative learning but can also be applied in face-to-face teaching (blended learning).

## III. THE ROLE OF TECHNOLOGY

Various Internet and web-based technologies exist, most of which are used to facilitate distance and collaborative learning, also referred to as e-learning. Today, we have terms like instructional technology, learning technology, Computer Aided Instruction or computer-based training. Online learning or education is however, specific to web-based learning. For learning through mobile applications, M-learning is used. E-learning, which is our focus in this paper, goes beyond the application of technology in learning but also encompasses the actual learning activities enabled by the use of these systems.

### A. *Aligning Educational Theories and Pedagogical Designs With E-Learning Technologies*

Distance learning has witnessed three major pedagogical evolution and many technological changes, yet, no single educational theory has been able to solely provide all the solutions demanded by e-learning designs as each has extended the work of others instead of developing a new prototype (Ireland, 2007). Each evolution demands a different approach to learning, different type of knowledge and contextual application, thereby requiring expertise on the mix of technology and pedagogy to be employed. In all three generations, the major players of teachers, learners, and content remain constant, however, the way in which these three relates vary; increasing from a learner-content interaction for cognitive-behaviourist models to learner-learner interaction in constructivism, and then to the deeply networked learner-content-teacher interrelationship propagated in connectivist pedagogies. In connectivist

theory, the student plays the role of teachers and the teacher plays the role of students through a mediated digital artefact interaction created by all. The scenario is more of where the teacher’s role changes from being a guide to a co-passenger or a role model, but no more as the sole creator or guide in the learning activity.

However, learning management systems (LMS) providers’ tend to focus more on technology without any pedagogical integration and this leaves most of these e-learning tools unusable or being used in a way contrary to pedagogical principles, which negatively affects the learning process let alone the waste of resources used in implementing these e-learning systems. Previous researches reveal that most e-learning solutions lack pedagogical background and present some serious deficiencies regarding teaching strategies [26]. With the evolution of technology, e-learning interfaces get more complex and overloaded with excess information and eye-catching graphics. This most times, has an adverse effect as the learner attention is distracted away from the content to technology. This can be avoided if the design interface is anchored on pedagogical principles, which will improve learner’s efficiency while making the learning activity more flexible. Though many learning theories abound, it is advisable to employ a combination of more than one learning theory in the design of e-learning systems as no one theory is specially tailored to support e-learning environments. Table I below shows a summary of learning theories principles.

From the findings of Mayes and de Freitas [27], we can cluster these pedagogical theories into three broad perspectives from the assumptions they make about learning: the associationist/empiricist perspective (learning as activity), the cognitive perspective (learning as achieving understanding), and the situative perspective (learning as social practice), as presented in Table II.

TABLE I. SUMMARY OF LEARNING THEORY PRINCIPLES

Learning Theories	Core Principles
Behaviourism	Brain is a black box
	External stimuli provoke reaction
	Observation of behaviours
Cognitive	Involvement of different kinds of memories
	Learning is an internal process
	Use of node patterns
Constructivism	Attribution of a personal meaning to information
	Building of personal understanding
	Active process of information
Active Learning	Active involvement
	Combining previous involvement and dialogue

TABLE II. MAPPING LEARNING THEORIES TO LEARNING OUTCOMES, ADOPTED FROM [27]

Learning Theory	Learning Outcome
Associative perspective	emphasises task analysis, defining sequences of component-to-composite skills. It provides a highly focused set of objectives, described as learning competencies.
Cognitive perspective	emphasises conceptual development, stressing the importance of achieving understanding of the broad unifying principles of a domain. This view also encourages us to frame learning outcomes in meta-cognitive terms, with the educational aim of achieving learning how to learn, and encouraging the development of autonomous learners.
Situative perspective	encourages the definition of learning objectives in terms of the development of disciplinary practices of discourse and representation. It also focuses on learning outcomes that are dependent upon the establishment of collaborative learning outcomes, and on learning relationships with peers. This perspective also encourages us to formulate learning outcomes in terms of authentic practices of formulating and solving realistic problems.

*B. Internet and Web-Based Groupware Technologies*

The first generation of complex groupware technologies was based on private networks and proprietary communication protocols and clients. Such technologies provided users within a limited area with rich user interfaces and customized functions. With the advent of the Internet, access to these groupware systems was extended to users in different areas, countries or even continents. The WWW in a weak sense can be regarded as a collaborative technology [28][29]. Actually, the Web provides a global platform for information sharing across what looks like a unique large file system [30]. This distinguishing feature of the WWW encouraged an increasing use of web-based collaborative tools especially in the field of project teams support [31]. Other reasons for this fast proliferation of the web-based technologies are the open network client standards characterizing Internet and the WWW [32][29]. Such standards enable any-place-any-time interaction, and make it possible to use standard browsers as clients (in the case of Web-based tools), freeing new users from the necessity to install new software. This interoperability is very useful to widely dispersed working groups, where the localization of people in different organizations and countries used to make deploying existing groupware technologies very hard [33]. Other positive consequences to the use of standard browsers are a reduction in the users’ need for training and set-up costs, and the Web’s suitability to be used as an interface as it combines platform independence and rapid distribution, two factors that are highly appreciated as interface characteristics. Despite all these assets, Internet-based and Web-based groupware tools also encounter a series of new problems and challenges. Among the disadvantages associated with Internet-based and Web-based groupware tools are slowness and unreliability, uncertainty and perceived insecurity. Advanced Web-based groupware has

now partially overcome these last limitations, thanks to relatively new technologies such as Java, XML and CORBA [34].

C. The Future of Distance and Collaborative Learning

E-learning is being increasingly viewed as an important activity in the field of distance and continuing education. Web-based courses offer obvious advantages for learners by making access to educational resource very fast, just-in-time and relevant, at any time or place. Increasingly, the WWW is used to support and facilitate the delivery of teaching and learning materials. This use has progressed from the augmentation of conventional courses through web-based training and distance learning to the web-based and e-learning education. E-learning is not just concerned with providing easy access to learning resources, anytime, anywhere, via a repository of learning resources, but is also concerned with supporting such features as the personal definition of learning goals and the synchronous and asynchronous communication and collaboration, between learners and between learners and instructors. One of the hottest topics in recent years in the Artificial Intelligence (AI) community, as well as in the Internet community, is the semantic web. It is about making the web more understandable by machines. It is also about building an appropriate infrastructure for intelligent agents to run around the web performing complex actions for their users. Furthermore, semantic web is about explicitly declaring the knowledge embedded in many web-based applications, integrating information in an intelligent way, providing semantic-based access to the Internet and extracting information from texts [35]. Figure 1 presents a simple knowledge-based e-learning system architectural framework, Figure 2 shows a wider integrated view of the learning model, while Figure 3 shows a typical e-learning system.

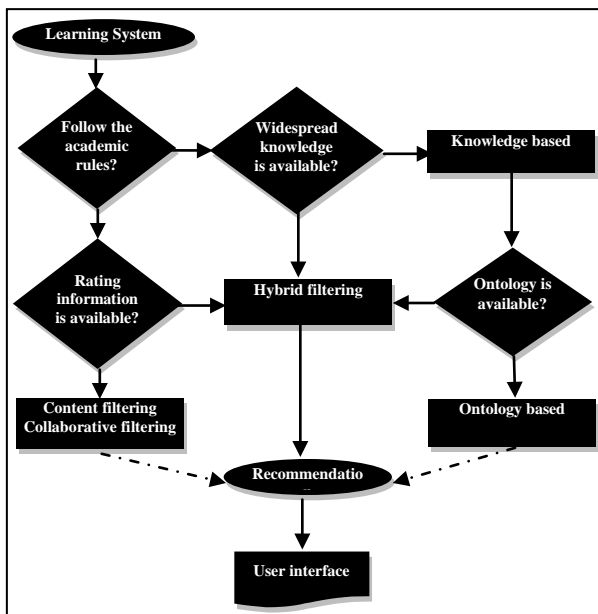


Figure 1. A simple Knowledge based e-learning systems framework. Adapted from [36]

From Figure 1 above, the learning system serves as the input to the system and its output is communicated to the learner through the user interface which may be in a blended learning, distance learning or traditional learning environments respectively (as shown in Figure 2). What the knowledge-based e-learning system in Figure 1 simply does is to extract user-specific information from a vast collection of available information in the learning system by subjecting the content to a set of decision rules. The output from the ontology based and content/collaborative filtering processes are then streamlined and recommended to the user based on those criteria. This process ensures that only user-specific content reaches the learner. Ultimately, semantic web is about how to implement reliable, large-scale interoperation of web services, to make such services computer interpretable – to create a web of machine-understandable and interoperable services that intelligent agents can discover, execute and compose automatically. The problem is that the web is huge, but not smart enough to easily integrate all of those numerous pieces of information from the web that a user really needs. Such integration at a high, user-oriented level is desirable in nearly all uses of the web. Unfortunately, the web was built for human consumption, not for machine consumption - although everything on the web is machine-readable, it is not machine-understandable.

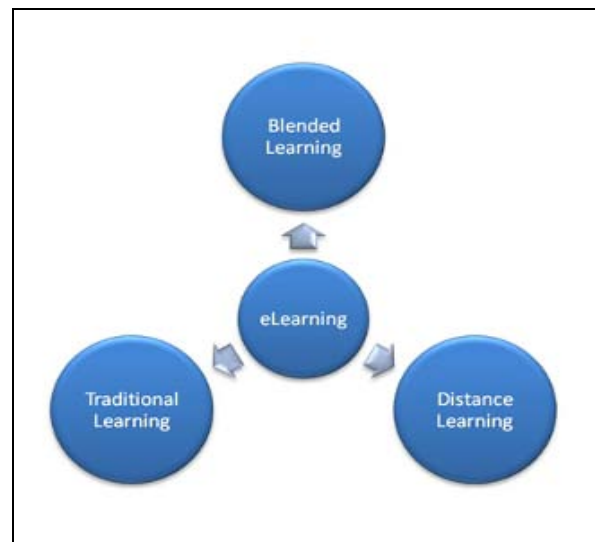


Figure 2. A Typical E-learning Model

D. Connecting for Lifelong Learning

Several studies abound in the field of Web-based technology application in learning [37][36]. Rokou et al. [38] distinguished three basic levels in every web-based application: the Web character of the program, the pedagogical background and the personalized management of the learning material. They defined a web-based program as an information system that contains a Web server, a network, HTTP and a browser in which data supplied by users act on the system's status and cause changes. The pedagogical background means the educational model that is



used in combination with pedagogical goals set by the instructor. The personalized management of the learning materials means the set of rules and mechanisms that are used to select learning materials based on the student's characteristics, the educational objectives, the teaching model and the available media.

Many works have combined and integrated these three factors in e-learning systems, leading to several standardization projects [39][9]. Some projects have focused on determining the standard architecture and format for learning environments, such as IEEE Learning Technology Systems Architecture (LTSC), Instructional Management Systems (IMS) and Sharable Content Object Reference Model (SCORM). IMS and SCORM define and deliver XML-based interoperable specifications for exchanging and sequencing learning contents, i.e., learning objects, among many heterogeneous e-learning systems. They mainly focus on the standardization of learning and teaching methods as well as on the modelling of how the systems manage interoperating educational data relevant to the educational process. IMS and SCORM have announced their content packaging model and sequencing model, respectively. The key technologies behind these models are the content package, activity tree, learning activities, sequencing rules and navigation model.

Their sequencing models define a method for representing the intended behaviour of an authored learning experience and their navigation models describe how the learner and system initiated navigation events can be triggered and processed. Quemanda and Simon [40] have also presented a model for educational activities and educational materials. Their model for educational activities denotes educational events that identify the instructor(s) involved and take place in a virtual meeting according to a specific schedule. Also [38] described the introduction of stereotypes to the pedagogical design of educational systems and appropriate modifications of the existing package diagrams of UML (Unified Modelling Language). The IMS and SCORM models describe well the educational activities and system implementation, but not the educational contents knowledge in educational activities. Most e-learning models add more pedagogical background by emphasizing educational contents and sequences using the taxonomy of learning resources and stereotypes of teaching models. But the educational contents and their sequencing in these models are dependent on the system and lack standardization and reusability. Thus, we believe that if an educational contents frame of learning resources can be introduced into an e-learning system, including ontology-based properties and hierarchical semantic associations, then this e-learning system will have the capabilities of providing adaptable and intelligent learning to learners.

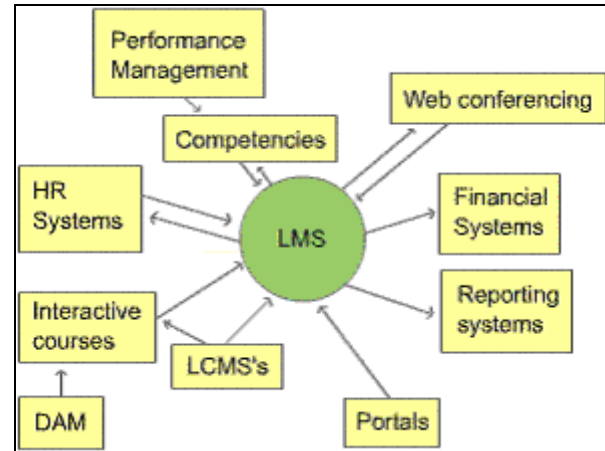


Figure 3. A Typical E-learning System

#### IV. THE IMPACT OF INTERNET AND WEB-BASED TECHNOLOGIES ON LEARNING

With the increasing importance of education for social development and improvement of human competencies, especially in the face of a rapidly changing technological advanced world that affects the living and working environment, the demands posed by e-learning on educational institutions, teachers and students is new and different from the traditional. This calls for new means of knowledge dissemination and skills acquisition, especially for the socially challenged and people with disabilities. The traditional way of learning has come short of meeting the needs of a society as new media and communication services and the ever busy demands placed by working, learning and family life schedules poses new challenges. To properly harness the impact of these Internet and web-based technological tools on distance and collaborative learning, we have decided to look at it from four perspectives:

- Applications of e-learning.
- Implications of e-learning.
- Benefits of e-learning.
- Guidelines for best practices of e-learning.

##### A. Applications

Positive effects have been recorded in all major subject areas, from preschool to higher learning, and for both regular students and those with special needs through the application of e-learning in education. Educational institutions must adopt new ways of disseminating information and acquiring skills and knowledge via e-learning in order to meet these challenges faced by today's advanced and knowledge-based society. Some of the ways in which e-learning is applied to impact on education are mentioned below.

1) *Preschool*: Different types of learning now take place among pupils even before they enroll into formal education. [41] posits that the impact of such e-learning is yet to be ascertained, however, parents are of the opinion exposing children to handheld devices or computers at an early age is a positive experience for the kids as it exposes them to

information that aids them when they enroll into schools. This claim is in agreement with [42][43].

2) *K-12*: Virtual schools, which make use of synchronous or asynchronous learning are e-learning applications adopted by public K-12 school. Classes can be attended either in conventional classrooms or from home or both. Students are provided computers and printers and other technological tools while in the school and reimbursed for home use. These cyber schools enable learners study at their own pace, select their courses, and create their schedules, thereby making learning very flexible and convenient. K-12 schools often make use of with innovative course delivery technology and administrative models [44].

3) *Higher Education*: In most developed and developing country, the application of e-learning in post-secondary education is predominantly on the increase [45]. With an increase in online education, some research universities now offer online doctoral programs [46]. Even though some of the massively-open online courses have limitations, research institutions like MTT, Stanford and Princeton University now offer non-credit courses to a wide range of global audience [46]. University course programs and a wide variety of courses can also be found on YouTube. Other such online higher education courses embarked by Universities include Udacity (which offers free computer classes), Coursera (an online-enrollment platforms that offers education for millions around the world) and several free online course websites in almost every field and discipline.

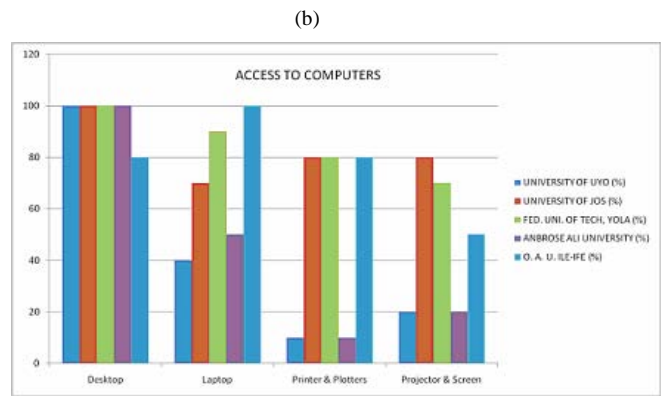
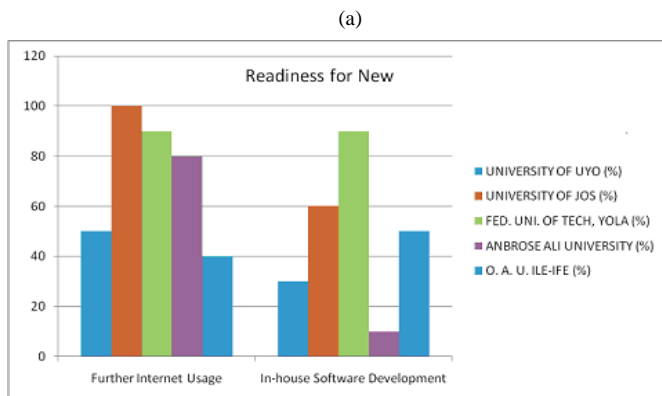
4) *Corporate and Professional*: The shift toward a Knowledge (K-economy) has affected the work setting, and created a need for knowledge workers with high intuition and ability to solve detailed and genuine problems confronted in the work environment. E-learning can adequately meet these requirement because employees can take advantage of online learning and online training, from their homes or offices, and at their pace and convenience. E-

learning has been adopted and used by various corporations and companies to overcome this challenge.

*B. Implications*

The difficulty in understanding the future implication of new technologies in society is seen in the growth of the Internet throughout the world, and in particular, in educational institutions. Despite the achievements recorded by e-learning systems and its considerable impact on education and training, it is still not presently utilised to its full capacity for the benefit of distance and collaborative learning. The high hopes raised by the application of these Internet and web-based technologies in learning are yet to be fulfilled.

1) *Misappropriation of Use*: The existence of any technological artefact is no guarantee that it will be used for its intended purpose, and even if computers are found in classrooms, they may not be used to their full potential [47]. To this effect, a survey of five selected faculties in different Nigerian Universities was conducted to show the readiness for new computerized functions and total integration of ICT in education. The survey reveals that the purpose for which these computers are being used were primarily for word processing, spread sheet, and power point presentations outside the leading role of using computers for e-learning. While all the universities investigated had Internet access, only a few percent of them use the Internet for the primary purpose of e-learning as most use the Internet mostly for emails and which has no direct impact on learning or teaching. Figure 4 below depict the findings of the survey. Such results are not surprising. [48][49] argued that previous classroom technologies such as television have not been widely adopted because teachers need to be convinced that challenges to the rituals of everyday classroom life are in their interests. Educational systems are inherently conservative and resistant to change. Departure from conventional practices and continuity are justified only if the intended innovation is simple, durable and reliable.



(c)

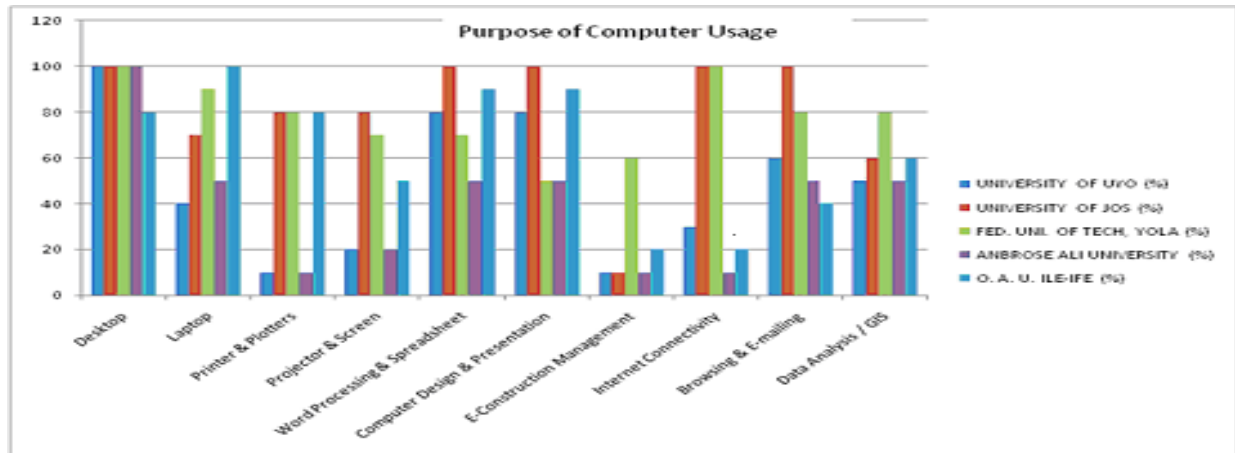


Figure 4. (a) Readiness for New Technology, (b) Access to Computers, and (c) Purpose of Computer Usage in 5 Different Nigerian Universities.

2) *Lack of timely access to user tailored information:* Nowadays, emphasis is not laid on accessibility to knowledge, but on its timeliness, relevance and usefulness to the user. Most e-learning systems tend to overload users with irrelevant information without giving attention to the user's specific needs. The true measure of e-learning is not in training and accessibility, but its ability to use this feature to train the right individual to obtain the required knowledge/skill when needed.

3) *Lack of Standards:* There is a need for users, vendors and developers of e-learning systems to adopt standards in order to facilitate interoperability and increase information sharing between different e-learning platforms. This can be achieved through an international body comprising of major stakeholders in the field of e-learning, by obtaining information on user requirements and drawing specifications from these collected data for standard definition and e-learning content architecture and protocol development. Cooperation among the e-learning community is beginning to be witnessed as user requirements are being validated and the approved specifications converted into standards.

4) *Lack of Interoperability Among E-learning Systems:* Another divergent problem arising from lack of standards is the lack of interoperability among various e-learning systems as noted by [50]. Whole and parts of different e-learning systems must co-operate and interoperate to facilitate full integration and wider accessibility. Institutions of learning should adopt the use of standards that cover every area of learning. Developers of e-learning tools must incorporate standards in their design, and any customised protocol or standards should be resisted in order to forstall the chaos presently witnessed by multimedia technologies.

### C. Benefits of E-learning

The use of Internet and web-based technology tools in learning enables disabled people carry out tasks, which prior to e-learning, was seen as difficult or impossible. These include the use of chat programs for the deaf, text-to-speech software for the blind, and diction software for the armless who cannot write [51]. Online courses are not limited by distance, location, or time as they can be easily accessed by learners from anywhere and at anytime.

Technology allows real-time modifications, accreditation of a continuously updated tool in accordance with the research literature. Foremost, the WWW can be used to dynamically transfer knowledge in real time around the globe. This will lead to higher-education opportunities for foreign students. Countries without "mass" university education can access universities in other countries through the Web. Finally, this technology is able to track use and activity, provide reports, and record information about every learner's performance. This feedback could be included in continuing education courses. Demand for individualized services, tools, interactive experiences, and open access to knowledge is growing. Learning is no longer expected to be paced so much by the teacher as it is by the student's capacity to grasp the material (student-focused learning). The speed at which students can progress through a course of instruction varies by factors of three to seven, even in classes of carefully selected students. In traditional training models, it is impossible to deliver individually customized learning solutions because of cost. The capacity of e-learning for real-time, on-demand adaptation can provide individualized learning at affordable cost. Technologies that allow collaboration, interactivity, simulation, and self-testing can help students acquire the skills being taught effectively and efficiently. It is thus possible to create a learning environment in which students become active participants, fully engaged in the learning process. Additionally, the educational topic selection can cater to a student's particular needs. Any given student may be studying any given topic at

any time, and progressing through that material at a pace appropriate to his or her learning ability.

These benefits mentioned above can be summarized into five key points in agreement with past research findings in this field:

- Wider access to education [52].
- Increased integration for disabled, part-time, and working students, especially in continuing education [52].
- Enhanced student-student and student-teacher interaction and collaboration [53].
- Promotes self-paced learning and independent problem solving by learners [53].
- Promotes the acquisition of ICT knowledge and development of ICT skills through usage of computers and web-based tools [43].

#### D. Guidelines for Best Practices

We have discussed in this paper the various evolution that has taken place in learning, particularly, distant and collaborative learning, in the face of changing technologies. It is clear that we are in a stage of rapid technological development and profound new discoveries of life and learning in connected contexts. The emergence of collective understanding formed by the selective use and analysis of the networks, sets, behaviours and activities within which we engage promises much deeper understanding of our knowledge construction and application.

However, to achieve an educational semantic web, and maximize the potentials of Internet and web-based technologies in distance and collaborative learning, two main issues must be addressed: interoperability among different e-learning systems, and the use of learning objects for content generation. While the earlier is subject to the standardization of these e-learning tools, the later can be achieved by using structured, unified and automated authoring tools to generate content. Thus in realizing the Semantic Web dream, it is necessary to consider the use of common standards in communication syntax, semantic concepts and ontologies, full integration/unification of educational content and authoring tools.

Below is a summarized set of best-practice guideline, which this paper proposes:

- 1) *Identify the current processes involved in designing an e-learning system.*
- 2) *Provide a proper definition of learners' competency.*
- 3) *Integrate the different design components.*
- 4) *Identify the nature of interaction among the various components of the e-learning framework:*
  - a) *the collaboration between student-student and student-teacher,*
  - b) *the physical components of the e-learning system,*
  - c) *the pedagogical framework, and*
  - d) *the educational setting; learning (activity, tasks, outcome, environment, and relationships).*

## V. CONCLUSION

The Information and communication technologies act as a catalyst for innovation in learning, providing access to contextualized high-quality content. With the development of e-learning and its ability to provide rich animated content rapidly to a wide audience, new methods for teaching have evolved. The authors of e-learning material can use multiple media to present ideas and concepts, combining traditional educational content (text, images, graphs, and diagrams) with interactive computer-based resources (sound, video, animation, image series) and over an inter-networked environment made possible by web-based technologies. Among all the "e" movements in computer sciences, e-learning is one of the fastest growing. However, its full impact is yet to be felt. Therefore, this paper has attempted to present a guideline for best practice which it believes will help enhance the efficiency and impact of these Internet and web-based technologies employed in distance and collaborative learning.

## REFERENCES

- [1] D. L. Webb, A. Metha, and K. F. Jordan, *Foundations of American Education*, 6th Ed. Upper Saddle River, NJ: Merrill, 2010, pp. 77-80,192-193.
- [2] B. Courts, and J. Tucker, "Using technology to create a dynamic classroom experience," *Journal of College Teaching & Learning (TLC)*, vol. 9(2), 2012, pp. 121-128.
- [3] J. Bocchi et al., "Technology-enhanced learning in industry and higher education: preliminary report on a "gap" analysis," *The Technology Source*, May/June 1999, doi:technologysource.org/article/technologyenhanced\_learning\_in\_in\_dustry\_and\_higher\_education.
- [4] M. G. Moore, and W. Anderson, *Handbook of Distance Education*, 2nd ed., Psychology Press, 2012.
- [5] B. Holmberg, "The evolution, principles and practices of distance education," *Studien und Berichte der Arbeitsstelle Fernstudienforschung der Carl von Ossietzky Universität Oldenburg [ASF]* (in German) vol. 11. Bibliotheks-und Informationssystem der Universität Oldenburg, 2005, p. 13.
- [6] M. G. Moore, and K. Greg, *Distance Education: A Systems View*, 2nd ed., Belmont, CA: Wadsworth, 2005.
- [7] P. Olszewski-Kubilius, and S. Corwith, "Distance education: where it started and where it stands for gifted children and their educators," *Gifted Child Today*, vol. 34(3), 2011, p. 16-24.
- [8] R. Preston, "Down to business: higher education is ripe for technology disruption," *Information Week (UMB)*, vol. 60, May 16, 2011.
- [9] P. Pumilia-Gnarini, "Didactic strategies and technologies for education," *Incorporating Advancements*, 2012, pp. 48-56.
- [10] E. K. Kahiigi, L. Ekenberg, H. Hansson, F. F. Tusubira, and M. Danielson, "Exploring the e-learning state of the art," *The Electronic Journal of e-Learning*, vol. 6(2), 2008, pp. 77-88.
- [11] D. Wolley, "PLATO: The Emergence of Online Community," *Thinkofit: Consultants in Online Communication*, David R. Woolley, Feb. 2013.
- [12] S. Hiltz, "Evaluating the virtual classroom," in Harasim, L. (ed.) "Online Education: Perspectives on a New Environment," New York: Praeger, 1990, pp. 133-169.
- [13] C. B. Whyte, "Student affairs-the future," *Journal of College Student Development*, vol. 30, 1989, pp. 86-89.

- [14] R. Mason, and A. Kaye, *Mindweave: Communication, Computers and Distance Education*, Oxford, UK: Pergamon Press, 1989.
- [15] A. Bates, *Technology, e-Learning and Distance Education*, London: Routledge, 2005.
- [16] H. M. Johnson, "Dialogue and the construction of knowledge in e-learning: Exploring students' perceptions of their learning while using Blackboard's asynchronous discussion board," *European Journal of Open, Distance and E-Learning*, vol. 5(1), 2007.
- [17] T. Ellis-Christensen, "What Is Virtual Education?." wiseGEEK: clear answers for common questions. Conjecture Corporation, Feb. 2013.
- [18] G. M. Farrell, *The Development of Virtual Education: A Global Perspective*, Vancouver: Commonwealth of Learning, 1999.
- [19] A. Di Iorio, A. A. Feliziani, S. Mirri, P. Salomoni, and F. Vitali, "Automatically producing accessible learning objects," *Educational Technology & Society*, vol. 9(4), 2006, pp. 3-16.
- [20] M. Learners, "CALCampus - About." Accredited Distance Learning Courses-CALCampus Online, CALCampus, Feb. 2013.
- [21] A. Bates, and G. Poole, *Effective Teaching with Technology in Higher Education*, San Francisco: Jossey-Bass/John Wiley, 2003.
- [22] OECD, *E-Learning in Tertiary Education: Where Do We Stand?*, Paris: OECD, 2005.
- [23] C. Baker, "Blended learning: Teachers plus computers equal success," *Desert News*, 31 March 2013, doi:deseretnews.com/article/865569876/Blended-learning-teachers-plus-computers-equal-success.html?pg=all
- [24] V. Strauss, "Three fears about blended learning," *The Washington post*, doi:washingtonpost.com/blogs/answer-sheet/post/three-fears-about-blended-learning/2012/09/22/56af57cc-035d-11e2-91e7-2962c74e7738\_blog.html, 31 March 2013.
- [25] Doi:en.wikipedia.org/wiki/E-learning. 17 May 2013.
- [26] B. Smith, P. Reed, and C. Jones, "Mode Neutral' pedagogy," *European Journal of Open, Distance and E-learning*, 2008.
- [27] T. Mayes, and Sara de Freitas, *JISC e-Learning Models Desk Study*, JISC Report, 15 April 2012, doi:jisc.ac.uk/uploaded\_documents/Stage%20%20Learning%20Models%20(Version%201).pdf
- [28] N. Brügger, "Historical perspective on the World Wide Web, including issues of culture, content, and preservation," ed. *Web History*, 2010, pp. 362.
- [29] T. Berners-Lee, T. Bray, D. Connolly, P. Cotton, R. Fielding, M. Jeckle, C. Lilley, N. Mendelsohn, D. Orchard, N. Walsh, and S. Williams, *Architecture of the World Wide Web, Volume One, Version 20041215, W3C*, 15 December 2004.
- [30] J. Lynn, "Internet users to exceed 2 billion ...," *Reuters*, 19 October 2010.
- [31] R. Mitnik, M. Recabarren, M. Nussbaum, and A. Soto, "Collaborative robotic instruction: a graph teaching experience," *Computers & Education*, vol. 53(2), 2009, pp. 330-342.
- [32] R. Grier, "Interoperability Solutions," *Interoperability, Catalyst Communications*, 28 May 2011.
- [33] T. Slater, "What is Interoperability?," *Network Centric Operations Industry Consortium - NCOIC*, 2012.
- [34] S. Britain, "A review of learning design. Concept, specifications and tools," *JISC report*, 2004.
- [35] L. Yu, *A Developer's Guide to the Semantic Web*. Springer. January 6, 2011.
- [36] S. Shishehchi, S. Y. Banihashem, N. Azan Mat Zin and S. A. M. Noah, "Ontological approach in knowledge based recommender system to develop the quality of e-learning system," *Australian Journal of Basic and Applied Sciences*, vol. 6(2), 2012, pp. 115-123.
- [37] D. Vlachopoulos, and N. Cabrera, "Building an inclusive definition of e-Learning: an approach to the conceptual framework," *The International Review of Research in Open and Distance Learning*, vol.13(12), 2012.
- [38] F. P. Rokou, et al., "Modeling web-based educational systems: process design teaching model," *Educational Technology Society*, vol.7, 2004, pp. 42-50.
- [39] C. Redecker, "Review of Learning 2.0 Practices: Study on the Impact of Web 2.0 Innovations on Education and Training in Europe," *JRC Scientific and technical report*, 2009, (EUR 23664 EN – 2009).
- [40] J. Quemanda, and B. Simon, "A use-case based model for learning resources in educational mediators," *Educational Technology Society*, vol. 6, 2003, pp. 149-163.
- [41] V. Rideout, E. Vanderwater, and E. Wartella, *Zero to Six: Electronic Media in the Lives of Infants, Toddlers, and Preschoolers*, Menlo Park, CA: The Henry J. Kaiser Family Foundation, 2003.
- [42] W. Buckleitner, "So young, and so gadgeted," *New York Times*, 2008.
- [43] K. Meidlinger, "Choosing media for children checklist," *San Francisco: Kids Watch Monthly, KQED.org* (adapted from Rogow, F.), 2011, doi:kqed.org/assets/pdf/education/earlylearning/kidswatch/kidswatch-choosingmedia-english.pdf
- [44] C. Cavanaugh, "Effectiveness of cyber charter schools: A review of research on learnings," *TechTrends*, vol. 53(4), July/August 2009, pp.28-31.
- [45] I. E. Allen, and J. Seaman, *Staying the Course: Online Education in the United States*, Needham MA: Sloan Consortium, 2008.
- [46] D. G. Hebert, "Five challenges and solutions in online music teacher education," *Research and Issues in Music Education*, vol. 5(1), 2007.
- [47] F. Chukwunonso, et al., "Challenges for the adoption of new ICTs in architectural education in Nigeria," unpublished.
- [48] D. Crow, S. Parsowith, and G. Wise, "The evolution of CSCW – past, present and future developments, 1997," doi:acm.org/sigchi/bulletin/1997.2/students.html.
- [49] R. M. Glass, and J. A. Putnam, "Cooperative learning in teacher education: a case study," *In Action in Teacher Education*, vol. 4, Association of Teacher Educators, 1988, pp. 47-52.
- [50] D. Namuth, S. Fritz, J. King, and A. Boren, "Principles of sustainable learning object libraries," *Interdisciplinary Journal of Knowledge and Learning Objects*, vol. 1, 2005, pp. 181-196. doi:ijklo.org/Volumel/v1p181-196Namuth.pdf
- [51] J. V. Asuncion, C. S. Fichten, and M. Barile, "Which forms of eLearning are accessible to Canadian postsecondary students with disabilities?," *Communiqué*, vol. 7(3), 2007, pp. 36.
- [52] Z. Ahmad, "Virtual education system (current myth & future reality in Pakistan)," *Entrepreneurial Tutors*, Nov. 2010, doi:ssrn.com/abstract=1709878.
- [53] C. Dalsgaard, "Social software: E-learning beyond learning management systems". *eurodl.org*. University of Aarhus, 2013.

# A New Approach Based on Computer Vision and Collaborative Social Networking for Environmental Preservation:

Theory, Tools and Results of Italian ACI Project

Maria Grazia Albanesi

Dept. of Electrical, Computer and Biomedical Engineering  
University of Pavia  
Pavia, Italy  
mariagrazia.albanesi@unipv.it

Roberto Albanesi

Thea s.r.l.  
Pavia, Italy  
ufficiostampa@albanesi.it

**Abstract**—The ACI project (from the Italian acronym **Antropentropia Comuni Italiani**, i.e., **Italian Municipality Anthropentropy**) is a collaborative social media based project whose main goal is to measure the impact of invasive presence of human settlements on the environment. The novelty of this approach is to propose a new indicator, which considers not only the area of regions which are occupied by human settlements (buildings, roads, recreational areas), but also their relative shape and contiguity, in order to include in the metric the negative effect of territory fragmentation. Land use and fragmentation is expressed by the Anthropentropy Factor (AF); it is computed by classical computer vision operators (morphological dilation). The paper describes the theory of the new metric, its application to the Italian territory and the involvement of social network and crowdsourcing for generating open data on the current state of land use. One of the aims of this contribution is to show how well known computer vision algorithms, applied to open data collected by the User Generated Content paradigm, can be useful to map the state of degradation of the Italian territory.

**Keywords**- *Land use, territory fragmentation, anthropentropy crowdsourcing, social network, UGC, morphological operators.*

## I. INTRODUCTION

There are plenty of aspects of the effects of human activities on environment, and one of the most important is preserving land use from the threat of an inappropriate urban expansion; this is a great challenge of European Community in the field of environmental protection, as documented by several studies and data sheets: the European Environment Agency stated [1] that annual land take in 36 European countries was 111 788 ha/year in the period 2000-2006. In the most urbanized countries, the annual land take increased by 9 % in the last years. One of the consequences of land use is soil sealing, a modification of the soil which prevents it from performing vital functions for the ecosystems. The phenomenon of soil sealing is particularly evident near urban and rural areas and along main road axes. Urbanization is not the only cause of these negative effects: intensive farming and touristic activities (especially for coastlines) contribute to a relevant increase of land use, as reported by the European Environment Agency.

In the case of Italy, the situation is dramatic: land take and soil sealing assessments have been reported by ISPRA [2], showing a rapid medium growth of the sealed surface areas of 6.3% for the period 1956-2006. The main effort of researchers is twofold: to search new indicators to express the land use, and to recall or update data on the current state of the soil. In this paper, both these problems are addressed.

In Section II, the motivations of our approach are explained, in particular with reference to the current state of the art, and the definition of the Anthropentropy Factor (AF) is given. In Section III, the proposed metric is explained and discussed, and the experimental results are reported for some significant cases of Italian territory. In Section IV, conclusions and considerations about future work are given.

### A. Related work

In literature, very few contributions are related to new indicators for land use and their automatic or semi-automatic computation. For example, indicators of land quality [3] and bio indicators, such as populations of ants [4] and bryophytes [5], have been investigated. The main limitation of these indicators is that they take into consideration not only land use, but also other aspects of environmental preservation, such as agro-biodiversity, water quality and land contamination/pollution. In fact, these complex indicators consider not only soil use and degradation, but also combined aspects of water, terrain, and biotic resources. Sometimes, these indicators show a weak correlation with land use, as they are more sensitive to other features of ecosystem degradation; for example, the maps of bio indicator lichens [6], which seem very similar to our anthropentropy maps (see Section III), are very sensitive to pollution and presence and directions of winds, rather than urbanization; even if the negative effects of the two facts can be correlated, it is difficult to find indicators which isolate the dependence on each single aspect of environmental degradation.

The second problem is to compute land use in a completely automatic way, for example by using image processing primitives: several contributions refer to the use of remote sensing [7], but the most significant results are obtained after a post processing of supervised classification, which requires prior knowledge of the ground cover in the

study area. Besides, the two main problems of these systems are the coarse map resolution and a limited number of classes. A manual supervised classification is the most effective method, as in Corine Cover Project [8], with 400 classes of land cover, whose data have been used in our approach.

In conclusion, contributions in literature underline that the problems of finding significant indicators for land use and their supervised or unsupervised computation are still an open challenge, with dramatic consequences on environment preservation.

## II. A NEW INDICATOR FOR LAND USE

In this research, the first step to measure the impact of human settlement on environment is the computation of a new two-dimensional global metric. The metric is based on the concept of *anthropentropy*: this term is a neologism [9] derived from the Greek term *Anthropos* (*ἄνθρωπος*) = man, and *entropy*. As entropy is the well-known measure of disorder of a system, the term anthropentropy refers to the “disorder” introduced in a virgin environment by the presence/penetration of human beings. The basic idea is to associate to a limited geographic area a new indicator, called *Anthropentropy Factor* (*AF* in the following), which expresses the degree of penetration of human settlement in the environment. This is not only based on classical well-known indicators, such as the percentage of soil occupied by human activities and urban expansion, but it takes into consideration also the *shape* of the areas subtracted to nature. For this reason, we have chosen the term two-dimensional metric, where the two dimensions refer to a two-dimensional plan description of the areas of human settlements and their relative shapes and spatial relationships, in particular their contiguity. This choice is supported by the results of several studies in ecology: many researches highlight that the shape of the land is important for wild animal species that inhabit it and that the fragmentation of the territory contributes greatly to limit some environmental key treats, for example biodiversity. In fact, the UN Convention on Biological Diversity [10] considers fragmentation as a major threat to habitats and species populations, because it has a direct impact (due to proximity and disturbance) in creating barriers to the wandering and spreading of animals. One of the most important causes of fragmentation of territory is a disordered urban expansion where the incremental areas (due to new human settlements) are distributed in the territory in such a way that new settlements are not contiguous to existing ones. This is potentially a great drawback, because it increases greatly fragmentation. To illustrate this concept with a simple schematic example, suppose that the initial situation of an area of 16 square km is that of a single urban settlement of square shape of 1 km (Fig. 1a): as a consequence of the urban expansion, this area gets wider, by rising the measure of the side to 2 km (Fig. 1b). The increase of the use of soil passes from 0.0625 to 0.25. Now let us suppose instead of switching to a new situation in which there are three other new settlement squares of 1 km each (Fig. 1c). The new occupation of the land use remains the same (0.25), because cases (b) and (c) have the same relative

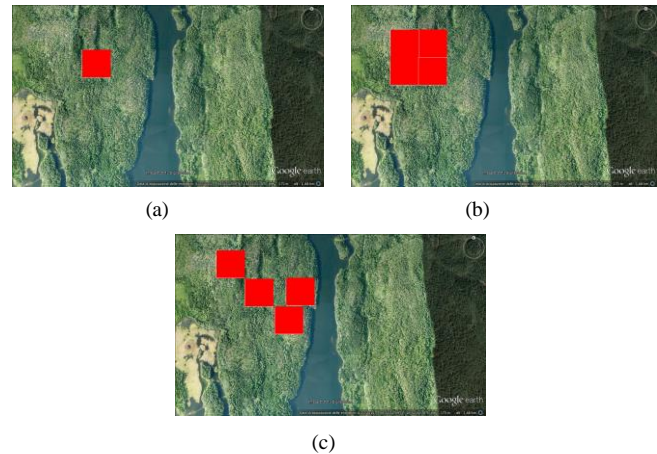


Figure 1. The sprawl of urban settlements causes a loss of continuity and consequent territory fragmentation. (a) original situation; (b) contiguous incremental areas (c) non contiguous incremental areas.

land take increment (+25%), but the situation, regarding the fragmentation of territory, is clearly more dramatic. Despite the calculated indicators are the same, the first scenario (b) is definitely preferable than the second (c) from the ecological point of view. A simple qualitative example is the construction of a road in the countryside: the impact on wildlife is not just due to the simple loss of land, most likely because the animals of the area will not only reduced their amplitude territory occupied by concrete and asphalt, but it is slightly larger, due to other aspects of environmental pollution (noise, vibration, impossibility to cross the road).

For this reason, by considering the increment of urbanized areas, it is important to notice that the relative numeric increment of land use is not a sufficient indicator to take into account. In [11] the problem of land fragmentation is analyzed in full details, but with two great limitations: data refers to 2000 and have no longer updated and (2) the fragmentation taken into consideration is the only one induced by great communication roads. However, if you want to measure the quality of life of individuals, the use of such macroscopic level data may not be sufficient. In order to overcome the limitations of the classical approaches of previous studies, and to improve our knowledge about environment preservation, it is mandatory (a) to update data to a more recent situation and (b) to take into consideration several urban settlements, both from the point of view of considering all their possible typologies, and also their spatial distribution, in order to minimize fragmentation. How you can put all this logic in an indicator that: a) takes into account the relative positions of the urbanized areas and b) is easy to compute and to be updated? The solution proposed here is the metric based on anthropentropy and the consequent definition and computation of indicator *AF* on a given geographic area in a collaborative environment, using the paradigm of User-Generated Content (UGC). The metric and the *AF* definitions answer challenge (a), while to answer the challenge (b), the second novelty of our research is adopted: *social crowdsourcing*, by involving in the project a social networking community in order to generate open data.

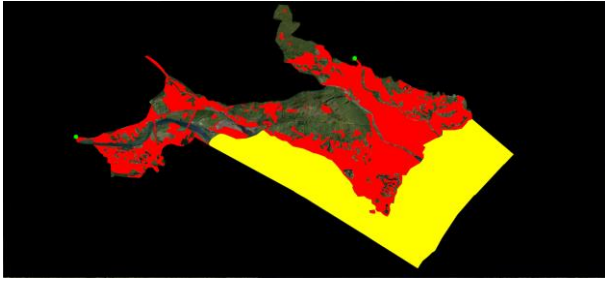


Figure 2. Example of anthropentropy mapping for municipality of Verbania (Italy): in red the anthropic places, in yellow the neutral area.

#### A. The AF definition

The solution consists in considering an urbanized area and then "enlarge" of a security zone by expanding its shape in the two dimensions and, consequently, by increasing its size. The extension has the purpose of taking into account the negative effects (due to noise and pollution) in the immediate vicinity of human urbanized areas; moreover, if the urbanized areas are sufficiently close each other, the enlargement causes an effect of dilation (where contiguous "holes" can be filled), which defines wider areas of individual settlements. The indicator must take into account these enlarged areas, instead of the original ones.

In order to investigate the human presence as the occupation of land by human resources that are not compatible with nature, the whole set of *anthropic places* are defined by four typologies (in the following, referred as *List 1*): *buildings of any kind* (housing, workplaces, schools, hospitals, if there is the actual presence of human activity, also in occasional mode, for example a settlement of voltaic solar panels), *paved roads, railways, and places of intensive agriculture* (such as greenhouses, nurseries, where the human presence is fairly continuous and has effectively ousted the wild).

In order to compute the *AF*, some considerations and definitions are necessary. Let us consider a generic geographic region bounded by recognized borders (e.g., a municipality, a county or a state) and let define *S* the area (in squared kilometers) of the region (assuming you have an appropriate two-dimensional scaled map and the numeric value of *S*). Within the region, we proceed to the identification of all the sub-regions occupied by the anthropic places of *List 1*. Sub-regions can be disjoint. Each sub-region contains at least one of the anthropic places listed in *List 1*. We define a *neutral sub-region* as a part of territory containing at least these two kinds of elements: (a) stretches of inland water (lakes, lagoons etc.) extending more than two squared kilometers (this choice is to be consistent with the coastline, which is normally considered as administrative boundary of the coastal municipalities) and (b) the land located more than 3,000 m above sea level. For the given region, the union of all the neutral sub-regions (if present), correspond to the *Neutral Zone*. Let define *NA* as the area (in squared kilometers) of the Neutral Zone. Obviously, not all the geographic region necessary contain a neutral zone, in that case *NA* is set to 0. In Fig. 2, an

example of anthropentropy mapping of the municipality of Verbania is shown. The portion of the lake is the yellow area, the red area are the anthropic places (for clarity, the territory outside the municipality is blackened). Each area occupied by anthropic places is enlarged (along the two Cartesian dimensions *X* and *Y*) with a factor of "charging" of 50 meters, to give rise to anthropic sub-regions; for example, suppose a generic sub-region is a square of side 1000 meters, then the corresponding anthropic sub-region will be a square of side 1100 meters. The choice of a "buffering" of 50 meters is based on the principle of reciprocity: in Italian laws, most particularly intrusive activities, e.g., hunting, shall be allowed only at a distance greater than 50 m from urban settlements. This distance then seems a good compromise between the two extremes of a too restrictive and a too permissive limit. We define the union of all the anthropic sub-regions as *Death Zone* of the region. Let define *DA* as the area (in squared kilometers) of the Death Zone. We define the *Anthropotropic Factor AF* as the ratio:

$$AF = DA / (S - NA) \quad (1)$$

The *AF* is a fractional number between 0 (completely uninhabited regions without human active settlements) and 1 (fully populated regions). In (1) the special case of *NA = S* is not considered, as it would mean that the entire geographic area is occupied by water or it is located above 3,000 m above the sea, thus it is not suitable to urban settlements and the computation of *FA* is meaningless.

#### B. Critical issues in the AF computation and their solutions

The main purpose of the ACI project is to map the complete Italian territory in such a way that, for each municipality, the *AF* factor (1) is computed according to the procedure described in Section II.A. The first problem is the availability of a description of land use in terms of anthropic places (see *List 1*), possibly with an accuracy of scale comparable with the size of the dilation (50 m) accomplished by the algorithm. At this purpose, the ACI project refer to the data-set created by the Corine Land Cover (CLC) project [8]; it was born in Europe specifically for the detection and monitoring of the characteristics of cover and land use, especially for environmental protection. The first realization of the CLC data-set refer to 1990 (CLC90) and subsequent updates refer to the year 2000 through the project Image & Corine Land Cover 2000. In 2000, 33 countries adhered to the project, including Italy.

Unfortunately, the first critical issue of the project is that Corine data set is not available for all the Italian territory, but only for 7 regions (over a total of 20 regions), namely for the 40,9% of municipalities (3311 over 8092). For the remaining municipalities, data of land use are not available at all, or they are not available free of charges. To fill the gap of data, an UGC approach has been use (see Section II.C). As our project stresses the importance of open data for society growth, we had to face this problem and we solved it by proposing a solution based on User Generated Content and social networking. Therefore, the results of the ACI project



derive from two distinct procedures: (a) the AF mapping based on Corine Land cover data set (due to the availability of Corine data set, this part can be developed for a total percentage of 40.9% of Italian municipalities), and (b) the AF mapping based on UGC and social networking open data.

In Section III, results for the two cases are described. First, the procedure for UGC and open data generation of our collaborative project has to be described.

### C. UGC and collaborative social networking for AF computation

The ACI project can count on a social network consisting of 7000 users [12]. All of them are potentially involved in the project of UGC. Collaborative users are asked to generate a map (UGC map) on Google Earth environment in which the limits of anthropic sub-regions and neutral zones (if any) are plotted. Obviously, each user will produce the map for the municipality in which he/she actually lives: in this way the knowledge of one's own territory, provided by everyday user experience, is coded in the map. A procedure for the generation of the UGC map of the territory is given to the users of the social network. The procedure must have the following characteristics: (a) it requires a *low level of expertise* in informatics: the only required skills are to be able to open files and do some basic operations such as selecting and tracking area boundaries filled with different colors (red for anthropic sub-regions and yellow for neutral zones), and (b) it is based on *open software and open data* (Google Earth maps and Gimp). The user is given a detailed procedure to download the Google Earth map and to fill areas with the proper set up of Gimp tools (types, fillings, colors). In Fig. 3a, an example of UGC map for the municipality of San Martino Siccomario (Pavia, Italy) is shown: red areas refer to the anthropic places, in black the limits of the municipality are shown. After the user has generated the UGC map of his/her own municipality, we received the map and processed it according to a simple computer vision algorithm, implemented in Matlab. It is composed of four steps. The first one is a calibration step, which is necessary to discover the scale of the UGC map. At this purpose a couple of points are selected on Google earth original map and their distance is measured in Google Earth with the Ruler tool. The same two reference points are selected in matlab interactive environment and the Euclidean distance is computed (see Fig. 3b, the green line is drawn for euclidean distance and computation of pixel dimensions). This gives the effective area represented by one pixel, which is assumed to be squared. In the second step, the red areas are enlarged by the well-known morphological operator of dilation [13], in X and Y dimensions, assuming a radius of dilation of 50 meters (see Fig. 3c). This step creates the Death Zone. By comparing Fig. 3b and Fig. 3c, it can be noticed that areas are enlarged and, in some points, "holes" are partially or completely filled (further examples are reported in Experimental results, see Fig. 6). In the third step, the areas of dilated zones and neutral zone (if any) are computed by simply counting the number of pixels of the Death Zone and of the Neutral zone (if any) and by

multiplying these numbers for the effective area of a pixel computed in the calibration step. This step gives rise to the values of DA and NA in (1). In the last step, the area of the entire municipality is read from official on line Istat database [14] and AF factor is computed according to (1).

### D. Discussion on applicability, merits and limitations

The first important question is: is AF a significant factor? This indicator gives, in an absolute scale (from 0 to 1), the degree of land use and it is very intuitive to understand. Consider the *paradox of cement*: if every Italian citizen occupied with his/her business (home, work, communication routes) an area of 70 X 70 meters, the whole Italian peninsula (including mountains!) would be populated and the factor AF would be equal to 1 everywhere. The distance from value 1 indicates how far we are from this tragic limit situation. The second important question is: in which aspects can the AF help in environment preservation? As the metric is based on an objective, scientific indicator and it is very simple to understand, it can be a very powerful tool to help legislators and local administrators to evaluate the present situation and to impose limits to urbanization.

Some limitations of the approach have to be underlined, especially in the part of UGC data set, in particular: (a) the precision of Death Zone and Neutral Zone definition varies from map to map, as it is related to the manual skill of selection by mouse of the user (this precision cannot be estimated in advance in a quantitative way), and (b) the knowledge of the land use of the territory varies from map to map, because it is related to the awareness of land use of the user. For the first limitation, several tests have been done to assure that the precision of manual selection by the mouse of the two reference points does not affect the computation of FA. Even with a simulated error of selection up to five pixels, the variation of the indicator AF is less than 0.01%. For the second limitation, a preliminary check on the quality of the UGC map is performed manually, by a strict cooperation with the social network users, in order to verify the level of their expertise. Obviously, the UGC values of AF cannot replace the same precision the computation based on Corine data set, but as the main goal is not replacing but filling the gap of open data, this method can be a valid solution: besides, the method based on UGC maps have three main advantages. First of all, it increases the awareness of the citizens on the state of the environment where they lives. Map reading and, above all, participation to complete them makes us more aware of the environmental richness that we are losing. And it is, after all, the ethical purpose of ACI project, to show that crowdsourcing and social technologies can help to improve society. Besides, the procedure for AF computation can be used as a simulation to study future urban expansions in a *scenario of What if?*. This suggests a possible very useful application of our system: to help local government in planning the annual Territory Government Plan where new urban expansions are decided. Finally, it can react rapidly to the local changes for land use, as the main source of knowledge is the citizen (a typical bottom-up knowledge). For what concern the comparison to a fully automatic image processing approach, as we pointed out in

Section I.A (Related work), a fully automatic method to recognize anthropic places is not possible, because general image processing characteristics, such as multispectral information, color and texture are not sufficient to code in a map the level of knowledge necessary to differentiate a nursery for intensive agriculture from a field of sorghum (extensive agriculture), a golf course from a simple lawn, or a complex of abandoned and inhabited houses from an habited one.

### III. THE METRIC AND EXPERIMENTAL RESULTS

#### A. The metric

After the computation of AF index, a simple numeric (and the corresponding visual representation) metric is proposed here: if AF is between 0 and 0.2, the area is considered a very low “anthrophized area” (the ideal condition for the environment). This condition is represented visually on the map by coloring the area into varying levels of green. If AF is between 0.2 and 0.4, the area is considered with a limited but worrying level of anthropentropy. This condition is represented visually on the map by coloring the area into varying levels of yellow. This type of area have to be monitored in time to control its evolution potentially to undesired higher level of anthropentropy. If AF is between 0.4 and 0.6, the area is considered with a serious level of anthropentropy. This condition is represented visually on the map by coloring the area into varying levels of red. In these areas, the presence of humans greatly impact negatively on environment. If AF is between 0.6 and 1, the area is considered with a very serious level of anthropentropy, in an irreversible environmental degradation. This condition is represented visually on the map by coloring the area into varying levels from violet to black. In these areas, the presence of humans has completely compromised the situation. Following the metric and the visual convention on visual mapping, for a given area of the Italian territory (a group of municipalities, i.e., region), it is now possible to generate an *Anthropentropy Map*. In Section III.B, significant results are shown and commented.

#### B. Experimental results

In Fig. 4, an example of Anthropentropy Map for region Trentino (Italy) is shown. The most compromised areas are near big cities, and this is, to a certain extent, a predictable result, however, the situation is quite good because no black areas are present even on the cities, and red areas are limited both in extension and dispersion. This can be quantitatively expressed by the statistics: the minimum value of AF is 0.01, the average is 0.14, and the standard deviation is 0.11. From the anthropentropy map, a low level of fragmentation of green territory is evident. In Fig. 5, the most compromised situation is shown. The black spot refers to the high urbanized area of Milan. The map shows clearly however that the worrying situation is also far off the big cities, in countryside, due to local uncontrolled urban settlement growth. In Fig. 6, some examples of UGC maps are presented, and the corresponding Death Zone computed by the computer vision algorithm based on morphological

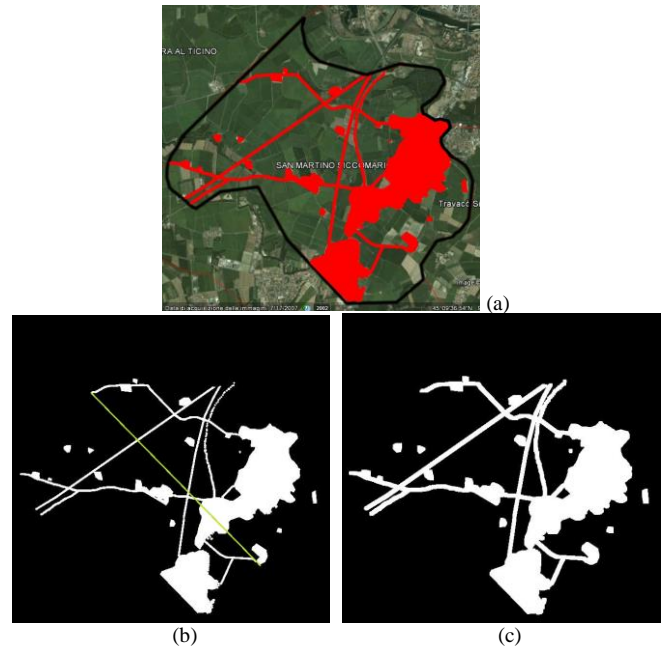


Figure 3. AF computation for the municipality of San Martino (Pavia, Italy): (a) original UGC map (b) calibration step (c) Death Zone.

dilation. The maps refer to the municipalities of Erice (Sicily) and Scarperia (Tuscany), with AF of 0.214 and 0.1637, respectively.

### IV CONCLUSIONS AND FUTURE WORK

In this paper, an innovative indicator to evaluate land use is proposed, basing upon the concept of anthropentropy. According to the definition, a method to compute the Anthropentropy Factor for a given territory is described, both on open data and on collaborative UGC maps. On the time of writing of the present paper, the ACI project covers a percentage of generation of Anthropentropy Maps for the whole set of Italian municipality (8092) equal to 36,7%, and new results are constantly being collected, both from the social collaborative approach and on new, recently available, open data. Future work includes the mapping of the entire Italian territory and the presentation of results to local and governmental administrations, in order to put in practice an innovative policy of environment preservation based upon a sustainable land use and to measure with some metric the efficacy of the method on real cases of environmental degradation.

#### ACKNOWLEDGMENT

The authors greatly appreciate for his collaboration Antonio di Gennaro [15], for AF and map generation from Corine data set.

#### REFERENCES

- [1] Overview of exiting Policies to reduce and mitigate soil sealing in the EU and Member States, [http://ec.europa.eu/environment/soil/pdf/sealing/4.%20Overview%20of%20existing%20Policies%20EU%20and%20MS%20\(A-H\).pdf](http://ec.europa.eu/environment/soil/pdf/sealing/4.%20Overview%20of%20existing%20Policies%20EU%20and%20MS%20(A-H).pdf) [retrieved: May, 2013].

- [2] Land Use, State and impact (Italy). <http://www.eea.europa.eu/soer/countries/it/land-use-state-and-impacts-italy> [retrieved: May, 2013].
- [3] J. Dumanski and C. Pieri, "Land quality indicators: research plan," *Agriculture, Ecosystems & Environment* 81.2, 2000, pp. 93-102.
- [4] J. D. Majer, "Ants: Bio-indicators of minesite rehabilitation, land-use, and land conservation," *Environmental Management*, Vol. 7, n. 4, Springer Verlag, 1983, pp. 375-383.
- [5] H.G. Zechmeistera, I. Schmitzbergera, B. Steurerb, J. Peterseila, and T. Wrkkaa, "The influence of land-use practices and economics on plant species richness in meadows," *Biological Conservation*, n. 114, Elsevier, 2003, pp. 165-177.
- [6] <http://www.arpa.veneto.it/temi-ambientali/aria/file-e-allegati/docup-2007/09-Isocrono.pdf> [retrieved: May, 2013]
- [7] J. Rogan and D.M.Chen, "Remote sensing technology for mapping and monitoring land-cover and land-use change," *Progress in Planning*, n. 61, Elsevier, 2004, pp. 301-325.
- [8] Corine Land Cover Project: <http://www.eea.europa.eu/publications/COR0-landcover>, [retrieved: May, 2013]
- [9] ACI project: <http://www.albanesi.it/Ambiente/ACI.htm> [retrieved: May, 2013]
- [10] The UN convention on biological diversity, <http://enrin.grida.no/biodiv/biodiv/cbd/eeafoll.pdf> [retrieved: May, 2013]
- [11] Fragmentation of ecosystems and habitats by transport infrastructure, <http://www.eea.europa.eu/data-and-maps/indicators/fragmentation-of-land-and-forests/eu-ac-fragmentation>. [retrieved: May, 2013]
- [12] Facebook page: <https://www.facebook.com/RobertoAlbanesi>, [retrieved: May, 2013]
- [13] R. C. Gonzales and R. E. Woods, *Digital Image Processing*, Pearson Prentice Hall, 2008, Chapter 9, "Morphological Image Processing".
- [14] ISTAT: [www.comuni-italiani.it](http://www.comuni-italiani.it) [retrieved: May, 2013]
- [15] LinkedIn page: <http://www.linkedin.com/in/adigennaro> [retrieved: May, 2013].

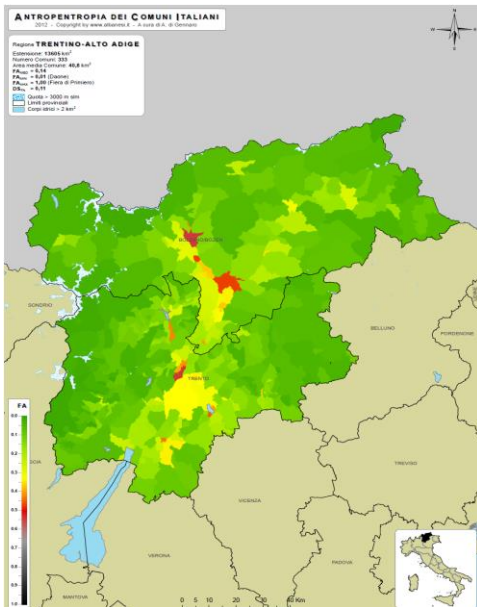


Figure 4. Anthropentropy Map of Region Trentino, Italy (land area: 13605 squared kilometers).

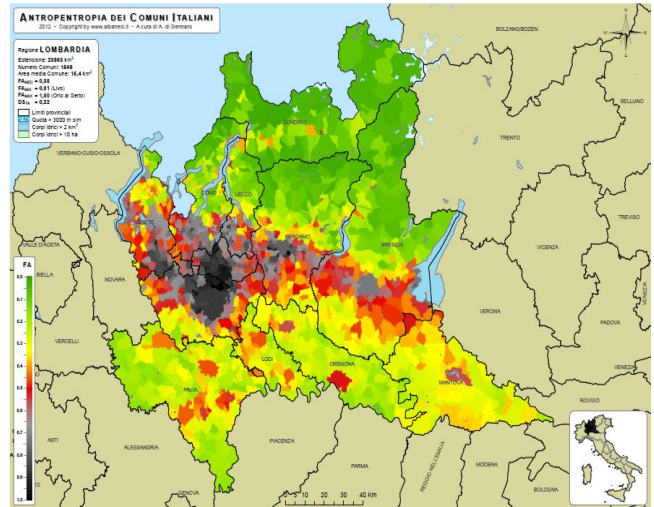


Figure 5. Anthropentropy Map of Region Lombardia, Italy (land area: 23863 squared kilometers).

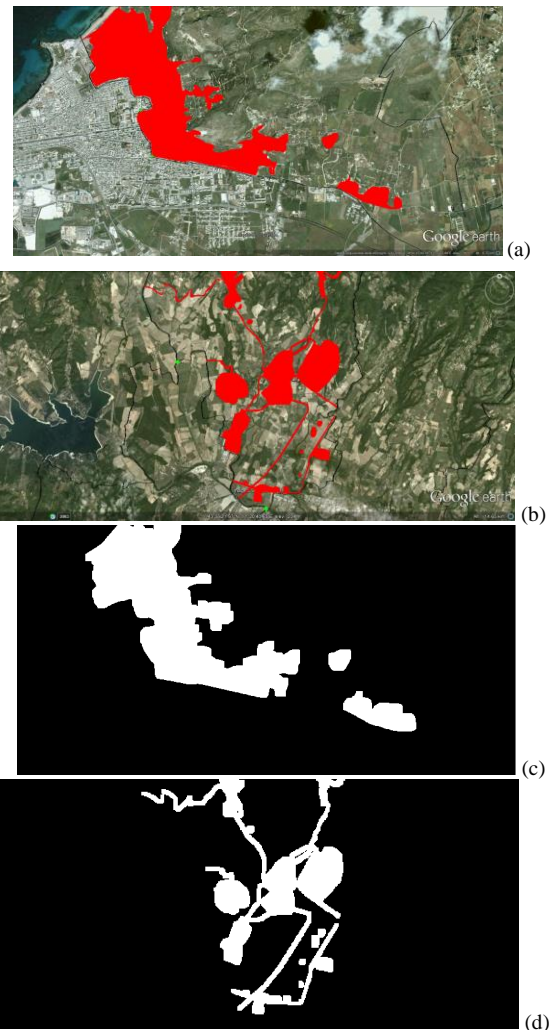


Figure 6. Example of UGC maps for the municipalities of (a) Erice (Trapani, Sicily) and (b) Scarperia (Florence, Tuscany) and their corresponding Death Zones (c and d).

# Impact of the Network Structure on the SIR Model Spreading Phenomena in Online Networks

Marek Opuszko

Department of Business Informatics  
Faculty of Economics and Business Administration  
Friedrich-Schiller-University of Jena  
Jena, Germany 07743  
Email: marek.opuszko@uni-jena.de  
Telephone: (0049) 3641-943314

Johannes Ruhland

Department of Business Informatics  
Faculty of Economics and Business Administration  
Friedrich-Schiller-University of Jena  
Jena, Germany 07743  
Email: j.ruhland@wiwi.uni-jena.de  
Telephone: (0049) 3641-943310

**Abstract**—We present an analysis of how the spreading phenomena, which includes the spread of information, innovations, ideas, trends, etc. is influenced by the structure of the underlying (social) network. We conducted spreading simulations using the SIR (Susceptible-Infected-Recovered) model on a large number of real-world and artificially generated network datasets. The results show that the network characteristics have a significant effect on the SIR diffusion. The results reveal that the network structure affects the diffusion in terms of the achieved diffusion, the course of the diffusion and the predictability of the diffusion. We could also show that the effect of other parameters, such as the impact of seeding (network targeting), significantly differs in relationship to the underlying network structure. With these findings, we are able to provide a more differentiated picture on previous contradictory findings, especially in the field of seeding strategies. We further provide useful recommendations for future research to make results more generalizable and comparable.

**Keywords**—online social networks, information diffusion

## I. INTRODUCTION

The diffusion phenomena in social networks, such as the Internet or the World Wide Web, has attracted much attention in recent years, both in the research and the industrial sector. Developments like the Web 2.0 and subsequent technologies facilitated a peer-to-peer spread of information besides the traditional one-to-many broadcasters like TV or radio stations. The emergence of virtual social networks on the Internet offers new data and insights into this phenomenon, since it has been challenging to access detailed and large-scale data in the past. New forms of marketing emerged, such as "viral marketing" as first coined by Steve Jurvetson in 1996 [1]. The metaphorical term *viral* directly leads the origin of the phenomenon in the research on the spread of epidemic diseases [2]. Although research helped to understand the spreading behavior of diseases like H1N1 [3], Severe acute respiratory syndrome (SARS) [4] and influenza A [5], the diffusion of violent topics in social media [6], or the success of product innovations [7], there are still unresolved issues regarding the influencing factors of a diffusion.

In an information diffusion or viral marketing scenario, three main components determine the (viral) process: (1) the behavioral characteristics of the members of the network, (2) the seeding strategy, and (3) the structure of the social network [8]. Since the seeding strategy, in other words the selection of

the first network nodes, is largely under the control of the marketer, this issue received a lot of attention. In particular, the computer science and marketing community addressed this issue [9][10][11][12][13]. Kempe, Kleinberg, and Tardos, for instance, formulated the theoretical approach of the diffusion maximization problem [9]. Still, questions about the impact of seeding and the best seeding strategy exist [10][14]. Due to the previous focus on this issue the third aspect, the structure of the underlying network, only recently received consideration. One reason could be the progress made in social network analysis in the last years. Nevertheless, although efforts have been made, many questions remain open.

The majority of the research is based on investigations on only very few network datasets [9][12][15] and it is unclear if the findings are applicable to any other network. Watts and Strogatz further showed that some types of networks facilitate an epidemic spread [16]. Hence, a deeper understanding is needed, what characteristics of a network influence a diffusion.

In this paper, we examine an information diffusion process based on diffusion simulations using the SIR diffusion model on a large number of both real-world and artificially generated networks. We will introduce several metrics to describe network characteristics and evaluate the relation between the metrics and the diffusion results. This also includes the relation between several possible seeding methods to answer the question whether and how seeding methods interact with different types of networks. The paper is organized as follows. The next sections describes the latest state of research and highlights important research gaps. Section III explains the simulation analysis including the datasets and network characteristics used in this paper. Section IV presents the analysis and the results. Finally, the paper is closed with our conclusion in Section V.

## II. DIFFUSION IN ONLINE NETWORKS

The research on the diffusion of information in online networks has mainly derived from the studies of infectious diseases and epidemics [2][17]. Although there has been research on product growth models prior to the groundbreaking work made in the field of epidemics [18], the knowledge gained in the field of information diffusion has enriched research in other contexts, such as the diffusion of (product) innovations [19][7]. Especially the boost of computer networks, the progress made

in social network analysis and the success of the Internet have caused a large spread of research on diffusion in networks in various contexts. Of particular interest is the research on electronic word-of-mouth marketing or viral marketing [20][21] and the research on the effect and the optimization of seeding strategies [22][9][10][23]. The seeding strategy determines an initial set of (network) nodes, usually to maximize the spread of information or to boost the effect of word-of-mouth. However, in the past, contradictory findings have been made in terms of the effectiveness of seeding [10][9][12][14][24]. Aral, Muchnik, and Sundararajan recently argued that conventional wisdom about seeding strategies should be questioned as the effectiveness of seeding might be overestimated if other factors like network characteristics or node similarities (homophily) are not taken into account [24].

In the past, different models have been introduced to theoretically describe the spreading of diseases or information. Beside approaches like threshold models [25] based on the idea of adoption or cascade models [21], the SIR model, and in particular, the network SIR model can be considered as a reference model in terms of information diffusion due to its numerous successful applications [26][27]. The SIR acronym is derived from the three states or conditions a network node can occupy: *susceptible*, *infected* or *recovered*. In the context of information diffusion the *susceptible* state would reflect nodes that have not received a circulating information artifact. The *infected* nodes are nodes which have already received the artifact and are able to spread it further to neighbor nodes, and *recovered* nodes (sometimes also referred to as *removed* or *inactive*) may unsuccessfully receive the circulating information without any consequences. Due to the dichotomous nature of the possession of an information artifact, the peer-to-peer transmission, in other words an infection, can be modeled as a Bernoulli game. The outcome is 1 if the transmission was successful and 0 if not. Hence, there is an expected infection probability  $p$  of a single node reflecting the likelihood to infect other nodes.

Past research showed that the characteristics of many (social) networks facilitate an epidemic spread, which also accounts for information diffusion and viral marketing [16]. Bampo, Ewing, Mather, Stewart, and Wallace compared a real-world spreading in a large peer-to-peer network with a simulated SIR modeled spreading on generated random networks [8]. They created random graphs comprising the same topological properties like density and degree distribution as the real observed network. Their findings showed that the social network had a significant impact on the performance of a viral marketing campaign. They further stated that clustered networks are not very efficient and that small-world networks generally temper the spread of messages. Shakarian and Paulo investigated on viral marketing diffusion simulations in numerous networks and observed three different types of networks, each type showing a very distinct diffusion pattern [13]. They grouped the networks based on the diffusion results into the groups *highly susceptible*, *susceptible*, and *diffusion hamper*. The results did, however, not provide an explanation what characteristics relate to a higher diffusion susceptibility. A more detailed examination was conducted by Opuszko and Ruhland [28], who investigated independent cascade and linear threshold diffusion simulation and showed that the diffusion strongly depends on the underlying network. They found that

the seeding impact is not present in every network, especially not for cascade-like diffusions.

### III. METHOD

To address the questions mentioned in the introduction, a simulation analysis was conducted using a set of 35 different network datasets (see next subsection for details). For every network dataset 200 *runs* of simulations with randomly chosen start parameters were conducted. For every *run* the following parameters were set randomly: the *number of start nodes* was set at random in the interval [1,50], the *seeding criterion* for choosing the start nodes (based on several network centralities, see below) was selected randomly and the nodes have been selected accordingly by calculating the metrics for all network nodes, the *infection probability* for all nodes was set randomly in the interval [0,1]. Every simulation run comprised 50 single simulations with the prior set parameters. This was done in order to assess the variation of a spreading under identical preconditions. We measured three simulation outcomes, the average number of *infected nodes*, in other words the achieved network diffusion, the average *number of circulations* and the *standard deviation of the diffusion* of one simulation run. The number of circulations reflects the discrete time steps a virus or a diffusion was present and circulating before it died out. The results of one simulation run were stored including all used parameters and the network metrics as one single case. All in all, 350,000 simulations were conducted. The resulting dataset comprised 7,000 cases for each simulation run including the network characteristics (see next subsections), the simulation parameters and the results.

#### A. Network Datasets

TABLE I. NETWORKS USED IN THE ANALYSIS

Name	Nodes	Edges
Dolphin social network [29]	62	159
Les Miserables character network [30]	77	254
Power grid network [16]	4,941	6,594
Student Network	471	926
Gnutella 2008 peer-to-peer network [31]	6,301	20,777
OCLinks social network [32]	1,899	20,297
NetScience coauthorship network [33]	1,461	2,742
Internet snapshot [34]	22,963	48,436
Hep-th coauthorship network [35]	7,610	15,751
Cond-mat 2003 coauthorship network [35]	30,460	120,029
Erdős collaboration graph [36]	6,927	11,850
Astrophysics coauthorship network [35]	16,046	121,251
Email messages network [37]	1,133	5,451
Jazz musician network [38]	198	2,742
PGP users giant component [39]	10,680	24,316
Barabási-Albert 1 [40]	60	177
Barabási Albert 2 [40]	80	237
Barabási Albert 3 [40]	1025	1024
Barabási Albert 4 [40]	1,400	2,798
Barabási Albert 5 [40]	5,241	15,714
Barabási Albert 6 [40]	20,000	39,998
Barabási Albert 7 [40]	30,000	119,996
Erdős-Rényi 1 [41]	868	1,040
Erdős-Rényi 2 [41]	914	12,683
Erdős-Rényi 3 [41]	1,000	14,902
Erdős-Rényi 4 [41]	1,000	25,362
Erdős-Rényi 5 [41]	6,290	19,996
Erdős-Rényi 6 [41]	6,917	23,767
Watts-Strogatz 1 [16]	60	177
Watts-Strogatz 2 [16]	80	240
Watts-Strogatz 3 [16]	60	177
Watts-Strogatz 4 [16]	1,400	2,800
Watts-Strogatz 5 [16]	6,927	14,994
Watts-Strogatz 6 [16]	20,000	40,000
Watts-Strogatz 7 [16]	30,000	120,000

Both real-world and artificially generated networks have been used as a source for the simulations. Table I shows the network datasets used for the SIR simulations. The real-world networks include some of the most common datasets used in social network analysis. One exception is the dataset *Student Network*. This network has been extracted in a former analysis. It comprises a Facebook friendship network of university freshmen after their first semester of study. It should be noted that prior to the analysis, all isolated nodes have been deleted from the graphs.

Three state-of-the-art algorithms have been used to generate the artificial networks: Erdős-Renyi game [41], Watts-Strogatz game [16] and Barabási-Albert game [40]. The artificial networks have been generated in order to represent the characteristics of the real-world datasets in terms of node and edge count. All calculations have been done using the *igraph* [42] package in the R software [43].

### B. Network Characteristics

To describe a network's structure and characteristics and to later evaluate diffusion predictors, we calculated several metrics to describe a network.

- *Number of network nodes*, usually the number of users.
- *Number of network edges*, the number of connections between the nodes.
- *Network density*, the relation between existing and possible edges.
- *Connected graph* (yes, no), a graph is connected if all nodes belong to one (giant) component and no individual clusters exist.
- *Average path length*, the average of all shortest paths between any two nodes of the network.
- *Number of components/clusters*, number of isolated components comprising at least two nodes.
- *Network/Graph diameter*, the diameter of a network is the length of the longest shortest path between two arbitrary nodes in the network.
- *Average node degree*, a normalized ([0,1]) metric of the degree centrality of all nodes.
- *Average node betweenness*, a normalized ([0,1]) metric of the betweenness centrality of all nodes.
- *Average node closeness*, a normalized ([0,1]) metric of the closeness centrality of all nodes.
- *Average node eigenvector*, a normalized ([0,1]) metric of the eigenvector centrality of all nodes.
- *Average clustering coefficient*, a normalized ([0,1]) coefficient of the clustering coefficient of all nodes.
- *Number of network communities*, communities are sub graphs or dense groups of nodes within a network that are sparsely connected to other groups. In contrast to components, these groups are not isolated from each other. We used the leading eigenvector community detection algorithm according to Newman [36].

- *Degree distribution power law fit*, since the degree distributions of network nodes often show a power law distribution, we fitted a power-law distribution with maximum likelihood methods as recommended by Newman [44] against the degree distribution of each network.

### C. Seeding Criteria for the Selection the Seeding Nodes

There are several criteria for selecting the initial *infected* nodes. The most simple is by randomly activating a set of nodes. A common method is to evaluate the centrality of every node in the network based on different centrality measures and to choose the most central nodes as those are supposed to be most influential [45]. Kempe et al. initially formalized this problem as the influence maximization problem [9]. We will evaluate the following centrality measures as a criterion for the selection of the seeding nodes. All centrality metrics have been calculated according to Wasserman and Faust [46] as well as Newman [47]. We refer to those publications for further details on the calculations. It should be mentioned that the presented list of seeding criteria is far from complete. The methods used are based on network centralities and cannot reflect node behavioral aspects like adoption propensities or node homophily.

- *Degree Centrality*, one of the most common centrality measures. Degree centrality reflects the number of ties (also known as neighbors or friends in the context of online social networks) of a node.
- *Betweenness Centrality*, this centrality reflects the probability of a node to lie on a shortest path between two randomly chosen nodes.
- *Closeness Centrality*, this reflects the inverse farness of a node to any other node in the network.
- *Eigenvector Centrality*, a natural extension of the degree centrality. The difference is that nodes also award "points" for the degree centrality of their neighbors. A node is central if it is connected to other important nodes.
- *Node Clustering Coefficient*, sometimes also referred to as transitivity. The clustering coefficient of a node is the relation of the number of pairs of neighbors that are connected to the number of pairs of neighbors. In online social networks this reflects to the connection among a users' friends.
- *PageRank Coefficient*, extension of the eigenvector centrality used by Google to rank the centrality of web pages [40]. The difference is that the centrality of a node is further divided by the out-degree of a node.
- *Random*, the initial start nodes are chosen randomly.

## IV. ANALYSIS

### A. Descriptive analysis

Fig. 1 shows the resulting SIR diffusion means of all networks in relation to the infection probability and boxplots for five different infection probability intervals. The plot highlights

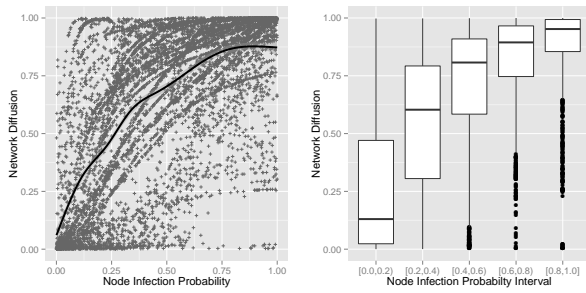


Fig. 1. Left Plot: Simulated diffusion of all networks depending on the infection probability of the network nodes. The figure includes a regression line (solid line). Right Plot: Boxplots of the network diffusion of different infection probability intervals

a high variance in the diffusion depending on the underlying network. Interestingly, the standard deviation of the network diffusion for all simulation runs (comprising 50 individual simulations using a fixed parameter setting) is only 0.022, showing that within a similar parameter setting, the diffusion is quite foreseeable.

Aside from very high values of the infection probability, we can observe that the diffusion percentage spreads over the whole range from almost no diffusion to nearly total diffusion, especially in the middle and lower area of the infection probability. The boxplots in Fig. 1 depict the high variation and uncertainty in the diffusion. In the node infection probability interval of [0.2, 0.4] the standard deviation is 0.306 having a mean of 0.543. When the networks are examined separately, the picture becomes more clear as seen in Fig. 2. Here we plotted the diffusion results and the number of circulations versus the node infection probability of four exemplary networks. As we can see, the curves significantly differ. As some diffusion curves seem to follow the typical S-shape (*power network*), some networks show an almost linear relation to the node infection probability (*Erdős Coauthorship Network*). We can further state a dramatic difference between some networks, especially in the node infection probability interval from 0.4 to 0.5. Comparing the *power* and the *jazz network* at a node infection probability of 0.4, we can observe a difference in the diffusion of almost 80%, having network diffusions around the mean of 0.08 for the *power network* and diffusions around the mean of 0.95 for the *jazz network*. This means that for identical SIR model settings, one network will usually reach diffusions of not even 10% of its network nodes, having other networks where almost the whole population (95%) is reached.

Similar results are highlighted by the graph showing the distribution of the number of circulations taken by a simulation. We can see characteristic differences for different networks, showing that some networks never reach the maximum number of circulations of 14 (*Erdős Coauthorship Network*), whereas other networks show a more than nine times higher maximum diffusion circulations of 127 (*Watt-Strogatz Network 7*). Moreover the plot shows that the maximum number of circulations is reached for totally different values of the node infection probability. The *Watt-Strogatz Network 7* reaches the maximum at values around 0.2 and the *power network* reaches the highest values at a node infection probability at around

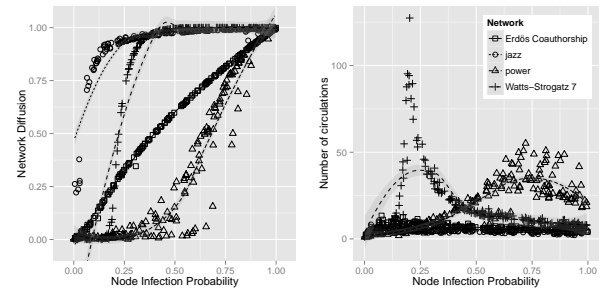


Fig. 2. Diffusion (left) and number of circulations (right) curves of four real-world networks given the infection probability. The figure includes smoothed regression lines.

0.75.

A first conclusion to draw from the figures is that the networks show tremendous variation in their diffusion behavior, both in terms of achieved diffusion (infected nodes) and the circulation time. To evaluate the influence of the network parameters on the diffusions we calculated the correlation between those parameters and the mean diffusion. The results revealed several significant correlations. It should be noted that all correlations reported hereinafter showed a *p*-Value of  $< 0.001$ , if not stated otherwise. Obviously, the infection probability plays an important role ( $r = 0.644$ ). Furthermore, the *average closeness* ( $r = 0.323$ ) had a moderate significant effect. The *network density* ( $r = 0.264$ ), the *average eigenvector* ( $r = 0.263$ ) and the *average degree* ( $r = 0.230$ ) of a network correlated significantly with the network diffusion showing weak relationships. On the other hand, the *network diameter* ( $r = -0.342$ ) and the *average path length* ( $r = -0.302$ ) of the network showed moderate negative relationships.

The picture is similar regarding the number of circulations of a diffusion. We can identify various strong positive relationships. The *average path length* ( $r = 0.588$ ) and the *diameter* ( $r = 0.561$ ) of a network correlated with the number of circulations. The *average closeness* ( $r = -0.421$ ) showed a strong, the *density* ( $r = -0.317$ ) a moderate, the *average eigenvector* ( $r = -0.260$ ) and the *average betweenness* ( $r = -0.226$ ) showed weak negative relationships with the number of circulations.

When evaluating how strong the diffusion varies within one simulation run, represented by a diffusion run's standard deviation, we can observe moderate relationships with the *average betweenness* ( $r = 0.343$ ) and the *density* ( $r = 0.315$ ) of a network. This indicates that networks with high values for these metrics are usually less predictable in terms of the network diffusion for a given fixed parameter setting.

Interestingly, the network size and the number of edges did not have a great effect on the diffusion. Highly connected and dense networks seem to lead to higher diffusion rates at comparable diffusion parameters. Another important factor appears to be the spatial dimension, represented by the average path length between two arbitrary nodes of the network and the network diameter. This directly relates to the *small world* property of networks, first discovered by Watts and Strogatz [16]. Small world networks are characterized by highly clustered nodes having small average path lengths. Many contemporary

online social networks like Facebook show very small average path lengths and a small diameter. This is interesting as these networks may contain hundreds of millions of nodes. Since Milgram's first estimate of 6 degrees (edges) between any two people in the world [48], this number is estimated to be 3.74 in the Facebook online social network in 2011, comprising, at that time, over 721 million users [49]. We can expect shorter path lengths and shorter spatial dimensions of networks in the future. According to the results presented, we can expect higher diffusions under the same preconditions for future networks.

### B. Effect of generated random networks

When conducting experiments, researchers or marketing campaigners sometimes make use of generated random networks to examine different spreading scenarios. We were interested whether these networks might show a significant difference in the diffusion compared to real-world networks. An ANOVA on the effect of the mean diffusion regarding an underlying real-world or generated network was conducted. The ANOVA revealed a significant mean difference:  $F(1, 6998) = 28.28, p < .001, \omega^2 = .003, d = 0.12$  for the groups *real-world* ( $\mu = 0.62, \sigma = 0.32$ ) and *random generated* ( $\mu = 0.67, sd = 0.35$ ) networks. Another ANOVA of the effect on the circulation length in real-world and generated networks showed a significant difference:  $F(1, 6998) = 10.712, p < .001, \omega^2 = .001, d = 0.07$  for the groups *real-world* ( $\mu = 9.28, \sigma = 6.38$ ) and *random generated* ( $\mu = 10.19, sd = 14.14$ ) networks. Although the difference is significant, the effect according to  $\omega^2$  [50] or Cohen's  $d$  [51] effect size is rather minor. We can conclude that random networks tend to a slight overestimation in the diffusion and a small overestimation in the circulation length.

### C. Influence of the Number of Seeding Nodes

To show whether the number of seeding nodes has a significant influence on the resulting diffusion we calculated the overall correlation between the mean diffusion of a simulation run and the number of seeding nodes used. We can state that, although significant ( $p < .001$ ), there is no overall effect ( $r = 0.07$ ) of the number of seeding nodes on the diffusion.

When calculating the correlation for every network separately, the picture changes dramatically. We can observe networks that show a strong effect regarding the number of start nodes: *dolphins network* ( $r = 0.45, p < .001$ ), *Watts-Strogatz network 1* ( $r = 0.34, p < .001$ ). A first interpretation would be that small networks, in terms of the number of nodes, show a high correlation, as would be expected. Interestingly, there are other networks with only a few number of nodes that do not show this effect, e.g. the *jazz network* ( $r = -0.03$ ). On the other hand, there are also quite big networks that show an effect like the *netscience network* ( $r = 0.29, p < .001$ ).

Since the effect differs so strongly from network to network, we were interested if there might be a metric that could explain this behavior. Therefore we calculated several regressions of the effect with the network metrics. We could identify two significant relationships. The *average betweenness* ( $p < 0.001, \beta = 0.78, R^2 = 0.59$ ) and the *network density* ( $p < 0.01, \beta = 0.52, R^2 = 0.25$ ) of a network showed a strong

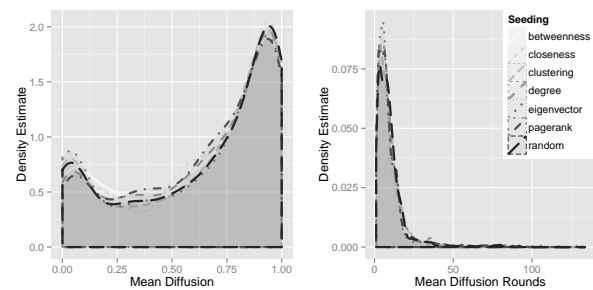


Fig. 3. Estimated densities of the mean diffusion and the number of circulations grouped by the seeding criterion

relationship with the effect of the number of seeding nodes. Hence, if the network is characterized by bridging nodes with a high betweenness and a high density, this parameter is of higher importance.

### D. Influence of Seeding Criterion

Fig. 3 shows the estimated density from a kernel density estimator for the achieved diffusion and the number of circulations grouped by the seeding criterion. Interestingly, the figure highlights a rather minor effect of this parameter on the diffusion and the number of circulations. Surprisingly, an ANOVA showed that there was no significant overall effect of the seeding criterion. We should recall at this point that the seeding methods used in this paper are not a complete list of possible metrics. The results, however, confirm some previous findings, for instance Opuszko and Ruhland [28] who found that diffusions based on cascades show a rather minor sensitivity to the seeding method. Nevertheless, we were interested if the effect is network specific so we conducted an ANOVA for each network separately. The results showed that there are in fact some differences among the networks. The *netscience network* ( $p < 0.001, \omega^2 = 0.103$ ), the *Barabasi-Albert network 3* ( $p < 0.01, \omega^2 = 0.066$ ), the *hep-th network* ( $p < 0.01, \omega^2 = 0.062$ ) showed a medium effect. Six other networks showed a weak effect. Still the number is quite low. We further conducted regressions to investigate if the effect is related to a network metric. Only the *number of components* showed a weak significant relationship ( $p < 0.05, R^2 = 0.107$ ) with the effect.

When evaluating the effect of seeding on the number of circulations the overall results are similar as there was no significant relationship. The picture changes, however, if the relationship is calculated for each network separately. ANOVA calculations for every network showed a diverse picture. 19 networks showed at least a weak relationship. Some networks showed a very strong relationship with the effect: *Barabási-Albert 2 network* ( $p < 0.001, \omega^2 = 0.303$ ), *Barabási-Albert 1 network* ( $p < 0.001, \omega^2 = 0.201$ ), *jazz network* ( $p < 0.001, \omega^2 = 0.184$ ). We can state that seeding has a much higher effect on the number of circulations than it had on the diffusion. Again we conducted several regressions to evaluate possible relationships of the effect to network metrics. The results reveal positive significant weak relationships with the *network density* ( $p < 0.01, R^2 = 0.206$ ), the *average clustering coefficient* ( $p < 0.05, R^2 = 0.152$ ) and the *average closeness* ( $p < 0.05, R^2 = 0.106$ ). We can conclude that dense



clustered and fragmented networks show a higher sensitivity to seeding in terms of the circulation time.

## V. CONCLUSION

The results presented in this paper lead to two main outcomes: First, we have shown that the underlying network has a tremendous effect on the simulated SIR spreading. We have seen that networks show significantly different behavior under comparable parameters. This also includes the variation of the diffusion. Some networks show significantly more variation and uncertainty in the diffusion than others. As a consequence, we question if results based on one specific network are transferable to any other network. Therefore we recommend for future investigations that authors either present further network characteristics other than the network size, since we have seen that the size is not a profound network metric, or they make the dataset publicly available. In this way other researchers are able to assess the network characteristics. If that is not possible, perhaps due to copyright or secrecy issues, the researchers might present parameters to generate a random network reflecting the examined real-world network's characteristics. As we have shown, the use of random generated networks is suitable for analyses.

The second main insight is that seeding must be seen on a differentiated basis. Besides the fact that the overall effect in the SIR diffusion was rather minor, we could identify significant differences among the networks. This could possibly explain some of the previous contradictory findings [24][10][20]. We can further state that seeding might not affect cascade-like diffusions as much as it effects adoption-like diffusions. Opuszko and Ruhland [28] found in their experiments with linear-threshold models that the effect of seeding is higher in adoption models. However, we have to emphasize the fact that the seeding methods used in this paper are based on network centralities and might omit other node specific factors not related to a centrality metric. Aral and colleagues [24] argued that the omission of these characteristics is a factor in the ongoing disagreement about the effect of seeding. Based on our results we can state that even with a state-of-the-art simulation analysis, the picture is diverse and the results are related to the underlying network characteristics.

The results presented are not tied to the research in information diffusion. They relate to numerous research fields like viral marketing, the research on epidemics, the flow of innovations and ideas in collaboration networks etc.

## REFERENCES

- [1] A. M. Kaikati and J. G. Kaikati, "Stealth marketing: How to reach consumers surreptitiously," *California Management Review*, vol. 46, no. 4, pp. 6–23, 2004. [Online]. Available: [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1394975](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1394975)
- [2] H. E. Tillett, "Infectious Diseases of Humans; Dynamics and Control," *Epidemiology and Infection*, vol. 108, no. 1, p. 211, 1992. [Online]. Available: <http://onlineibrary.wiley.com/doi/10.1111/j.1753-6405.1992.tb00056.x/abstract>
- [3] L. Y. L. Yulian, "Investigation of Prediction and Establishment of SIR Model for H1N1 Epidemic Disease," in *Bioinformatics and Biomedical Engineering iCBBE 2010 4th International Conference on*. IEEE, 2010, pp. 1–4. [Online]. Available: [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=5517654&tag=1](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5517654&tag=1)
- [4] Q. Chen, "Application of SIR model in forecasting and analyzing for SARS," *Beijing da xue xue bao Yi xue ban Journal of Peking University Health sciences*, vol. 35 Suppl, no. Suppl, pp. 75–80, 2003.
- [5] R. Casagrandi, L. Bolzoni, S. A. Levin, and V. Andreasen, "The sirc model and influenza a," *Mathematical Biosciences*, vol. 200, no. 2, pp. 152 – 169, 2006. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0025556405002464>
- [6] J. Woo, J. Son, and H. Chen, "An SIR model for violent topic diffusion in social media," in *Proceedings of 2011 IEEE International Conference on Intelligence and Security Informatics*. IEEE, 2011, pp. 15–19.
- [7] V. Mahajan, *New-Product Diffusion Models (International Series in Quantitative Marketing)*. Springer, 2000.
- [8] M. Bampo, M. T. Ewing, D. R. Mather, D. Stewart, and M. Wallace, "The Effects of the Social Structure of Digital Networks on Viral Marketing Performance," *Information Systems Research*, vol. 19, no. 3, pp. 273–290, 2008. [Online]. Available: <http://isr.journal.informs.org/cgi/doi/10.1287/isre.1070.0152>
- [9] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '03*. New York, New York, USA: ACM Press, Aug. 2003, p. 137. [Online]. Available: <http://dl.acm.org/citation.cfm?id=956750.956769>
- [10] O. Hinz, B. Skiera, C. Barrot, and J. Becker, "Seeding Strategies for Viral Marketing: An Empirical Comparison," *Journal of Marketing*, vol. 75, pp. 55–71, 2011.
- [11] M. Kimura, K. Saito, and R. Nakano, "Extracting influential nodes for information diffusion on a social network," in *AAAI'07 Proceedings of the 22nd national conference on Artificial intelligence - Volume 2*, Jul. 2007, pp. 1371–1376. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1619797.1619865>
- [12] M. Kimura, K. Saito, R. Nakano, and H. Motoda, "Finding Influential Nodes in a Social Network from Information Diffusion Data," *Social Computing and Behavioral Modeling*, pp. 1–8, 2009. [Online]. Available: <http://www.springerlink.com/content/p54n387114265075/>
- [13] P. Shakarian and D. Paulo, "Large Social Networks can be Targeted for Viral Marketing with Small Seed Sets," in *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, no. 2, 2012, pp. 1–8. [Online]. Available: <http://arxiv.org/abs/1205.4431>
- [14] B. C. Thompson, "Is the Tipping Point Toast ?" *California Management Review*, vol. 43, no. 4, pp. 44–63, 2008. [Online]. Available: <http://m6d.com/wp-content/themes/m6d/documents/is-the-tipping-point-toast.pdf>
- [15] T. Sun, W. Chen, Z. Liu, Y. Wang, X. Sun, M. Zhang, and C.-Y. Lin, "Participation Maximization Based on Social Influence in Online Discussion Forums," in *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media ICWSM'11*, 2011.
- [16] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks." *Nature*, vol. 393, no. 6684, pp. 440–2, Jun. 1998. [Online]. Available: <http://dx.doi.org/10.1038/30918>
- [17] R. M. Anderson, R. M. May, and B. Anderson, *Infectious Diseases of Humans: Dynamics and Control (Oxford Science Publications)*. Oxford University Press, USA, 1992.
- [18] F. M. Bass, "A New Product Growth for Model Consumer Durables," *Management Science*, vol. 15, no. 5, pp. 215–227, Dec. 1969. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1245920.1245934>
- [19] E. M. Rogers, *Diffusion of Innovations*. Free Press, 1995.
- [20] J. Leskovec, L. A. Adamic, and B. A. Huberman, "The dynamics of viral marketing," *ACM Transactions on the Web*, vol. 1, no. 1, pp. 5–es, May 2007. [Online]. Available: <http://arxiv.org/abs/physics/0509039>
- [21] J. Goldenberg, B. Libai, and Muller, "Using complex systems analysis to advance marketing theory development," *Academy of Marketing Science Review*, vol. 9, 2001.
- [22] P. Domingos and M. Richardson, "Mining the network value of customers," in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '01*. New York, New York, USA: ACM Press, Aug. 2001, pp. 57–66. [Online]. Available: <http://dl.acm.org/citation.cfm?id=502512.502525>
- [23] J. Yang, C. Yao, W. Ma, and G. Chen, "A study of the spreading scheme for viral marketing based on a complex network

- model,” *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 4, pp. 859–870, Feb. 2010. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0378437109009042>
- [24] S. Aral, L. Muchnik, and A. Sundararajan, “Engineering Social Contagions: Optimal Network Seeding in the Presence of Homophily,” *Forthcoming in Network Science*, February 18 2013. [Online]. Available: <http://dx.doi.org/10.2139/ssrn.1770982>
- [25] M. Granovetter, “Threshold models of collective behavior,” *American journal of sociology*, vol. 83, no. 6, pp. 1420–1443, 1978. [Online]. Available: <http://www.jstor.org/stable/10.2307/2778111>
- [26] C. Moore and M. Newman, “Epidemics and percolation in small-world networks,” *Physical Review E*, vol. 61, no. 5, pp. 5678–5682, May 2000. [Online]. Available: <http://arxiv.org/abs/cond-mat/9911492/>
- [27] M. E. J. Newman, “The structure and function of complex networks,” *SIAM REVIEW*, vol. 45, pp. 167–256, 2003.
- [28] M. Opuszko and J. Ruhlman, “Effects of the network structure on the dynamics of viral marketing,” in *Proceedings of the 11th International Conference on Wirtschaftsinformatik. Internationale Tagung Wirtschaftsinformatik (WI-2013), 11. February 27 - March 1, Leipzig, Germany*, R. Alt and B. Franczyk, Eds. Leipzig, Germany: Universitt Leipzig, Leipzig, 2 2013, pp. 1509–1524.
- [29] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson, “The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations,” *Behavioral Ecology and Sociobiology*, vol. 56, pp. 396–405, 2003.
- [30] D. E. Knuth, *The Stanford GraphBase: A Platform for Combinatorial Computing*. Reading: Addison Wesley.
- [31] R. Matei, A. Iamnitchi, and P. Foster, “Mapping the Gnutella network,” *IEEE Internet Computing*, vol. 6, no. 1, pp. 50–57, 2002. [Online]. Available: <http://dl.acm.org/citation.cfm?id=613352.613670>
- [32] T. Opsahl and P. Panzarasa, “Clustering in weighted networks,” *Social Networks*, vol. 31, no. 2, pp. 155–163, May 2009. [Online]. Available: <http://dx.doi.org/10.1016/j.socnet.2009.02.002>
- [33] M. E. J. Newman, “Finding community structure in networks using the eigenvectors of matrices,” *Physical Review E*, vol. 74, no. 3, p. 22, Sep. 2006. [Online]. Available: <http://arxiv.org/abs/physics/0605087>
- [34] M. Newman, “Network data,” [retrieved: 11 2012], <http://www-personal.umich.edu/~mejn/netdata/>, 2011.
- [35] M. E. Newman, “The structure of scientific collaboration networks.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 2, pp. 404–9, Jan. 2001. [Online]. Available: <http://arxiv.org/abs/cond-mat/0007214/>
- [36] V. Batagelj and A. Mrvar, “Pajek datasets,” [retrieved: 11 2012], <http://vlado.fmf.uni-lj.si/pub/networks/data/>, 2006.
- [37] R. Guimer, L. Danon, A. Diaz-Guilera, F. Giralt, and A. Arenas, “Self-similar community structure in a network of human interactions,” *Physical Review*, vol. E 68, no. 065103, 2003.
- [38] A. Arenas, L. Danon, A. Díaz-Guilera, P. Gleiser, and R. Guimerá, “Community analysis in social networks,” *The European Physical Journal B - Condensed Matter*, vol. 38, no. 2, p. 8, 2004.
- [39] M. Boguñá, R. Pastor-Satorras, A. Díaz-Guilera, and A. Arenas, “Models of social networks based on social distance attachment,” *Physical Review E*, vol. 70, no. 5, Nov. 2004.
- [40] A. Barabási and R. Albert, “Emergence of Scaling in Random Networks,” *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999. [Online]. Available: <http://www.sciencemag.org/content/286/5439/509.abstract>
- [41] P. Erdős and A. Rényi, “On random graphs,” *Publicationes Mathematicae (Debrecen)*, vol. 6, pp. 290–297, 1959. [Online]. Available: [http://www.renyi.hu/~p\\_erdos/Erdos.html#1959-11](http://www.renyi.hu/~p_erdos/Erdos.html#1959-11)
- [42] G. Csardi and T. Nepusz, “The igraph software package for complex network research,” *InterJournal*, vol. Complex Sy, p. 1695, 2006. [Online]. Available: [retrieved:112012],<http://igraph.sf.net>
- [43] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2010. [Online]. Available: [retrieved:112012]<http://www.r-project.org>
- [44] M. E. J. Newman, “Power laws, Pareto distributions and Zipf’s law,” *Contemporary Physics*, vol. 46, no. 5, pp. 323–351, May 2005. [Online]. Available: <http://dx.doi.org/10.1080/00107510500052444>
- [45] V. Junapudi, G. K. Udgata, and S. K. Udgata, “Study of diffusion models in an academic social network,” in *Distributed Computing and Internet Technology*, ser. Lecture Notes in Computer Science, T. Janowski and H. Mohanty, Eds., vol. 5966. Berlin, Heidelberg: Springer Berlin Heidelberg, Feb. 2010, pp. 267–278. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2127870.2127905>
- [46] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, ser. Structural analysis in the social sciences, 8, M. Granovetter, Ed. Cambridge University Press, 1994, vol. 8, no. 1.
- [47] M. Newman, *Networks: An Introduction*, M. E. J. Newman, Ed. Oxford University Press, 2010.
- [48] S. Milgram, “The small world problem,” *Psychology Today*, vol. 1, no. 1, pp. 60–67, 1967. [Online]. Available: [http://measure.igpp.ucla.edu/GK12-SEE-LA/Lesson\\_Files\\_09/Tina\\_Wey/TW\\_social\\_networks\\_Milgram\\_1967\\_small\\_world\\_problem.pdf](http://measure.igpp.ucla.edu/GK12-SEE-LA/Lesson_Files_09/Tina_Wey/TW_social_networks_Milgram_1967_small_world_problem.pdf)
- [49] B. B. C. BBC, “Facebook users average 3.74 degrees of separation,” BBC [retrieved: 05 2013], <http://www.bbc.co.uk/news/technology-15844230>, 2011.
- [50] S. Olejnik and J. Algina, “Generalized eta and omega squared statistics: measures of effect size for some common research designs.” *Psychological Methods*, vol. 8, no. 4, pp. 434–447, 2003. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/14664681>
- [51] J. Cohen, “A power primer,” *Psychological Bulletin*, vol. 112, no. 1, pp. 155–159, 1992. [Online]. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3153748&tool=pmcentrez&rendertype=abstract>

# Using Quick Response Codes For Student Interaction During Lectures

Robert Law

Computer, Communications and Interactive Systems  
Glasgow Caledonian University  
Glasgow, Scotland  
e-mail: robert.law@gcu.ac.uk

**Abstract—** This paper will present an ongoing project to encourage student interaction during lectures through the use of Quick Response (QR) codes and Google forms to generate rapid response polls and quizzes. The use of Google applications (Apps) software and the students' own mobile phone presents a free alternative to the current clicker systems. The pedagogical issues associated with such a project will be investigated and an attempt made to incorporate these into the student experience. The overall process from the creation of the software to the roll out and use of the software in an interactive lecture, the issues encountered and participant feedback will also be described.

**Keywords-** QR Codes; Student Interaction; Feedback.

## I. INTRODUCTION

Smartphones in the UK, and elsewhere, have seen a surge in popularity over the last few years evidenced by recent figures showing "Over half of the British population (50.3%) now owns a smartphone" [1]. Edinburgh University recently conducted a survey of their student population determining that 67% had ownership of a smartphone "an increase of seventeen percent from those students surveyed seven months earlier" [2]. This uptake in smartphone ownership within the student population opens a new dimension for interaction.

A tangible increase and use of Quick Response (QR) codes by many companies as a form of marketing has ensued on the back of this increase in smartphone popularity. This has allowed many companies to develop new and engaging ways for their customer base to interact with products in situ reinforcing brand presence. This potential has already been harnessed by Education to extend learning materials through the use of QR codes. Learning materials have been enhanced by providing "just in time support materials" [3] such as videos, explanatory text, Uniform Resource Indicators (URI) and staff details.

Using this as a platform to build from the next logical step is to combine the technologies to allow students to interact during lectures through quick multiple choice based questions. The students' responses can be compiled to show the result in a timely manner [4].

The remainder of this paper is organized as follows: Section II gives information about pedagogical issues related

to interactive lectures, Section III introduces the technology used for implementing interactive lectures; this covers both hardware and software. Section IV discusses the authors experience of implementing interactive lectures, while Section V discusses issues encountered during the interactive lectures. Section VI offers a summary of our experience of interactive lectures, and concludes the paper giving proposals for future work.

## II. PEDAGOGICAL ISSUES

Take a typical lecture; what does this encompass? Information is imparted upon the student in a relatively one way passive communication format. This traditional didactic approach is a format that has been used for centuries. This research intends to explore the possibility of improving and enhancing the lecture experience through the use of technology, and in particular, Audience Response Systems (ARS). The ability of such systems to encourage active learning through student participation and engagement provides an opportunity for enhancing the passive lecture format by introducing two way interactions with the student audience [7].

Murphy and Sharma further suggest that the research literature available for the topic of interactive lectures and the related pedagogical issues are "almost non-existent, with major issues waiting to be examined... inadequate research on the pedagogical implications of the emerging interactive forms of learning" [7]. With this in mind there appears to be an opportunity to examine and suggest how ARS technology could be used to not only enhance lecturer-student interactions, but also develop the underlying pedagogical issues inherent with lectures.

ARS technology provides a means for the lecturer to engage and interact with the students using the responses to offer the student audience feedback. This should lead to further discussion and the opportunity for student reflection. Other research, reported in Murphy and Sharma [7], identifies two pedagogical aspects of interactive lecturing: dialogic form of learning and active learning.

The project intends to examine these issues and the resultant effects that they have on the student audience. The primary concern is that the interactive lecture will stimulate engagement and interaction with the student audience and

the lecturer through the use of instant feedback. This feedback will engender in both the student audience and the lecturer the need for reflection on many aspects of the material delivered and possibly the module in general. Through the use of relevant and targeted questions the students can be cajoled into discussions that will help expand their understanding of the topic area. Through these discussions both the students and the lecturer will be able to better understand the level of the students understanding of the subject area.

Gannon-Leary et al. [5] reported a number of other positive aspects to arise from interactive lectures including improved concentration, greater enjoyment and improved attendance. Simpson et al. [8] also noted that the anonymity provided by ARS technology played an important part in encouraging students to contribute to answering questions but suggests, as does Gannon-Leary et al. [5], that the design of the questions is very important to the process.

Saravani and Clayton [9] have developed a conceptual framework referred to as A.C.E. This framework is composed of the three A's: Awareness, Action, and Accomplishment; three C's: Context, Content, and Capability; and the three E's: Enabled, Engaged, and Empowered. The three E's aspect of the framework fits the concept driving interactive lectures as the use of mobile technology enables, engages and empowers both the student body and the lecturer.

Ramsden [3] suggests, amongst other things, that Quick Response (QR) codes can be used for "just in time information in a face to face lecture"; drawing on this point allows for the expansion of the concept to include feedback for both the student and the lecturer. Mooted by Law and So [11] is the idea that QR codes can facilitate a "trinity of "location independence," "time independence" and "meaningful content"". Of interest in this "trinity" is the idea of "location independence"; being able to deliver and receive feedback to and from the students in the lecture hall.

### III. TECHNOLOGY

ARS systems are available in many forms and price points. A typical classroom package can cost \$1000 or more for software, receiver and 12 clickers [12]. Some systems require the student to purchase a clicker and yearly registration at a cost of \$20/\$15, respectively [12]. This project developed a "no cost in-house" system that was based on three key components: smartphones, QR codes and a Google spreadsheet. No proprietary software for the phone is required simply a standard browser and bar code reader. The "back-end" is relatively simple to create as the implementation interface is supplied by Google.

#### A. Hardware

The student participants referred to in Section IV were surveyed to determine the spread of handset manufacturers and phone operating systems. Around 50 students were surveyed.

Figure 1 shows the spread of handset manufacturers and Figure 2 shows the spread of phone operating systems with the surveyed population.

With such a range of manufacturers and operating systems an "app" based solution would be time consuming and prohibitive. With further investigation, it was discovered that third party barcode reading software was available for all the platforms, thus allowing the students to use their own phones for participation in the lectures.

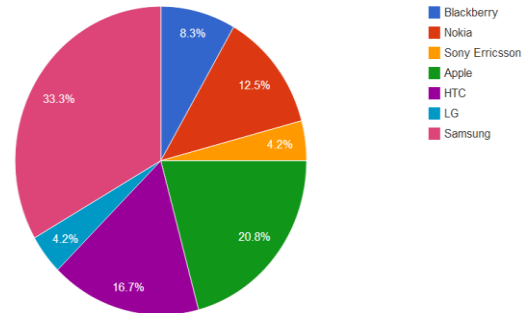


Figure 1. Spread of manufacturers

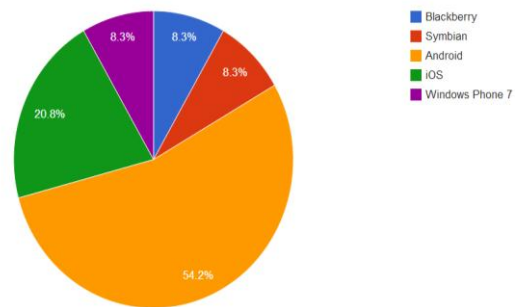


Figure 2. Spread of operating systems

#### B. Software

The software can be split into three categories: third party barcode reading software, third party browser software and the development of the interactive engine software using Google spreadsheets. It was up to the students themselves to decide on a suitable third party barcode reading software, although some phones did have such software preinstalled.

Three components of Google spreadsheets were used in the creation of the software application: the spreadsheet, the form and Google Script.

The spreadsheet itself is used as a repository for the student responses and also to house a summary sheet which keeps a running total of the number of responses for each possible answer to the question. The work horse of the system is the form and the scripts generated to process the responses at the back end. When a spreadsheet is created using Google Drive, a unique identifier is generated to identify the spreadsheet. When the subsequent form is created for the spreadsheet, another unique identifier is generated.

Although Google forms can handle a number of different inputs, the decision was taken to keep the question to a simple multiple choice style question, thus presenting the student with two or three QR codes per question. To create the QR codes requires the compilation of an http

request based on the URI for the Google spreadsheet and the data to be sent to the spreadsheet. Once the http request was constructed and tested an online QR code generator was used to generate the required QR codes. These QR codes were subsequently saved as image files for insertion into the lecture slides.

Google script was used to create code that processed the student responses as they were received to generate a response summary that was visible to the student audience.

C. Process

The students used their smartphones to scan the QR code of their choice, submitting the data request via their phones browser. This, in turn, populated the spreadsheet with the students' choices, activating the script, allowing the results to be observed in near real time.

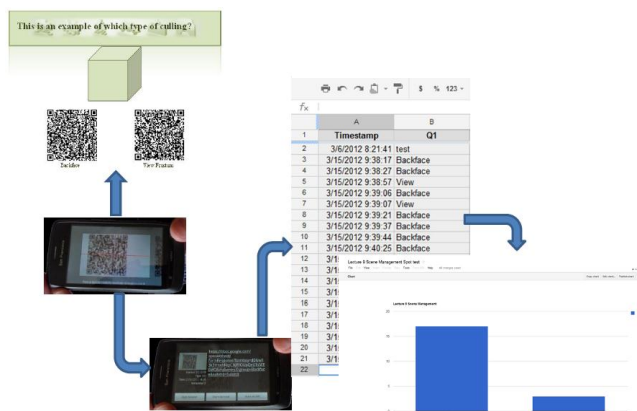


Figure 3. The interactive process

The QR Code is an encoded representation of the Google URI and data that will be sent to the spreadsheet when student scans it. As can be seen in Figure 3 above, once scanned, the information encoded in the QR code is decrypted and becomes visible to the phone's user. At this point, the participant has the ability to accept or decline the invite to send the data request. When the participant accepts the request to send the data, the next step is to invoke the phone's browser (this can work in different ways, depending on the phone/operating system) which will send the http request to Google for processing. Once processed, a "thank you" message is displayed in the browser indicating the data request has been received.

Once the data request has been received, the data is placed in the spreadsheet; the data will be based on the URI encoded in the QR Code. An example is shown below in Figure 4.

It is noticeable from Figure 4 that the spreadsheet date and time stamps the entries as it receives them. The only data that the spreadsheet records is the data encoded in the QR Code and the date and time stamp it generates on receipt of the data. Hence, all entries are anonymous.

	A	B
1	Timestamp	Which one is true?
2	3/22/2012 9:25:08	DAG
3	3/22/2012 9:26:13	DAG
4	3/22/2012 9:26:17	DAG
5	3/22/2012 9:26:20	DAG
6	3/22/2012 9:26:25	DAG
7	3/22/2012 9:26:33	BST
8	3/22/2012 9:27:01	DAG
9	3/22/2012 9:28:00	DAG
10	3/22/2012 9:28:01	DAG
11	3/22/2012 9:29:18	DAG
12	3/22/2012 9:33:01	DAG
13	3/22/2012 9:34:50	Octree
14	3/22/2012 9:35:43	Two
15	3/22/2012 9:35:47	Two
16	3/22/2012 9:35:53	Quadtree
17	3/22/2012 9:37:15	Two
18	3/22/2012 9:42:17	Two

Figure 4. Data sent to the spreadsheet

	A	B
1	Rast	13
2	Geo	6
3	Vertex	0
4	Stencil	12

Figure 5. Summary sheet

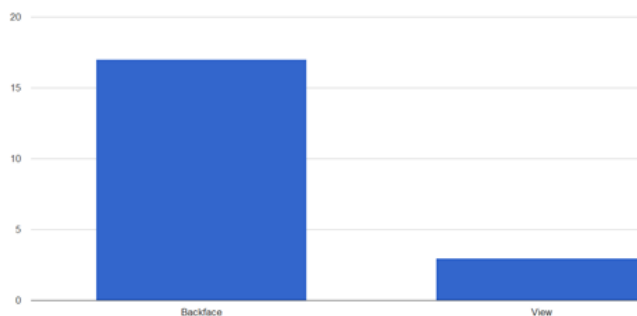


Figure 6. Generated bar chart

As each entry is received, a script is triggered, which counts the entries based on the predefined data set and populates a summary sheet which is used to generate the "near real time" bar charts.

An example of a summary sheet is shown in Figure 5 and an example of the bar chart displayed to the students is shown in Figure 6.

IV. EXPERIENCE

A set of initial tests were developed to examine the viability of the technology and gauge the reaction of the

students to the use of this technology within the lecture environment. The aim of the tests was to introduce the interaction concept in a gradual staged manner that wouldn't over burden the student or detract from the lecture.

#### A. Test Setup

The process used to create this interactive lecture was based on designing a set of suitable questions that could be used to strategically punctuate the lecture to gain maximum benefit for the students [13].

The desired effect was to integrate the technology within the lecture while stimulating interaction with the students [13]. As such, the first set of tests was built to increase in a systematic manner the number of questions within the lecture and the number of QR codes within the questions.

The structure of the first set of tests was designed to build from one question with one QR code in the lecture to four questions with three codes per question in the lecture. Suitable points within the lecture were identified such that the questions could be inserted to maximise their impact. An attempt was made to define suitable points through natural break points within the lecture e.g. end of a topic, end of the lecture, worked example. Using this principal, questions could be deployed with the aim of giving both the student and the lecturer instant feedback on the comprehension of the material delivered.

#### B. Participants

Both sets of participants were studying on the Games Software Development Degree. The first group to undertake the interactive lectures was a second year cohort of around 30 students and the second group to undertake the interactive lectures was a final year cohort of around 20 students.

The second year cohort had three consecutive lectures. The first of the three lectures had one question with two QR codes positioned at the end of the Lecture. The second Lecture had three questions, each with two QR codes positioned at appropriate points within the Lecture and the final Lecture had four questions, each with three QR codes again positioned at appropriate points within the Lecture.

The final year cohort had two lectures which were non-consecutive. The first of the two lectures had two questions each with three QR codes positioned at appropriate points within the Lecture and the second Lecture had two questions each with three QR codes positioned at appropriate points within the Lecture.

At the appropriate point in the lecture, the slide would be displayed. To help minimize issues with scanning, paper copies of the slide were also distributed. This allowed for the difference within the quality of phone cameras to focus on the projected QR codes. It was fully explained to the students the nature of the experiment and the procedure which should be followed to correctly participate in the experiment.

#### C. Feedback

Initial feedback from both test groups has been positive and very informative. Feedback ranged from the ease of operation of the process to the size of the QR codes. In general a "buzz" was created within the participant groups

generating a positive reaction from the students. This reaction must be tempered by the fact that the students are open to the "Hawthorne effect" [14] and, as such, more trials will need to be run to assess and define a clearer picture of the students reaction.

## V. ISSUES

Although the overall outcome of the experiments was positive, a number of issues were highlighted that require further polishing prior to subsequent use.

Currently, the whole process required to create and integrate the QR codes into a lecture are quite cumbersome and may prove challenging for a non-computing subject specialist. An ongoing project has been put in place to integrate and automate the creation of the QR codes through a Web based Google spreadsheet application.

An issue flagged up by the participants centered on the size and positioning of the QR codes as this can have an impact on the accuracy of the scanning process.

The student participants indicated that the size of the QR codes on a projected slide proved difficult to scan directly. Not all students were able to scan it directly. This had been anticipated and paper based copies of the slides were issued to counter that problem. The issue appears to lie with the quality of the camera supplied with the phone.

Positioning of the QR Codes on the slide raised debate with the participants as some indicated that the barcode scanner software could find it difficult to focus on the required code. This seemed to be more prevalent when the codes were positioned side by side horizontally rather than stacked vertically. The number of QR codes on the slide also influenced the ease of scanning, with three codes per slide proving more challenging for the phone's barcode scanner. This was not insurmountable, but merely added a small time overhead as the participants positioned their camera phone to optimize the scan.

During one trial, a small number of participants encountered an issue with the barcode scanner software they were using. One student managed to automatically scan and send the same QR Code a number of times which did skew the graph for that particular question. On the lighter side, it did add a slice of humor, as the graph seemed to be growing by more than there were participants! On investigation, the student concerned indicated that the problem had arisen due to a setting in the third party barcode scanner software.

An intermittent issue, which arose on one occasion, was the apparent delay/failure of the chart to update itself as the data was received. This led to the chart being manually refreshed which did have an effect on the desired impact of the interactivity of the lecture.

Time management of both the interactions and the subsequent discussions should be built into the lecture timings allowing for leeway should anything go awry with the technology.

This approach relies on all the students in the lecture having a smartphone and it is conceivable that a small percentage of the student audience may in fact not have a smartphone or even a mobile phone. The perception would be that this student would be disadvantaged by not being

able to take part in the interaction. This would be, of course, true. However, it could be mooted that the student is still engaged in the wider discussion that will come from viewing the generated graph. Another possibility is to rely on the goodwill of a fellow student to share access to their phone.

## VI. CONCLUSION AND FUTURE WORK

The overall trials were generally well received by the students serving the purpose of generating positive interaction between the lecturer and the students.

Students appeared to enjoy the break in the lecture and the feedback and discussion generated by the visual charting of their responses. It also created a focus point for the students to reflect on their understanding of the material taught and to apply that understanding. By the same token, it proved to be beneficial to the lecturer indicating the level of understanding of the delivered material to the students.

The anonymity of the whole process was cited by a number of students as positive. They felt comfortable with the fact that they could answer the questions freely, getting them wrong and not feeling awkward in front of their peers.

The project is ongoing and the positive feedback received from the students indicates that it is a worthwhile pursuit for both the lecturer and the students.

With regard to performance, this prototype system works well, producing the column chart of responses in near real time. More evaluation of the systems performance against other ARS systems is required. Initial use suggests promise with this cloud based system. Advantages this system offers is the fact it is free, flexible, easily tailored to suit the lecturer's needs and platform independent.

A further avenue for investigation will be the correct utilization and positioning of the interactive lectures within the overall module lecture delivery schedule. Over or under use will have an impact on their effectiveness.

Further investigation will be made with regard to the sizing, positioning and visibility of QR Codes from projected and paper based slides.

The use of the technique within the tutorial/seminar setting to encourage more debate on theoretical and social subjects is path that will be followed.

Further investigation will be undertaken into the relative pros and cons of storing complex responses in the spreadsheet, as evidenced in figure 4, and simplistic responses in the form A, B, C, etc. The outcome of this

investigation will have an impact on the future development of the software.

A significant proportion of future work will be involved in developing a user friendly interface to the software to allow cross discipline use. The script code will be hidden from the user allowing them to concentrate on the development of their question bank.

## REFERENCES

- [1] Kantar Worldpanel, The smartest way to communicate: over half of the GB population owns a smartphone. Available: <http://goo.gl/Z1n7u>. [retrieved: July, 2013].
- [2] EUSA, Campus app now available for download. Available: <http://www.eusa.ed.ac.uk/news/article/6001/290/>. [retrieved: July, 2013].
- [3] A. Ramsden, "The use of QR codes in Education: A getting started guide for academics." Working Paper. University of Bath., 2008.
- [4] R.Law, Using Quick Response Codes for Student Interaction During Lectures, ITiCSE, ACM 978-1-4503-1246, 2012.
- [5] P. Gannon-Leary, C. Turnock, and M. McCarthy, "RECAP series paper 15", 2007.
- [6] R.C. Lowery, "Teaching and learning with interactive student response systems: A comparison of commercial products in the higher-education market", 2005
- [7] R. Murphy, and N. Sharma, "What Don't we know about interactive lectures?", Seminar. net-International journal of media, technology and lifelong learning, 2010, pp. 111.
- [8] V. Simpson and M. Oliver, "Electronic voting systems for lectures then and now: A comparison of research and practice", Australasian journal of educational technology, vol. 23, no. 2, 2007, pp. 187.
- [9] S.A Saravani and J.F. Clayton, A conceptual model for the educational deployment of QR codes. In Same places, different spaces. Proceedings ascilite Auckland 2009.
- [10] V. Yfantis, P. Kalagiakos, C. Kouloumperi, and P. Karampelas, "Quick response codes in E-learning", Education and e-Learning Innovations (ICEELI), 2012 International Conference, 2012, pp. 1.
- [11] C.Y. Law, and W.W.S. So, "QR codes in education", Journal of Educational Technology Development and Exchange, vol. 3, no. 1, 2010, pp. 85-100.
- [12] E. Bojinova and J. Oigara, "Teaching and learning with clickers: Are clickers good for students", Interdisciplinary Journal of E-Learning and Learning Objects, vol. 7, 2011, pp. 169-183.
- [13] A.P. Fagen, C.H. Crouch, and E. Mazur, "Peer instruction: Results from a range of classrooms", Physics Teacher, vol. 40, no. 4, 2002, pp. 206-209.
- [14] S.R. Jones, "Was there a Hawthorne effect?", American Journal of Sociology, 1992, pp. 451-468.

# Administration of Knowledge Assessment at Riga Technical University

Natalija Prokofjeva, Alla Anohina-Naumeca, Olga Lebedeva

Faculty of Computer Science and Information Technology

Riga Technical University

Riga, Latvia

e-mail: Natalija.Prokofjeva@rtu.lv, Alla.Anohina-Naumeca@rtu.lv, o.lebedeva@inbox.lv

**Abstract**—This paper is devoted to administration of students' knowledge assessment at the Faculty of Computer Science and Information Technology of Riga Technical University. The faculty uses a number of computer-based software systems for this purpose. The goal of the research is to evaluate them from the perspective of their usefulness for administration of students' knowledge assessment. This paper provides description of each system and the results of the analysis made on the basis of the experts' survey.

**Keywords**- computer-based learning; knowledge assessment; experts' survey

## I. INTRODUCTION

Due to the rising speed of scientific and technical progress, the volume of knowledge and skills needed by contemporary specialists is continuously growing. This substantially increases requirements in relation to organization of the study process in educational institutions and especially to selection of an appropriate knowledge assessment strategy. Nowadays, computer-based knowledge assessment is of special interest because it allows administration of knowledge assessment activities at all stages of the study process, it substantially reduces both workload of teachers and time spent for preparation and evaluation of students' assignments, as well as it assures the objectivity of assessment process.

The Faculty of Computer Science and Information Technology of Riga Technical University (RTU) is not an exception. Its staff is making research in the field of computer-based learning already for 40 years. Today, both traditional methods of knowledge assessment and computer-based software systems (electronic textbooks, the learning management system, the Moodle-based e-learning environment, and the concept map based intelligent knowledge assessment system) developed for this purpose are used in the educational process.

The goal of the research presented in this paper is to evaluate the developed computer-based software systems from the perspective of their usefulness for administration of students' knowledge assessment. This paper contains description of the mentioned systems, as well as results of their analysis and comparison.

This paper is organized as follows. Section II presents a brief overview of related works. Section III contains descriptions of the software systems used in the study process at the faculty. Section IV presents research methodology and results of the analysis made. Conclusions are given at the end of this paper.

## II. RELATED WORK

Many international scientific conferences (such as IEEE ICALT, IADIS e-Learning, IASTED CATE, etc.) and e-journals [5][6][7] are dedicated to research in the field of computer-based learning and knowledge assessment.

Researchers in this field are developing and implementing new methods of knowledge assessment [1][2][3][4]. The implementation of these methods can be found in modern e-learning systems. Latest research indicates that the use of such systems has a positive effect on students' progress in studies, it increases their motivation, and they appreciate this kind of approach to learning process [16][17][18].

However, the evaluation of e-learning systems from the perspective of students' benefits doesn't let to evaluate the usefulness of these systems for students' knowledge assessment. That's why this paper aims to compare e-learning systems that are used in RTU and to determine whether they are close to the benchmark – the model of an ideal system, which is based on the features chosen by experts.

## III. SOFTWARE SYSTEMS USED IN THE STUDY PROCESS

### A. The Electronic Textbook for HTML Language

The electronic textbook "Learn HTML from scratch" is a computer-based learning program, which includes eight topics on HTML language: structure of a document, text, graphics, references, tables, frames, styles, and dynamic HTML [8]. Each topic contains theoretical material supplemented by several examples and a series of assessment questions. Fig. 1 displays the structure of the electronic textbook.

Three modes of operation (self-assessment, testing, and reference mode) and a variety of knowledge assessment methods and models [9] are implemented in the electronic textbook. Two types of questions (multiple choice questions and text/numerical input) are available for assessment activities.

A method called "Strict sequence" is used in the mode of self-assessment. It means that a set of assessment tasks is prepared prior to the test and tasks are offered to all students in the form of the same fixed set. In this mode, students answer questions and receive a short overview of their performance (number of correct answers, time for completing a specific task, and average score).



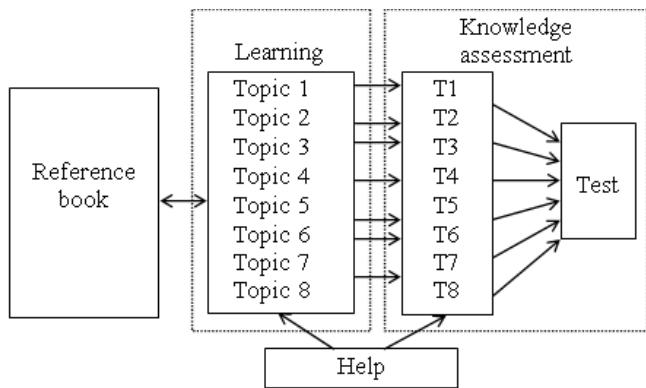


Figure 1. Structure of the electronic textbook.

In the testing mode, a method called “Random selection” is used for creating and issuing a set of assessment tasks. Therefore, the set contains  $n$  tasks, which are selected in a random fashion from a task pool.

A test for final examination is implemented in the textbook. It is composed of questions from all eight topics. Therefore, it allows assessing of students' knowledge in relation to the whole course. A set of assessment tasks is formed using the method “Strict sequence”.

Student's score is the average score obtained taking into account the student's number of correct answers and completion time of assessment tasks.

The electronic textbook „Learn HTML from scratch” was applied as a learning tool in the course "Development of Web Applications for Internet". Results of the research made in 2003 showed that students actively used the textbook both for learning and for knowledge assessment [10]. However, the electronic textbook had a number of significant drawbacks: a) it could be used only for one study course, b) learning materials (content of topics, sections, examples, and a set of assessment tasks) were embedded into the book, thus, making difficult their modifications and changing, and c) the electronic textbook was available for use only from RTU local network. However, the major disadvantage was related to usage of inappropriate methods („Strict sequence” and „Random selection”) for forming a set of assessment tasks for continuous and final assessment [11].

**B. The Universal Electronic Textbook**

The universal electronic textbook was developed in 2005. The drawbacks of the previous book were eliminated and its functionality was extended [12]. Fig. 2 shows the structure of the universal electronic textbook.

The textbook allows the teacher a) to develop new courses by determining their structure - number of sections and topics, b) to change and/or to add learning materials of a course, c) to create a set of assessment tasks, d) to choose a topic or topics for knowledge assessment, and e) to obtain assessment results of a student or a group of students. The main advantage of the universal electronic textbook is a teacher's possibility to administer different types of knowledge assessment: prior, continuous, and final assessment.

In assessment of prior knowledge, the methods “Strict order” and “Random selection” are used for creating and issuing a set of assessment tasks. They were selected on the basis of recommendations concerning usage of different methods for creation of a set of assessment tasks in different types of knowledge assessment. Recommendations were acquired through the experts' survey presented in [11]. In the universal electronic textbook, the teacher can perform continuous and final knowledge assessment according to the sequence, in which learning material was mastered by preparing a test from assessment tasks related to one or more topics.

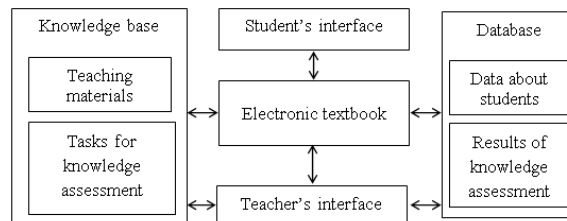


Figure 2. Structure of the universal electronic textbook.

With the help of the universal electronic textbook a student can master a chosen study course (learning mode), as well as he/she can assess knowledge in a particular topic or in the entire course (testing mode). Two types of questions are used: multiple choice questions and text/numerical input. However, comparing with the first electronic textbook, this book allows the teacher to specify the difficulty level of an assessment task (minimal, average, or maximal).

The mentioned textbook was used in the study course „Development of Web Applications for Internet” from 2005 till 2007.

**C. The Learning Management System**

Learning Management System (LMS) [13] includes two modes of operation that are directly related to knowledge assessment: random selection of tasks and/or questions for assessment and training mode (Fig. 3). Both modes can be used for knowledge self-assessment as well.

The mode of the random selection uses the non-adaptive assessment method. In this case the number of questions for knowledge assessment in a group of students is defined by the teacher, who also determines the number of tasks of different difficulty level (maximal, average, or minimal), which should be included in each set of assessment questions. Moreover, the teacher specifies type of comments, which will be provided as a reaction to student's answers: short (right, partly right, or wrong) or detailed with explanation of a mistake made. In case of self-assessment, the student chooses and sets these parameters by him/herself.

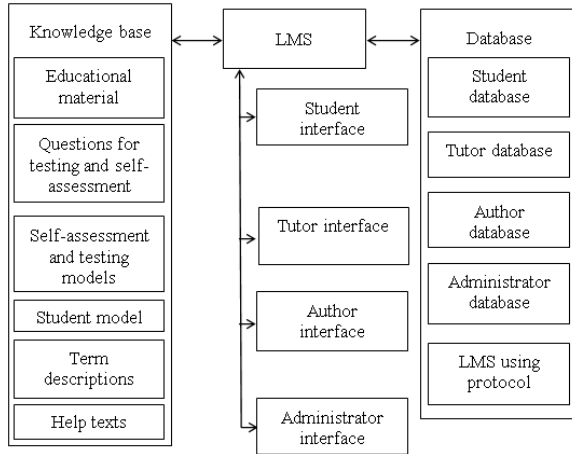


Figure 3. Structure of the LMS

Depending on the choice, the training mode performs non-adaptive assessment, when the student completes either all or a selected number of assessment tasks, or partly adapted assessment (taking into account student's answers), when the number of tasks provided to the student depends on his/her overall success. In this mode, type of comments, which will be provided by the system, can also be chosen.

After completion of the test, the student gets a grade, which takes into account both correctness of answers and difficulty level of questions.

At present, the LMS includes 9 e-courses related to the study courses „Software Engineering”, “Software Metrology and Planning”, “Software Development Tools and Environments”, and “Programming Languages”. During the period from 2004 till 2006, students had possibility to use the courses included in the system for self-assessment, training, and/or learning. Since 2007, students' knowledge assessment in the mentioned courses is compulsory and is usually performed during practical assignments in class. Usually, each student receives from 8 to 10 assessment tasks. At the same time, in order to improve the grade in a test, students are allowed to take the test repeatedly in 2 days' time. In this case the grade is determined by the teacher. Students can also use other modes of system's operation at any time.

#### D. *ORTUS Portal*

RTU has developed the portal *ORTUS*, which provides administrative, scientific, and educational support for students and university staff. One part of this portal is a Moodle-based e-learning environment that allows creation of e-learning courses starting from uploading of different learning materials and finishing with developing tests and administering knowledge assessment activities.

*ORTUS* portal has been used in the study process since 2008. In relation to knowledge assessment, the teacher can create a test, set the time for its completion and the number of attempts. The test can be completed in class (one attempt) or at a distance (more than one attempt). It is possible to create tests from different categories and types of questions. All together, 10 types of questions are available. The teacher has possibility to assign number of points for each question

taking into account its difficulty level. Comments for all kinds of questions are envisaged. The only drawback of the *ORTUS* e-learning environment is availability of only one method of knowledge assessment, namely, “Random selection”.

#### E. *IKAS system*

*IKAS* is a web-based intelligent knowledge assessment system, which is intended for assessment of students' structural knowledge through the use of concept maps [14]. It has been developed since 2005 and, therefore, the system is described in many publications, for example [3][15][16]. As a result below only the general overview of the system is given.

The main goals of the system are the following: a) to promote students' structural knowledge self-assessment and b) to support teachers in improvement of study courses through systematic assessment and analysis of students' knowledge structures. Knowledge self-assessment is supported by automatic evaluation of students' concept maps and provision of informative and tutoring feedback. Systematic knowledge assessment is based on the possibility to extend an initially created concept map for other assessment stages. Statistics on differences between students' and teacher's concept maps allow teachers to improve their courses.

*IKAS* supports three categories of users: a) an administrator, b) a teacher, and c) a student. The administrator prepares the system for use by other users and manages its default parameters and data related to knowledge assessment process and its participants. Activities directly related to knowledge assessment are split between the teacher, the student, and the system and they include: 1) creation of a concept map by the teacher; 2) reproduction of the teacher's concept map by the student during completion of concept map based tasks; 3) comparison of the teacher's and student's concept maps by the system; 4) generation of feedback by the system.

Six tasks of different degrees of difficulty are implemented. Four of them are ‘fill-in-the-map’ tasks where an obligatory component is the structure of a concept map, which must be filled-in by students using available concepts and/or linking phrases. Last two tasks are ‘construct-the-map’ tasks where the student constructs a concept map from provided concepts and/or linking phrases. Ten transitions between tasks are realized: five of them increase the degree of task difficulty and another five reduce it.

The system provides rich student's support and includes a number of adaptive, adaptable, and intelligent features. Student's support is comprised of possibility to change the degree of task difficulty, to insert in automatic way some concepts into the provided structure of the concept map, to explain a concept, to check a created proposition, as well as quantitative data (like score received, time spent, number of help used, etc.) and qualitative data (concept mastering degrees, individual study plan). The main task of the system

is to perform automatic evaluation of students' concept maps. This is done in intelligent way by using the teacher's concept map as a reference map and a comparison algorithm that is based not only on the isomorphism of both graphs representing concept maps, but which is sensitive to the arrangement and coherence of concepts in students' concept maps and is capable to recognize partly correct patterns of a student's solution. Operation of the IKAS is based on interpretation of values of parameters available in a student model. The student model supports four adaptation operations in the IKAS: selection of the initial degree of task difficulty at the first assessment stage and its changing at next assessment stages, as well as setting and changing priorities of types of concept explanation. Adaptable features of the system available to students include the following ones: 1) adjusting the degree of task difficulty by directly changing the knowledge level of the course in the student model or reducing the degree of task difficulty during the completion of CM based tasks; 2) adjusting settings of the user interface such as language, theme, etc.; 3) changing approach for receiving of explanations and priorities for different types of concept explanation.

Since 2005, IKAS has been evaluated in more than 20 courses. Typically the system is used in content-rich courses, which include a lot of closely interrelated concepts, for example, "Fundamentals of Artificial Intelligence", "Methods of Systems Theory", "Discrete Structures in Computer Science". As a rule, a teacher prepares a concept map for a logically completed part of a course (module or topic block) and gives it to students at the end of this part. After completion of concept map based tasks by students, the teacher examines statistical data about students' incorrect or missing relationships and makes corrections in further curriculum of the course. Therefore, during the course students work with concept maps 3-4 times.

#### F. Summary of the Described Systems

Table 1 specifies the main characteristics of the previously described software systems. These systems are successfully used at the Faculty of Computer Science and Information Technology of RTU for knowledge assessment and self-assessment in the following study courses: "Data Structures", "Programming Languages", "Software Engineering", "Software Metrology and Planning", "Fundamentals of Artificial Intelligence", "Methods of Systems Theory", "Discrete Structures in Computer Science", etc. Results of the research presented in [16][17][18] show how use of the systems improves progress of students, motivates students, and increases quality of their knowledge at a stage of the final knowledge assessment (examination).

### IV. ANALYSIS OF THE SYSTEMS

The next sections discuss the methodology and results of the research.

#### A. Research Methodology

During the research presented in this paper the experts' survey was performed with aim to determine features of e-learning systems, which are most significant for knowledge assessment. The group of experts was selected by studying literature in the field of e-learning and through direct communication between the authors of this paper and experts at international workshops and conferences devoted to problems of computer-based knowledge assessment. As a result, nine experts participated in the survey.

The questionnaire used in the survey asked experts to evaluate importance of different features of e-learning systems, e.g. the processing of student's answer given in natural language, the management of knowledge assessment on the bases of mathematical models, etc., in relation to administration of students' knowledge assessment in such systems. Thirty features were included. Authors offered the following scale for evaluation of importance of each feature:

- 10 – an important and useful feature, but it is possible that it cannot be implemented at the moment (such features should not be in a large number);
- $8 \div 6$  – a feature, which can be implemented and should be implemented in the system;
- 4 – a useful, but not so important feature (moreover, it can be easily implemented);
- $1 \div 2$  – a feature, without which the quality of the system will not suffer significantly;
- 0 – a useless feature.

Data acquired through the questionnaire were processed in 2 stages. First of all, the coefficient of concordance  $W$  characterizing agreement between experts was calculated taking into account the methodology presented in [19].

After that, importance ( $S_j$ ) of each feature of e-learning systems was identified and the list of features in the decreasing order (the most important feature has the smallest  $S_j$  value) was created.

At the end, the ordered list of features was used for evaluation of software systems described in this paper. For this purpose the following indicator was introduced:

$$Z = \sum_{j=1}^n S_j C_j, \quad (1)$$

where  $S_j$  – importance of  $j$ -th feature;  $C_j$  – the coefficient characterizing extent of implementation of the feature in a system under evaluation, the value of this coefficient is determined by the reviewer of the system and depends on the level of feature implementation, but the range of values is permanent:  $C_j = [0..1]$ .

The authors of this paper determined the following levels of feature implementation:

- if a feature is fully implemented in the system, then  $C_j = 1$ ;
- if a feature is implemented, but it has some shortcomings, then  $C_j = 0.8$ ;
- if a feature is under implementation, then  $C_j = 0.4$ ;

- if a feature is not implemented, then  $C_j = 0$ .  
Further an indicator of relativity  $z$  was calculated:

$$z = \frac{Z}{Z_{\text{benchmark}}} \quad (2)$$

It characterizes the degree of compliance of a system under evaluation to a benchmark. The benchmark is calculated when values of the coefficient for all features are equal to 1. The value of the coefficient  $z$  is in the range from 0 to 1. The closer this coefficient is to 1, the closer the system is to the ideal system.

**B. Results of the survey**

The range of values for the coefficient of concordance  $W$  can vary from 0 to 1. In our case  $W = 0.64$ , so we can conclude that the agreement between experts is high. The involved experts considered the following features as the most important ones:

- Management of the knowledge assessment process on the basis of mathematical models;
- Possibility to analyze different types of answers (word, phrase, hotspot on an image, etc.);
- System’s openness for further improvements and development;

TABLE I. SUMMARY OF SYSTEMS

	Electronic Textbook	Universal Electronic Textbook	LMS	ORTUS	IKAS
<b>Users</b>	Student	Student; Teacher	Student; Teacher; Author; Administrator	Student; Teacher	Student; Teacher; Administrator
<b>Courses</b>	HTML language	Any course	Any course	Any course	Any course
<b>Mode</b>	Testing; Reference; Self-assessment	Learning; Testing	Random selection of tasks and/or questions for assessment; Training; Self-assessment	Training; Knowledge assessment	Knowledge assessment; Knowledge self-assessment with tutoring component
<b>Methods</b>	Fixed sequence	Fixed sequence; Random selection	Fixed sequence; Random selection	Random selection	The student starts with teacher’s specified degree of task difficulty. Further the degree can be changed on the basis on student’s request or in adaptive manner
<b>Adaptive knowledge assessment</b>	No	No	No	Yes	Yes
<b>Task type</b>	Multiple choice questions; Text/numerical input	Multiple choice questions; Text/numerical input	Multiple choice questions; Multiple response questions; Matching; Text/numerical input	Description; Essay; Multiple choice questions; Multiple response questions; Matching; True/False; Text/numerical input; Equation input; Sentence input	“Fill-in-the-map”; “Construct-the-map”

- Possibility of using data from a domain model during knowledge assessment;
- Possibility of performing adaptive knowledge assessment;
- Support of different types of users (student, author, teacher, administrator, etc.).

As the most unimportant features the experts mentioned the following ones:

- System’s functioning in a network (Internet, a local network, etc.);
- Provision of the certain response time to users’ actions;
- Possibility of using of probabilistic assessment models.

Table II presents results of evaluation of the software systems taking into account the coefficients  $Z$  and  $z$ .

TABLE II. RESULTS OF EVALUATION

	Electronic Textbook	Universal Electronic Textbook	LMS	ORTUS	IKAS	Benchmark
$Z$	64.8	136.4	209.6	298.4	242.6	465
$z$	0.14	0.29	0.45	0.64	0.52	1

The systems “Electronic Textbook” and “Universal Electronic Textbook” received the smallest evaluation ( $z=0.14$  and  $z=0.29$  accordingly), but ORTUS portal is closest to the benchmark ( $z=0.64$ ).

**CONCLUSION**

Taking into account experts' opinion about the most important features of e-learning systems for students’ knowledge assessment, the methodology for the usefulness

evaluation of such systems were developed. This methodology can be used in a combination with other methodologies for the evaluation of any e-learning system, as well as for comparison of this kind of systems between themselves. Five systems for administering students' knowledge assessment were evaluated using the developed methodology. Evaluation results showed that ORTUS and IKAS are the closest to the benchmark. These systems implements methods of adaptive assessment of students' knowledge, and they are open for further improvement and development. Therefore, the further work will be related to the improvement of ORTUS and IKAS considering the list of the most important features for knowledge assessment, determined by experts.

#### ACKNOWLEDGMENT

Travel costs and participation fee for this conference are financially supported by ERDF project „The development of international cooperation, projects and capacities in science and technology at Riga Technical University” Nr.2DP/2.1.1.2.0/10/APIA/VIAA/003.

#### REFERENCES

- [1] C. Gütl, K. Lankmayr, J. Weinhofer, and M. Höfler, "Enhanced Automatic Question Creator – EAQC: Concept, Development and Evaluation of an Automatic Test Item Creation Tool to Foster Modern e-Education", *Electronic Journal of e-Learning*, Vol. 9, Issue 1, 2011, pp. 23 – 38.
- [2] P. N. K. Lau, S. H. Lau, K. S. Hong, and H. Usop, "Guessing, Partial Knowledge, and Misconceptions in Multiple-Choice Tests", *Journal of Educational Technology & Society*, Vol. 14, Issue. 4, Special Issue on "Advanced Learning Technologies", 2011, pp. 99 – 110.
- [3] A. Anohina-Naumeca, J. Grundspenkis, and M. Strautmane, "The Concept Map Based Assessment System: Functional Capabilities, Evolution, and Experimental Results", *International Journal of Continuing Engineering Education and Life-Long Learning*, Vol. 21, Issue. 4, 2011, pp. 308-327.
- [4] R. Z. Cabada et al., „EDUCA: A Web 2.0 Collaborative, Mobile and E-learning Authoring System”, *Proceedings of IEEE International Conference on Advanced Learning Technologies (ICALT-2009)*, Riga, Latvia, Jul. 2009, pp. 287 – 289.
- [5] The Electronic Journal of e-Learning (EJEL) [Electronic resource] / – <http://www.ejel.org/main.html> [retrieved: May, 2013].
- [6] *Journal of Educational Technology & Society* [Electronic resource] / – English version: <http://www.ifets.info/>, Russian version: <http://ifets.ieee.org/russian/periodical/journal.html> [retrieved: May, 2013].
- [7] International Academy, Research, and Industry Association (IARIA) [Electronic resource] / – <http://www.iariajournals.org/> [retrieved: April, 2013].
- [8] L. Zaitseva, N. Prokofjeva, and V. Popko, *Electronic Textbook «Study HTML» // Proceedings of TECHNOMAT&INFOTEL 2004. Materials, methods and technology. Scientific articles. - Vol.1. - Bulgaria, 2004. - p. 63 - 70.*
- [9] N. Prokofjeva, "Computer knowledge control models and methods", *Proceedings of First International Conference "Information Technologies in Education for All", ITEA - 2006, Ukraine, IRTC – Kiev, 2006, p. 231 – 240 (in Russian).*
- [10] V. Popko and N. Prokofjeva, "Elektroniskās mācību grāmatas izstrāde", *RTU zinātniskie raksti. 5. sērija. Datorzinātne. Lietišķās datorsistēmas. 18. sējums. Rīga: RTU, 2004, - 103. - 109. lpp.(in Latvian)*
- [11] N. Prokofjeva, "The Methodical and Technical Aspects of Students Knowledge Control", *Proceedings of the 13th Education and Virtuality International Conference, VIRT - 2011, Ukraine – Yalta, 2011, pp. 263 – 272. (in Russian).*
- [12] L. Zaitseva and J. Bule, "Electronic Textbook and E-Learning System in Teaching Process", *Proceedings of E-learning conference'06 Computer Science Education, Coimbra, Portugal, Sept. 2006, pp. 189 – 192.*
- [13] L. Zaitseva, J. Bule, and U. Kuplis, "Advanced e-learning system development", *Proceeding of International Conference "Advanced Learning Technologies and Applications", ALTA-2003, Kaunas, Lithuania, 2003, pp. 14-18.*
- [14] J. D. Novak and D. B. Gowin, "Learning How to Learn", *Cambridge University Press, 1984, 199 p.*
- [15] A. Anohina-Naumeca and J. Grundspenkis, "Concept Maps as a Tool for Extended Support of Intelligent Knowledge Assessment". *Proceedings of the 5th International Conference on Concept Mapping, Malta, Sept. 2012.*
- [16] J. Grundspenkis, "Concept Map Based Intelligent Knowledge Assessment System: Experience of Development and Practical Use", *Multiple Perspectives on Problem Solving and Learning in the Digital Age, Springer, pp. 179-198.*
- [17] L. Zaitseva, "E-courses using Results and Efficiency", *Proceedings of the 11th IASTED International Conference "Computers and Advanced Technology in Education", CATE – 2008, Crete, Greece, Sept. 2008, pp. 90 – 93.*
- [18] L. Zaitseva and N. Prokofjeva, "Knowledge control during the preparation of IT specialists in Riga Technical University", *International Conference on Information Technologies (ICIT 2012): Information and Communication Technologies in Education, Manufacturing and Research, Russia, Saratov, Jun. 2012, pp 1-10.*
- [19] S. D. Beshelev and F. G. Gurevich, *Mathematical and statistical methods of expert evaluation. Moscow: Statistica, 1980 (in Russian).*

## ICT Related Tasks and Challenges In The New Model of Technical Teacher Training

András Benedek

Department of Technical Education  
Budapest University of Technology and Economics  
Budapest, Hungary  
e-mail: benedek.a@eik.bme.hu

György Molnár

Department of Technical Education  
Budapest University of Technology and Economics  
Budapest, Hungary  
e-mail: molnar.gy@eik.bme.hu

**Abstract**— As constant changes become a commonplace of everyday life, the clearly perceptible processes by which learning spaces are expanding are releasing new teaching and learning potentials and creating the opportunity to build knowledge networks. The symptoms of this development can be seen in the behaviour of learners, the formation of e-learning habits and the spread of community-based forms of communication. The effect of these changes can be perceived in the transformation of the roles of both teachers and students, and of the learning environment. The results of the survey conducted on a micro-scale in connection with the performance appraisal of the new elements may be relevant to the new educational and graduation requirements for technical teachers currently being formulated, and may help to guide future syllabus design and teaching material development. The results of the survey informs future teachers on how the new benchmark analysis is implemented in micro environment and can help appoint future training requirements and develop the appropriate curriculum.

**Keywords**—*ICT; learning network; Web 2.0 communication; e-portfolio.*

### I. INTRODUCTION

If the development of the learning environment is analysed in terms of changes in human behaviour, it can be seen that a new generation is developing whose members are devotees of online education. This generation moves more easily in the info-communication space than previous generations and it is becoming ever more informed and organized. With these competencies, the members of this generation get more information and support from each other than from the various institutions.

Forecasts predict that the role of learner communities will grow. These new communities are characterised primarily by their common areas of interest, where the learners interact with each other, learn together and amass shared reserves of information sources [1].

This nascent practice does not, however, preclude learning opportunities in a system of higher education in a period of transition. In this dynamized “learning space”, the classic roles (teacher-student) are scarcely perceptible, as thanks to the altered architecture of the net, an architecture which is based on collective knowledge sharing and content

generation with simple user interfaces, “every downloader will potentially become an uploader” [2].

First, the paper presents the items in the digital learning, then the E-portfolio support system, and, finally, shows an empirical research.

### II. ICT IN LEARNING ENVIRONMENTS

#### A. Items in the Digital Learning

ICT (Information and Communications Technology) and the development of forms of e-learning is characterised by ubiquitous computing, in which it is necessary to take into consideration the technical and societal qualities of a learning space shaped by virtual reality. In this space, an organic learning environment, an educational theoretical basis may be provided by e-learning. In this learning paradigm the key differentiating features are interactivity, the ability to manage time and space and asynchronous learning. All this is understandably in tension with traditional forms of education and instruction—teaching and learning which in some cases has remained unchanged for decades, cloistered in a closed world of classrooms. The forms of learning, which are taking shape under the influence of new technology are characterised by a focus on devices for individual use, alongside constant striving for interactivity and the application of networks.

The learning network is not merely a pedagogical theory, but an environment where the application of a new pedagogical theory of learning effectively supports the acquisition and generation of knowledge. Currently, the main features of the interaction between education and information are:

- Developed forms of human-machine interaction
- Spatial and temporal independence
- The ubiquity of mobile telephony
- The opportunity to create a complex multi-media “learning space”

The Web 2.0 [4] concept does not only refer to new forms of websites and services, but also to the technologies they are based on which allow community content development, and to the ability of these social networks to democratically locate the individual. Pedagogically speaking, the Web 2.0 implies openness to constructive forms of communication, which include free or low-cost

services with well-developed ICT support. Currently, such services can be said to include (and this list is far from complete) social networking services (Facebook, iWiW), photograph sharing sites (Flickr), video sharing sites (YouTube, Videa) and various types of blogs. Wikipedia and other open freely editable information resources have significant potential for content development, and auction websites (Vatera, eBay) can also be counted as part of Web 2.0, Twitter, various social bookmarking sites (delicious), forums, online office suites (Google Docs), web (news)feeds (RSS) and file hosting services (Dropbox, Google Drive). In the Web 2.0 environment, the fundamental orientation is towards community participation and creative application of methods which are based on community content development. From a pedagogical standpoint, it is particularly important to recognize that in the Web 2.0 space the users jointly create the content and share the knowledge thus generated with each other [2]. It is, however, typical of web 2.0 services—and in this the progressive institutions of higher education are pioneers—for the institution to provide a technological framework and organizational structure where the users upload and develop content themselves, and where the system facilitates the further development of knowledge, knowledge sharing and the expression of opinions. All this requires the establishment of technical and ethical norms, which will have a defining role in the development of these systems.

The technical criteria of individual-oriented learning are also changing in step with the spread of information technology, broad-band internet access and access to mobile networks. The greatest challenge for traditional pedagogical thinking in the coming years will come in the shape of the spread of nonformal learning. It is becoming a universal tendency in developed countries for this form of learning to play an ever more important role in comparison with institutional study. This is also the case when this form of learning is not typified by formal qualifications or degrees, although learning goals and content are still important in their absence. Mobile learning is growing in significance for learners/students in initial training, as well as in later periods of study.

Beyond those learning forms which function outside of formal education (in the school system): informal learning, atypical learning encompasses all the methods (distance learning, open learning), which exercise an influence on the intellectual and physical development of the individual [2].

Digital pedagogy has arisen as a response to the questions raised by the interplay of new technology and education described above. Its aim is to make as wide reaching a study as possible of all related challenges and opportunities which affect active participants in the teaching and learning process. As the media environment has transformed, the teaching objectives and roles need to be reformulated too. It is worth considering the ways and means in which interactive solutions and online and mobile support can be incorporated when planning courses in

various contexts. The aspects of the latest incarnation of the Web (wikis, blogs) make ICT device based education usable on an everyday level.

Grouped from a pedagogical perspective, work on the following content types, activities and processes is especially applicable to Web 2.0 communication:

- Varying levels of education, competence and knowledge. Raising awareness, boosting motivation. Communication management implementation in learning
- Computer literacy; development of required abilities and attitudes. Giving recognition of an individual's level of knowledge
- Formation of learning communities. Collaborative and co-operative learning
- E-learning as the application of distance learning in a virtual educational environment
- The use of blogs in education: course-blog, podcast, videoblog as (homework) assignments [4].

#### B. *E-portfolio support system—Mahara*

Portfolio is a word of Italian origin meaning a dossier or folder of documents, or an expert's dossier, or in more prosaic terms it refers to a collection of a student's work. Many definitions have been proposed, the one used here is that cited by Iván Falus and Magdolna Kimmel: "The portfolio is a purposeful, systematic collection of works completed by a student on one or more courses". It has two objectives: on the one hand it is for assessment purposes, on the other hand to aid in learning [5].

Three basic types of portfolios can be distinguished: [6][7]:

- The Working portfolio (progress documenting collection): The student collects all documents pertaining to a particular learning process in this portfolio. This type of portfolio fulfils the role of demonstrating the presence and level of development of teaching competences. The student can then receive regular feedback on the basis of these documents, and evaluate her own progress.
- Showcase (Best works) Portfolio: Only the best work of the student is included in the showcase portfolio. This work is selected by the student himself from among the documents in the working portfolio which in his opinion best reflect him and his achievements. Naturally in order to do this the student must be able to judge which his best pieces of work are, and it is necessary for the student to have access to a clear evaluative framework.
- Assessment portfolio (evaluation of results): this is a type of alternative assessment, measured against external assessment criteria. Every document has to meet a certain standard of achievement, and so the assessment criteria are the boundary lines for this type of portfolio. The main aim is for the teacher to evaluate the student's achievement not by means of traditional grades but holistically, with the help of representative samples of the student's work.

Portfolios have a long history in Hungarian public education, as when, for instance, portfolio-based methods were tested in a number of primary schools as part of the Hungarian Institute for Educational Research and Development's "MAG" ("seed") project. The results of this initiative are presented in the OKI (National Institute of Public Education)'s publication "Magtár-Ötletek tanítóknak a fejlesztő értékeléshez és az adaptív tanulásszervezéshez" (Seed-store-Ideas for educators towards formative evaluation and adaptive learning management) [8]. The goals of the portfolios which were created in this project were decided upon by the teachers involved, and consequently the aims and content of the portfolios were extremely varied. Some educators set out to assess all of the students' work over a school year, while others focused on the students' progress in a single subject or course. Some teachers focused on the achievements of only one student, while others dealt with whole classes or larger groups.

In 2009, the BME Department of Technical Education and the BME Institute for Applied Pedagogy and Psychology (APPI), recognising the difficulties of implementing a traditional portfolio system drawing on experience garnered from the final examinations of the first graduating year, while bearing in mind environmental protection, and the growing power of electronic forms of learning, introduced an electronic portfolio framework employing the Mahara electronic environment integrated into the Moodle system, which has been in use since 2006 [9]. The applications of an electronic learning environment and computer based learning have become widespread in Hungarian higher education [10][11].

The e-portfolio, that is, an electronically-based portfolio, aims to face the challenges of the constantly and rapidly evolving ICT environment and fulfil the learning needs of the digitally savvy "generation Z" as well as the need to optimize the process of gathering work for assessment. The e-portfolio is especially suitable for the following tasks:

- Management of multifaceted complex dossiers
- Handling large quantities of documents in a single system
- Continuous preparation and monitoring (by mentor teachers)
- Gaining experience in the use of current technology-centred theories of learning
- Excellent ICT support

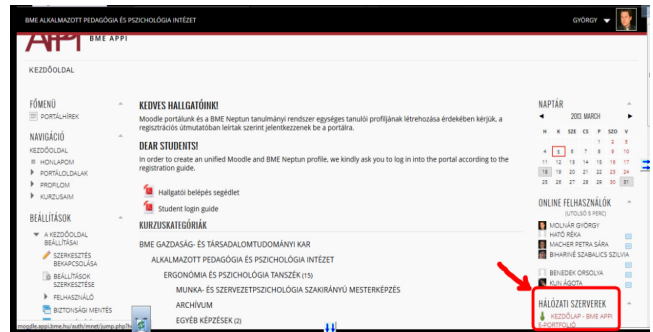


Figure 1. E-portfolio in the BME Moodle.

The Mahara system can be directly accessed at the [12]. It can also be reached through the Moodle system via a link in the bottom right hand corner of the page (See Figure 1.). The figure below shows the main options page ("dashboard") after logging on:

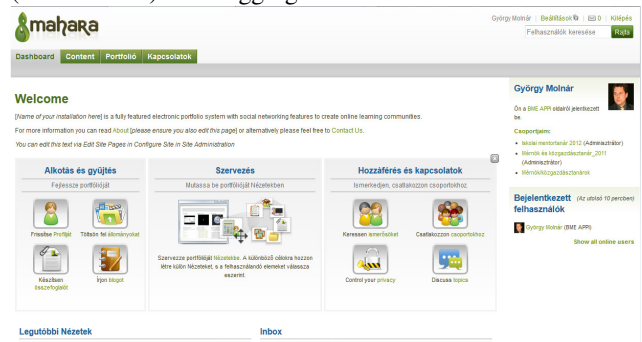


Figure 2. The Mahara interface after logging on.

Using this interface is relatively simple for users, and the "drag and drop" technique can be used to edit the user profiles. Uploading files and blog posts is also straightforward, by means of the browser function and text box editing (See Figure 3.) [13].

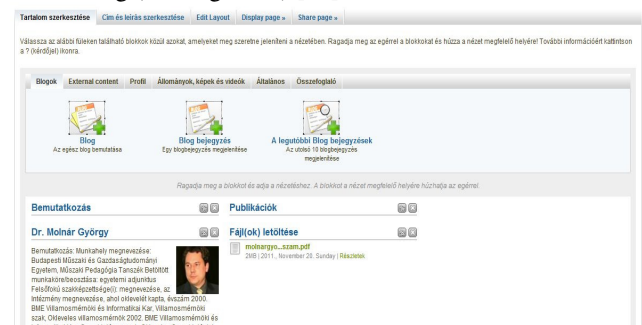


Figure 3. Editing a Mahara profile.

The e-portfolio system facilitates the creation of community forums and groups, and it even has an instant messaging function. Employing this model, for example, the BME Department of Technical Education established a group for mentor teachers and students pursuing engineering teacher training, for which all participants had to register and create their own profiles. The members of the groups



could thus directly contact each other through the system, as well as the mentor teacher and a maximum of 4-5 trainee teachers assigned to him. The interface also facilitated the tracking, monitoring and evaluation of work added to the portfolios over the course of the semester.

The uploading of documents takes place by means of the file management interface on the first page of the site. Trainee teachers belonging to the group are able to edit the folders containing work they have already uploaded, which they can then share to make them accessible to university students and mentor teachers. The structure of the portfolio, in accordance with the theoretical elements of the relevant legislation, is organized into 3 main folders. A student's e-portfolio created in this way is shown in the following screen capture image (See Figure 4.)[14].

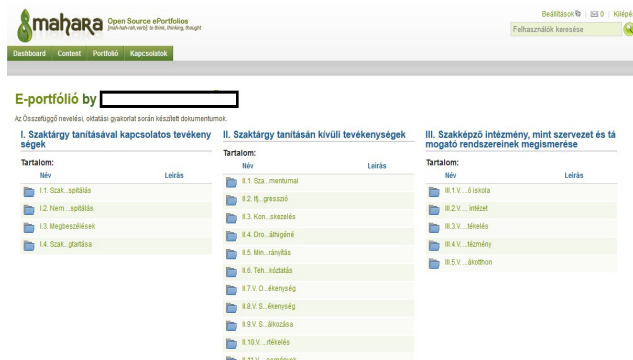


Figure 4. A Trainee Teacher's portfolio in Mahara.

C. Description of the research

The main hypothesis of the research is that, in order to provide high quality education, there is just as much of a need for future teaching professionals who are well-versed in methodology, learning technology and their own subject area at school and vocational college level as there is in higher education. At the same time, the recent changes present student teachers with significant challenges along with the increased burdens related to admission and graduation requirements and new types of tasks.

The numbers of university students graduating and beginning courses under the Bologna system were initially rather modest. The figures obtained retrospectively after 4 years were:

- 1st generation–6 students (2009)
- 2nd generation–18 students (2010)
- 3rd generation41 students (2011)
- 4th generation57 students (2012)

A cross-sectional performance appraisal was conducted in autumn 2012 to gauge participants' impressions of and satisfaction with the new education format. The investigation employed a well-established approach in pedagogical practice: the online questionnaire. This comprised 24 closed-ended multiple choice questions or Likert-scale type items. Number of the answers are N=116.

The survey used a closed-ended type survey engine to collect the respondents' answers [15].

The figure below shows the distributions of several typical results of the survey.

The area where students had the greatest difficulty in creating their teacher portfolios was assembling and organizing suitable documents to include.

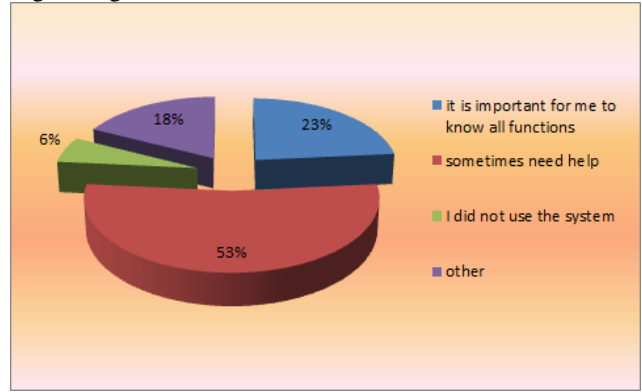


Figure 5. Competencies in using Mahara.

53% of the respondents were satisfied with their ability to use the Mahara system, and only rarely required assistance with it (See Figure 5.).

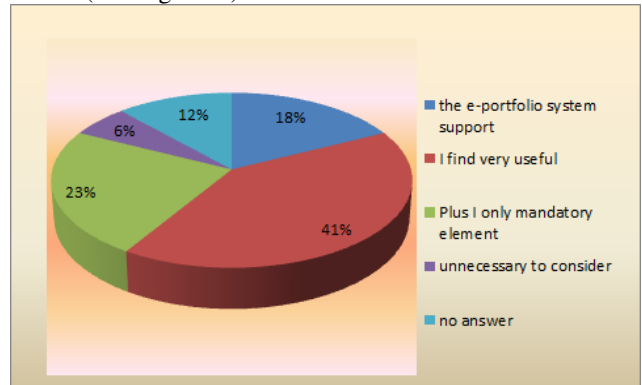


Figure 6. The role of Mahara in preparing teacher portfolios.

41% of the respondents rated the Mahara system as very useful for preparing the e-portfolio (See Figure 6.).

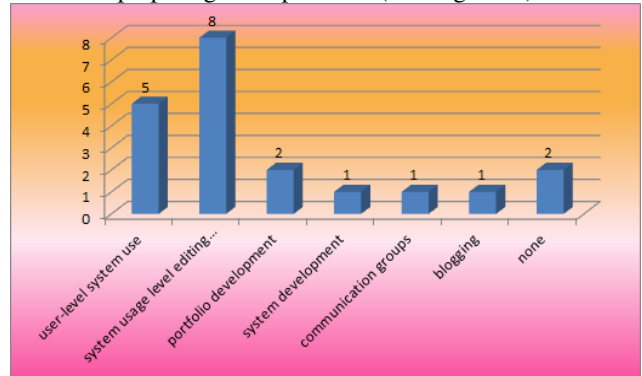


Figure 7. Further reasons for using Mahara.

More than half of the respondents required assistance connected to editing their portfolios, and a quarter needed help to use the system at all (See Figure 7.).

### III. CONCLUSION

The evaluation indicates that 35% of those asked found the preparation of the portfolio the most difficult part of the assessment, while 28% found teaching practice the hardest, and the final examination was considered to be the least challenging of the tasks they faced. This may be explained by the need for students to familiarize themselves with elements of the course which were new to them, and to gradually learn the necessary skills by the time they took the final examination. The majority of respondents rated the Mahara portfolio management system as very useful, and only occasionally required minor assistance in using it, and would willingly use it in the future for other applications. In addition the respondents wished to learn more about other portfolios created by the community, and would welcome reinforcement of the system's approach. The majority would support the incorporation of material regarding the e-portfolio into one of the course modules. According to one written opinion, "Wide-scale use of (the system) becomes increasingly essential as the pedagogical work progresses." [15].

The new, currently still evolving teacher training model raises numerous important questions and will replace current courses by 2016 at the latest. The road to that destination is a long and rather winding one. The results of the survey conducted on a micro-scale in connection with the performance appraisal of the new elements may be relevant to the new educational and graduation requirements for technical teachers currently being formulated, and may help to guide future syllabus design and teaching material development.

### REFERENCES

- [1] <http://index.hu/tech/net/web1214>, last acces: 05.09.2012.
- [2] Piet Kommers, P.A.M.: ICT as explicit factor in the evolution of life-long learning. *International journal of continuing engineering education and life-long learning*, 20 (1/2010), pp. 127-144., last access: 15.08.2012
- [3] <http://enc.phil-inst.hu/1enciklopedia/fogalmi/ped/atiptan.htm>, last acces: 15.09.2012.
- [4] András Benedek (ed.): *Digital pedagogy 2.0 – Typotext Budapest 2013.*, pp. 18-23
- [5] Kalimkova: *Portfolio kak sztredestzvo szamoorganyizacii i szamorazvityija licnsoszyi*, 5.2002 pp. 23-25
- [6] De Fina, Allan A.: *Portfolio assessment, getting Started*, New York, 06.1992
- [7] Helen C. Barret: *Create Your Own Electronic Portfolio*, In: *Learning & Leading with Technology*, 04.2000., pp. 15-21
- [8] <http://mag.ofi.hu/magtar-otletek/magtar-otletek>, last acces: 22.06.2013.
- [9] András Benedek and György Molnár: The empirical analysis of a Web 2.0-based learning platform, In: Constantin Paleologu, Constandinos Mavromoustakis, Marius Minea (ed.): *ICCGI 2011, The Sixth International Multi-Conference on Computing in the Global Information Technology*, Luxembourg, June 19-24, 2011., ISBN: 978-1-61208-008-6, 06.2011 pp. 56-62, last access: 05.2012
- [10] Péter Tóth: *Adaptive Online Learning Environment and Web Usage Mining*, IEEE 8th International Symposium on Applied Computational Intelligence and Informatics (SACI), Timisoara, Romania, 2013. pp. 61-66
- [11] I. Simonic, "eLearning and Presentation Techniques." *Óbuda University e-Bulletin*, Vol. 1 (1), 2010, pp. 211-217.
- [12] <http://e-portfolio.appi.bme.hu/>, last acces: 22.06.2013.
- [13] György Molnár: *Collaborative Technological Applications with Special Focus on ICT based, Networked and Mobile Solutions.*, *Wseas Transactions on Information Science and Applications* 9:(9)2012 pp. 271-281
- [14] György Molnár: *Flashes or steady light? Or the potentials of developing networked learning*, In: Miguel Baptista Nunes, Maggie McPherson (ed.): *Proceedings of the IADIS International Conference e Learning, IADIS international conference E-learning 2011, Volume II*. Rome, Italy, July 20-23, 2011, ISBN: 978-972-8939-38-0, 07.2011 pp. 405-408
- [15] <http://appi.bme.hu/survey/admin/admin.php>, last acces: 22.06.2013.

# The Impact of Requirements on Software Quality Across Three Product Generations

John Terzakis  
Intel  
Hudson, MA USA  
john.terzakis@intel.com

**Abstract**— In a previous case study, we presented data demonstrating the impact that a well-written and well-reviewed set of requirements had on software defects and other quality indicators between two generations of an Intel product. Quality indicators for the second software product all improved dramatically even with the increased complexity of the newer product. This paper will recap that study and then present data from a subsequent Intel case study revealing that quality enhancements continued on the third generation of the product. Key product differentiators included changes to operate with a new Intel processor, the introduction of new hardware platforms and the addition of approximately fifty new features. Software development methodologies were nearly identical, with only the change to a continuous build process for source code check-in added. Despite the enhanced functionality and complexity in the third generation software, requirements defects, software defects, software sightings, feature commit vs. delivery (feature variance), defect closure efficiency rates, and number of days from project commit to customer release all improved from the second to the third generation of the software.

**Keywords**— requirements specification; requirements defects; reviews; software defects; software quality; multi-generational software products.

## I. INTRODUCTION

This paper is a continuation of an earlier short paper [1] that presented quality indicator data from a case study of two generations of an Intel software product. The prior case study compared the quality metrics for a first generation software product (“Gen 1”) developed without a requirements specification (e.g., Product Requirements Document or Software Requirements Specification) versus a second generation product (“Gen 2”) developed with one as its foundation. This paper includes the background and validation results from a third generation product (“Gen 3”) that was designed and coded utilizing the set of requirements for the Gen 2 product as its basis. All three products were developed using traditional (i.e., waterfall) software development methodologies.

This paper is organized into eight sections. Section I provides an introduction. Section II gives the backgrounds on the three product generations. Section III presents the requirements defect rates in the three product requirements documents by revision. Section IV analyzes the predicted defect potential for the three products. Section V presents the test results for the first versus the second generation products. Section VI presents the test results for the second

versus the third generation products. Section VII describes conclusions based on the data. Section VIII discusses possible future work.

## II. PRODUCT BACKGROUNDS

The requirements for Gen 1 that existed were scattered across a variety of documents, spreadsheets, emails and web sites and lacked a consistent syntax. They were under lax revision and change control, which made determining the most current set of requirements challenging. There was no overall requirements specification; hence reviews were sporadic and unstructured. Many of the legacy features were not documented. As a result, testing had many gaps due to missing and incorrect information.

The Gen 1 product was targeted to run on both desktop and laptop platforms running on an Intel processor (CPU). Code was developed across multiple sites in the United States and other countries. Integration of the code bases and testing occurred in the U.S. The Software Development Lifecycle (SDLC) was approximately two years.

After analyzing the software defect data from the Gen 1 release, the Gen 2 team identified requirements as a key improvement area. A requirements Subject Matter Expert (SME) was assigned to assist the team in the elicitation, analysis, writing, review and management of the requirements for the second generation product. The SME developed a plan to address three critical requirements areas: a central repository, training, and reviews. A commercial Requirements Management Tool (RMT) was used to store all product requirements in a database. The data model for the requirements was based on the Planguage keywords created by Tom Gilb [2]. The RMT was configured to generate a formatted Product Requirements Document (PRD) under revision control. Architecture specifications, design documents and test cases were developed from this PRD. The SME provided training on best practices for writing requirements, including a standardized syntax, attributes of well written requirements and Planguage to the primary authors (who were all located in United States). Once the training was complete, the primary author submitted early samples of his requirements to the SME for review and constructive feedback. The requirements were then rigorously reviewed by both technical content experts and the SME at each major revision of the PRD.

The Gen 2 software product shared many of the same characteristics of the first product: it ran on similar platforms, was developed across multiple sites and had a two

year SDLC. However, it was far more complex than the first in that the software had to run with and implement functionality for a next generation Intel processor with a new microarchitecture. In addition, the multiple code bases were combined into a single release.

The Gen 3 software utilized the final set of requirements for Gen 2 as a basis for the initial PRD. The requirements SME remained with the team and worked with a new primary requirements author and later with four additional authors, who were located outside of the United States. No other requirements methods or practices changes were made. With the exception of adding support for a new CPU and the functionality it enabled (approximately 50 new software features), the basic attributes of the Gen 3 software were similar to those of Gen 2. The software development process did change slightly; the team switched to a continuous build for source code check-ins.

### III. REQUIREMENTS DEFECT RATES

As mentioned previously, requirements for the first generation product were spread across documents, emails and web sites. Since the reviews were infrequent and informal, no data was captured to quantify the requirements defect levels. Furthermore, there was no one on the team able to objectively assess and measure defect levels.

For the second generation product, the SME reviewed each revision of the PRD, logged the defects and calculated the defect rate (measured in defects/page or DPP) for each revision. Requirements were evaluated using various checklists including the ten attributes of a well written requirement (complete, correct, concise, feasible, necessary, prioritized, unambiguous, verifiable, consistent and traceable), an ambiguity checklist (including vagueness, subjectivity, passive voice, and weak words) and a checklist to determine if a non-functional requirement is verifiable. Non-conformance to any of the checklist items constituted a requirement defect.

Initial requirements defect levels were high as this was the first formal set of requirements written by the author. However, with mentoring, peer reviews and stakeholder feedback, the requirements defect density for the PRD was reduced from about 10 DPP in an initial revision (0.3) to less than 1 DPP in the final revision (1.0), a reduction of approximately 98%. The results, published initially in another short paper [3], appear in Table I below.

TABLE I: GEN 2 REQUIREMENTS DEFECT DENSITY

PRD Revision	# of Defects	# of Pages	Defects/Page (DPP)	% Change in DPP
0.3	312	31	10.06	-
0.5	209	44	4.75	-53%
0.6	247	60	4.12	-13%
0.7	114	33	3.45	-16%
0.8	45	38	1.18	-66%
1.0	10	45	0.22	-81%
Overall % change in DPP revision 0.3 to 1.0: <b>-98%</b>				

The defect rate for each revision of the third generation software PRD was also assessed and recorded by the SME. In addition to a new primary author, four other authors contributed after the initial revision. Their impact can be observed at the release of revision 0.5, which shows an increase of 20% in the number of defects per page. Due to budgetary restrictions, these new authors had not been previously trained in writing requirements like the U.S. based author. They were mentored via phone by the SME for subsequent revisions. From initial to final revision, there was an approximate 80% decrease in the document defect levels. The details are presented in Table II that follows.

TABLE II: GEN 3 REQUIREMENTS DEFECT DENSITY

PRD Revision	# of Defects	# of Pages	Defects/Page (DPP)	% Change in DPP
0.3	275	60	4.58	-
0.4	350	78	4.49	-2%
0.5	675	125	5.40	+20%
0.7	421	116	3.63	-33%
0.75	357	119	3.00	-17%
1.0	115	122	0.94	-69%
Overall % change in DPP revision 0.3 to 1.0: <b>-79%</b>				

Defect prevention practices involved early inspections of requirements and a cross-functional review process. For both Gen 2 and Gen 3, the primary authors submitted initial samples of the requirements to the SME for inspection. The SME provided detailed feedback on the requirements defects and then mentored the authors on how to rewrite them so that they have a clear, common and coherent understanding amongst all stakeholders. Next, requirements were reviewed by peers for technical correctness, completeness and technical feasibility. Once the peer review was complete and the changes incorporated, the PRD was circulated for cross-functional feedback. To ensure better response rates, a "differences" document was circulated (generated from the RMT) that listed the changes between revisions. Also, several review meetings were held to discuss key sections of the PRD and obtain direct feedback on any issues.

This emphasis on minimizing requirements defects was driven by industry data linking requirements defects to software defects. Depending on the study [4], [5], [6], requirements defects are responsible for between 50% and 75% of the total number of software defects. The Chaos Reports by the Standish Group, including the one from 2009 [7], have identified requirements as one of the leading causes of project failures. Also, there is industry data from multiple sources [8], [9] indicating that the cost to fix a defect is at least 100 times more expensive in production than in the requirements definition phase. Consequently, there was considerable effort expended on the Gen 2 and Gen 3 products to focus on defect prevention rather than the traditional defect detection done by testing teams.

From a requirements perspective, the primary author for the Gen 2 product would have continued to create

requirements at a rate of about 10 DPP (or higher based on the defects the initial samples) without mentoring from the SME. Hence, the final revision of the PRD would have had approximately 450 defects as opposed to the 10 defects actually identified. Many of these defects would eventually have been coded into the product. Similarly, the Gen 3 author would have potentially introduced about 560 defects into the software without mentoring (vs. the 115 defects in the final revision of the PRD). The key question became whether this focused effort on requirements defect prevention would have any impact on software defects and other quality indicators.

#### IV. PREDICTING DEFECT POTENTIAL

To predict the defect potential [10] across the three generations of the software, we must analyze the various factors that impact the number and severity of software defects including: maturity of the team (development and validation), number of new features, the complexity of the new features, test coverage and stability of the code base at the start of the project. Comparing the three software development efforts, the teams were of about equal size and maturity and their development methodology was identical, specifically traditional waterfall. The validation teams were also of similar size and maturity. There was some overlap of personnel between projects. Overall, team maturity was consistent and thus should not influence software defect levels.

From a feature perspective, each new generation added features upon the base set of requirements from the previous generation. These new features were more complex, as they had to enable functionality in the newer Intel CPU that the software ran on. The number of requirements continued to grow since no requirements were removed for subsequent versions of the product. Test coverage of the software increased due to the introduction of formalized requirements starting with Gen 2. These factors would normally contribute to a rise in software defect rates.

The final factor to analyze is the stability of the code bases. The first version of the software consisted of multiple code bases with differing source code control systems (SCCS) and build processes. Code stability across each of the components was good. These code bases were merged into a single, unified software release at the start of the second project. After an initial period of instability due to integration issues associated with the move to a common SCCS and build process, the software should have reached the same stability as the original components since the code itself did not change. The impact of the continuous build process introduced in the third generation project also needs to be assessed. The smaller, incremental builds are likely to improve factors such as defect closure efficiency, feature variance and time from commit to product delivery since the team is able to get new builds in a much more timely manner (hours vs. days). As a result, issues can be resolved faster and this could enhance code stability. However, the more frequent build cycles can also lead to a higher number of sightings and defects because the test team can perform more

testing. Thus, from a code stability perspective, software defect and sighting rates should be relatively unaffected by the code merge in Gen 2 and the continuous build process in Gen 3.

When all factors are taken into consideration, the defect potential should be *higher* for the second generation than the first and for the third generation versus the second. The key driver is number of features. Each of the subsequent products has many more features than its predecessor, those features are much more complex and the software runs on a more advanced version Intel CPU. The other factors are neutral relative to the defect potential.

#### V. TEST RESULTS: GEN 1 VS. GEN 2

The following data presents a comparison of the number of software defects at release, requirements volatility at major milestones, feature variance at major milestones, and defect closure efficiency at release between the first generation and second generation software products. One additional set of data is now available: time from committed product release date to actual product release date. Note that all of this data was collected for the products on two similarly configured mobile platforms only. The primary difference between them was the generation of the processor.

The most impressive set of test results is presented in Table III: the total number of software defects by type per product at the end of validation testing. SW defects rates impact not only development times and efficiency, but also customer satisfaction levels and brand reputation. The results are dramatic: each severity of software defect demonstrated a very precipitous decrease between generations. Of particular note are the 86% reduction in critical defects and over 50% drop in total defects. These results are extraordinary given the increased complexity of newer software.

TABLE III: NUMBER OF SW DEFECTS BY TYPE

Defect Type	Gen 1	Gen 2	Delta
Critical	21	3	-86%
High	137	69	-50%
Medium	111	62	-44%
Low	24	6	-75%
Totals:	293	140	-52%

Table IV compares the requirements volatility (1) per product at key milestones during development. Requirements volatility is a measure of how much the requirements are changing due to additions, modifications or deletions. A stable product will have a lower volatility index. At release, the second generation product had almost half the volatility of the first. An analysis of the Gen 1 requirements volatility by the development team revealed that most of it was due to changes needed to resolve

customer reported defects. Customer-driven change is evident from the tripling in the requirements volatility index from Alpha to Beta and the more than fourfold increase from Alpha to Release. For the Gen 2 product, customer defect levels were much lower and hence the volatility index increase from Alpha to Release was slightly over double.

TABLE IV: REQUIREMENTS VOLATILITY

Milestone	Gen 1	Gen 2	Delta
Alpha	0.4	0.4	0%
Beta	1.2	0.7	-42%
Release	1.7	0.9	-47%

$$\text{Volatility} = \frac{\# \text{ of added+changed+deleted requirements}}{\text{Total \# of requirements}} \quad (1)$$

Feature variance (2) per product at key milestones during development is displayed in Table V. This metric shows how well the features delivered in final product matched what was committed by the team to be delivered. At each milestone, the second generation product team was able to deliver between 1.67 times and 3 times as many supplementary features as the first generation product team. This is likely due to the efficiency gain of the Gen 2 team by not having to debug as many defects as the Gen 1 team (Table III).

TABLE V: FEATURE VARIANCE

Milestone	Gen 1	Gen 2	Delta
Alpha	0.05	0.15	+3.00x
Beta	0.15	0.25	+1.67x
Release	0.15	0.35	+2.33x

$$\text{Feature Variance} = \frac{(\text{Current} - \text{Planned Features})}{\text{Planned Features}} \quad (2)$$

Table VI shows the software Defect Closure Efficiency (DCE) (3) at the end of validation testing for the first software release versus the second. DCE is measured by dividing the number of SW defects closed by the number of SW defects submitted. The goal is to have this percentage approach 100% (all defects closed) by release. A lower DCE is an indication that the development and validation teams are spending more time identifying, researching and correcting software defects (likely due to high defect levels). In this table, DCE at product release increased from about 69% in Gen 1 to about 87% in Gen 2, an improvement of over 25%.

TABLE VI: SW DEFECT CLOSURE EFFICIENCY

Milestone	Gen 1	Gen 2	Delta
Release	69%	87%	+26%

$$\text{DCE} = \frac{\text{Cumulative SW defects closed}}{\text{Cumulative SW defects submitted}} \quad (3)$$

Finally, Table VII provides a comparison of the number of days between the official project commit date and the actual customer release date for the two products. As part of project planning, the development team submits a full plan consisting of the features, resources, schedule (including release date), costs and risks. Once this project plan is approved, the project is committed. Factors such as inaccurate estimates, changing customer requirements, and technical problems can cause the actual delivery date to slip. The data shows that the second generation product was released 123 days earlier (or about 22% faster) than the first generation. The cost savings are substantial given the size of the project team (in excess of 100 people).

TABLE VII: PROJECT COMMIT TO RELEASE TIMES

Milestone	Gen 1	Gen 2	Delta
Project Commit to Release	564 days	441 days	-22%

In summary, all five quality indicators displayed substantial improvements at release to customer from Gen 1 to Gen 2 of the software:

- Total software defects: -52%
- Requirements volatility: -47%
- Feature variance: +2.33x
- Software DCE: +26%
- Time from project commit to release: -22%

## VI. TEST RESULTS: GEN 2 VS. GEN 3

The next set of data compares the total number of software defects, total number of sightings, feature variance at release, defect closure efficiency and the number of days from project commit to product release between the second and third generation software products. This data was gathered from testing for the two products on all mobile and desktop platforms. This is different from the data presented in Section V, which was for the Gen 1 and Gen 2 products on two specific mobile platforms. Unfortunately, there is no way to extract the data for mobile platforms only from the Gen 3 data as only the combined results for all mobile and desktop platforms were available for that product. Access to this would have allowed a direct comparison among the Gen 1, Gen 2, and Gen 3 products. However, the findings that follow do demonstrate continued improvements on similarly configured hardware.

The total number of software defects and total number of sightings (open issues reported by the test team, not all are

defects) at release for the second and third generation software are shown in Table VII. The reduction in overall defects that started from Gen 1 to Gen 2 continued from Gen 2 to Gen 3, with a decrease of 35%. Total sightings dropped by 31% between releases. These figures are noteworthy due to the increased functionality in the Gen 3 product. Despite an almost tripling in the revision 1.0 PRD length (122 pages vs. 45 pages), total software defects and sighting still declined by sizable percentages.

TABLE VII: TOTAL # SW DEFECTS AND SIGHTINGS AT RELEASE

Milestone	Gen 2	Gen 3	Delta
Total Defects	1,060	690	-35%
Total Sightings	3,800	2,640	-31%

Table VIII displays the feature variance per product at release. The Gen 3 product team was able to deliver about 1.23 times as many supplementary features as the Gen 2 product team at release. Some of this gain is attributable to the continuous build process (and being able to respond to and fix defects faster), but a good percentage is due to the stability of the initial set of requirements. The Gen 3 team could spend more time on newly arriving customer requirements requests.

TABLE VIII: FEATURE VARIANCE AT RELEASE

Milestone	Gen 2	Gen 3	Delta
Release	0.35	0.43	+1.23x

Table IX presents the software Defect Closure Efficiency at first customer shipment for the second generation in comparison to the third generation software. In this table, DCE increased from slightly (about 7%) from 87% in Gen 2 to about 93% in Gen 3. This small gain is not totally unexpected given already high DCE for the second generation product. The quality of the requirements and continuous build process both appear to be positively affecting these figures.

TABLE IX: SW DEFECT CLOSURE EFFICIENCY

Milestone	Gen 2	Gen 3	Delta
Release	87%	93%	+7%

The next set of data in Table X provides a comparison of the number of days between the project commit date to delivery of the released product to the customer. Again, improvements were made from the second to the third generation as 84 days were removed from the schedule (a reduction of about 19%). If the data from the first generation product is included, there is a 207 day or an overall 37% decrease in time from commit date to customer delivery from Gen 1 to Gen 3. Again, this contributed to considerable cost savings based on the team size.

TABLE X: PROJECT COMMIT TO RELEASE TIMES

Milestone	Gen 2	Gen 3	Delta
Project Commit to Release	441 days	357 days	-19%

One final set of data involved customer satisfaction levels. Within the first six months of release, customers were asked to rate the quality of the software delivered according to a four point scale: poor (0), fair (1), good (2) and very good (3). Scores for Gen 1 software averaged between the “fair” to “good” range. For the Gen 2 product, they had moved into the “good” to “very good” range. The Gen 3 product was also in this range, slightly higher than Gen 2.

To summarize, five key quality indicators displayed continued improvements at customer release from Gen 2 to Gen 3 of the software:

- Total software defects: -35%
- Total sightings: -31%
- Feature variance: +1.23x
- Software DCE: +7%
- Time from commit to delivery: -19%

VII. CONCLUSIONS

The key quality result from this extended case study is that the dramatic reductions in total software defects observed from the first to the second generation software on mobile platforms (-52%) continued for the second to the third generation software on the combined mobile and desktop platforms (-35%). A number of factors could have had some impact in these results. They include applying lessons learned from one project to the next, augmented developer experience and maturity, enhanced code review practices and more rigorous unit testing prior to the start of validation. No doubt these factors had some influence on improving software defect levels. However, given the increased complexity of the second and third generation products, these factors should have had a minimal effect on total software defect density levels. Some other factor was playing a dominant role in these extraordinary quality results.

Based on these observations, the improved requirements were the major contributing factor to these reductions in software defects. The project participants noted that the focus on requirements defect prevention appreciably minimized requirements ambiguity, subjectivity and misinterpretation. In addition, non-functional requirements (quality and performance) were written to be more verifiable. The net result was fewer requirement defects propagating to downstream work products like architecture specifications, design documents, code, and test cases. As a result, the development team was able to release code to the test team with fewer defects despite considerable increases in functionality and complexity for the newer versions of the software.

The focus on requirements defect prevention also dramatically impacted other quality indicators including feature variance, defect closure efficiency and project duration. Since less time was spent fixing defects, the developers could spend more time adding feature requests arriving late in the software lifecycle. This is reflected in the increases of feature variance from the first to the second generation (2.33 times) and from the second to the third generation (1.23 times). These late feature additions allowed the teams to be more responsive to changing market conditions, competitive pressures and customer requests.

Defect closure efficiency increased by 26% and 7% respectively from Gen 1 to Gen 2 and from Gen 2 to Gen 3. There are likely three dynamics here: having requirements in a searchable database, better requirements quality and the continuous build process. Determining the source of the defect (e.g., requirement, code, or test case) was facilitated by these good requirements engineering practices.

Another important quality measure improvement was time from project commit to product delivery. From Gen 1 to Gen 3, a total of 207 days were removed from the schedule. Several factors probably influenced these numbers including the quality of the requirements, the requirements database, the merge of the code bases and the continuous build process. For a development team size of over a 100 people, the cost savings are measured in the millions of dollars.

While more subjective, customer satisfaction levels improved from Gen 1 to Gen 3. Scores for Gen 1 software averaged between the “fair” to “good” range. For both the Gen 2 and Gen 3 products, they had moved into to the “good” to “very good” range. This increase in customer satisfaction levels indicates that the focus on requirements produced a software product that more closely met the customer needs and expectations.

An analysis of the data presented in this paper confirms that a well-written and well-reviewed set of requirements is a major factor in overall software quality. They decrease the total number of software defects, minimize rework, reduce wasted effort, improve schedule predictability, and increase team velocity and efficiency. Since test and development teams are spending less time identifying and correcting defects in the code, they can focus more time on productive tasks such as adding functionality and being more responsive to changing customer needs.

### VIII. FUTURE WORK

A fourth generation product is currently in development. The number and complexity of the new features continues to grow and almost none of the existing Gen 3 functionality will be removed. It will run on both desktop and mobile platforms (including the new Ultrabook™ products) incorporating a more advanced Intel processor. Software development methodologies remain essentially unchanged, although the team is currently in the process of investigating Scrum.

Similar to its predecessors, this project will leverage the final set of requirements from the previous software (Gen 3). The requirements engineering process is basically

unchanged: the requirements SME is still assigned to the team, a requirements management tool is being utilized and the PRD will undergo frequent and comprehensive reviews. The largest variation in methods for this project is in the number of contributors to the PRD. The Gen 3 product had five primary authors, while this product already has over twenty authors. These additional authors are scattered across different sites and countries, making training and mentoring more complicated. To date, the length of the PRD has already tripled. How these requirements changes influence the overall software quality will provide good data for a subsequent paper.

### REFERENCES

- [1] J. Terzakis, “The impact of a requirements specification on software quality and other quality indicators”, 19th IEEE International Requirements Engineering Conference (RE '11), 2011, Trento, Italy.
- [2] J. Terzakis, “Requirements defect density reduction using mentoring to supplement training”, Seventh International Multi-Conference on Computing in the Global Information Technology (ICCGI 2012), 2012, Venice, Italy.
- [3] T. Gilb, *Competitive Engineering: A Handbook For Systems Engineering, Requirements Engineering, and Software Engineering Using Planguage*, Butterworth-Heinemann, June 25, 2005.
- [4] S. V and G. Nair T.R., “Defect management strategies in software development” from *Recent Advances in Technologies*, edited by M. Strangio, InTech, 2009
- [5] T. King and J. Marasco, “What is the cost of a requirement error?”, StickyMinds.com, <http://www.stickyminds.com/sitewide.asp?Function=edetail&ObjectType=ART&ObjectId=12529&tth=DYN&tt=siteemail&iDyn=2>
- [6] G. Tassej, “The economic impacts of inadequate infrastructure for software testing”, prepared by RTI (project Number: 7007.011), 2002
- [7] “CHAOS summary 2009”, The Standish Group, 2009.
- [8] B. Boehm and V. Basil, “Software defect reduction top 10 list”, IEEE Computer, January 2001
- [9] D. Reifer, “Profiles of level 5 CMMI organizations.” *Crosstalk: The Journal of Defense Software Engineering*, January 2007
- [10] C. Jones, *Software Quality: Analysis and Guidelines for Success*, International Thomson Computer Press, June 14, 2000.



# A Review of Domain-Specific Modelling and Software Testing

Teemu Kanstrén

Security and Testing Technologies  
VTT, Oulu, Finland  
teemu.kanstren@vtt.fi

**Abstract**—Domain-specific modeling is an approach of using customized, domain-specific languages tailored for the domain as a basis for modeling the target system. The intent is to provide a means for domain experts to work with tools and a language closer to their domain knowledge, while abstracting away excess detail. This should provide more effective communication and ease the work done by providing a higher abstraction level. In the test automation domain, this means providing the domain experts with means to effectively create test cases based on their domain knowledge, and to communicate with the test automation experts. Despite the potential benefits and its applications, this viewpoint domain-specific modeling has received little consideration so far in test automation research. This paper reviews different approaches to applying concepts from domain specific modeling to test automation to provide a basis for further work in the area.

**Keywords**—domain-specific modelling; software testing; test automation

## I. INTRODUCTION

Testing is generally considered to be one of the biggest cost factors in software development. The testing process requires collaboration between several stakeholders, large investments in test infrastructure and continuous efforts in maintenance and evolution. The test infrastructure needs to be built to be able to address verifying both low-level details and high-level requirements. Domain experts need to be able to effectively communicate with the testers to ensure what needs to be implemented is implemented and is implemented correctly. Optimally, this means test automation needs to be built in layers to enable test engineers and software developers to verify the low level details, while providing domain experts the means to work and understand what is implemented and verify it, while working together to improve the resulting product.

When discussing concepts at a local level, where only a single team at a single organization is involved, having a common understanding becomes quite naturally. The people can sit down at common face-to-face meetings and quickly reach a common understanding. When teams become geographically and organizationally distributed, more difficulties arise. Different backgrounds and limited communications contribute to long delays in reaching a common basis for discussion between the different parties.

This applies to all works, not just software testing. Reaching a common understanding and maintaining that understanding requires that people can communicate using a shared terminology. Agreeing such a domain terminology is

an obvious requirement for applying domain specific modeling and creating domain specific languages. A less obvious requirement is the need to first agree on what the different parties mean when they talk about domain specific languages and modeling in general.

A domain-specific model (DSM) is expressed in terms of a domain-specific language (DSL). These languages can be hugely diverse and take completely different representations, typically with “domain specific” referring to the language being specific to a company and its application(s), each language being highly customized to a specific purpose [1]. A domain specific language works best in a domain that has a lot of variation that can be expressed by the language, leading to possibilities for cost-effective application vs. the initial language design costs. In software testing, test cases describe the behavior of a system, through a common base language while each test case can be seen as a variant expressed over that language. This makes testing a great domain for application of DSM.

Domain specific languages in the context of test automation can take different forms. In our experience, some people prefer shell scripts as their domain-specific languages. Others prefer to create their own textual scripting languages, such as keywords over test frameworks. Some prefer to create graphical modeling languages, such as those presented in [2] and [3]. This can be influenced by different factors such as the expert background, test requirements, and the target domain. Often there is no clear understanding of the involvement of a domain-specific language design process, which leads to less optimal results.

In our experience the language choice and design is heavily influenced by the people’s background. Someone with a strong background in Unix scripting wants to write everything in shell scripts (e.g., termed previously as “little languages” [4]). Someone with a strong background in graphical modeling languages will only consider those when talking about domain-specific languages. While all these different factors contribute to what is a suitable solution for test modeling in different contexts, a better understanding of domain specific language concepts in the testing domain, and knowledge of different options provide a basis for making more informed decisions. This paper contributes to this basis by reviewing current work and approaches for domain-specific modeling and test automation.

The following section II presents different types of domain specific languages we have observed in our work on building and applying test automation systems. Section III

presents examples of these using an example of a calendar application. Finally, conclusions end the paper.

## II. LANGUAGES FOR TEST AUTOMATION

Various tools exist specifically intended for designing domain-specific languages. In this paper the MetaEdit+ [5] is mainly used to illustrate the concepts but various others are also available. These tools can provide good mechanisms for defining language concepts and transforming these into different types of artefacts (e.g., test scripts). The tools intended to create these languages are in general not intended to build languages with features required to express all the low-level details of the systems in the domain-specific models. They work best when the transformations for them can be written to target higher level abstractions.

For these reasons, it is typically a useful approach to build the support for the domain-specific test languages in different layers. **Figure 1** illustrates these different layers. Test frameworks (TF) are in essence programs written using general purpose programming languages (GPPL) such as Java and Python. General purpose programming languages allow for freely expressing different computational concepts, providing good support and existing libraries for writing a test platform that allow one to express the required test concepts at different levels. The test framework takes care of connecting to the actual system under test and executing the concrete test cases.

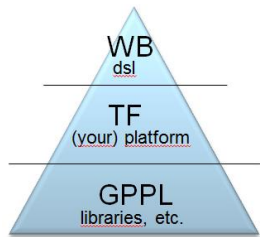


Figure 1. DSL layers.

On top of this test framework, the higher level representations and test languages can be created. In the terminology of domain-specific languages the tools used to create these languages are often called DSL workbenches (WB). Depending on the tools used and the type of test language targeted, this top layer can also be integrated with the test framework layers in a form such as a keyword driven test framework, where the keywords form the test DSL.

It is our experience that a domain specific test solution is often best build from bottom up. That is, having a good test platform available first to create and execute test cases, and using this to create the required support for testing in general. Once this support is in place it makes sense to start designing and providing the higher level DSL support on top of this existing platform. This also lends itself well to support cost-effective decision making when the extension of the support towards the top layers can be made when requirements and needs are identified.

### A. Common Language Elements

While there are different approaches to building domain specific test languages, and using those to model test cases,

these share a number of common language elements. The target domain needs to be expressed in terms of the domain terminology and as such a set of domain objects needs to be defined for the language. In the test automation domain, we are typically interested in expressing the various ways that the different actors involved in the system behavior can interact, and how the results of these actions should be considered (correct or failed).

Testing can be seen to represent a number of different concepts. However, in general testing is about exercising the different relevant aspects of system behavior in different ways and evaluating the results. A basic element of the test models is then the ability to explore the flow of execution in the system. In some cases, such as textual scripting languages this can take the form of implicit expression through the ordering of the script elements. In graphical notations the tools may allow one to connect the different elements as best seen fit. In test generation modeling languages this may take the form of expressing constraints over the possible combinations of the different test elements. The following subsections will discuss each of these types of test expressions.

### B. Scripting Languages

A typical approach to test automation is to have a scripting language that the user can use to write regression test cases. These scripting languages can take different forms and abstraction levels.

An example of a low-level test scripting language is the TTCN-3 (testing and test control notation version 3 [6]), which is a scripting language intended for testing communication systems. While it is a low-level language requiring a lot of effort and expertise in its use to write test cases, it is something designed for testing of systems in a particular (communications) domain. For example, it has specific support for features such as ports and messages. The benefit of this type of a standardized industry domain language is also the ability to exchange information between different partners in an executable format, in a well-defined and formalized terminology.

Examples of higher level scripting languages are those defining a set of keywords for writing test cases. This is supported by keyword-driven test frameworks such as Robot Framework [7]. These keyword-based languages provide textual domain-specific testing language. Examples of their application include telecommunication systems [8] and enterprise systems [9]. These are very domain-specific and different in each case.

### C. Graphical Languages

As noted, in the domain-specific modeling community, various approaches exist for creating the domain-specific languages. Besides using general tools to build textual scripting languages, there are also tools specifically intended for creating graphical modeling languages. These are typically used to build a graphical notation on top of a framework that is implemented using a general purpose programming language, or a lower level scripting language. For example, in the test automation domain, we can build a

graphical modeling language on top of a specific test scripting or keyword-based language.

For example, a graphical test modeling language built on top of TTCN-3 was presented in [2]. Highlighting the typical benefits to application of domain-specific modeling, significant benefits were reported in allowing a domain expert to create relevant test cases at a high level, communicate results and test intents with different stakeholders such as management, and in providing cost savings in focusing the variation modeling at a high-level on the most important aspects. At the same time, lower-level details could still be expressed in the underlying scripting language where required.

Other application examples of graphical domain-specific test languages include digital libraries and information systems [10]. Sometimes the distinction is also not so clear, for examples, with a focus on textual elements and formalisms with some graphical elements (an example of safety-critical systems in the railway domain) [11].

#### D. Model Based Testing

A related concept to domain-specific modeling in the software testing domain is model-based testing (MBT). In our experience, model-based testing for different people can be defined in many ways, such as using a mental model as a basis to manually write test cases, or using test stubs to model the system environment. However, a commonly used definition that we use here is from [12] as “generation of test cases with oracles from a behavioral model”.

The models in model-based testing are typically different forms of state-machines, defining the potential test steps of interest and their possible combinations. The models in MBT are traditionally not considered from a viewpoint of a domain-specific language as they are hand-crafted model-programs (term used, e.g. in [13]) used to generate test cases, not to manually model test cases. However, as shown in [14], the act of modeling the potential test steps and their possible orderings also provides a basis for the definition of domain-specific test modeling language. That is, the potential test steps in the model program define the model elements (along with the state variables of the model), and the guard statements defining the possible ordering of the test steps for the generator provide a definition of the possible execution flows that can be created from these elements.

Model-based testing approaches by their nature lead to creating test models for specific domains. Examples of these application domains include smartphones user interface testing [15], automotive systems [16], and healthcare systems [17]. These approaches typically report good results when applied in a suitable context (i.e., choice of right abstraction level, addressing of high variability, and so on). However, only a few works discuss these in relation to domain-specific modeling concepts (for smartphones in [15] and generally in [14]). As widespread adoption of model-based testing has long been an elusive goal in practice, providing more synergies in this area to make it more approachable to domain experts while also making it more cost-effective to tie into other different testing techniques holds a lot of potential.

### III. CALENDAR EXAMPLE

This section illustrates the principles discussed above with an example of applying them on a calendar example. This example is available online on the OSMO Tester MBT tool website [18]. The calendar is an example application where the user can create meetings and invite other people to those meetings. The user can also create tasks that are only visible to himself. Several users each have their own calendar instance.

The following subsections show an example of defining a domain-specific modeling language for this application. It starts with examples presented using a graphical notation built with the MetaEdit+ domain-specific modeling workbench [5], and proceeds to show different ways to create the underlying implementations of the test frameworks including the use of a keyword-driven framework and a model-based test tool. Possibilities for combining these different options are also discussed.

#### A. Terminology

Any process of applying DSM needs to begin with a definition of a common terminology. Sometimes this can be the biggest step in getting started and producing a useful and accepted solution. One might expect this to be simple for a calendar application, which is a widely used tool and concept. Yet domain-specific models are commonly defined for internal use at a company, where over time custom terms will have been adopted for effective communication between workers.

Here we use the calendar as an example for the readers of this paper, who can only be assumed to have a varying background. In a global context, many people will not be native English speakers, while the language commonly used to communicate in this context is English. Thus different mappings from the organization and personal language have an effect on how to approach building the basic blocks of a DSL. This makes it much more difficult to stay in line with the target audience and intended use of the language, as the terminology should be generally understandable and not just for the (paper) author(s).

A calendar is a very general entity and the base functionality of a calendar application as discussed in this paper is to add and remove events. As this example was originally devised it was influenced by the background of the author(s), which led to simply using the names “event” and “task” for the calendar entries involving several or just one person respectively. Yet, an event is an overloaded term in the English language, and using it in this way easily leads to confusion on what type of an event it is. Thus “event” was later renamed “meeting”, while the underlying platform(s) use the terms varyingly. This simple example shows how the language design needs to consider many factors.

Besides calendar entries, we also need to fix the terminology related to the actors using the calendars. The users are people, who have certain roles in the system (organizers, participants), and perform certain actions on the different elements (create, remove, invite).

**B. Defining the Language Elements**

From the terminology, we can already pick a set of language elements as a starting point. We need to be able to model the properties of calendar users (people), meetings and tasks. We also need to be able to model actions of creating and removing meetings and tasks, as well as inviting people to the meetings. We start with these elements and the basic sequential test flow. **Figure 2** shows an example of a test case where a user named “bob” creates and removes a task for the date first of January, 2012.

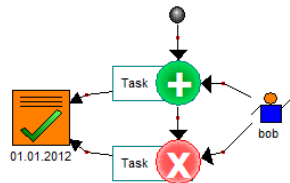


Figure 2. Two-step test flow.

Figure 3 shows the same scenario but with error notification where a second user (“john”) has been added and a test case is created where he tries to delete a task created by “bob”. However, as he is not in the “owner” role for the referenced task, he is not allowed to delete it. This also illustrates a design choice that needs to be made in considering a test language in general. In this case the choice has been made to allow using the language to model only the “correct” behavior of the system, and notifying when errors in the test models are observed by the modeling tool.

In other cases, it can be meaningful to allow for also the creation of test cases that exercise invalid flows of operation on the target system to test error handling behavior. This, however, requires creating different language elements as a different type of a test oracle needs to be bound to the test flow in this case (to define the correct expected response from the system). In our experience, the use of test cases and the creation of the language also serves a very useful purpose in facilitating communication between the different parties working on the system, where defining what is allowed explicitly also helps communicate the different expectations over the system behavior.

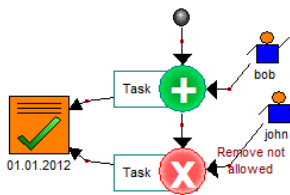


Figure 3. Embedded error checking.

**Figure 4** shows examples of some of the other elements in our test language. The meetings are represented by the people shapes in three different colors (red, green, blue). Actions for working with the meetings are shown similar to how the user would interact with the task objects. The action of inviting people to meetings is not explicitly visible here, but is shown by the different types of lines used to connect the people to the meeting object(s). The organizer is connected by a solid line (the main actor on the top of the sequence flow line), whereas the participants are connected

with a dashed line along the sequence path. Several concurrent flows of users working on their calendars are shown as parallel flows.

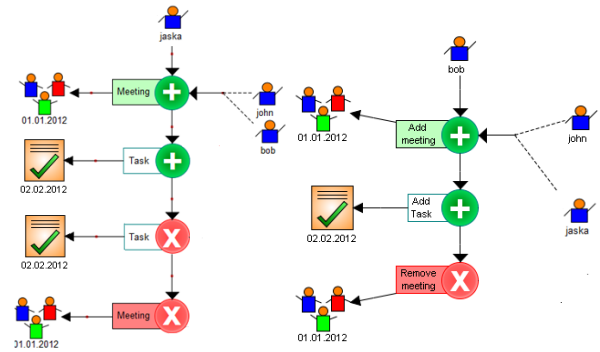


Figure 4. Parallel test flows.

**Figure 5** shows an example of defining a set of model building blocks and using those as templates for building test models for the calendar application. In this snippet, we have the actions for creating and removing tasks and meetings shown. Although not shown here, similar approaches can be used to model the calendar users as well, by creating templates for the person objects and allowing copying and modifying these as required.

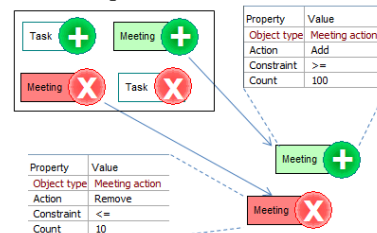


Figure 5. Model building blocks.

Finally, **Figure 6** shows an example of creating a test generator configuration for a MBT tool using these same model elements. In this case, we have configured the test generator to produce test cases where four different people are involved in different roles for performing actions on creating and removing meetings.

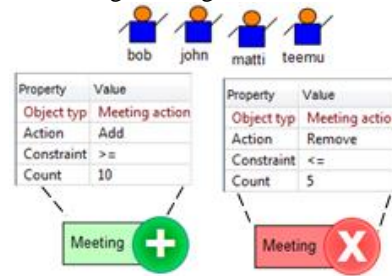


Figure 6. Generator configuration.

**C. Scoping the Language**

To scope the calendar test language, we need to consider who it is intended for and what the users (domain experts) are intended to use it for. In this case, the main purpose of the language is to allow the users (who are expected to be familiar with calendar concepts) to use it to model their basic interactions and functionality of the calendar.

They do not need to be able to model boundary conditions over the possible characters and strings used to express names, dates and other variables for the model elements. These types of low-level details are best handled by test experts who have more direct access to the low-level test platform functionality. The domain experts in this case just need an easy and effective way to express the different objects, fill in valid values and compose them to express their ideas of how the calendar should work, and how the different elements can relate to each other. The test cases generated from their test models and executed against the target system can then be used to validate how well these assumptions hold.

Thus the scope of the language in this case is defined to be exactly what is shown in the example figures in the previous subsection. There is no need to create any more complex properties or model hierarchies for the elements to achieve the set goals.

#### D. Test Platforms

The test platform for the calendar as described here and provided at [18] is based on different layers as discussed before. The bottom layer is based on a general purpose programming language. It is used to create a basis for the keyword driven layer, which is based on the Robot Framework (RF) test platform. This layer already allows writing test cases using a keyword based language for the calendar application, as shown in **Figure 7**. It allows one to manually compose test cases using such terms as “Add Event” and “Remove Event”. However, composing all these elements together manually is still error prone and not an intuitive approach that a domain-expert with no programming background is typically interested to use.

Test Case	Action	Argument	Argument	Argument	Argument	Argument	Argument
Test1	\$(Event1)= Add Event	\$(bob)	18.08.2007 at 21:03:10 EDT	19.08.2007 at 00:57:28 EDT	event1	location1	
	Remove Event	\$(bob)	\$(Event1)				
	\$(Task1)= Add Task	\$(alice)	27.04.2002 at 02:48:35 EDT	task1			
	Remove Task	\$(alice)	\$(Task1)				
	\$(Task2)= Add Task	\$(alice)	26.06.2005 at 04:16:20 EDT	task2			

Figure 7. Calendar script in RF.

A snippet of a test model for the calendar as implemented on top of the OSMO Tester MBT tool is shown in **Figure 8**.

```

@Guard("RemoveTask")
public boolean allowRemove() {
    return tasks.size() > 0;
}
@TestStep("RemoveTask")
public void doRemove() {
    ModelTask task = tasks.next();
    state.remove(task);
    scripter.removeTask(task);
}
    
```

Figure 8. OSMO MBT model snippet.

This shows the part for generating a “Remove Task” test step, basically stating that this step is allowed when some tasks exist to be removed, and when it is taken, an existing

task is chosen and removed both from the model state and the system under test state (through the scripter). As described in [14], the names given to these test steps in this type of a test model typically represent domain concepts in domain terminology, and as such provide a basis for a domain-specific test language. In this case, the scripter also generates scripts for the robot framework similar to those shown in **Figure 7**.

The final layer to provide on top of this is the graphical language described in the previous subsections. It can make use of both the model-based testing tool and directly the keyword based language, according to what is available and what is preferred. The main point to take away is that these different layers can be created as required and as seen cost-effective and useful. Different experts with different backgrounds can then use the different layers as they best see fit for their purposes.

#### IV. DISCUSSION

In general, it is our experience that it typically makes sense to create test languages at different abstraction levels using the different techniques described here when best seen useful. In some cases, it may be enough to just stay at the lowest level and only write unit tests using a general purpose programming language, augmented with some manual testing at the highest level. For example, this is the approach applied with the OSMO Tester MBT tool also mentioned before. It is mainly tested with extensive unit tests and by building a set of test models and test cases manually on top of it. This is possible when domain experts are also technical experts and comfortable with programming tools and techniques. However, this is different in large organizations and with a large number of different stakeholders (managers, customers, domain experts,...) who have a close interest to see and understand and work with the test artifacts.

In some cases, it can be enough to create a simple keyword driven test language to test the applications when it allows expressing all the test cases needed in sufficient detail. Finally, a model-based testing layer can be used on top of the keyword driven layer to provide variation in the generated test cases. However, creating a model-based testing layer or a domain-specific language layer rarely makes sense directly on top of low-level test support such as provided directly by general purpose programming languages (or even general purpose testing languages such as TTCN-3). These modeling tools generate test cases using specific transformations from a source model to a target model. These transformations are made much simpler when built on top of a higher abstraction such as that provided by a keyword driven test framework. By having a layer in between that allows generating test cases as a form or configuration for this (i.e., keyword combinations) makes it much simpler to generate these from the models, simplifies the required transformations, and makes for much better maintenance of the modeling infrastructure. Having a working middle layer (such as keywords based) also allows for writing manual test cases directly on this layer where desired.

For the final part of the domain-specific test languages, it can be built either on top of the model-based testing tools and their test models, or on its own. The model-based testing tools typically do not offer much added abstraction for the domain-specific tools but rather provide a lot of added power in expressing domain variance and generating higher test coverage automatically. When model-based testing tools are used with domain-specific models, it is useful to include them in the loop when possible in order to reduce the costs of maintaining and evolving several different models. That is, using the domain-specific workbench to generate a configuration for the model-based testing tool, which can then generate the actual test scripts from this configuration. This of course depends on having the required support available in the different tools for this type of functionality. Better understanding their relations in a domain-specific context as presented in this paper helps achieve these goals.

## V. CONCLUSIONS

This paper presented an overview of domain-specific modeling in the context of software testing. While various approaches for applying test automation in different domains exist, few explicitly consider these two concepts together and aim for most benefit. The overview presented here provides a basis for building better support for making these concepts benefit from each other.

The graphical modeling language definition for the calendar example is available as a MetaEdit+ project on the OSMO Tester MBT tool website [18]. The test platform code can also be accessed as part of the OSMO Tester MBT tool examples code repository at [18].

Overall, although different definitions and approaches are presented here for both domain-specific modeling and test automation, as noted, many different interpretations for these terms exist, and any definition should suffice, as long as it helps the people involved perform the task at hand and the stakeholders are able share the common definition(s). In our experience, being more explicit in the domain terminology and building the test frameworks based on this helps achieve this.

In the future we look forward to more extensive case studies on applying different DSM approach to testing, and how it affects and benefits the different stakeholders.

## VI. REFERENCES

- [1] S. Kelly, and J.-P. Tolvanen, "Domain-Specific Modeling: Enabling Full Code Generation", Wiley-Blackwell, 2008.
- [2] O.-P. Puolitaival, T. Kanstrén, V.-M. Rytty, and A. Saarela, "Utilizing Domain-Specific Modelling for Software Testing," in 3rd International Conference on Advances in System Testing and Validation Lifecycle (VALID 2011), 2011, pp. 115-120.
- [3] T. Kanstrén, O.-P. Puolitaival, V.-M. Rytty, A. Saarela, and J. Keränen, "Experiences in Setting up Domain-Specific Model-Based Testing," in IEEE International Conference on Industrial Technology (ICIT 2012), 2012, pp. 319-324.
- [4] J. Bentley, "Programming Pearls: Little Languages," Communications of the ACM, vol. 29, no. 8, 1986, pp. 711-721.
- [5] MetaCase, "MetaEdit+ Domain-Specific Modeling (DSM) environment," MetaCase, [Online]. Available: <http://www.metacase.com/products.html>. [Accessed January 2013].
- [6] ETSI, "Testing and Test Control Notation version 3," European Telecommunication Institute (ETSI), [Online]. Available: <http://www.ttcn-3.org/>. [Accessed January 2013].
- [7] NSN, "Robot Framework - A generic Test Automation Framework," [Online]. Available: <http://code.google.com/p/robotframework/>. [Accessed January 2013].
- [8] S. Stresnjak, and Z. Hocenski, "Usage of Robot Framework in Automation of Functional Test Regression," in The 6th International Conference on Software Engineering Advances (ICSEA 2011), 2011.
- [9] S. Wiczorek, and A. Stefanescu, "Improving Testing of Enterprise Systems by Model-Based Testing on Graphical User Interfaces," in 17th IEEE International Conference and Workshops on Engineering of Computer Based Systems (ECBS 2010), 2010, pp. 352-357.
- [10] S. Bärish, "Domain-Specific Model-Driven Testing," Vieweg+Teubner Verlag, 2009.
- [11] J. Kloos, and R. Eschbach, "A Systematic Approach to Construct Compositional Behaviour Models for Network-structured Safety-critical Systems," Electronic Notes in Theoretical Computer Science, vol. 263, 2010, pp. 145-160.
- [12] M. Utting, and B. Legeard, "Practical Model-Based Testing: A Tools Approach", Morgan Kaufman, 2006.
- [13] W. Grieskamp, N. Kicillof, K. Stobie, and V. Braberman, "Model-Based Quality Assurance of Protocol Documentation: Tools and Methodology," Journal of Software Testing, Verification and Reliability, vol. 21, no. 1, 2011, pp. 55-71.
- [14] T. Kanstrén, and O.-P. Puolitaival, "Using Built-In Domain-Specific Modeling Support to Guide Model-Based Test Generation," in 7th Workshop on Model-Based Testing (MBT 2012), 2012, pp. 58-72.
- [15] A. Jääskeläinen, M. Katara, A. Kervinen, M. Maunumaa, T. Pääkkönen, T. Takala, and H. Virtanen, "Automatic GUI test generation for smart phone applications - an evaluation," in Proceedings of the Software Engineering in Practice track of the 31st International Conference on Software Engineering (ICSE 2009), 2009, pp. 112-122.
- [16] A. Pretschner, W. Prenninger, S. Wagner, C. Kühnel, M. Baumgartner, B. Sostawa, R. Zölch, and T. Stauner, "One Evaluation of Model-Based Testing and its Automation," in Proceedings of the 27th International Conference on Software Engineering (ICSE 2005), St. Louis, Missouri, USA, 2005, pp. 392-401.
- [17] M. Vieira, X. Song, G. Matos, S. Storck, R. Tanikella, and B. Hasling, "Applying Model-Based Testing to Healthcare Products: Preliminary Experiences," in Proceedings of the 30th International Conference on Software Engineering (ICSE 2008), Leipzig, Germany, 2008, pp. 669-672.
- [18] T. Kanstrén, "OSMO Tester Home Page," May 2012. [Online]. Available: <http://code.google.com/p/osmo>. [Accessed May 2012].

## Lack of Software Engineering Practices in the Development of Bioinformatics Software

Dhawal Verma, Jon Gesell, Harvey Siy, Mansour Zand

Department of Computer Science

University of Nebraska at Omaha

Omaha, Nebraska 68182

Email: {dverma,jgesell,hsiy,zand}@unomaha.edu

**Abstract**—Bioinformatics is a growing field in the software industry. However there is very little evidence that sound software engineering practices are being applied to bioinformatics software development. As bioinformatics is a merging of the disciplines of biology and computer science, it would appear very odd that this, particularly important aspect of the computer science field would be absent, however that is the case. This paper will attempt to go into the reasons for this, as well as propositions that others have put forward to remedy this issue. We finally propose an approach towards resolving the software / requirement engineering challenges by comparing four methodologies Agile, SSADM, UP and Domain Engineering, and select the best approach that can help resolve the software / requirement engineering issues while developing bioinformatics software.

**Keywords**—Requirements engineering; bioinformatics; agile; UP; SSADM; domain engineering

### I. INTRODUCTION

The field of bioinformatics is a relatively recent; however, rapidly growing field, and one that spans both computer science and biology. It is focused on making cutting-edge scientific discoveries through sophisticated analysis of biomolecular data such as DNA, RNA, and protein sequences. It is an interdisciplinary field in which computer technology and computer science techniques, including software, hardware and algorithms, are applied to solving problems arising in biology. The primary stakeholders are biologists rather than computer scientists. As such, it presents a unique situation for the field of software engineering, as it presents challenges and opportunities that are not typically present during the normal engineering process [1]. For instance, stakeholders may be more inclined to sacrifice program structure to get something that works. While the field itself is indeed very different from the typical software engineering situation, with generally much tighter restraints on budget and timetables, as well as less time allotted for verifying and testing, the end goal of creating accurate and reliable scientific software is no less critical since incorrect results would greatly compromise the validity of the discovery. Furthermore, developing an easily maintainable and functional, as well as well-documented piece of software is still ever-present. Just as in the more general field of computer science, the practices of requirements and software engineering should be introduced in the academic lives of

those involved. As this field is part of computer science, and is included in the computer science departments of many universities, requirements engineering would be assumed to be part of the curriculum that these students would enroll in; however, as Umarji, et al. [1] discuss, this is not the case at all. In their studies of the syllabi of over 50 universities and colleges across the United States, Umarji et al. found that, while the students were always well instructed in general computer science topics, such as “design and analysis of algorithms, databases and programming languages in all the bioinformatics programs, however little or no training given to these students on basic software engineering principles.” [1]. Part of this may stem from the fact that “only recently have studies on end-user programming and information activities in bioinformatics started to emerge; there is still a large gap in our understanding problems in bioinformatics software development” [2]. Their particular study into this lack of application of the software engineering discipline is particularly compelling, as it includes roughly 50% of persons from the computer science discipline, and 50% from the molecular biology and biochemistry disciplines, so that it shows the influences that both sides have had on each other in this process of discovery.

#### A. Problem Statement

Bioinformatics, as a field is almost unique in regards to how requirements engineering is concerned. As Umarji, et al. [3] point out, within this field, unlike most others, there is a relatively large percentage of the practitioners of the field (i.e., bioinformaticians) who are doing the programming themselves, and have been left to their own devices in terms of software development and documentation. While science itself, and the experimental model teaches that documentation of every detail is important, this would appear to stop when it comes to the development of the tools that these scientists would use, and documentation is very limited, if it exists at all. This would appear to match up with their findings that, when asked about where these scientists learned what they knew about the software development process, “84% of [126] respondents indicated that they had learned through self-teaching alone, or that self-teaching was one of their main modes of learning.” [1]. This statistic in itself shows a considerable void in the formal training of the scientists involved with bioinformatics

or computational biology, let alone the formal training in software engineering.

## II. LITERATURE SURVEY

The study from Umarji, et al. [1] clearly suggests a lack in implementation of software engineering and requirements engineering methodologies in the development of bioinformatics tools. It would fit in with their earlier findings in regards to the number of university programs where bioinformatics students who had taken software engineering or requirements engineering as part of their degree program, and this also would make sense given that “70% [of those same respondents] responded that they had taken some computer science courses. Ten percent of the respondents completed certification programs to gain proficiency in software development.” The fact that so few of these respondents have any formal training in requirements engineering is indeed a disquieting figure: the implications are that there is a lot of software that is developed by an individual, and that once that particular individual is gone, the maintenance and efficiency of that software, even just the simple understanding of its function, will also be lost. While, unlike in a traditional business setting, this loss may not be financial in nature, the loss still could have far-reaching repercussions on the advancement or understanding of a particular experiment. In addition, the majority of these same respondents also identified themselves as bioinformatics specialists with some programming ability (57%), as opposed to programmers first (35%). Finally, most of these respondents indicated that they worked on mid-sized projects of about 5-20 KLOC, and that there was a broad range of team sizes. In short, this is clearly a discipline that covers a very broad range of experience in terms of programming ability, but which shows a very inconsistent level of education in the merged field, and one which is primarily graduate-level. This would indicate that though they completed many years of advanced computer science courses, in addition to those in biology, one of the most fundamental of these graduate courses, software engineering, requirements engineering or architecture were never covered.

To address this, Umarji, et al. propose a curriculum that fits both the demands of the field, and which emphasizes proper software engineering disciplinary techniques. Instead of adding a new class or requirement, their approach spreads the importance of disciplined software engineering in current classes, avoiding adding time to an already lengthy training process. Chilana, et al. [2] also bring up the lack of standardization since bioinformatics is such a new field. As software engineering is in a similar state when compared to other engineering disciplines, this should be a problem that software engineers themselves are familiar with.

Another problem pointed out by Chilana, et al. is that while bioinformatics is a cross-disciplinary field, it is one in which the two disciplines do not even speak remotely the same language. For example, they discuss how the two groups in their study, the computer science-oriented and the

molecular biology and biochemistry-oriented, do not even approach the problems in the same manner: the computer science group “often used command-line interfaces and programming languages that they were comfortable with and did not find it challenging to locate any related technical information.” [2]. Meanwhile, the more bioinformatics-oriented group, developed applications which “were simple at the beginning to match their research purposes the participants in this category had primarily self-taught programming skills and often sought information ... or obtain additional help from colleagues in implementing a solution.”

Most bioinformaticians and computational biologists believe that good bioinformaticians build up their own toolbox, and are aware of and use existing tools to do more powerful work. The software development practices mostly surround the notion of “Don’t reinvent the wheel” which essentially refers to the use of existing frameworks and to take advantage of large existing projects like BioPython [4], which contains a lot of ready to go code for practically everything.

This would indicate that while those in the bioinformatics background are less likely to have a software engineering or requirements engineering course to teach them the necessity of requirements specification, software maintenance and development, they are more willing to branch out than those who were educated in a far more rigid, single discipline environment. An ideal solution would, therefore, include a cross-over: better training of those in bioinformatics with regards to requirements engineering, and a cross-discipline research course in the computer science field.

## III. SOFTWARE AND REQUIREMENTS ENGINEERING CHALLENGES

The term bioinformatics has a range of interpretations, but the core activities of bioinformatics are widely acknowledged: storage, organization, retrieval, and analysis of biological data. A bioinformatician works to provide services to the scientific community in the form of databases and analytical tools. The serious challenges facing bioinformatics over the next decade include integration and presentation of the enormous and ever expanding size data. In order to make use of the relatively weak signals present in a single data source, it is necessary to integrate data from different views of the same system. Of course, integrated data is still just a mountain of data. Researchers need tools that present the data in a comprehensible fashion, annotated with context, estimates of accuracy and explanation. Another challenge that bioinformaticians face while developing bioinformatics software is that in most cases the tool or collection of tools was written in an ad hoc manner to be used for an experiment. If the experiment is successful, this tool evolves into a large scale project and sometimes even considered to be commercialized. In such situations, little emphasis is paid on the organization and requirement gathering process in the early stages of the software. A bioinformatician gathers information from the biologists and takes an evolutionary approach towards building the software and in most cases the



approach is so raw in the sense that absolutely no importance is given to the software and requirement engineering process simply because the main objective is to create a tool/software “quickly”, that only a selective set of users who are involved in the research can use [5].

We put forward an approach to resolve the aforementioned challenges. We considered four different methodologies, Agile [6], SSADM (Structured Systems Analysis and Design Method) [7], UP (Unified Process) [8] and DE (Domain Engineering) [9]. Table I compares the four methodologies in terms of how they address the technical challenges of bioinformatics software engineering. After an exhaustive study of the pros and cons of each methodology for software development in the field of bioinformatics we came to the conclusion that an integrative methodology of UP, SSADM and DE would be the best candidate for software development in the field of bioinformatics. The following section explains the reasons why.

#### IV. EVALUATION OF METHODOLOGIES

##### A. Evaluation: Agile methodology

One of the fundamentals of agile methodology is user involvement. However, in our case, user involvement was expected to be low for the following reasons:

- Most of the projects are research oriented, the users are mainly the biologists of the research group and they are usually unavailable due to scheduling constraints and only play a small part in the knowledge transfer of the domain to the development team.
- The development team members are mostly themselves part of the research group, so they perceive (correctly or not) that they understand the system requirements and do not feel the critical need to involve the users.

Requirements although expected to change along with the changes in research objectives for this system, they are still stable for a phase due to the following reasons:

- There are not many stakeholders involved (only the research group).
- Existing work process is to be automated
- User stories are not that complex since most bioinformatics tools developed are for analytics and hence it is the algorithm that takes preference.
- Existing process/work-cycle can be demonstrated easily in short span of time thereby giving us enough insight into how the system should work and create plan. On an average, it takes bioinformaticians about 3-4 hours to obtain a small subset of results using theoretical methods/algorithms.

In practice, the lightweight nature of agile methods affords a lot of flexibility to the development process, but

makes agile methods difficult to implement in a disciplined manner without coaching [10]. An undisciplined application of agile methods leads to a “patch and go” attitude. Agile methods are commonly used in the development of scientific software (e.g., [11], [12]). However, the use of agile methods by programmers with no formal training in software engineering increases the likelihood of undisciplined application of the process, leading to lack of documentation, and code that is difficult to maintain or reuse. Moreover, the common quality assurance practice employed in agile processes, test-driven development, is very difficult to apply to scientific software, where it is not always clear what the expected output should be. In addition, the fact that the stakeholders, developers and users are generally all the same person does not help in this regard, as they are less likely to seek outside help, and to view documentation as a waste of time, as they are the only persons involved. For these reasons, agile is not a good fit for creating maintainable scientific software.

##### B. Evaluation: Unified Process

Unified Process (UP) [8] views software development as consisting of four iterative phases:

- 1) *Inception*: the need for the software is justified and cost-benefit and risk analyses are conducted.
- 2) *Elaboration*: this is where requirements are elaborated and the software architecture is created.
- 3) *Construction*: the design is further detailed and the software is incrementally built.
- 4) *Transition*: the developed software is delivered to the client.

UP provides an extensible framework to manage software development: one of its major features is that it considers software development to be an iterative and incremental process whereby each stage has a specific area of focus, which is just the kind of approach required in a research-oriented field of bioinformatics.

The iterative and incremental process divides the project into smaller chunk called increment and each increment refines the functionality of intended system by undergoing several iterations, following the complete process of traditional waterfall method [6]. In UP, the system development starts with little knowledge about the objective and as projects continues the knowledge about the system increases, which is exactly the kind of needs a research project calls for. While agile processes have a similar iterative development cycle, a key distinction is the discipline imposed by the heavyweight process associated with UP, particularly the explicit requirements engineering activities carried out during the elaboration phase.

Unified Process handles risk well throughout the development process. With deliverables at each stage of the development there is a go/no-go decision to be made analyzing each deliverables. [13]. This helps in identifying and analyzing the risk at early stage making it possible to decide the steps to be taken in course of time.

TABLE I. COMPARISON OF METHODOLOGIES

Methods	Challenges		
	<i>Explosion in Data Sources</i>	<i>User Centric Design</i>	<i>Other</i>
Agile	Lack of documentation makes it difficult to understand how code deals with data formats. Lack of upfront design modeling makes it harder to adapt to unanticipated formats.	Requires more customer participation time than users (biologists) can commit to.	Advantage in dealing with rapid change is not applicable due to relatively stable requirements.
UP		Well-documented incremental development makes it easier for users to add more feature requests over time.	Risk management process facilitates review at every stage of development.
SSADM	Strong data modeling emphasis fits well with managing diverse data needs.	Too rigid for incremental development.	
DE	Domain analysis provides understanding of the commonalities and variabilities of the various data formats.	Reduces dependency on domain expert in later phases of project or future similar projects.	Develops infrastructure, generic architecture and common "assets". Facilitates development of taxonomy and ontologies.

### C. Evaluation: SSADM

We consider SSADM [7] to be an important candidate for software development in the field of bioinformatics because of the three most important techniques used in SSADM:

- Logical data modeling: the process of identifying, modeling and documenting the data requirements of the system being designed; the data are separated into entities (data required to record information) and relationships (the associations between the entities).
- Data Flow Modeling: the process of identifying, modeling and documenting how data moves around a system; this process examines processes (activities that transform data from one form to another), data stores (the holding areas for data), external entities (what sends data into a system or receives data from a system), and data flows (routes by which data can flow).
- Entity Behavior Modeling: the process of identifying, modeling and documenting the events that affect each entity and the sequence in which these events occur.

Since dealing with huge amount of data is the major part of the problem in bioinformatics, the explicit data modeling techniques inherent in SSADM are suitable for the problems bioinformaticians deal with the most.

While the overall SSADM process is too rigid [14] for the incremental user-centered design approach that is ideal for development of bioinformatics software, there are certain aspects that are still applicable. Of the different stages involved, we select the ones that are essential for software development in the field of bioinformatics. These stages are:

- 1) *Feasibility study*: A feasibility study is effectively a condensed version of a fully blown systems analysis and design, to investigate the goals and implications of a research project before committing resources to it.
- 2) *Investigation of the current environment*: This is one of the most important stages in the software development methodology; although the new environment may be radically different, the concepts

underlying bioinformatics software will remain the same and hence it is critical to investigate some form of the current system.

- 3) *Requirement Specification*: This is probably the most complex stage in the methodology. Using the requirements developed in stage 2 and working within the framework of the selected module, the analyst must develop a full logical specification of what the new system must do.
- 4) *Technical system options*: This is the first stage towards a physical implementation of the system and a large number of options for the implementation are generated such as the hardware architecture, software to use, cost of implementation, staffing required, physical limitations and constraints such as a space occupied by the system, the distribution including any network which that may require and the overall format of the human computer interface.
- 5) *Logical design*: This level concentrates on the requirements for the human computer interface, which is something important when it comes to research projects having potential of being widely used, such as a Genome Browser or an interface for querying a disease database. The logical design specifies the main methods of interaction in terms of menu and command structures.
- 6) *Physical design*: This is the final stage where all the logical specifications of the system are converted to descriptions of the system in terms of real hardware and software.

### D. Evaluation: Domain Engineering

Domain engineering [9] provides a systematic process for analyzing a family of similar applications in order to produce a common extensible framework. It consists of the following activities:

- 1) *Domain analysis*: In this stage that is unique to domain engineering, the domain is studied to understand what the different applications have in common and how they vary. This also leads to the definition of the scope of the domain, i.e., which applications to consider part of the domain and which ones to exclude.

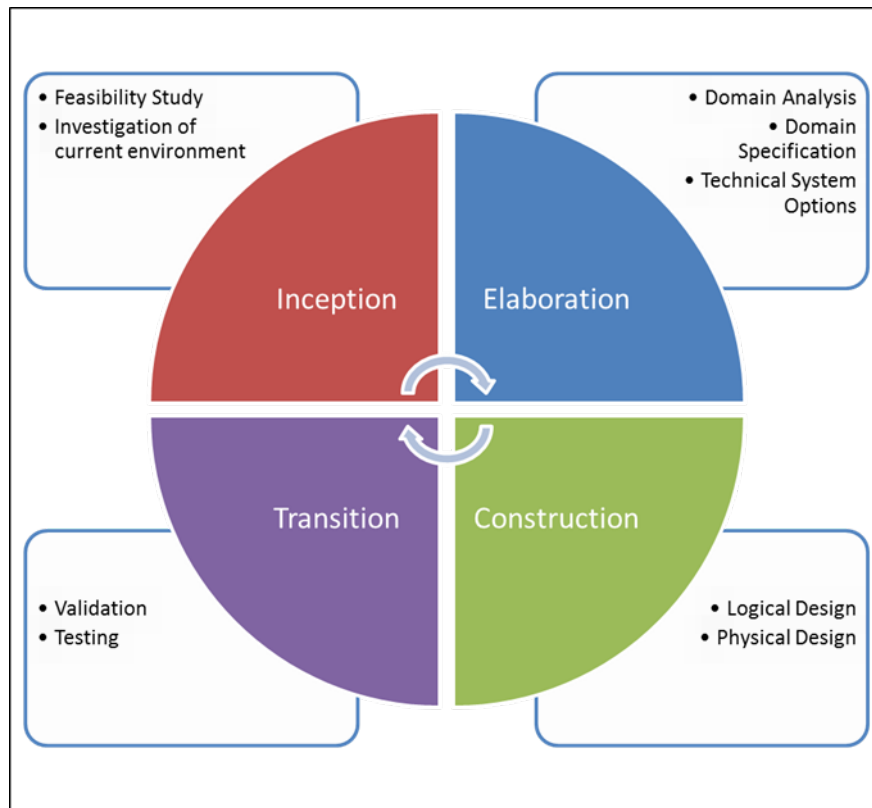


Fig. 1. An integrative approach towards a suitable Software/Requirement Engineering Methodology for Bioinformatics Projects

- 2) *Domain specification:* In this stage, the artifacts are refined, producing a domain-specific requirements document followed by a reference architecture that defines the common components and reusable assets as well as where application-specific components and modifications can be added.
- 3) *Domain implementation:* This stage uses existing software development processes to take the domain-specific requirements and architecture and create the framework and other core assets that can be used for developing additional applications.

The domain engineering process is well-suited to analyze and model the diverse data sources and organize the multiplicity of scripts that perform similar analyses on bioinformatics data. Domain analysis studies the commonalities and variabilities of the various data sources and formats, enabling developers to understand how they vary and anticipate additional variations.

The outcome of domain engineering activities facilitates further development of a product line of similar bioinformatics applications. Development of a product line tool or workbench makes it possible to easily develop add-ons to existing products. It makes every function and feature of the family of products available for use, extensions, and possible adaptation of developed/abstracted features. This can be

accomplished with the creation of bioinformatics-specific designer tools that simplify many of the programming tasks for bioinformaticians.

#### E. Integrative Approach

We present here an integrative software and requirement engineering approach for the domain of bioinformatics where applied research is the brainchild for development of software in this unique discipline. Research can be perceived differently depending on viewpoint, academic or industrial. In this paper, we define applied research as an activity that gains intellectual leadership that ultimately leads to commercial reward for a company by combining the classical viewpoints. We measure this in terms of the combination of intellectual leadership (knowledge) together with the ability to demonstrate new ideas through proof of concept prototypes.

From our survey, we observed that since a number of the methodologies used in software development today have their origins in the waterfall process and tend to separate activities into distinct phases of design, coding, testing and integration. These activities have been found to occur repeatedly on each cycle or iteration. Iterative design processes such as the Unified Process (UP) have become widely used over the past decade. Since, research is considered to be an incremental and iterative process; we chose UP

to be the backbone of our methodology and incorporated different stages of SSADM and domain engineering into the four stages of UP methodology. Figure 1 shows the integration of these methodologies. The feasibility study and investigation of current environment activities are adopted from SSADM to guide the inception phase, justifying the business case. During the elaboration phase, domain analysis and specification are adopted from DE to guide requirements analysis activities that can be shared across several related bioinformatics applications. A study of the technical systems options are used to understand the requirements from the systems engineering standpoint. In the construction phase, logical and physical design activities model the structural and behavioral aspects of the software. Finally, the transition phase includes system testing and validation.

## V. CONCLUSION AND FUTURE WORK

As was stated before, the lack of any formal training in requirements engineering principles has led to a major problem in the field of bioinformatics: many programs with no requirements specifications, no maintenance plans, and which, more often than not, have only a single user who will ever understand these issues. While general education in the field of requirements engineering and architecture would certainly assist in making sure that these principles are reinforced while in academia, the uniqueness of the field in practicum presents challenges that are not seen in most other areas. To this end, several techniques that are discussed in a typical requirements engineering class are on the table. The first of these, agile, is actually very common in the field itself, however, as it is practiced almost to the point of exclusion, it has contributed to the problem almost as much as the lack of education in the requirements engineering. Instead, the more practical option would be a combination of the Unified Process, SSADM or Domain Engineering methods that were discussed, as they would ensure a more robust architecture thru requirements specification, an actual architectural design, and a maintenance plan put in place to ensure that others actually know what the developed software is for and how it is to be used, beyond simply those using it in the here and now.

Future work includes empirical validation of the effectiveness of the integrated approach. Validation of any new process or methodology is always difficult, particularly during the formative years, partially due to the subjective nature of evaluation but primarily due to the limited experience in using the technique. As every research project is different, it is difficult to exactly compare productive gains; however many approaches for examining the credibility, reusability and the efficiency of performing research have been discussed at length [14], [15].

In addition, we will investigate aspects of other software development processes that can be used to further support informally trained software developers in building reliable systems.

## REFERENCES

- [1] M. Umarji, C. Seaman, A. G. Kuru, and H. Liu, "Software engineering education for bioinformatics," in *Proceedings of the Conference on Software Engineering Education and Training (CSEET 09)*, 2009, pp. 216–223.
- [2] P. K. Chilana, C. L. Palmer, and A. J. Ko, "Comparing bioinformatics software development by computer scientists and biologists: An exploratory study," in *Proceedings of the ICSE Workshop on Software Engineering for Computational Science and Engineering*, 2009, pp. 72–79.
- [3] M. Umarji, M. Pohl, C. Seaman, A. G. Kuru, and H. Liu, "Teaching software engineering to end-users," in *Proceedings of the 4th International Workshop on End-User Software Engineering*, 2008, pp. 40–42.
- [4] P. J. Cock, et al., "Biopython: freely available Python tools for computational molecular biology and bioinformatics," *Bioinformatics*, vol. 25, no. 11, pp. 1422–1423, 2009.
- [5] K. Pavelin, J. A. Cham, P. de Matos, C. Brooksbank, G. Cameron, and C. Steinbeck, "Bioinformatics meets user-centered design: A perspective," *PLOS Computational Biology*, vol. 8, no. 7, pp. 1–4, 2012.
- [6] I. Sommerville, *Software Engineering (9th Ed)*. Addison-Wesley, 2010.
- [7] J. S. Hares, *SSADM for the Advanced Practitioner*. John Wiley & Sons, Inc., 1990.
- [8] Rational Software, *The Rational Unified Process, version 5.0*, Cupertino, CA, 1998.
- [9] D. M. Weiss and R. Lau, *Software Product-Line Engineering: A Family-Based Software Development Process*. Addison-Wesley, 1999.
- [10] M. M. Muller and W. F. Tichy, "Case study: Extreme programming in a university environment," in *Proceedings of the International Conference on Software Engineering (ICSE 01)*, 2001, pp. 537–544.
- [11] O. Chirouze, D. Cleary, and G. G. Mitchell, "A software methodology for applied research: extreme researching," *Software: Practice and Experiences*, vol. 35, no. 15, pp. 1441–1454, 2005.
- [12] W. A. Wood and W. L. Kleb, "Exploring XP for scientific research," *IEEE Software*, vol. 20, no. 3, pp. 30–36, 2003.
- [13] B. W. Boehm, "A spiral model of software development and enhancement," *IEEE Computer*, vol. 21, no. 5, pp. 61–72, 1988.
- [14] B. Kitchenham, S. L. Pfleeger, D. C. Hoaglin, K. El Emam, and J. Rosenberg, "Preliminary guidelines for empirical research in software engineering," *IEEE Transactions on Software Engineering*, vol. 28, no. 8, pp. 721–734, 2002.
- [15] C. Robson, *Real World Research*. Backwell Publishers: Oxford, 2002.

## Pinpoint Analysis of Software Usability

Divya K. V. Dasari, Dan E. Tamir,  
Oleg V. Komogortsev, Gregory R. LaKowski  
Department of Computer Science  
Texas State University  
San Marcos, Texas USA  
{dd1290, dt19, ok11, gl1082}@txstate.edu

Carl J. Mueller  
Texas A&M University Central Texas  
Killeen, Texas, USA  
muellercj@ct.tamus.edu

**Abstract**—The effort-based model of usability aids in evaluating user interface (UI), development of usable software, and pinpointing software usability defects. In this context, the term pinpoint analysis refers to identifying and locating software usability issues and correlating these issues with the UI software code. In this paper, the underlying theory of the effort-based model along with pattern recognition techniques are used to produce a framework for pinpointing usability deficiencies in software via automatic classification of segments of video file containing eye tracking results. This allows developers to harness their effort and focus on excessive effort segments that need attention. To verify the results of the pattern recognition procedures, the video is manually classified into excessive and non-excessive segments and the results of automatic and manual classification are compared. The paper details the theory of effort-based pinpoint analysis and reports on experiments performed to evaluate the utility of this theory. Experiment results show more than 40% reduction in time for usability testing.

**Keywords**-Software Development; Software Usability; Human Computer Interaction; Pinpoint Analysis.

### I. INTRODUCTION

The *Effort-based model* of usability [1-4] aids in evaluating user interface, development of usable software, and pinpointing software usability defects. It is developed using the principle that the usability is an inverse function of effort. The model is used for comparison of different implementations of the same application. The results of several experiments conducted on the effort-based model show strong relationship between effort and usability [1-4].

The underlying theory of the effort-based Model is used to produce a framework to identify usability deficiencies in the software. Identifying and locating software usability issues and correlating these issues with UI software code is referred to as Pinpoint Analysis [3,4]. For example, users who are in a state of confusion, and users that are not sure how to use the software, tend to look around the screen to figure out the best way to accomplish a task. This behavior is referred to as an excessive effort [3,4]. Identifying and pinpointing excessive effort behavior helps UI designers rectify numerous usability related issues.

This research attempts to evaluate the utility of pinpointing user interface deficiencies using pattern

recognition techniques for identifying excessive effort in segments of software interaction session records. Segmentation of user's software interaction session is done using the time slice between two consecutive mouse/keyboard clicks. Automatic identification of segments showing excessive effort behavior helps the UI designers to reduce the time required for analysis and rearranging the interface at the pinpointed time snapshot.

The pattern recognition methods used in this work include feature selection, principal component analysis, K-means clustering, and threshold based classification [5-7]. Several experiments were conducted to evaluate the new framework for pinpointing software usability issues. Experiment results show more than 40% reduction in time for usability testing.

The rest of this paper is organized as follows. Section II contains background information. Section III summarizes the related work. Section IV details the experimental setup. Section V details the experiments. Section VI presents experiment results and Section VII contains results evaluation. Section VIII concludes the paper with a summary of our findings and proposals for further research.

### II. BACKGROUND

#### A. Software Usability

According to the International Organization for Standardization/International Electro Technical Commission (ISO/IEC) 9126 standard, software usability is: "The capability of a software product to be understood, learned, used, and be attractive to the user when used under specified conditions." There are several characteristics that play an important part in defining software usability: understandability, learnability, operability, and attractiveness [8,9].

Cognitive modeling involves creating a computational model to estimate how long it takes the users to perform a given task [10-13]. It involves one or more evaluators inspecting a user interface by going through a set of tasks by which understandability and ease of learning are evaluated. The user interface is often presented in the form of a paper mock-up or a working prototype; but, it might be a fully developed interface. Cognitive models are based on

psychological principles and experimental studies to determine times for cognitive processing and motor movements. They are used to improve user interfaces or predict problem areas during the design process.

### B. The Effort-based Usability Model

Several studies indicate that many system users associate the “physical” effort required for accomplishing tasks with the usability of the software [1-4]. The effort-based model for software usability stems from the notion that the usability is an inverse function of effort. For example, an eye tracking device is used to measure the effort expended by the user in navigating through the user interface of software. In the case of interactive computer tasks, it is possible to calculate effort as a linear combination or a weighted sum of metrics such as the number of mouse clicks, number of keyboard clicks, eye path traversed as well as other eye activity measures, and mouse path traversed.

Eye trackers acquire eye position data and enable classifying the data into several eye movement types useful for eye related effort assessment [11,12]. The main types of eye movements are: 1) *fixation* – eye movement that keeps an eye gaze stable with regard to a stationary target providing visual pictures with high acuity, 2) *saccade* – rapid eye movement from one fixation point to another, and 3) *pursuit* – stabilizes the retina with regard to a moving object of interest [1,11,12]. Usually, the Human Visual System (HVS) does not exhibit pursuits when dynamically moving targets are not a part of the interface [1,11,12].

In this research, the following metrics are used as a measure of the *physical* effort 1) Average fixation duration, 2) Average saccade amplitude, 3) Number of saccades, and 4) Average eye path traversed [1-4,11,12].

The effort-based software usability evaluation is divided into three phases: *Measurement, Analysis, and Assessment* [3,4]. In the *measurement phase*, a group of users executes a set of identical independent tasks, which emerge from a single scenario. These tasks differ in key parameters, which prevent the users from memorizing a sequence of interaction activities. Throughout the interaction process, certain user actions such as eye movement, time on task, keyboard activities, and mouse activities are logged.

The *analysis* phase involves accumulating data for several metrics such as the number of saccades, average saccade amplitude, number of fixations, average fixation duration, and average eye path traversed, that relate to user effort. Another metric is the time on task. The average task completion time is compared to a learning curve, which reflects users’ mastery of software.

The final step is the *assessment*. Using the above steps, the learnability of software systems is assessed and the point of users’ mastery of the software is identified. The same model is applied to obtain operability and understandability of various software systems or different groups of users

using the same system. The effort-based metrics provides interface designers with means to evaluate their designs [1].

### C. Pinpoint Analysis

Software usability testing is considered one of the most expensive, tedious, and least rewarding tests to implement [1,2]. This perception is likely to change if the usability testing is made less expensive and more rewarding. This requires accurate means through which an engineer can identify and pinpoint issues in the software or the interface. This process is called pinpoint analysis. Pinpoint analysis is one of two types; inter-pinpoint analysis deals with identifying issues with tasks performed by the users in a specific system, whereas intra-pinpoint analysis refers to identifying issues within tasks in a specific system. For example, outlier tasks might be identified through inter-pinpoint analysis and used for intra-pinpoint analysis. This analysis also helps graphical user interface (GUI) designers to make decisions about element placement on displays and determine the level of effort that is related to different widgets [3,4].

#### 1) Inter-pinpoint Analysis

Inter-pinpoint analysis involves detecting tasks that present anomalies and identifying the reasons for these anomalies at a high level. The mouse is used as an example to understand inter-pinpoint analysis. In a particular task, the right mouse button helps users complete a task effectively; however, some of the users are unaware of it. It is possible that anomalies like this can be identified in inter-pinpoint analysis [3,4].

Inter-pinpoint analysis helps identifying alternative methods to perform a task effectively with less effort; however, it does not provide users with a hint of the alternative method. Other issues like the necessity of help facilities in software can be identified by the high level analysis of tasks that present anomalies.

#### 2) Intra-pinpoint Analysis

A more detailed method for analyzing tasks and identifying specific issues with the software is intra-pinpoint analysis. Intra-pinpoint analysis can be done manually by watching all the video recordings of the users’ interactions with software, obtained from an eye tracking device. This review helps identifying interaction issues and areas where the user has difficulty while performing tasks. For example, the analysis might reveal that most of the users go into a state of confusion in a specific part of a task and are looking around the screen to identify the best way to proceed with the task. This might prompt the designers to rearrange the interface at the relevant time snapshot. Clearly, this option is tedious and potentially expensive. An alternative is to use a semi-automatic method applying pattern recognition technique. This method eliminates the need for a person to watch the entire video in order to identify interaction issues thereby cutting down the cost and time. It enables automatic identification of areas where the user has difficulty and

marking these areas for further evaluation. For this reason, we refer to the process as semi-automatic.

#### D. Pattern Recognition

One of the applications of pattern recognition is the assignment of *labels* to a given input value, or *instance*, according to a specific algorithm. An example of pattern recognition is classification, which attempts to assign each input value to one of a given set of classes. Pattern recognition is generally categorized according to the type of learning procedure used to generate the output value. *Supervised learning* assumes that a set of training data (the training set), consisting of a set of instances that have been properly labeled by hand with the correct output, has been provided. Next, a learning procedure generates a model that attempts to meet two sometimes conflicting objectives: Perform as well as possible on the training data, and generalize as well as possible to new data. On the other hand, *unsupervised learning* assumes the availability of training data that has not been hand-labeled and attempts to find inherent patterns that are used to determine the correct classification value for new data instances [5-7].

### III. LITERATURE REVIEW

Usability is a highly researched topic with much literature available [8-15]. Nevertheless, extensive review did not reveal any research papers related to pinpointing usability issues. There are some papers on effort-based usability evaluation that are discussed below.

Tamir et al. [2] conclude that effort and usability are related but they did not address pinpointing issues. Mueller et al. [16] use effort metrics to evaluate software usability. Their method allows comparison of two or more implementations of the same application, but does not identify where exactly the problem lies. Hvannberg et al. [17] describe the design and test of a defect classification scheme that extracts information from usability problems, but is limited since it does not define the causes underlying usability problems. Nakamichi et al. [18] investigate the relations between quantitative data, viewing behavior of users, and web usability evaluation by subjects. They conclude that the moving speed of the gazing points is effective in detecting low usability. Makoto et al. [19] use a Web-Tracer to evaluate web usability. Web-Tracer is an integrated environment for web usability testing that collects the operation log of users on the Web pages. The data collected is used to determine the usability of the Web pages. However, the reasons for low usability are not identified using this approach. This paper thoroughly addresses and resolves all of the issues listed above.

### IV. EXPERIMENTAL SETUP

#### A. Manual Input Devices

The subject performs the tasks on a computer using a standard keyboard and a mouse as input devices. An event driven logging program is used to obtain details of mouse

and keystroke activities from the operating system event queue. The program saves each event along with a time stamp into a file. The logged events are: mickeys (mouse pixels), keystrokes, mouse button clicks, mouse wheel rolling, and mouse wheel clicks.

The eye tracker used for the experiments is Tobii X120 Eye Tracker, version 2.2.5 [20]. The Tobii device is a standalone eye tracking unit designed for eye tracking studies. It measures unfiltered and spontaneous human reactions, responses along with gaze and other real-time data. The data collected by the eye tracker is logged to a file, which is referred to as a *data file* in this paper. The eye tracker also records video version of the user interaction session and is referred to as a video file, which is very helpful in verifying experiment results.

#### B. Software Environment for Analysis

A software program developed in MATLAB is used to perform data analysis of the experiments performed in this paper [21]. The parameters of the program are detailed in the respective sections.

#### C. Test Procedure

Experiments conducted to evaluate the capability of pattern recognition techniques to identify software usability issues are done using the steps depicted in Fig. 1.

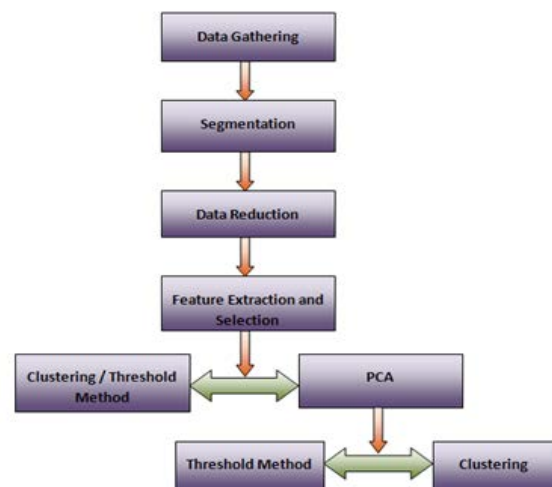


Figure 1. Experiment Procedure.

As the figure shows, the main steps are: data gathering, segmentation, data reduction, feature extraction and selection. These actions are followed by several different classification techniques. The sequence of actions depicted in the figure is further described in three subsections: data gathering, data reduction and identification of excessive effort segments.

### 1) Data Gathering

A group of five users executes a set of seven identical independent tasks, which emerge from a single scenario. Throughout the interaction process, certain user activities such as eye movement, time on task, keyboard, and mouse activities are logged using an eye tracking device. According to the learnability-based usability model, the point at which the user's effort reaches the acceptable level is called the learning point. Based on this model, it is assumed that the users' effort reaches the acceptable level by the time they perform task 5. Hence, in this paper, task 5 of each subject is used for conducting experiments.

### 2) Data Reduction

Phase 2 includes activities such as segmentation, data reduction, and feature extraction. The data logged throughout the user interaction session, i.e., the data file is used for event based segmentation where the events are consecutive keyboard/mouse clicks. Metrics such as: (a) segment duration (for event based segmentation), (b) the average fixation duration, (c) the average saccade amplitude, (d) the number of fixations, (e) the number of saccades, (f) the standard deviation of the fixation duration, (g) the standard deviation of the saccade amplitude, and (h) the eye path traversed are inferred for each segment. These metrics are used to generate a feature set, which is obtained by applying data reduction programs to the data file. The features data is calculated for all features within each segment and this data is useful to identify excessive effort segments.

### 3) Identification of Excessive Effort Segments

Pattern recognition techniques are applied to the feature set obtained from the data reduction process to identify segments that exhibit excessive effort. The techniques used and applied on the feature set are briefly explained below.

**Thresholding** - a threshold value is calculated for each feature in the feature set. For a given feature, all the segments that have a feature value that is less than the threshold value are classified as non-excessive segments and vice-versa.

**K-means clustering** - the segments are grouped into clusters. Based on the value of cluster centers, the cluster is classified as excessive or non-excessive. All segments that fall in the excessive cluster are segments exhibiting excessive effort behavior and vice versa.

**Principle component analysis (PCA)** - the first, the second, and the third principal components are obtained for the feature data. The threshold classification is applied on the first principal component and K-means clustering is applied on the first, second, and third components to classify the segments into excessive or non-excessive.

By the end of phase 3, the excessive effort segments are identified by the software program. To verify the results, the video file is carefully watched segment by segment and classified into excessive or non-excessive segments manually. The manual classification process of the video file is described in the following section.

### D. Manual Classification

The manual classification process involves event based segmentation on the entire video file. Each segment is carefully watched and classified into the following categories:

**Idle behavior segments;** idle behavior is due to system response. Waiting for a progress bar to complete; or waiting for a page to load are examples of idle behavior. Segments with such behavior are classified as idle behavior segments.

**Excessive effort segments;** segments without any useful user actions are classified as excessive effort segments. A subject looking at different components on an interface instead of the actual target component, which help in accomplishing the task is an example for excessive effort behavior. Such behavior can be eliminated without sacrificing task completion quality.

**Non-Excessive effort segments;** segments with useful action that result in task completion are classified as non-excessive segments.

**Off screen behavior segments:** Intervals of time where the subject's view is not within the screen for more than one second, with no meaningful user action, are classified as off screen behavior segments.

**Attention segments;** segments with frequent off screen behavior, frequent mouse/keyboard clicks are classified as attention segments.

Once the video file is classified into one of the above five segment categories, the manual classification results are ready for comparison with the automatic classification results.

### E. Result Verification

The number of Excessive vs. Excessive, Excessive vs. Non-Excessive, Non-Excessive vs. Excessive and Non-Excessive vs. Non-Excessive segments as well as related error rates are calculated for each result file and graphs are plotted to visualize the results and enable comparing the performance of different methods and features. During the verification of results, the attention segments are not considered as they are not clearly distinguished as excessive effort or non-excessive effort. Non-Excessive vs. Excessive segments are regarded as false positive or type-I error segments. It is assumed that all the segments classified as *excessive effort segments* are due for manual evaluation. Hence, in the case of type-I error, the software program is highlighting extra segments for further review, but is not missing any segments that need attention.

On a similar note, segments that show excessive effort per manual classification but identified as non-excessive effort segments by the software program are regarded as false negative or type-II error segments. These segments require extra attention as the software program has missed identified segments that require manual inspection. The total time of segments classified as excessive by the software program is also referred as inspection time. It is the sum of the time interval of each excessive effort segment. In this



paper, type-II errors and inspection time are considered as the most important factors for analyzing experiment results.

## V. EXPERIMENTS

In this paper, the automatic part of the process is used to analyze five data files by applying the different pattern recognition techniques discussed. Each of the files is a text file that contains the entire data collected by the eye tracker throughout each experiment. The following is a list of the experiments performed:

1. Applying the threshold method
  2. Applying heuristic feature selection and K-means clustering
  3. Using principal component analysis
  4. Applying K-means clustering on principal components.
- Each experiment procedure is discussed in detail in the following sections.

### A. Experiment 1: Applying the threshold method

In this experiment, event based segmentation is applied to the video and data file generated by the eye tracker. Next, a feature set is generated for the data file. All the segments are classified into excessive or non-excessive effort segments by the software program, which applies the threshold method on the following features: 1) number of fixations, 2) average fixation duration, 3) number of saccades, 4) average saccade amplitude, and 5) eye path traversed.

### B. Experiment 2: Applying heuristic feature selection and K-means clustering

Due to the fact that evaluating all the possible subsets of the feature set is prohibitively time consuming, we have adopted a heuristic feature selection method. The following subsets are selected: 1) Number of fixations, 2) Number of saccades, 3) Eye path traversed, 4) Number of fixations, number of saccades, eye path traversed, and 5) Number of fixations, number of saccades, eye path traversed, average fixation duration and average saccade amplitude.

### C. Experiment 3: Using principal component analysis

In this experiment, the feature set is transformed into principal components by a program that implements PCA. Here, only the first principal component is considered, as it carries the most significant information related to the feature set. The first principal component is subjected to the threshold method for identifying segments exhibiting excessive effort and non-excessive effort.

### D. Experiment 4: Applying K-means clustering on principal components

In this experiment, K-means clustering is applied to different combinations of principal components for identifying segments exhibiting excessive effort and non-excessive effort. The following constitute the feature set for this experiment: 1) 1<sup>st</sup> principal component, 2) 1<sup>st</sup> and

2<sup>nd</sup> principal components, and 3) 1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup> principal components.

## VI. RESULTS

In this section, the results obtained from the experiments are discussed. The results of each data file in the experiments are shown in [4]. A sample of these results, concentrating on applying the threshold method using data file 1, is presented here. For clarity, the notation used for the feature values in the graphs is presented below: 1) # Fix – denotes number of fixations, 2) Avg. Fix Dur. – denotes average fixation duration, 3) # Sacc – denotes number of saccades, 4) Sacc Amp. – denotes average saccade amplitude, 5) Eye Path denotes eye path traversed, and 6) FPC - denotes first principal component:

The video file corresponding to data file 1 is 6.09 minutes in length. Fig. 2 shows the results of an experiment using the threshold method on data file 1.

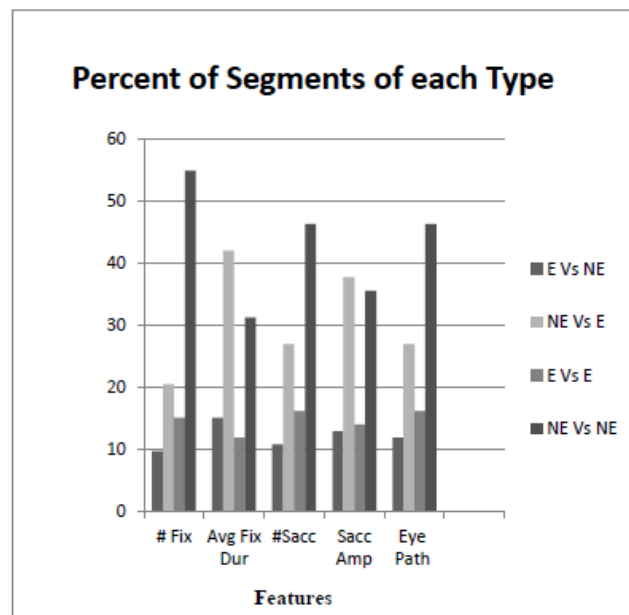


Figure 2. Percent of segments of each type.

When the graph in Fig. 2 is extrapolated and as seen from the E vs. NE bars, the feature value, number of fixations demonstrate a small percentage of E vs. NE segments. This shows that the number of fixations has the least number of type-II errors. Number of saccades and eye path traversed follow number of fixations in terms of type-II errors.

Fig. 3 shows the total time of segments classified as excessive by the software program and the manual process after the threshold method is applied on each of the following features: 1) number of fixations, 2) average fixation duration, 3) number of saccades, 4) average saccade amplitude, and 5) eye path traversed.

The light black bars in Fig. 3 represent the total time of video recorded by the eye tracker. Manual classification of the video file, depicted by the dark black bars, shows 1.71 minutes of excessive effort. The average fixation duration and average saccade amplitude show a relatively low value for time of segments classified as excessive by the software program when compared with the total video time. This is depicted by the bright bars present in the figure. From Fig. 3, it is observed that the percentage of type-II errors is 15.05% for average fixation duration and 12.9% for average saccade amplitude. However, the feature value with a reasonable type-II errors and lower percentage of time of segments classified as excessive is average saccade amplitude.

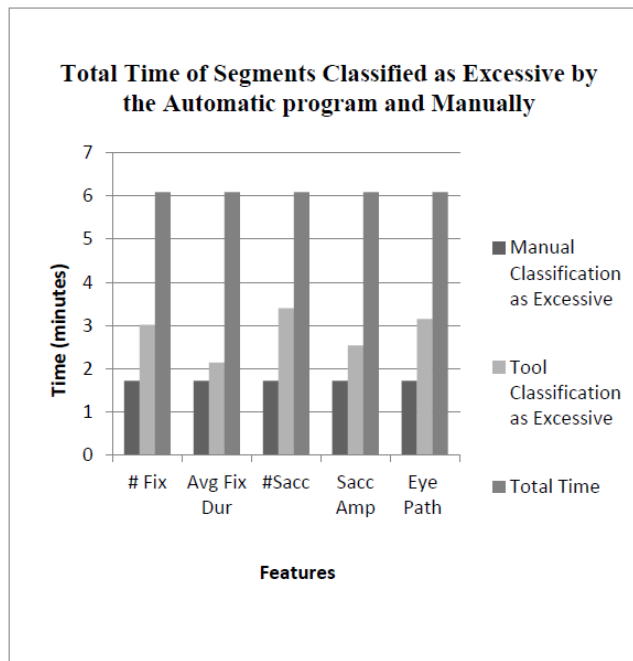


Figure 3. Total time of excessive effort segments.

The set of experiments include three more experiments the results of these experiments are detailed in [4]. The experiments are:

- Identifying excessive effort segments using heuristic feature selection and K-means clustering.
- Identifying excessive effort segments using principal component analysis
- Identifying excessive effort segments using K-means clustering on principal components.

## VII. RESULT EVALUATION

In this section, we evaluate and discuss the results of the experiments conducted in this work. Our criteria for success are based on 1) The number of type-II errors and 2) A minimal time to investigate the usability issues with an acceptable level of type-II errors. Based on discussions with several engineers in the company sponsoring this work and

other companies, we are assuming that 15% of error of type-II is the upper bound for being considered as acceptable. This is also consistent with a two-step approach where after a first pinpoint analysis stage, which allows for high rate of errors but provides significant reduction in evaluation time, the errors identified are fixed; leading to a more rigorous pinpoint analysis with lower error bound. The results are evaluated based on the performance of each pattern recognition method on individual features. In addition, the overall performance of each pattern recognition method is evaluated.

Tables I to IV (attached at the end of this paper) summarize the results of the experiments. An additional set of tables, which contains the entire results, can be found in [4]. The items listed in Tables I to IV are:

### A. Applying the threshold method

The following observations are derived from Table I.

- The results of Table I show that the threshold method on the feature value, *number of fixations*, gives good results in terms of type-II errors but, the average inspection time is relatively high when compared to other feature values. The average value of type-II errors for number of fixations is 3.3%. Average saccade amplitude and eye path traversed follow the number of fixations in terms of type-II errors.
- A threshold on *average fixation duration* performs well in terms of minimal inspection time with an acceptable value of 9.8% for type-II errors.
- A feature value with minimum number of total errors is *eye path traversed*. This feature value is a good choice when inspection time is not taken into account.
- The inspection time is not completely correlated to type-I errors. In the case of average fixation duration, the inspection time is 1.67 minutes with 29.4% of type-I errors. On the other hand, the average saccade amplitude with almost the same percentage of type-I errors has higher inspection time than average fixation duration.
- The values of the average number of excessive effort segments for all features are in close proximity to each other. However, the percentage of type-I and type-II errors differs invariably. This portrays that the segments classified as excessive are different for each feature value.
- Despite the fact that the percentages of total errors for each feature value are in close proximity to each other, the inspection time varies. This delineates that the segments classified as excessive are different for each feature value.

### B. Applying heuristic feature selection and K-means clustering.

The following observations are derived from Table II:

- The results from Table II show that the K-means clustering on the *feature subset - number of fixations*,

*number of saccades, eye path traversed, average fixation duration, and average saccade amplitude*, gives good results in terms of type-II errors with an average value of 5.4%. But, the average inspection time is relatively high when compared to other feature values. The number of fixations follows the above identified feature value in terms of type-II errors.

- Clustering on the *eye path traversed* performs well in terms of minimal inspection time with an acceptable value of 10.1% for type-II errors.
- A feature value with minimum number of total errors is the number of fixations. This feature value is a good choice when inspection time is not taken into account.
- The average number of excessive effort segments for number of fixations and the feature subset with the following features - number of saccades, eye path traversed, average fixation duration, average saccade amplitude are the same. However, the inspection times vary. This portrays that the segments classified as excessive are different for each feature value.
- Unlike the results of the threshold method, the percentages of total errors for each feature value vary by a wide margin when applying the K-means clustering on the feature subsets.

#### C. Using principal component analysis

The results summarized in Table III are compared with the results obtained from Experiment 1 to analyze the performance of the threshold method on the first principal component with other features such as: 1) number of fixations, 2) average fixation duration, 3) number of saccades, 4) average saccade amplitude, and 5) eye path traversed. Experiment 1 result evaluation shows that the feature value, number of fixations, gives good results in terms of type-II errors. The average percentage of type-II errors for number of fixations is 3.3%, whereas it is 4.1% for first principal component. Initially, average saccade amplitude and eye path traversed succeeded number of fixations in terms of performance. However, the new results place a threshold on the *first principal component* after the number of fixations with respect to type-II errors. The inspection times for the first principal component and for the average fixation duration are 2.7 and 1.6 minutes respectively. A threshold on *average fixation duration* performs better than first principal component in terms of lower inspection time and an acceptable 9.8% for type-II errors.

#### D. Applying K-means clustering on principal components.

Table IV shows the average values of all the features used in experiment 4 over five data files. The average type-II error is very high when using the K-means on the principal components. The average inspection time is only 1.96%. When taking type-II errors also into consideration, this method is not suitable to identify excessive effort segments.

## VIII. CONCLUSIONS AND FUTURE RESEARCH

The framework presented in this research enables software developers to efficiently identify the usability issues thereby optimizing the time spent on software-usability testing. Excessive effort segments, which typically relate to the usability issues, are identified by applying pattern recognition techniques such as K-means clustering algorithm, thresholding, principal component analysis, and feature selection. The analysis of the experiments conducted in this paper shows that the time taken for software usability testing can be reduced by 40% or more.

Of all the pattern recognition methods used, a threshold on number of fixations yields the best results in terms of type-II errors and is followed by a threshold on the first principal component. The K-means clustering on feature subset with the features: number of fixations, number of saccades, average saccade amplitude, average fixation duration, and eye path traversed ranks third.

When the inspection time is taken into consideration while also confirming that type-II errors are within a reasonable limit, the K-means clustering on the number of saccades yields the best results and precedes the threshold method on average fixation duration in performance.

With time and resources at one's disposal, there is a scope to enhance the definition and implementation of pattern recognition techniques in identifying usability issues in software.

In this research, the time between two consecutive keyboard/mouse clicks by user is considered as a segment and this has served as the basic pattern for pattern recognition techniques. Equal time slicing of user's software interaction session can be used instead and the performance results can be analyzed and compared with the results from this research.

Further refinement of pattern recognition techniques can be pursued to minimize errors and inspection time. Also, more focus can be given to the criteria for manual classification of video segments thus allowing excessive effort segments to be identified more accurately.

Another direction for future research is to automate some of the manual steps in this process. This can include software that automatically log users' software interaction session data, manipulate data, and without human intervention lists the start and end times of all the excessive effort segments. This can significantly reduce time taken for the usability testing.

In this work, we have concentrated on pattern recognition techniques that do not rely on human intelligence. Hence, results are generated using non-supervised learning procedures. A surrogate approach can use supervised learning procedures to produce the output. This involves conducting experiments using training data sets to manually arrive at an archetype that can be applied on any data set to generate the output.

Another direction for further research is to consider information fusion. Information fusion combines different

techniques of pattern recognition classification to achieve more accurate results. An additional approach that can be researched is to arrive at a formula or function that can be applied to any data set by which the usability deficiencies can be identified.

#### ACKNOWLEDGMENT

This research was funded in part by Emerson Process Management [22], an Emerson business.

#### REFERENCES

- [1] O. C. Komogortsev, C. J. Mueller, D. E. Tamir, and L. Fledman, "An Effort Based Model of Software Usability," 2009 International Conference on Software Engineering Theory and Practice, FL, 2009, pp. 75-84.
- [2] D. E. Tamir, O. V. Komogortsev, and C. J. Mueller. "An Effort and Time Based Measure of Usability," 6th Workshop on Software Quality, 30th International Conference on Software Engineering, Leipzig, Germany, 2008, pp. 35-41
- [3] D. E. Tamir, et al. "Detection of Software Usability Deficiencies," International Conference on Human Computer Interaction, FL, 2011, pp. 528-536.
- [4] D. K. V. Dasari, Pinpoint analysis of software usability, Thesis Report, Texas State University, Computer Science, December 2012.  
<https://digital.library.txstate.edu/handle/10877/32/browse?value=Dasari+Kali+Venkata%2C+Divya&type=author>, retrieved June 2013.
- [5] J. T. Tou and R. C. Gonzalez, Pattern Recognition Principles, Reading, MA: Addison-Wesley Publishing, Inc., 1974.
- [6] R. O. Duda, P. E. Hart, and D. G. Stock, Pattern Classification, 2nd Ed., Indianapolis, Willey International, 2001.
- [7] D. E. Tamir and A. Kandel, "The Pyramid Fuzzy C-means Algorithm," International Journal of Computational Intelligence in Control, 2 (2), 2012 pp. 65-77.
- [8] ISO/IEC 9126-1: 2001, Software Engineering-Product Quality, Part-1, Quality Model, Geneva, Switzerland: International Standards Organization, 2001.
- [9] ISO/IEC 9126-1: 2001, Software Engineering-Product Quality, Part-2, External Metrics, Geneva, Switzerland: International Standards Organization, 2001.
- [10] A. Poole and L. J. Ball, Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects, Encyclopedia of Human Computer Interaction: Idea Group, 2004.
- [11] M. A. Just and P. A. Carpenter, "Eye Fixation and Cognitive Processes," Cognitive Psychology, vol. 8, 1976, pp. 441-480.
- [12] J. S. Dumas and J. C. Redish, "A Practical Guide to Usability Testing," OR, USA, Intellect Books., 1999.
- [13] J. Rubin and D. Chisnell, Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests, Indianapolis, Wiley Publishing, Inc., 2008.
- [14] H. Ebbinghaus, Memory: A Contribution to Experimental Psychology, 1885,  
<http://psychclassics.yorku.ca/Ebbinghaus/memory3.htm>, retrieved July 2013.
- [15] J. Nielsen, Usability Engineering, San Francisco, Boston, Academic Press, 1993.
- [16] C. J. Mueller, D. E. Tamir, O. C. Komogortsev, and L. Feldman, "Using Designer's Effort for User Interface Evaluation," *IEEE International Conference on Systems, Man, and Cybernetics*, Texas, USA, October 11, 2009, pp. 480-485.
- [17] E. T. Hvannberg and C. L. Lai, "Classification of Usability Problems (CUP) Scheme," Nordic conference on Human-computer Interaction, Oslo, Norway, 2006, pp. 655-662.
- [18] N. Nakamichi, S. Makoto, and S. Kazuyuki, "Detecting Low Usability Web Pages using Quantitative Data of Users' Behavior," Proceedings of the 28th international conference on Software engineering, New York, NY, 2006, pp. 569-576.
- [19] S. Makoto, N. Noboru, H. Jian, S. Kazuyuki, and N. Nakamichi. "Webtracer: A New Integrated Environment for Web Usability Testing," 10th Int'l Conference on Human - Computer Interaction. Crete, Greece, June 2003, pp. 289-290.
- [20] Anonymous, "Tobii X60 & X120 Eye Trackers: User Manual," Tobii, 2013,  
[http://www.tobii.com/Global/Analysis/Downloads/User\\_Manuals\\_and\\_Guides/Tobii\\_X60\\_X120\\_UserManual.pdf](http://www.tobii.com/Global/Analysis/Downloads/User_Manuals_and_Guides/Tobii_X60_X120_UserManual.pdf), retrieved June 2013.
- [21] Anonymous, "MATLAB Product Help," MATLAB, 2013,  
<http://www.mathworks.com/help/>, retrieved June 2013.
- [22] Anonymous, "Emerson Process Management," Emerson, 2013, <http://www.emersonprocess.com>, retrieved June 2013.

TABLE I. AVERAGE VALUES OF EXPERIMENT 1 RESULTS

Feature value	avg. # of excessive effort segments	avg. total no of segments	avg. % type- I errors	avg. % type- II errors	avg. % of total errors	avg. Inspection time	avg. Inspection time as a % of total time
# Fix	17.2	95	28.4	3.3	31.7	2.7	62.1
Avg. Fix Dur.	18.2	95	29.5	9.9	39.4	1.6	37.4
#Sacc	32	95	21.8	10.5	32.2	2.9	64.1
Sacc Amp.	17.6	95	29.1	4.6	33.7	2.5	56.4
Eye Path	17.8	95	25.7	5.1	30.8	2.6	57.7

TABLE II . AVERAGE VALUES OF EXPERIMENT 2 RESULTS

Feature value	avg. # of excessive effort segments	avg. total no of segments	avg. % type -I errors	avg. % type -II errors	avg. % of total errors	avg. Inspection time	avg. Inspection time as a % of total time
#fix	29.1	95	27.2	6.6	33.9	2.4	56.2
#sacc	23.5	95	17.8	8.9	26.7	2.0	45.1
eye path	19.7	95	18.0	10.1	28.1	1.6	37.5
#fix, #sacc, eye path	23.2	95	18.3	8.6	26.9	1.9	44.5
#fix, #sacc, eye path, avg. fix dur., avg. sacc amp.	29.2	95	32.6	5.4	38.0	2.5	56.3

TABLE III. AVERAGE VALUES OF EXPERIMENT 3 RESULTS

Feature value	avg. # of excessive effort segments	avg. total no of segments	avg. % type -I errors	avg. % type- II errors	avg. % of total errors	avg. Inspection time	avg. Inspection time as a % of total time
1st principal components	16.6	95	27.5	4.1	31.6	2.7	61.2

TABLE IV. AVERAGE VALUES OF EXPERIMENT 4 RESULTS

Feature value	avg. # of excessive effort segments	avg. total no of segments	avg. % type- I errors	avg. % type- II errors	avg. % of total errors	avg. Inspection time	avg. Inspection time as a % of total time
1st, 2nd & 3rd principal components	28.6	95	24.4	12.6	37.0	2.0	43.6

# Gender Classification of Face with Moment Descriptors

Wen-Shiung Chen, Po-Yi Lee

Dept. of Electrical Engineering

National Chi Nan University

Puli, Nantou, Taiwan

e-mail: {wschen,s97323529}@ncnu.edu.tw

Lili Hsieh

Dept. of Information Management

Hsiuping University of Science and Technology

Taichung, Taiwan

e-mail: lily@mail.hust.edu.tw

**Abstract**—Moments may effectively describe geometrical properties of an object and thus are widely used in image analysis and pattern recognition, except gender classification. In this work, we propose a modification of eigenmoment and apply to classify sex from facial features. We investigate and compare four moment descriptors with nearest neighbor and SVM in three different feature regions on face. The results show that the features used in gender classification can be effectively represented by moment descriptors.

**Keywords**—Biometrics; Moment; Face Recognition; Gender Classification.

## I. INTRODUCTION

Gender classification [1][2], detecting someone's sex from a human face, is used for personal authentication and multimedia interaction. Gender classification is really a difficult problem since there is no clear definition or rule on how to distinguish between males and female faces. Many different kinds of feature extraction methods have been applied in an attempt to solve the gender classification problem [3]. Literature shows that shape reveals a better effectiveness on classifying gender than texture and color [3]. Although moments were applied in computer vision for a long time [4]-[6], they have not been utilized so far in gender classification. In this paper, we address four different moment descriptors for extracting the features from facial images.

Most works on face recognition and gender classification use the face regions without hair. However, Lapedriza, *et al.* referred to the features of face regions by considering the external region (i.e., with hair) and the internal region (i.e., without hair) [7]. Makinen and Raisamo also used two different face regions and discussed their influences on performance [8]. We thought hair is worthy to be considered since sometimes it is difficult to determine someone's gender either when a man has long hair or a woman has short hair. Three different face regions will be considered in this work, namely, the external feature region, the general feature region and the internal feature region. The external feature region is a facial image with as much hair as possible. Contrarily, the other regions contain a face or faces without hair. The general feature region contains the forehead, the cheek and the chin. The internal feature region is a smaller region between the eyebrow and the lip. This work discusses the effectiveness of different regions on gender classification based on moments.

## II. THE PROPOSED METHOD

### A. System Overview

The proposed gender classification system consists mainly of three modules: pre-processing, feature extraction, and classification, as shown in Figure 1. First, the pre-processing module employs face detector to detect and localize the positions where faces are (i.e., face zone) from the input image. A cropped region, containing a face or faces, is the output. It performs four major tasks including feature region selection, face detection, image resize and histogram enhancement for the input face images. Next, the feature extraction module adopts moment descriptors to generate the feature vectors. Finally, the classification module employs nearest neighbor and the support vector machines to recognize the face images by comparing the feature vectors with what enrolled in databases.

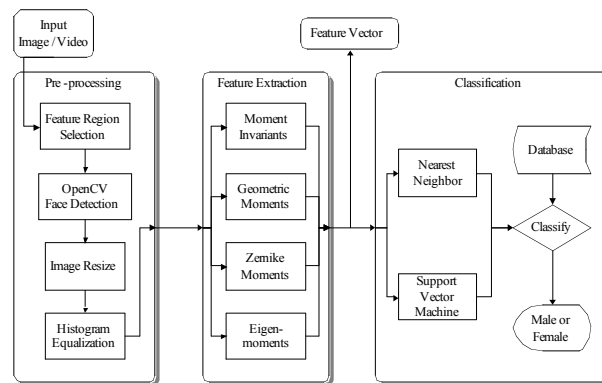


Figure 1. Diagram of the proposed face gender classification system.

### B. Pre-processing

Three different face regions, external feature region, general feature region and internal feature region, shown in Figure 2, are considered. First, we extract a face image scaled to the size of  $24 \times 24$  pixels, and called general feature region. The minimum face size in OpenCV detector [9] is set to  $24 \times 24$  pixels and the sub-image is scaled after scanning by multiplying the current sub-image size by 1.25 times (i.e., 1st scan sub-image size:  $24 \times 24$  pixels, 2nd:  $30 \times 30$  pixels, and so on).

We also extract the larger detected face area from the same images, similarly. We enlarge the width on both sides

by 10% (20% in total), the top by 40% and the bottom by 12%, of the image since this region usually contains hair in addition to face, but the remaining area of the image as little as possible, for example, the background. This region is called external feature region. In some cases the region could not be grown as much as intended since image borders were encountered, thus this kind of images were removed from the data. Other cases when images were considered of bad quality and were removed from the data included images in which objects such as a hat, a hand, or some other object, existed in front of the hair or the face. We extract the regions by decreasing the detected face area by 10% on both sides (20% in total), 20% on the top but unchanged to the bottom of the image since the bottom of default detected face area reach lip for many images. This feature region is called internal feature region. It is smaller than general feature region, so no image is removed.

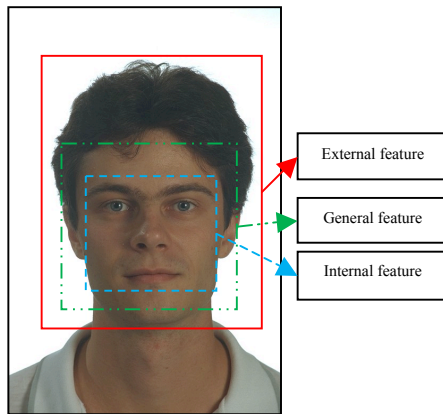


Figure 2. Three different feature regions.

We used  $32 \times 40$  image size (instead of  $24 \times 24$ ) when external feature region is considered since it is closer to the image size. Similarly, the image with internal feature region is resized to  $20 \times 20$  and transformed to gray-level, and then followed by histogram equalization. Moreover, Makinen and Raisamo achieved the best results without automatic alignment [8]. Manual alignment is a possible way to obtain better results but difficult to implement for a large database. Hence, we do not normalize any image in experiments. Finally, the histogram equalization is performed once before feature extraction.

### C. Feature Extraction

The geometric moment (GM) has the form of the projection of  $f(x, y)$  onto the monomials  $x^p y^q$ , where  $f(x, y)$  is intensity of an image. Unfortunately, the basis set  $\{x^p y^q\}$  is non-orthogonal, so that these moments are not orthogonal. Yap and Paramesran proposed an improvement of GMs by solving an eigenvalue problem in the moment space, called eigenmoment (EM) [10]. So far, eigenmoment is not applied on pattern recognition, especially gender classification, thus we propose to apply it in gender classification in this paper. Assume that our signal

dataset is composed of  $m$   $n$ -D column vectors  $s_i \mathbf{S}_i$ , where  $i = 1, 2, \dots, m$ . Let the  $n \times mn \times m$  matrix  $\mathbf{S}$  represent the signal dataset, a transformation matrix  $\mathbf{X}$ , constructed by GMs, converts the signal into the moment space. The dataset in moment space is represented as  $\mathbf{S}^T \mathbf{X}$ .

The original dataset, extracted by GMs, is information redundant since the kernel of GM is not orthogonal. Now the concept of principal component analysis (PCA) [11] helps to solve this problem. The goal is now to find a unity vector  $\mathbf{w}$  such that the squared sum of the dataset's projection onto this direction is maximal. The squared sum of the total projection of  $\mathbf{S}^T \mathbf{X}$  onto  $\mathbf{w}$  is a function of  $\mathbf{w}$ , denoted by:

$$L(\mathbf{w}) = (\mathbf{S}^T \mathbf{X} \mathbf{w})^T (\mathbf{S}^T \mathbf{X} \mathbf{w}) = \mathbf{w}^T \mathbf{X}^T \mathbf{S} \mathbf{S}^T \mathbf{X} \mathbf{w}. \quad (1)$$

To maximize  $L(\mathbf{w})$  under the constraint  $\|\mathbf{w}\|=1$ , solving

$$\max_{\mathbf{w}} L(\mathbf{w}) = \max_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{A} \mathbf{w}}{\mathbf{w}^T \mathbf{w}} \quad (2)$$

where  $\mathbf{A} = \mathbf{X}^T \mathbf{S} \mathbf{S}^T \mathbf{X} \mathbf{A} = \mathbf{X}^T \mathbf{S} \mathbf{S}^T \mathbf{X}$  is a real symmetric matrix (or covariance matrix). Equation (2) is called Rayleigh quotient. We use the Lagrange multiplier to form a new objective function:

$$\tilde{L}(\mathbf{w}) = \mathbf{w}^T \mathbf{A} \mathbf{w} + \lambda (1 - \mathbf{w}^T \mathbf{w}) \quad (3)$$

where  $\lambda$  is a Lagrange multiplier. The stationary points of  $\tilde{L}(\mathbf{w})$  occur at  $\nabla_{\mathbf{w}} \tilde{L}(\mathbf{w}) = 0$ . Then, we obtain

$$L(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{A} \mathbf{w}}{\mathbf{w}^T \mathbf{w}} = \lambda \frac{\mathbf{w}^T \mathbf{w}}{\mathbf{w}^T \mathbf{w}} = \lambda. \quad (4)$$

Therefore, the eigenvectors  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m$  of  $\mathbf{A}$  are the critical points of the Raleigh quotient and their corresponding eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_m$  are the stationary values of  $L(\mathbf{w})$ . Based on principal components analysis, we can arrange the eigenvalues of  $\mathbf{A}$  into a descending order  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ . Then, the maximal value of  $L(\mathbf{w})$  is  $\lambda_1$  occurring at  $\mathbf{w} = \mathbf{w}_1$ , while its minimum is  $\lambda_m$  occurring at  $\mathbf{w} = \mathbf{w}_m$ . Since  $\mathbf{A}$  is a symmetric matrix, its eigenvalues are positive and the eigenvectors are orthogonal to one another.

The original signal projected onto moment space via GMs can be represented as  $\mathbf{S}^T \mathbf{X}$ . However, the kernel of GMs is not orthogonal, it makes information redundant. Then we use PCA to find a vector  $\mathbf{w}$  and the projection  $\mathbf{S}^T \mathbf{X}$  onto  $\mathbf{w}$  can be represented as (1). Notice  $\mathbf{w}$  is an orthonormal basis obtained by solving (2). Transformation with orthogonal moment kernels into the moment space is equivalent to the projection of the signal onto orthogonal basis. Thus, we improve the information redundancy problem and still keep the low computational complexity of GMs. Furthermore, the eigenvalues of  $\mathbf{A}$  close to zero means low representation of dataset in the moment space. We can discard these features

to compact information. In the experiments, we discuss its influence as discarding 53, 27 and zero features. The concept of the proposed method is illustrated in Figure 3.

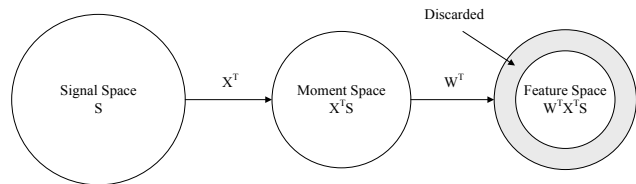


Figure 3. Conceptual chart of eigenmoments [10].

The features of the tested images are extracted by moment descriptors. These features form a high dimension feature vector as input to a classifier. TABLE I summarizes the number of features in our experiments.

TABLE I. # OF FEATURES IN THE EXPERIMENTS

	GMs	Hu moment invariants	Zernike moment	Eigen moment
Order	10	--	10	10
# of features	63	7	36	63

D. Classification

In this paper, we adopt *k*-nearest neighbor (*k*-NN) [11] and support vector machine (SVM) [12] as the classifier. We used LIBSVM [11] to train SVM with RBF kernel and predict classification rates for four moment descriptors. The parameters *Γ* (gamma) and *C* (cost) will be automatically determined by LIBSVM in experiments.

III. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed gender classification system, we implemented and tested the proposed schemes on two different databases. The first is a simple database collected by ourselves and containing 100 (50 males and 50 females) face images which are frontal with different expressions. The second is the well-known FERET database. In our experiments, only frontal images (FA and FB classes) are used. These classes contain 2,722 images. Duplicated images of the same person are removed, so only one image per person is kept. Next, we pick out 900 (450 males and 450 females) images from the rest. In order to compare fairly, this procedure is the same as [13]. Figure 4 depicts the flowchart of the procedure.

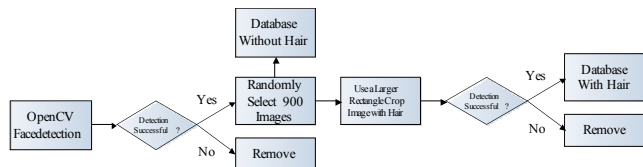


Figure 4. Flowchart of selecting tested images.

We take 900 images for general feature. As described in pre-processing, we increase the detected area for external

feature region. In some cases the area could not be grown as much as intended since image borders came across. We removed the images from the data. Zernike moments need to extract a square region transformed into a unit circle disk. Thus, a larger detected face area is used to extract a square region. Those images with crossing borders are removed. TABLE II summarizes the number of experimented images.

TABLE II. IMAGES OF DATABASE IN EXPERIMENTS

Region	Internal feature	General feature	External feature	External feature for ZM
size	20 × 20	24 × 24	32 × 40	40 × 40
Simple database	100	100	100	100
FERET database	900	900	862	600

We implement four moment descriptors and two classifiers for each feature region. The training and testing strategy is five-fold cross validation. The test sets are not overlapped with their respective training sets. TABLEs III and IV summarize the experimental results on simple database and FERET database, respectively.

TABLE III. THE RESULTS ON SIMPLE DATABASE

Simple database	Classification rate (%)					
	External feature 32 × 40		General feature 24 × 24		Internal feature 20 × 20	
Methods	NN	SVM	NN	SVM	NN	SVM
MI (7)	92	96	82	89	66	75
ZM (36)	90	93	91	95	77	85
GM (63)	93	98	91	97	77	86
EM (10)	93	100	89	98	77	87
EM (36)	82	97	82	95	68	75
EM (63)	67	91	62	90	58	67

TABLE IV. THE RESULTS ON FERET DATABASE

FERET database	Classification rate (%)					
	External feature 32 × 40		General feature 24 × 24		Internal feature 20 × 20	
Method	NN	SVM	NN	SVM	NN	SVM
MI (7)	71.2	77.8	69.8	78.9	61.8	67.8
ZM (36)	82.4	82.5	81.1	83.4	75.3	81.0
GM (63)	76.0	82.4	76.2	80.9	64.7	78.6
EM (10)	75.0	82.8	78.7	79.9	69.1	74.0
EM (36)	78.2	83.3	80.6	84.8	71.6	78.8
EM (63)	76.9	82.7	81.3	84.8	71.2	80.0

The results show that the performances of eigenmoment are not as well as expected when the features are not discarded. Simple database is a small dataset of only 100 images. The eigenvalues close to zero mean a bad representation of the dataset in the moment space. Accordingly, we can discard these features to compact information. The eigenvalues explicate the proportion of distribution of dataset obtained by using PCA theory. Figure 5 shows the proportion of accumulative eigenvalues in which



the accumulation of first 10 eigenvalues exceeds 99.5%, so the performance is improved significantly by discarding the other features.

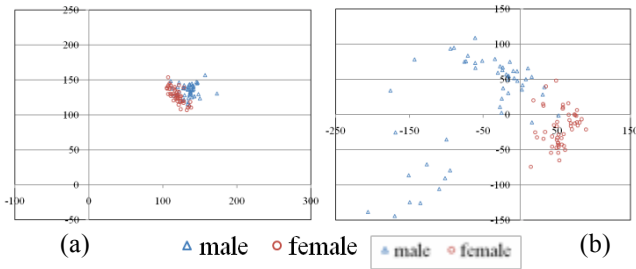


Figure 5. The distribution of first two features. (a) Geometrical moments. (b) Eigenmoments.

In Figure 5(a), we set  $M_{20}$  to  $x$ -axis and  $M_{02}$  to  $y$ -axis. In Figure 5(b),  $x$ -axis and  $y$ -axis represent the first and second data distribution eigenvalues of eigenmoment, respectively. It is easy to explain that the performances of eigenmoment overcome those of GMs. The performances get worse in the experiments on FERET database since the images are influenced by all kinds of variations. It is worth to notice that the performances of eigenmoment are not as well as expected again. This results from the variation of number of images. Simple database only contains 100 images while FERET database contains 900 images in our experiment. It is not enough to describe 900 images if we still use 10 features, so we increase to 36 and 63 features. The results become better when we increase the number of features. However, the improvement is not obvious between 36 and 63 features.

One the other hand, we test three feature regions in the experiments. According to the results, the performances of external feature have the best classification rate in simple database. However, to FERET database, the influence of background makes the performances of external feature worse. Overall, the general feature is better for moment descriptors on gender classification problem. The results in [13] are summarized in TABLE V. Comparing our results with those in [13], moment descriptors still have good enough classification rate which is above average level in FERET database.

TABLE V. EXPERIMENTAL RESULTS ON FERET DATABASE WITHOUT NORMALIZATION IN [12]

Methods	Classification rate (%)		
	Without hair (24 × 24)	With hair (32 × 40)	Average classification rate
Neural network	83.89	90.07	86.98
SVM	84.44	72.85	78.65
Threshold Adaboost	82.22	83.44	82.83
LUT Adaboost	80.56	87.42	83.99
Mean Adaboost	76.67	87.42	82.05
LBP + SVM	75.56	72.19	73.88
Average rate	80.56	82.23	81.40

#### IV. CONCLUSION

In this paper, we present a gender classification system based on moment descriptors. The results show that the features used in gender classification can be effectively represented by moment descriptors. The classification rate up to 100 % can be achieved when we test 10 features using eigenmoment with SVM on simple database.

#### REFERENCES

- [1] B. Moghaddam and M. Yang, "Learning gender with support faces," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp. 707-711, May 2002.
- [2] W.-S. Chen, W.-J. Chang, L. Hsieh, Z.-Y. Lin, "Component-based gender classification based on hair and facial geometry features, IJCCI, 2012, pp. 626-630.
- [3] W. Zhao, R. Chellappa, A. Rosenfeld and P. J. Phillips, "Face recognition: A literature survey," ACM Computing Surveys, vol. 35, no. 4, pp. 399-458, December, 2003.
- [4] M. K. Hu, "Visual pattern recognition by moment invariants," IEEE Transactions on Information Theory, vol. 8, no. 2, pp. 179-187, February 1962.
- [5] H. Shu et al., "Moment-based approaches in imaging, Part 1: Basic features," IEEE Engineering Medicine & Biology Magazine, 2007.
- [6] A. Khotanzad and Y. H. Hong, "Invariant image recognition by Zernike moments," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 5, pp. 489-497, May 1990.
- [7] A. Lapedriza, et. al., "Gender recognition in non-controlled environment," IEEE International Conference on Pattern Recognition, vol. 3, pp. 834-837, 2006.
- [8] E. Makinen and R. Raisamo, "Evaluation of gender classification methods with automatically detected and aligned Faces," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 3, pp. 541-547, March 2008.
- [9] "OpenCV 1.0, Open Source Computer Vision Library," <http://www.intel.com/technology/computing/opencv/>.
- [10] P. T. Yap and R. Paramesran, "Eigenmoments," Pattern Recognition, vol. 40, no. 4, pp. 1234-1244, April 2007.
- [11] R. Jang, "Data clustering and pattern recognition," <http://mirlab.org/jang/books/dcpr/index.asp>.
- [12] C. C. Chang and C. J. Lin, LIBSVM: A library for support vector machines, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm> (retrieved May 2011).
- [13] E. Makinen and R. Raisamo, "An experimental comparison of gender classification methods," Pattern Recognition Letters, vol. 29, no. 10, pp. 1544-1556, July 2008.

# Artificial Immune Memory Formed by Dynamics of Antibody Networks

Chung-Ming Ou

Department of Information Management  
Kainan University  
#1, Kainan Rd., Luchu, Taiwan 33857  
e-mail: cou077@mail.knu.edu.tw

Chung-Jen Ou

Department of Electrical Engineering  
Hsiuping University of Science and Technology  
#11, Industry Rd., Taichung, Taiwan 41280  
e-mail: crou@mail.hust.edu.tw

**Abstract**—Immune memory can be regarded as an equilibrium state of immune network system with nonlinear dynamical behavior. The rapid response of immune systems to the secondary antigen is owing to the stable structure of memory state forming by a closed idiotypic immune network. Internal image of an antigen is defined while memory state is formed via such network. Antibody chain based on tree structure is proposed which explains how the memory state is formed in the immune network. We also propose a network dynamics model of idiotypic immune network based on cross-reactive correlation matrix to explain the artificial immune memory.

**Keywords**—immune network; immune memory; antibody chain; internal image

## I. INTRODUCTION

It is well known that the injection of a given amount of some antigen into a mammal's body will stimulate the production of antibodies directed against that antigen, if the antibody is with a high affinity for that antigen [1]. The immune system of the animal has thus learned to produce high quantities of the antibody directed against that very antigen, which is called vaccinated. Therefore, the biological immune system can learn itself and memorize the characteristics of invading antigens.

Memory in a physical system refers to the ability of the system to preserve information of its environment at some previous time. Memory mechanism of biological immune systems is still a mystery. From computational biology viewpoints, memory mechanisms in immune systems can be regarded some physical systems. There are basically two theories to explain the immune memory [1]. The first one is the theory of memory cells, which are generated after the B-cell proliferations. These cells will remain for an immune memory in the human body until the death of the individual. On the other hand, immune memory mechanisms can be explained through Jerne's immune network theory [2] by investigation them as complex adaptive systems. Immune memory belongs to a class of sparse and distributed associative memory [3]. The concept of artificial immune memory is the following. Complex Adaptive Systems (CASs) have to deal with constant change of environment. Based on memory mechanisms, CASs can respond immediately to the same or similar environment. Here, we are adopting immune memory mechanisms in CASs. According to Fernandez et al. [4], it is an ongoing research topic to exploit relationship between immune memory and internal image. The internal

image can be regarded as a portion of immune memory of antigen [5].

Immune memory mechanisms can also be modeled from immune network theory [2][6]. Jerne also proposed that once the foreign antigen is removed, the immune system would restore some information of such antigen by some memory mechanism. Debates of whether memory cell theory or associative property is true have been discussed. The immune memory can be explained by the following network viewpoint. Assuming that an antibody  $Ab_1$  is produced by the stimulating antigen. Then the production of  $Ab_1$  is increased in the presence of another type of antibody  $Ab_2$ . The population of T-helper cell  $TH_1$  specified by  $Ab_1$  is also increased. In this way,  $Ab_2$  can be considered as some "internal image" of this antigen. This image will be remained after the antigen is removed. The interactions can be a long chain with length greater than 2.

Many idiotypic network models focus on the interactions between antibodies and antigens. The network interactions provide dynamical memory mechanism, by keeping the concentrations of antibodies, in particular those internal images [7]. However, how immune systems recalls similar antigen is still unknown. From computations viewpoints, it is worthwhile considering state transition, which represents the network dynamics of such unfamiliar pathogen. If this antigenic state converges, then immune systems gain some control and activate some (associative) recall process.

The major goal of this paper is to study the associative memory based on statistical immunodynamics inspired by [8]. This network structure will proved to be major contributions to the immune memory mechanism. The related research can be referred to Abbattista et al. [9]. The arrangement of this paper is as follows. In Section II, some preliminary knowledge of immune memory is introduced. In Section III, dynamical behavior of idiotypic immune network is described. Simulation analysis of immune network memory is also given.

## II. BACKGROUND KNOWLEDGE

### A. Immune Responses

While a specific antigen invades human body, the immune system will respond by producing some antibodies, which can eliminate this invaded antigen. It has three phases, namely, first immune response, second immune response and

cross-reactive response. For the first invaded antigen, the immune systems will massively produce the antibody, which binds to  $Ag$ . Therefore the amount of antigen will be reduced tremendously after the peaking and tended to some constant value. We can say that immune systems are turned to the memory state. For the second phase, the same antigen invades and the antibody will take much less time than that of first immune response to reach the population peak. Therefore this antigen cannot proliferate tremendously. This is the reason why we don't get sick at this stage. On the other hand, if a similar antigen invades, then the same antibody will also proliferate soon and eliminate population. We will analyze the network transitions between each phase. Once the first immune response has activated, we are particularly interested in the third stage, namely the associate memory mechanism for the immune response to similar antigens.

### B. Idiotypic Immune Network

Idiotypic network theory implies that immune systems will mimic the presence of the antigen even after it is destroyed [3]. This way, the antibody and receptor of the lymphocytes can recognize each other. The epitope of antibody molecule is called an "idiotope". An epitope of antigen  $Ag$  is recognized by the antibody molecule  $Ab_1$  and by the receptor molecule on the lymphocyte of  $LU_1$ . The antibody  $Ab_1$  and the receptor of  $LU_1$  have the idiotope, which is recognized by antibody  $Ab_2$  and the receptor on the lymphocyte of  $LU_2$ . Continuously, we reach an antibody  $Ab_N$ , while the antibody  $Ab_1$  and the receptor on the lymphocyte of  $LU_1$  also recognize idiotopes on antibody  $Ab_N$ .  $Ab_N$  constitutes an internal image of the antigen  $Ag$  (see Figure 1). The idiotypic network theory has been proposed as a central aspect of the associative properties of immune memory [3][10]. However, computational aspect of this paradigm needs to be further explored.

For simplicity, it is reasonable that we consider antibodies rather than LUs of IINs from both network dynamics and population dynamics perspectives. Some immune network models can be contributed to this antibody dynamics such as [11]. Such antigen-antibody interactions can be a long sequence structure, namely, an antibody chain, which is formerly defined as follows.

**Definition 1.** For an idiotypic immune network  $\langle \{LU_i\}_{i=1}^N, M \rangle$ , an antibody chain  $AC = \{Ab_1, Ab_2, \dots, Ab_N\}$  is defined as follows.

- $Ab_i \in LU_i$ , for all  $i = 1, 2, \dots, N$ .
- $Ab_i$  can recognize  $Ab_{i+1}$ , namely,  $Ab_i \rightarrow Ab_{i+1}$ , for all  $i = 1, 2, \dots, N-1$ .
- $Ab_N \rightarrow Ab_1$ .

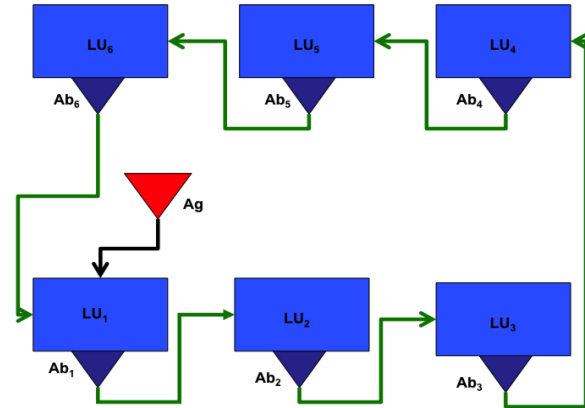


Figure 1. Schematic diagram of a closed idiotypic immune network.

The final element of the antibody chain  $Ab_N$  interests us. It can stimulate the antibody  $Ab_1$  even  $Ag$  is eliminated. In this way, population of  $Ab_1$  will be stable in small amount. In particular, its role in immune memory can be analyzed by network dynamics.

### C. Immune Memory

An immune system will react rapidly to the same or similar antigens which had invaded the same human body before. This phenomenon implies that immune system can "memorize" associatively the formations of previously invaded antigens. An evidence for immune memory is that it is strongly affected by the populations of soluble antibodies in the blood. Therefore, some variable quantified the immune memory can be correlated to the amount of antibodies. In fact, this is a major inspiration for this research, but different angle from computational biology.

The immune memory mechanism is not fully understood so far; according to Smith et al. [3], at the end of an immune response, when the antigen is cleared, the B cell population decreases, leaving a persistent sub-population of memory cells. The newer view of memory cells is that they are not intrinsically longer-lived than virgin cells, but that memory depends instead on the persistence of antigen [12]. However, some researches, especially those related to immune network theory, imply the immune memory is formed by a cyclic idiotypic network rather than specific memory cell [11]. Immunologists have discovered the vaccination mechanism for human immune systems for a long time. This takes advantages of so-called the associative memory of immune systems. The associative memory mechanism can be explained as follows [13]. If the secondary antigen is "similar" to the primary one, the set of antibodies activated by this antigen will overlap with the one activated by the primary antigen. The similarity between antigens can be defined by affinity of molecules. This memory is able to store and recall patterns when immune systems need. Associative recall is a general phenomenon of immune memory [3].

### III. MAIN RESULTS

#### A. Tree Structure of Idiotypic Immune Network

Perelson [14] suggested a tree structure with varied levels for idiotypic immune networks.  $Ag$  is the root node. Level 1 is the collection of all antibodies which are complementarily matching with  $Ag$ . Level 2 is the collection of all antibodies which are complementarily with those at level 1. This way, we may construct a tree structure for antibody chain (see Figure 2). We also define the directed edge  $\rightarrow$  from antibody at higher level to the one at lower level. Based on this structure  $T = \{A_g, AC, \rightarrow\}$ , all possible internal images for a given antigen can be searched within some antibody populations. We also observe the existence of internal image can be guaranteed if and only if second level is inward and outward according to some affinity relation  $\rightarrow$ . The following algorithm is the detailed description for the tree constructions for antibody chains.

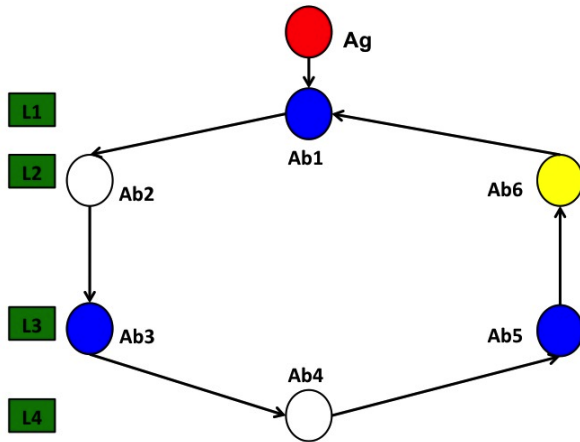


Figure 2. Tree Structure for an Antibody Chain ( $N=6$ ).

**Proposition 1.** The tree structure of idiotypic immune network  $T = \{A_g, AC, \rightarrow_\varepsilon\}$  satisfies the following properties:

- (1) Each antibody in  $AC$  can be located at most one level;
- (2) The edge can only exist between adjacent level;
- (3) If  $AC$  is closed, then its length is even;
- (4) If  $AC$  is closed, then number of level of  $T$  is equal to  $N/2 + 1$ .
- (5) Internal image must locate at the level 2.
- (6) The internal image exists if and only if  $Level(AC) = N/2 + 1$ .

This is obvious by definition of tree structure. If  $Ab_i \in AC$  is located to Level  $i$  of  $T$  and  $Ab_i \rightarrow Ab_j$ . Then  $Ab_j$  has to be located to Level 2. If  $Ab_i \rightarrow Ab_j$ , then  $Ab_j$  must be located at level  $i + 1$ . If the length of the left  $AC$  is equal to  $m$ , then the number of level is also  $m$  by the closeness of  $AC$ . The right  $AC$  is  $N - m$ , which is also even. However,  $m$  and  $N - m$  must have difference equal

to 2 by the closeness of  $AC$  and the fact that  $Ab_N$  must be located at level 2. Therefore the length of  $AC$  is equal to  $m + (m - 2) = 2m - 2$ , which must be even. According to (3),  $N$  must be even.  $Ab_N$  and  $Ab_2$  are at the level 2;  $Ab_{N-1}$  and  $Ab_3$  are at the level 3; continuing in this way,  $Ab_{N/2+1}$  is the unique antibody located the level  $N/2 + 1$ . If  $Level(AC) < N/2 + 1$  then  $length(AC_j) < N/2 + 1$ . Therefore,  $length(AC) \leq N/2 + N/2 - 1 = N - 1$ , which is a contradiction. Therefore  $Level(AC) = N/2$ .

If an internal image is highly complementary match with  $Ab_1$ , then it can be regarded as an reasonable stimulus for  $Ab_1$  even  $Ag$  is eliminated.  $Ab_1$  plays a pivotal role for internal image. If the  $\varepsilon$  of the  $\varepsilon$ -complementary match between  $Ab_1$  and  $Ag$  is higher, then the internal image and  $Ag$  are more similar. From mathematical viewpoint, it is still a question that  $Ag$  and its internal image may be of low similarity.

The network-layered structure for CIINs provide some evidences that the population dynamics cannot provide, such as recall process of immune memory. Every closed antibody chain  $T(AC)$  can be represented as a  $k$ -level tree structure, where  $k = N/2 + 1$ . Furthermore, internal image  $Ab_N$  must locate at level 2.

**Theorem 1.** For idiotypic immune network  $\rightarrow$  as the relation of "being recognized", the internal image of an antigen exists if and only if

- The tree structure of  $AC$  one antibody path  $P_1$  with length  $N/2 + 1$  and the other one  $N/2 - 1$ , for some even integer  $N$ .
- $P_1(end) \rightarrow P_2(1)$ .
- $P_2(end)$  is an internal image.

#### B. Network Dynamics based on Autocorrelation Matrix

From immunology viewpoints, a better antibody chain has the following two aspects: the first is that it will respond fast to the second-time antigenic invasion; the second one is that it will act adaptively to the similar antigenic invasion. The question is how to build some mathematical model to connect the concepts of tree structure with the dynamics of immune responses such as immune memory and recall. Most importantly, response to a similar antigen rather than second. This is related to the associative memory mechanism [15] [16].

The attractor networks proposed by Morita [17] and defined by tree structures can generate a dynamics. It is a transition of states represented by nodes of antigens. A network dynamics  $F$  can be defined on states of network  $S$  as follows. An antigen will induce an antibody chain, which will generate a network dynamics for interpreting immune

memory formation and recall process. Moreover, this attractor network dynamics can interpret the associative property of immune memory while population dynamics cannot. Let  $F$  be a function defined on state space  $S$ . For a given state  $S(t)$  at time  $t$ , the network dynamics can be defined as a transition of state  $S$  from time  $t$  to time  $t+1$ , namely,  $S(t+1)=F(S(t))$ . According to the basic concept of dynamical systems, we may define an attractor  $\mu$  of the network dynamics  $F$ , if there exists a state  $S$  such that  $F^n \rightarrow \mu$ , as  $n \rightarrow \infty$ , where  $t$  is a temporal variable and  $S$  is the state variable. Now we define the associative memory based on stability analysis of equilibrium points (states) of such network dynamics.

**Definition 2 (Associative Memory)** For an immune network  $\{AC, F\}$ , if there exists a set of  $p$  antigenic patterns,  $\{\mu^1, \mu^2, \dots, \mu^p\}$  which are attractors of the network dynamics  $F$ , then we say the immune network can memorize  $p$  patterns of antigens associatively.

The stable states of immune networks can guarantee that if an antigenic format fallen into the basin of attraction of  $\mu_k$ , then by memory recalling process, this antigen can invoke the same antibody proliferation immediately. The following question arises: what is the mathematical model of network dynamics  $F$ , if an antigen-antibody chain  $(Ag, AC)$  is given. We will first concentrate on some binary-valued function  $F$  to illustrate the dynamics of immune memory, rather than on discussing some complex and high-dimensional dynamical systems.

This mechanism is named autocorrelation matrix memory (ACMM) inspired by [17]. For an antigen  $Ag$ , there exists an antibody chain  $AC$  such that it is a CIIN. This  $AC$  will induce a autocorrelation matrix  $W$  as

$$W = \frac{1}{N-1} \sum_{i=1}^{N-1} Ab_i^T * Ab_{i+1} \quad (1)$$

Given a pattern  $S$  such as an antigen as an initial state  $X(0)$ , the network dynamics of  $w$  is given by  $X(t+1) = sgn(W \cdot X(t))$ , where  $sgn(x_i(t)) = 1$ , if  $x_i(t) > 0$ ;  $sgn(x_i(t)) = -1$ , if  $x_i(t) < 0$ . If  $X(t) \rightarrow S$ , as  $t \rightarrow \infty$ , then we call  $S$  a memory format (MF). Therefore, an antigen activates an antibody chain whose network dynamics can generate antigenic memory format.

*C. Associative Memory formed by the Immune Systems*

How immune memory pertains associativity is still a mystery for scientists even for the advances and success of vaccinations for more than 300 years. However, we propose a network dynamics model to describe this mechanism from computational immunology viewpoints. An immune network  $\langle \{LU_i\}_{i=1}^N, M \rangle$  is equipped with with associative memory, if for any  $\epsilon > 0$ , there exists  $\alpha > 0$ , such that whenever a new antigen  $Ag'$  with  $d(Ag, Ag') < \alpha$  implies that  $Ag' \rightarrow_{\epsilon} Ab_1$ .

If a new antigen  $Ag'$ , which is very similar to the previous invaded antigen  $Ag$ , invades human body, then the antibody  $Ab_1$  which binds  $Ag$  can also bind  $Ag'$  and invokes another immune response.

For an antigen  $Ag$  activating the antibody chain  $AC$ , the network dynamics defined on pattern space for memory formation and recall with  $s(t+1) = W \cdot s(t)$  must induce attractors, where  $W$  is some state transition matrix. This means, starting from an initial configuration  $s$  which is sufficiently closed to (or overlapped with) one  $Ab_i$ , the system will flow to a fixed point of the dynamics, which is either the pattern itself or the configuration with high overlap with that pattern.

*Memory Recall Process based on Network Dynamics*

Once the same or "similar" antigen to  $Ag$  invades immune systems again, the memory recall process of the immune systems will be activated by comparing the memory format generated and stored in the previous antigenic invasion of  $Ag$ . From system dynamics, we can define such similar antigens by basin of attraction. Suppose the same antigen  $Ag$  invades the immune systems again, then network dynamics  $F$ , namely,  $F^n(Ag)$  will immediately converges to memory format of  $Ag$ , say  $M(Ag)$ . On the other hand, if some similar antigen  $Ag'$  invades, the network dynamics  $F^k(Ag')$  converges to the same  $M(Ag)$  if  $Ag' \in BA(Ag)$ . Therefore, basin of attraction of  $Ag$  is the major criterion that  $Ag'$  will activate the  $Ab_1$  and the same antibody chain.

The third question we are interested here is the following. For a similar antigen  $Ag'$ , whether the original antibody chain  $AC$  activated by  $Ag$  can also produce immune response effectively to this mutated  $Ag'$ ? In this way, the immune network reflects a decent associative memory. If the internal image  $Ab_N$  is similar to  $Ag$ , then it is also similar to  $Ag'$ . If for initial condition  $X(0)$ , we have the following network dynamics of recall process inspired by [16] and [18].

$$X(i) = sgn\left(\frac{1}{N-1} \sum_{i=1}^{N-1} (Ab_i^T X(i-1)) Ab_{i+1}\right) \quad (2)$$

Therefore, if  $X(0) = Ag$  and  $\{Ab_k\}_k$  is a set of antibodies with sufficiently high dimension, then  $Ab_1$  is recall at  $t=1$ , as  $X(1) = Ab_1$ . In the same way,  $X(2) = Ab_2$ ,  $X(3) = Ab_3$ , or the stored antibody sequence of  $Ag$  is recalled. Once the same antigen  $Ag$  invades,  $Ab_1, Ab_2, \dots, Ab_N$  are activated successively. On the other hand, If a similar antigen  $Ag'$  invades, then the recall process can be computed as follows.

$$X(1) = sgn\left(\frac{1}{N} \sum_{i=0}^{N-1} (Ab_i^T Ag') Ab_{i+1}\right) \approx (Ag \cdot Ag') Ab_1$$

$$= \alpha_1 \cdot Ab_1 \quad (3)$$

where  $\alpha < 1$  is a small positive number. In the similar fashion,  $X(k) = \alpha_k \cdot Ab_k$ . Therefore this similar antigen  $Ag'$  will activate antibodies  $\alpha_1 Ab_1, \alpha_2 Ab_2, \dots, \alpha_N Ab_N$ .

Based on (3), there exists  $\lambda > 0$  such that  $d(Ag, Ag') < \lambda$  implies that  $d(F^k(Ag'), Ab_k) < \lambda_k$ . If  $\lambda_k \approx 1/k$ , then this network dynamics can guarantee that  $F^k(Ag')$  is approximately equal to internal image  $Ab_N$ . Therefore, CIINs, from computation viewpoints, is stable for memory recall. However, we are interested in the critical value for  $\lambda$  such that network dynamics will deviate the antibody chain.

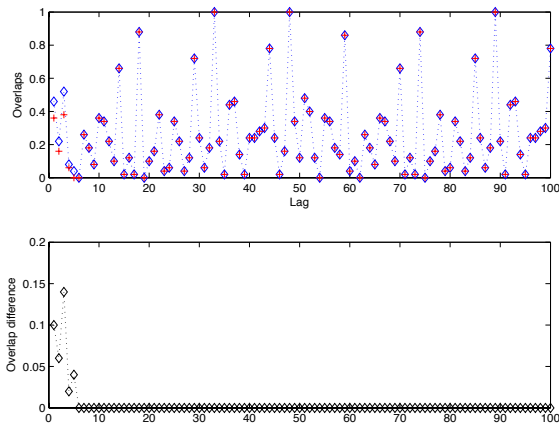


Figure 3. Stable Network Dynamics for Immune Memory Recall.

Figure 3 is a simulation that a particular network dynamics activated by a specific antibody chain could recall a mutated antigen completely. The parameters are set as follows.  $N=100$ , the antibody population is 20,000. The length of (randomly generated) antibody chain is  $N=4$ .

#### IV. DISCUSSIONS

Overlapping difference function shown in Figure 3 is stable. However, it is very often that such function is unstable for varied parameter values, for example,  $\varepsilon < 0.6$ . In fact, simulations have shown different unstable behaviors, from simple to complex ones. Anyway, Figure 3 shows that the mutated antigen  $Ag'$  can invoke the exact memory for the original antigen  $Ag$ ; and antibody  $Ab_1$  can eliminate this  $Ag'$ . Another issue is the existence of CIINs for invaded antigen  $Ag$ . If the affinity threshold  $\varepsilon \geq 0.6$ , the simulations have also shown that it is generally difficult to form CIINs.

#### V. CONCLUSION AND FUTURE WORK

We proposed an immune memory mechanism, based on the closed idiotypic immune network. The latter is analyzed by some tree structure. The formation of immune memory can be deduced by close loop of the cell's interactions, in particular, by antibody dynamics such as internal image

recognition. Network dynamics which describes the immune memory formation and recall process is modeled based on cross-reactive correlation matrix of antibody chains. This mechanism is associative by analyzing the state transitions of mutated antigens.

There are some issues for future research. As mentioned in Section VI, observation leads to the research of adopting weak connection principle to CIINs forming. This is inspired by the small-world network structure; it is reasonable that antibody-antigen binding should be based on high affinity threshold while lower affinity threshold for the antibody-antibody binding.

#### ACKNOWLEDGMENT

The authors would like to thank support from National Science Council of Taiwan under the grant number NSC-100-2221-E-424-003 and NSC-101-2221-E-424-006.

#### REFERENCES

- [1] A. de Castro, "A network model for clonal differentiation and immune memory," *Physica A*, vol. 355, 2005, pp. 408–426,
- [2] N. Jerne, "Towards a network theory of the immune system," *Ann. Immunol. (Inst. Pasteur)*, vol. 125C, 1974, p. 373.
- [3] D. Smith, S. Forrest, and A. Perelson, "Immunological memory is associative," *Artificial Immune Systems and Their Applications, The international Conference on the Multi-Agent Systems Workshop Notes*, Kyoto, 1996, pp. 62–70.
- [4] J. Fernandez, G. Acostra, and M. Mayosky, "From network-to-antibody robustness in a bio-inspired immune system," *BioSystems*, vol. 104, 2011, pp. 109–117.
- [5] C. Bona, "Internal image concept revisited," *Proceeding of the Society for Experimental Biology and Medicine*, vol. 213, 1996, pp. 32–42.
- [6] G. Parisi, "A simple model for the immune network," *Proceedings of the National Academy of Sciences*, vol. 87, 1990, pp. 429–433.
- [7] A. Barra and E. Agliari, "Stochastic dynamics for idiotypic immune networks," *Physica A*, Vol.389, 2010, pp. 5903–5911.
- [8] S.-I. Amari and K. Maginu, "Statistical neurodynamics of associative memory," *Neural Networks*, vol. 1, 1988, pp. 63–73.
- [9] F. Abbattista, G. D. Gioia, G. D. Santo, and A. Fanelli, "An associative memory based on the immune networks," in *Proceedings of the IEEE International Conference on Neural Networks*, 1996, pp. 519–523.
- [10] J. Farmer, N. Packard, and A. Perelson, "The immune system, adaptation, and machine learning," *Physica*, vol. 22D, 1986, pp. 187–204.
- [11] T. Sonoda, "Formation and stability of a memory state in the immune network," *J. Physical Society of Japan*, vol. 81(4), 1992, pp. 1408–1424.
- [12] P. Matzinger, "Memories are made of this ?" *Nature*, vol. 369 (23 June) 1994, pp. 605–606.
- [13] B. Borowik, B. Borowik, J. Kucwaj, and S. Laird, "Associative memory of artificial immune systems," *Annales UMCS Informatica AI X*, vol. 2, 2010, pp. 111–122.
- [14] A. S. Perelson, "Immune network theory," *Immunological Reviews*, vol. 10, 1989, pp. 5–36.
- [15] H. Gutfreund, *The Physics of Neural Network: Spin Glasses and Biology*, ed. D. Steinl. World Scientific Publishing, 1992.
- [16] M. Morita, "Associative memory with nonmonotone dynamics," *Neural Networks*, vol. 6, 1993, pp. 115–126.

- [17] T. Kohonen, *Self-Organization and Associative Memory*. Springer-Verlag, 1983.
- [18] M. Oda and H. Miyajima, "Autocorrelation associative memory using refractory period of neurons," *Electronics and Communications in Japan, Part 2*, vol. 87(8),2004, pp. 41–52.

# High Performance Grid Environment for Parallel Multiple Biological Sequence Alignment

Plamenka Borovska  
Computer Systems Department  
Technical University of Sofia  
Sofia, Bulgaria  
pborovska@tu-sofia.bg

Veska Gancheva  
Programming and Computer  
Technologies Department  
Technical University of Sofia  
Sofia, Bulgaria  
vgan@tu-sofia.bg

Nikolay Landzhev  
Computer Systems Department  
Technical University of Sofia  
Sofia, Bulgaria  
landzhev@yahoo.com

**Abstract**— The parallel in silico simulations based on methods and algorithms for analysis of biological data using high-performance distributed computing is essential for accelerating the research and reduce the investment. This paper presents a high-performance Grid environment integrating various services and middleware to facilitate access to distributed resources for carrying out scientific experiments in the area of bioinformatics. This environment enables parallel computer simulations increasing the efficiency of the computations and allowing scientists easy and user friendly access. Web portal provides as services access and execution of parallel program implementation based on an algorithm for comparative analysis of biological data. An innovative algorithm called MSA\_BG for multiple sequences alignment based on the concept of Artificial Bee Colony metaheuristics is designed. Experimental simulations on the basis of parallel implementation of MSA\_BG algorithm for multiple sequences alignment through the high-performance Grid environment have been carried out for the case study of the influenza virus variability.

**Keywords**—artificial bee colony; bioinformatics; GRID computing; multiple sequence alignment; portal.

## I. INTRODUCTION

In silico biological sequence processing is a key investigation in the area of bioinformatics. This scientific area requires powerful computing resources for exploring large sets of biological data. The huge amount of biological sequences accumulated in the world nucleotide and protein databases requires efficient parallel tools for structural genomic and functional analysis. The parallel implementation of methods and algorithms for analysis of biological data using high-performance distributed computing is essential.

Multiple sequences alignment involves more than two biological sequences, generally protein, DNA, or RNA [1]. Multiple sequence alignment is computationally difficult and is classified as a NP-Hard problem [2][3][4]. ClustalW is a widely used multiple sequence alignment algorithms for Deoxyribonucleic acid (DNA) or proteins and implements a progressive method for multiple sequence alignment [5]. It calculates the best match for the selected sequences and lines them up so that the identities, similarities and differences can be seen. The basic algorithm behind ClustalW proceeds in

three stages: pairwise alignment (PA), guide tree (GT) and multiple alignment (MA). The ClustalW phases are relatively independent. Each of the phases produces intermediate data which is used as an input for the next one, but the calculations are independent.

Several parallel ClustalW algorithms for multiple sequence alignment have been reported in recent years. ClustalW-MPI is a distributed and parallel implementation on distributed computer clusters and on the traditional parallel computers and uses a scheduling strategy called fixed-size chunking where batches of tasks of one fixed size are to be allocated to available processors [6]. The algorithm ClustalW\_MPI is not so effective in the case of highly parallel implementations. The analysis of the ClustalW\_MPI algorithm show that it has not regular structure, does not consist of replica modules and is characterized by significant computational imbalance, because it is based on the parallel algorithmic paradigm “master-slave” [7]. Increasing input file size is limited due to allocated memory size (Juqueen’s memory allocation per core is 1GB [8]). The maximum number of input sequences is approximately 10000 (depending on the sequences length).

The aim of our work is to provide a generic high level user friendly heterogeneous high-performance grid based environment for massively parallel in silico biosimulations. The solution allows scientists easy customization, parameterization, execution and analysis of complex simulation scenarios utilizing bioinformatics tools.

An innovative parallel algorithm MSA\_BG for multiple sequences alignment based on the concept of Artificial Bee Colony (ABC) metaheuristics and the concept of algorithmic and architectural spaces correlation is designed.

The ABC algorithm is an optimization algorithm based on the intelligent foraging behaviour of honey bee swarm [9]. In the ABC model, the colony consists of three groups of bees: employed bees, onlookers and scouts.

The choice of the ABC algorithm is based on the fact that, in essence, it is a hybrid metaheuristics - a combination of methods based on populations (scouts generate a number of possible solutions simultaneously) and a method based on trajectories (employed bees perform the local searches around the decisions of the scouts, seeking to improve the decisions quality).



This paper is structured as follows. Section II explains the scientific application gateway. The design and the algorithmic framework of the algorithm MSA\_BG are explained in Section III. Section IV presents the experimental framework. The experiments, performance evaluation and results analysis are discussed in Section V. We present the conclusion and future work aspects in Section VI.

## II. SCIENTIFIC APPLICATION SPECIFIC GATEWAY

The objective of an academic scientific portal gateway is to allow a large number of users to have transparent access to available distributed advanced computing infrastructure, software tools, visualization tools and resources in the area of High Performance Computing (HPC).

The portal provides user-centric view of all available services and distributed resources. The Web-based environment is organized to provide a common user interfaces and services that securely access available heterogeneous computing resources, data and applications. It also allows many academic and scientific users to interact with these resources.

In order to achieve these requirements, the execution of computing tasks on these resources needs to be managed. Therefore, resource management and security services are needed. The required execution management can be described in terms of three classes of services: Resources, Job Management and Monitoring, and Resource selection. End-users (e-scientists) will be able to: enable the access to the various components of the e-science infrastructure via specific, user community views; execute the published applications with custom input parameters by creating application instances using the published applications as templates; store and share applications; observe and control application execution; support user's collaboration via sharing applications and databases. Application developers will be able to develop applications by the portal (application porting) and publish the completed applications into the infrastructure in order to end-users to use them. Portal developers will be able to develop user interfaces as portlets for end-users of the portal and associate the user interfaces to the portal using an API (Application Programming Interface).

The portal provides a customized application specific gateway, and can be summarized as follows:

- Identifying and clarifying the specific requirements, scenarios and real needs of target groups such as scientists, researchers, educators and students.
- Defining of specific views for the user community according to their requirements and development of user scenarios for the specific applications.
- Identifying software packages, testing these packages and selecting the suitable ones.
- Creating the relevant biological database and data extracting.
- Designing and developing application-specific portlets corresponding to high-level user scenarios in

order to execute the user scenarios and to allow users to define their own input parameters, execute and monitor the experiments, gather the results and visualize them.

- Experimental studying and evaluating of the specific application portlets and defining the usability of portlets.

Depending on the user community needs, the scientific gateway may provide any of the following features:

- High-performance computation resources (local clusters).
- General or domain-specific analytic and visualization software.
- Collaborative interfaces for sharing interfaces and database.
- Job submission and monitoring tools and performance analysis tools.
- Specific user community views.

The major user scenarios can be summarized as follows:

- Supporting users with uniform access to heterogeneous distributed high-performance computation resources such as local clusters.
- Storing, sharing and parameterising of ready-to-use applications.
- Monitoring and controlling of the application execution, collecting and visualizing the results.
- File transfer (transfer based on user request and transfer based on subscription/notification).
- Accessing collaborative tools for data management, search tools, and analysis tools.
- Running new Distributed Computing Infrastructure (DCI) applications based on the developed gateway.
- Enable the integration of other services.
- Enable the store and share of half-made applications, application templates.

The basic contributions of the portal are as follows:

- Developing of application-specific gateway.
- Establishing and providing a production service for the user communities in the area of high performance computing.
- Increasing the number of specific portlets and applications in the portlet and application repositories, intended for user communities.
- Support user's community with uniform access to high-performance computation resources such as local clusters.

The proposed high-performance Grid-based infrastructure as services for carrying out remote simulations consists of HPC resources, learning management system, resource database, knowledge resources as services, software tools and web portal (Fig. 1). The Grid environment is based on the guse/ws-pgrade system [10]. Building the portal enables users to manage their tasks through the interface and administrators to easily manage users.

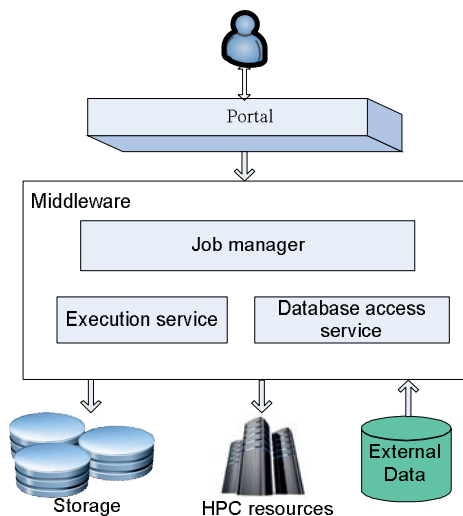


Figure 1. High performance infrastructure.

The web portal provides also: user profile management, e.g., different views for different user; personalized access to information, software tools and processes; getting information from local or remote data sources, e.g., from databases, transaction systems, or remote web sites; aggregated the information into composite pages to provide information to users in a compact and easily consumable form. In addition, the portal also includes applications like execution of some software tools, etc.

### III. MULTIPLE SEQUENCE ALIGNMENT ALGORITHM MSA\_BG BASED ON ARTIFICIAL BEE COLONY

The concept is to synthesize a parallel iterative algorithm with a regular computational and communication system based on data parallelisms and replica code, which is executed on all computing nodes. The parallel paradigm is Single Program Multiple Data (SPMD) and data decomposition. The granularity is hybrid - coarse granule computing for each node (multithreaded process) that runs multithreading (fine granule) of the cores within the computing node. In the case of hybrid granularity in order to effectively use the resources of supercomputers is appropriate to use hybrid parallel implementations.

A new parallel algorithm MSA\_BG, in which each process simulates the behaviour of a beehive and each hive contains multiple swarms is proposed. The parallel algorithm is designed based on the methodology for the synthesis of parallel algorithms, which is based on the correlation of the parameters of the algorithmic and architectural spaces. The threads simulate the behaviour of many bees in the swarm.

Scout bees in the swarms round certain subregions in the searching space and construct a potential solution. Once the scout bees obtain possible (feasible) solution, they return to the hive and begin to dance in the hive.

Onlookers watch the dance of employed bees, choose one of their solutions and evaluate them.

The employed bees select one of the solutions and make attempts to improve it based on local search. The quality of obtained solutions is determined by the grade of sequences similarity. As much as this evaluation is greater, the quality of obtained alignment is higher, i.e., the criterion of optimality is a maximum similarity score.

In case of higher quality of solution generated by a scout the possibilities to include it in the list of elite solutions increased. As next step this solution will be improved based on a local searching by the employed bees.

#### A. Metaphor

The food source is presented by possible sequence alignments. The quality of solutions (nectar amount) is evaluation function of the similarity between the sequences for a particular alignment.

The proposed algorithm is based on the concept of swarms and hives. The allocation of computing supercomputer resources is as follows. The entire supersystem simulates the behaviour of a colony of beehives, and the number of hives is equal to the number of computing nodes. Each computing node simulates the behaviour of a hive. In a hive are included  $q$  swarms, where  $q$  is the segments number of the supersystem. The swarms within a hive work on common lists of best temporary solutions and elite solutions. Each hive has a queen bee, which gets the elite decision of all swarms on the hive. Scout 0 reads the sequences from the memory and stores it in sharing memory of the computational node (hive). Scouts generate initial solutions thought sequence alignment including gaps. Random number generator Mersenne Twister is used. Scouts evaluate quality of the alignment using the following method.

An assessment by columns is done – in case of nucleotide sequences the numbers of symbols – A, G, C and T are counted. The numbers of symbols are compared and are selected the nucleotides, which occur mostly in the different columns. It is called “nucleotide-favourite” sequence ( $f_{ij}$ ). After the calculation of assessments in columns is formed the so-called sequence-favourite contained in each position the respective favourite in the column, i.e., the nucleotides form the so-called sequence-favourite. Compared sequences along with the sequence-favourite form the so-called working set sequences.

A scoring matrix is built up where for sequences (rows in the matrix) are stored (by columns) the values of the evaluation function  $S$ . For sequence (row)  $i$  in position  $j$  (column): nucleotide  $a_{ij}$  and nucleotide-favourite  $f_{ij}$ :

$$S_{ij} = 0 \text{ in case of } a_{ij} = \text{gap}$$

$$S_{ij} = 1 \text{ in case of } a_{ij} = f_{ij}$$

$$S_{ij} = -1 \text{ in case of } a_{ij} \neq f_{ij}$$

The elements of the scoring matrix are calculated for each sequence  $S_{p,n+1}$ , where  $n$  is the sequences length and  $p$  – the number of aligned sequences.

The overall assessment of the quality of alignment  $S$  is formed as:

$$S = \sum_{i=1}^p \sum_{j=1}^n S_{i,j} \tag{1}$$

The additional column (similarity counters) of the scoring matrix  $S_{n,n+1}$  consists of the similarity evaluation for each sequence compared with the sequence-favourite, i.e.,

$$S_{i,n+1} = \sum_{j=1}^n S_{i,j} \quad i \in \{1,p\} \tag{2}$$

The next example shows the working set. The sequence-favourite is marked in red.

<b>A</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>T</b>
A	A	T	C	G	A	T
A	G	T	C	A	T	T
A	G	-	G	A	A	G

The next example illustrates the calculation of scoring matrix  $S$ . The scoring column of the counters is marked in red and consists similarity scores for each sequence and the sequence-favourite.

1	-1	1	1	-1	1	1	<b>3</b>
1	1	1	1	1	-1	1	<b>6</b>
1	1	0	-1	1	1	-1	<b>2</b>

The scouts write down the scores of the working sets in the list of working solutions that is sorted in descending order by the total alignment scores. For each sequence (row) is generated dynamic data structure containing the indexes of the gaps in the sequences.

The onlookers select the best quality working set and perform local search – make minor changes in the working set and evaluate the quality of the modified alignment. In case of quality improvement, the modification is accepted and the working set is stored in the list of "best temporary solutions". Otherwise, the new alignment is ignored.

The following approach for modification of the aligned working set of sequences is proposed. The column with counters of the scoring matrix  $S$  is reviewed and is selected the row (sequence) with the lowest counter value (sequence that differs most from the favourite). Using a random generator is selected an index for insertion a gap (INS) and an index for deletion a gap (DEL) from the list of empty positions (to keep the length of the sequence) by obligatorily  $INS \neq DEL$ . Both indexes are compared. If  $DEL > INS$ , then all characters in positions between INS and DEL are shifted one position to the right (shift\_right), otherwise if  $DEL < INS$  – are shifted one position to the left (shift\_left).

An example for the sequence is shown below:

A-TGC-GGTA-CCGT-G

The list of gaps indexes is: 1, 5, 10, 15.

Based on random generator, the position for insertion  $INS=4$  and the position for deletion  $DEL=15$  are selected. In this case,  $DEL > INS$ , mandatory  $DEL \neq INS$ , and the operation is shift\_right.

A-TG**C**-GGTA-CCGT-G

A-TG-**C**-GGTA-CCGTG

The value of similarity counter for sequence compared with the sequence-favourite is calculated in case of improving the modification of the working set is accepted and is stored in the list of temporary best solutions; otherwise, the modification is ignored.

In case of inserting position  $INS=12$ , and deleting position  $DEL=5$ ,  $DEL < INS$  and the operation is shift\_left.

A-TGC-GGTA-CCGT-G

A-TGCGGTA-C-CCGT-G

After number of modifications, employed bees suspend the processing of the current working set and write down the best solution in the list of elite solutions that is sorted in descending order.

The condition for termination of the parallel algorithm is the number of iterations; then, the mother bee shall inform the queen bee of the colony and sends it the quality of the best solution (elite solution).

### B. Algorithmic framework for parallel multiple alignment

The algorithmic framework for parallel multiple alignment of biological sequences MSA\_BG on the basis of ABC algorithm is shown in Fig. 2.

```

Parallel For
For m=1, Q // For every hive
For r=1, q // For every swarm (node)
Input_sequences //Scout 0
Parallel Sections
Parallel For k=0,15 /*Scout_bees
Generate_Random_Alignment
Evaluate_Score_By Columns
Construct_Favorite_Sequence
Evaluate_Column of Counters
Save_Working_Set_in_Optimization list
End Parallel For
Parallel For k=0,15 /*Onlookers_bees
Select_Random_Working_Set_Out_of_Optimization_List
Select_Sequence_of_MIN_Counter
Select_INS(Random)
Select_DEL, such that DEL≠INS (RANDOM)
If DEL>INS then shift_right else shift_left
Evaluate_Column of Counters
If higher_quality then Save_In_Optimization_List
else skip
End Parallel For
End Parallel For
Sort_Optimization_List_by_Quality // Scout 0
Reduce_Elite_Alignment_to_Mother_of_Hive // Scout 0
Reduce_MAX_Elite_Alignment_of_Hive //Mother of Hive
End Parallel For
Reduce_MAX_Elite_Alignment_of_Colony //Queen of Colony
Output_Best_Alignment_Obtained_So_Far //Queen of Colony
    
```

Figure 2. Algorithmic framework for parallel multiple alignment of biological sequences MSA\_BG on the basic of ABC algorithm.

The conceptual model of the MSA\_BG method for parallel multiple alignment of biological sequences on the basis of ABC algorithm is shown in Fig. 3.

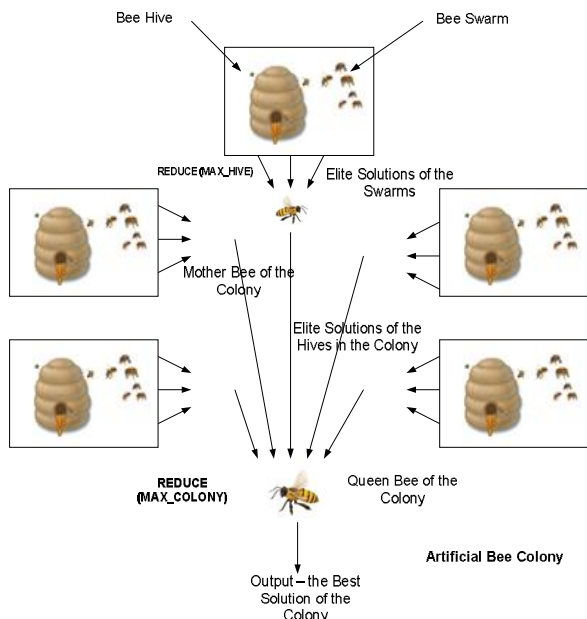


Figure 3. Conceptual model of the MSA\_BG method for parallel multiple alignment of biological sequences on the basis of ABC algorithm.

The colony queen through collective communication reduction and operation MAX gets the quality of the elite decisions by hives mother bees. Thereafter the colony queen determines and sends the best one.

IV. EXPERIMENTAL FRAMEWORK

The experimental framework is based on a heterogeneous computer cluster of 10 nodes including 8 servers AMD Opteron 64 Dual Core Processor 1.8 GHz, RAM: 2GB 800 MHz, HDD: 2x160GB Hitachi SATA in RAID 0 and 2 servers CPU: 2x Intel Xeon E5405 Quad Core Processor 2 GHz, RAM: 4GB 800MHz, HDD: 2x 146 GB Hitachi 10000 RPM in RAID 0. All nodes are interconnected via gigabit Ethernet switch. The operating system is 64 bit Scientific Linux 5.3. Message passing is based on the MPICH2 1.1.1p2 distribution of the MPI standard. The resource manager is SLURM.

V. PARALLEL PERFORMANCE EVALUATION AND RESULTS ANALYSIS

The objective of the experiments is to estimate experimentally parallel performance parameters and make profiling of the program implementation on the basis of MSA\_BG algorithm. Similarity searching between Ribonucleic acid (RNA) segments of various influenza viruses A/H1N1strains obtained from Genbank [11] has been carried out based on the parallel program MPI-based implementation of MSA\_BG algorithm for multiple sequence alignment on heterogeneous distributed high-performance resources.

The experiments are conducted on grid infrastructure using various numbers of cores and various numbers of iterations. Experimental results are shown in Table I.

TABLE I. EXECUTION TIME OF MSA\_BG ALGORITHM USING VARIOUS NUMBERS OF CORES AND VARIOUS NUMBERS OF ITERATIONS

Cores	Execution time				
	Iterations				
	1000	10000	100000	1000000	10000000
1	4 sec	41 sec	6,42 min	62 min	624 min
2	2 sec	20 sec	3,22 min	31 min	312 min
8	1 sec	6 sec	0,51 min	8 min	82 min
16	1 sec	3 sec	0,26 min	4 min	42 min
28	1 sec	2 sec	0,15 min	2 min	24 min

The speedup is evaluated as a ratio of the sequential execution time to the parallel execution time. The experimental results for the speedup of MSA\_BG algorithm on using various numbers of cores and various numbers of iterations are shown in Table II and Fig. 4.

The efficiency is evaluated as a ratio of the achieved speedup to the number of cores. The results for the efficiency of the MSA\_BG algorithm gains in the case of various number of cores and number of iterations are shown in Table III and Fig. 5.

TABLE II. SPEEDUP OF MSA\_BG ALGORITHM USING VARIOUS NUMBERS OF CORES AND VARIOUS NUMBERS OF ITERATIONS

Cores	Speedup				
	Iterations				
	1000	10000	100000	1000000	10000000
2	2	2,05	1,99	2,00	2,00
8	4	6,83	12,59	7,75	7,61
16	4	13,67	24,69	15,50	14,86
28	4	20,50	42,80	31,00	26,00

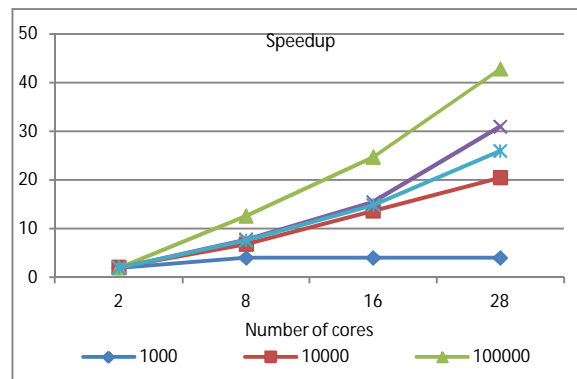


Figure 4. Speedup of MSA\_BG algorithm using various numbers of cores and various numbers of iterations.

TABLE III. EFFICIENCY OF MSA\_BG ALGORITHM USING VARIOUS NUMBERS OF CORES AND VARIOUS NUMBERS OF ITERATIONS

Cores	Efficiency				
	Iterations				
	1000	10000	100000	1000000	10000000
2	1,00	1,03	1,00	1,00	1,00
8	0,50	0,85	1,57	0,97	0,95
16	0,25	0,85	1,54	0,97	0,93
28	0,14	0,73	1,53	1,11	0,93

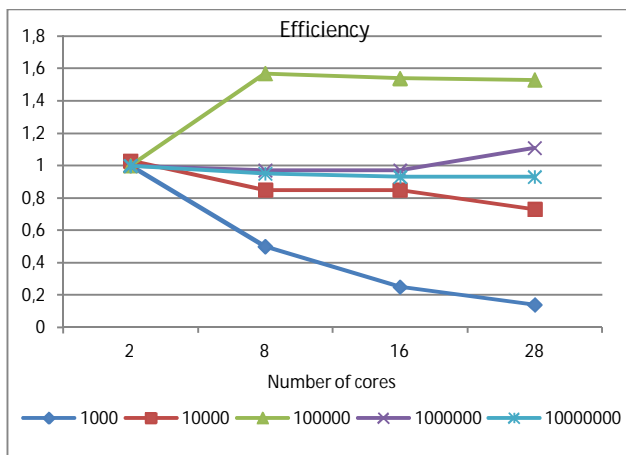


Figure 5. Efficiency of MSA\_BG algorithm using various numbers of cores and various numbers of iterations.

Obviously, the parallel system shows good scalability with respect to both the cluster size and the workload size. In the case of small number of iterations the speedup obtained is not sufficient. Increasing the number of iterations results in speedup improvement because of the better utilization of the cores.

## VI. CONCLUSION AND FUTURE WORK

We presented an environment enabling secure access to the grid based services as follows: security, parallel program implementation execution and database access on distributed heterogeneous high-performance grid infrastructure. Web portal provides as services access and extraction of biological data and execution of parallel program implementations based on algorithm for comparative analysis of biological data. The proposed portal is verified experimentally for the case study of investigation the influenza virus variability.

An innovative parallel algorithm MSA\_BG for multiple alignment of biological sequences that is highly scalable and locality aware has been designed. The MSA\_BG algorithm is iterative and is based on the concept of Artificial Bee Colony metaheuristics and the concept of algorithmic and architectural spaces correlation. The metaphor of the ABC metaheuristics has been constructed and the functionalities of the agents has been defined. The conceptual parallel computational model has been designed. The algorithmic framework of the designed parallel algorithm has been constructed.

MSA\_BG algorithm has a hierarchical structure, which enables observing the principle of locality (independent calculations), and very high scalability (hives and swarms), so it is expected high efficiency implementations for petaflops supercomputers.

Parallel performance evaluation and profiling of multiple sequence alignment on the basis of MSA\_BG algorithm heterogeneous distributed high-performance computation resources have been proposed in this paper.

The case study is investigating viral nucleotide sequences and finding out consensus motifs and variable domains in the different segments of influenza virus.

Parallel performance parameters, such execution time and acceleration, have been estimated experimentally. The performance estimation analyses have shown that the parallel system is well balanced both in respect to the workload and machine size.

## ACKNOWLEDGMENT

This work was supported in part by the Grant DFNI-M01/005 and in part by the Grant DCVP 02/1.

## REFERENCES

- [1] H., Carrillo and D. Lipman, "The Multiple Sequence Alignment Problem in Biology," *SIAM Journal of Applied Mathematics*, vol. 48, no. 5, 1988, pp. 1073-1082.
- [2] L. Wang and T. Jiang, "On the complexity of multiple sequence alignment," *Journal of Computational Biology*, vol. 1, no. 4, 1994, pp. 337-348, doi:10.1089/cmb.1994.1.337.
- [3] W. Just, "Computational complexity of multiple sequence alignment with SP-score," *Journal of Computational Biology*, vol. 8, no. 6, 2001, pp. 615-23.
- [4] S. Sze, Y. Lu, and Q. Yang, "A polynomial time solvable formulation of multiple sequence alignment," *Journal of Computational Biology*, vol. 13, no. 2, 2006, pp. 309-319, doi:10.1089/cmb.2006.13.309.
- [5] J. Thompson, D. Higgins, and T. Gibson, "ClustalW: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice," *Nucleic Acids Research*, vol. 22, no. 22, 1994, pp. 4673-4680.
- [6] K. Li, "ClustalW-MPI: ClustalW Analysis Using Distributed and Parallel Computing," *Journal of Bioinformatics*, vol.19, no. 12, 2003, pp. 1585-1586.
- [7] P. Borovska, V. Gancheva, and S. H. Ko, "Scaling of Parallel Multiple Sequence Alignment on the Supercomputer Juqueen," the 7th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, 12-14 September 2013, Berlin, Germany, accepted.
- [8] JUQUEEN – Configuration, [http://www.fz-juelich.de/ias/jsc/EN/Expertise/Supercomputers/JUQUEEN/Configuration/Configuration\\_node.html](http://www.fz-juelich.de/ias/jsc/EN/Expertise/Supercomputers/JUQUEEN/Configuration/Configuration_node.html), retrieved: 30.06.2013.
- [9] D. Karaboga, "An Idea Based On Honey Bee Swarm for Numerical Optimization," Technical Report-TR06, Erciyes University, Engineering Faculty, Computer Engineering Department, 2005, [mf.erciyes.edu.tr/abc/pub/tr06\\_2005.pdf](http://mf.erciyes.edu.tr/abc/pub/tr06_2005.pdf), retrieved: 20.05.2013.
- [10] G. Lewis, G. Sipos, F. Urmetzer, V. Alexandrov, and P. Kacsuk, "The collaborative P-GRADE grid portal," *Proceeding of ICCS'05 Proceedings of the 5th international conference on Computational Science - volume part III*, Springer-Verlag Berlin, Heidelberg, 2005, pp. 367-374.
- [11] GenBank, <http://www.ncbi.nlm.nih.gov/Genbank/>, retrieved: 20.05.2013.

# A New Method of Vehicle Initiative Safety: Heart Sound Acquisition and Identification Technology

Cheng Yu-han

College of Internet of Things, Nanjing University of  
Posts and Telecommunications,  
Nanjing, Jiangsu, 210003, China  
chengyh@njupt.edu.cn

Ma Yong

School of Computer Science and Technology, Nanjing  
University of Science and Technology,  
Nanjing, Jiangsu, 210094, China  
mayong@163.com

**Abstract**-This paper presents a method based on heart sound analysis of the characteristics of the vehicle initiative safety technology, to explore the feasibility and implementation method of monitor the pilot health condition using the heart sound signal. The article analyzes the characteristics of heart sound signal and vehicle background noise firstly, and proposes the model of the heart sound in vehicle internal environment; according to the model, a kind of acquisition device is designed; then, it discusses the heart sound classification and identification method based on independent sub-band function, especially the rules and algorithms of independent sub-band function construction, as well as how to make independent sub-band function as a new statistics characteristic parameter. In order to quickly determine a heart sound signal which separated from the heart sound acquisition apparatus, the paper introduces the new concept, i.e., degree of heart sound signal certainty. Finally, efficiency and feasibility are verified through the heart sound acquisition, classification and identification experiments.

**Keywords**-Vehicle initiative safety; Heart sound acquisition and identification; Independent sub-band function; Degree of heart sound signal certainty

## I. INTRODUCTIONS

In order to improve the automobile intelligence and security, contemporary cars have already gradually entered into active safety design from the past passive security settings. Now, except for the seat belts, airbags, bumpers and so on, we have to take active safety design into more consideration. The car can take the initiative to take measures to avoid accidents in advance. For example, the driver drowsy warning system, automobile anti-collision radar, GPS (Global Position System) system, tire pressure monitoring warning system, engine fire alarm prediction system, headlight automatic adjustment system, blind area monitoring system, vehicle information transmission system, automatic braking system, SOS parking system and automatic fire extinguishing system. All of these are to actively avoid a variety of accident factors that may be triggered by people or the vehicles. But studies into the various accident factors triggered by drivers is not a very active research topic; for example, traffic accidents caused by driver fatigue and cardiac sudden death is not uncommon, which contribute the appearance of driver on-site health

monitoring mode that is based on the driver's physiological state detection. In recent years, the aging population is a growing trend and aging problems are spreading between car drivers who take a certain proportion of the underlying heart disease; when a sudden heart attack occurs when driving a vehicle, current active safety facilities cannot detect it timely, which creates an obvious danger [1-3].

Some fundamental research about heart sound shows that heart sound can reflect the health status of different people [4-5]. When we make a combination of the existing driver fatigue detection system and automotive active safety systems, a more comprehensive unity of car reliability and security will be achieved. Therefore, the article analyzes the characteristics of heart sound signal and vehicle background noise firstly, and proposes the model of the heart sound in vehicle internal environment; according to the model, a kind of acquisition device is designed; then, it discusses the heart sound classification and identification method based on independent sub-band function, especially the rules and algorithms of independent sub-band function construction, as well as how to make independent sub-band function as a new statistics characteristic parameter.

In order to quickly determine a heart sound signal, which is separated from the heart sound acquisition apparatus, the paper introduces the new concept, i.e., the degree of heart sound signal certainty. Finally, efficiency and feasibility are verified through the heart sound acquisition, classification and identification experiments.

## II. CHARACTERISTICS OF HEART SOUND SIGNAL AND AUTOMOTIVE BACKGROUND NOISE

The ventricular and the atria's filling and contraction leads to heart beat, while heart beat information after cardiothoracic body reaches the surface and form the "heart sounds" that we can hear. Heart sound signals have the following main characteristics [5-6]:

In most cases, we only can hear the first heart sounds  $s_1$  and the second heart sounds  $s_2$ ;  $s_1$  is relatively low and long, lasting about 0.15s, while  $s_2$  is relatively high and short, lasting about 0.12s. Heart sound signals are mainly distributed in the first heart sound and second heart sound where a dominating place for us to extract heart sound parameters. However, the third heart sound and forth heart

sound are inconspicuous, so they are not underlined. Frequency of the heart sound is mainly concentrated on the range of frequency from 0~600Hz. Frequency of the first heart sound is between 50 and 150 Hz, while the second heart sound is concentrated in the range of 50 ~ 200 Hz. They have obvious change on amplitude while the other frequency range tends to zero. Figure 1 shows the heart sound signal spectrum.

The heart sound signal is periodic, so we perform the analysis of the heart sound in one cycle. Although the waveform of each cycle is different and there are various interferences, we can still deal with the heart sound signal by regarding it as a repeated stable signal.

Horn, the sound of a motor, brake sound, voice (audio), the wind, the sound of tire on the road and other noise are the main background noise in the automotive environment; in most cases, this noise can be divided into the interior noise and exterior noise. In this work, we focus on the interior noise. The typical car background noise has the following characteristics [1, 2]:

The sound of a motor, we also mean the sound of engine noise, which is one of the main sources of interior noise, typically including gasoline-combustion noise and mechanical vibration noise. When the car at low speed or go idling gasoline combustion noise is greater than the mechanical vibration noise, on the contrary, the later is greater than the former. Frequency of the engine noise is between 0 and 200 Hz, if the sound of a motor is smooth, there will a peak point in frequency 100HZ. Figure 1 (b) shows the frequency spectrum diagram of the motor sound. The car acceleration sound whose spectrum showed obvious ultra-low frequency and with a typical low-pass band spectrum, that is one special sound of a motor. Figure 1 (c) shows the acceleration sound spectrum.

The voice is the sound of the voices of the passengers in the car or stereo, and it is an active interior noise, which was relevant with the occupant's physiological characteristics, emotional and language content. Voice pitch frequency is a range of 130 Hz to 350Hz, while the harmonics of the frequency range up to 130 Hz to 4000 Hz, but the energy is mainly concentrated in the pitch range. Figure 1 (d) shows the sound spectrum; the music signal has a wider range of the spectrum and contains a richer, higher harmonic.

The brakes sound is that when there is a car braking, brake pads hugging tire, tire and ground cause a dramatic friction, then, the tire will make a harsh braking sound; the faster, heavier quality of the car and the sharper the brakes, the louder the noise; sometimes, the brakes sound will be up to 70-90 dB. Figure 1 (e) shows the acceleration sound spectrum diagram, which is energy-frequency mainly concentrate between 0 Hz and 200Hz, but there are also obvious treble component; the brakes sound is an interior noise with wider spectrum.

Horn, a peculiar language among car, people and vehicles to talk to each other, is indispensable in automotive safety systems. The horn is a steady-state signal with a good point

and timbre; in addition, the baseband frequency of horn general around 400Hz and its spectrum showed baseband and several octaves, according to the spectrum. We cannot see significant fluctuation in the sound stage. Figure 1 (f) shows the spectrum of horn. The noise in the range of 0 to 200Hz clearly coincides with the spectrum of the heart sound signal.

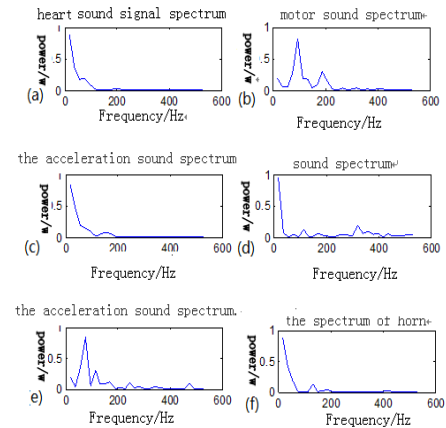


Figure 1. Heart sound with the typical curve of the spectral characteristics of the interior noise

### III. A HEART SOUND ACQUISITION DEVICE FOR AUTOMOTIVE ACTIVE SAFETY

In order to achieve real-time monitoring of cardiac health condition of the driver, the heart sound acquisition device was designed as an automobile active safety. The device configuration is installed on the car safety belt. The device not only can collect but also can finish the pretreatment of heart sound signal and classify the cardiac physiological state of the driver. According to such equipment, an alarm will alert when the physiological state of the heart is abnormal. Heart sound acquisition probe arrays are installed in the wall of a deformable elliptical acoustic cavity while the signal processing circuit, sound and light alarm were fixed at the top of the noise cavity; but what was configured in the cavity was an embedded lithium rectangular nylon sleeve, which cuffs can be installed on the car safety belt. Such kind of sleeves can be adjusted up and down along the seat belt position so effortlessly that it is possible to collect the heart sound signal without much effort. Recently, the device has applied for a Chinese invention patent, as shown in Figure 2.

When the driver plugs the seat belt, the heart sound monitoring device will changed into working state; once the monitoring device does not detect heart sounds heart sounds, the light will shine, the driver is required along the seat belt up and down to adjust the position of the heart sound monitoring device in the chest; when heart sounds are detected, the light does not shine. The device provides real-time monitoring of the cardiac physiological state of the driver; with this, we can make rapid analysis of the collected heart sound signals though the internal processor, or

transmitted to the onboard computer processing. According to the classification of the heart sound, we can determine the health status of the driver. Once the abnormal sound is found, lights will flash, an audible alarm will appear and the abnormal sound signal will be transmitted to the automotive active safety systems for further processing. The device is not only used in lithium battery-powered, but also ultra-low-power processor applied internally to improve working time of the signal acquisition. Once the seat belt is not properly plugged, the device will power off, reducing the power consumption.

In order to improve the reliability of the classification and identification of heart sound signals, that means we need to eliminate automotive environmental noise to the greatest extent as well as the rapid and effective algorithm to correctly identify the classification of heart sounds.

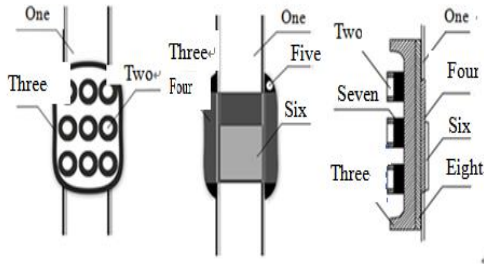


Figure 2. Heart sound acquisition monitoring device for an automobile active safety

#### IV. HEART SOUND SIGNAL EXTRACTION METHOD

##### A. Heart sound signal model in the automotive environment

Order the heart sound signal as:

$$S(t) = s_1(t) + s_2(t) + s_3(t) + s_4(t) = \sum_{i=1}^4 s_i(t) \quad (1)$$

is the car background noise, including trumpet sound, motor sound, brakes sound, voice and other noise, etc, then  $gN(t) = gN_1(t) + gN_2(t) + gN_3(t) + \dots$ , the model of the heart sounds in the automotive environment is:

$$\sum_k x_k(t) = \sum_{i=1}^4 s_i(t) + \sum_j gN_j(t) \quad (2)$$

##### B. Heart sound signal separation method based on ICA

For heart sound signals and interior noise characteristics, in this paper, we use the Independent Component Analysis (ICA) [6] method to separate heart sound signal from the interior noise, according to the mixed model of formula (2), and let  $X = [x_1, x_2, x_3, \dots, x_k]^T$  be its deviation matrix is [7]:

$$D = \sum_{j=1}^K (X^T - \frac{1}{N} \sum_{j=1}^N x_j^T) \quad (3)$$

We can get the covariance matrix:  $C = \frac{1}{N} \sum_{j=1}^N DD^T$ , the orthogonal eigenvector is  $V = [v_1, v_2, \dots, v_m]$ , there are  $m$  largest Eigen values  $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_m$ , then, whitening matrix  $P$  can be obtained:

$$P = V \left( \frac{1}{N} \Lambda \right)^{\frac{1}{2}} = \sqrt{NV} \Lambda^{\frac{1}{2}} \quad (4)$$

$\Lambda = \text{diag}(\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_m)$ , and  $P^T C P = I$ ,  $\bar{y} = P^T X$ , using these data in the use of albino to calculate the inverse matrix  $W = [w_1, w_2, \dots, w_m]$  that meeting statistical independence conditions, ICA can usually be estimated by using steepness:

$$\text{kurt}(\bar{w}^T \bar{y}) = E[(\bar{w}^T \bar{y})^4] - 3[E[(\bar{w}^T \bar{y})^2]]^2 + \lambda(1 - E[(\bar{w}^T \bar{y})^2]) \quad (5)$$

Inverse matrix  $W$  can be obtained by changing the steepness of the maximization. Then, the isolated heart sound signal is:

$$\begin{bmatrix} s' \\ gN'_1 \\ \vdots \\ gN'_j \end{bmatrix} = W \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} \quad (6)$$

##### C. Heart Sound Signal Certainty Degree (HSSCD)

In order to determine the signal separated from  $X(t)$ , which is the heart sound signal and which is the car background noise, we have introduced a new parameter-“Heart sound signal certainty degree (HSSCD)”.

Order the probability density function of separated signal  $S_k(t)$  as:

$$P(S_k) = \int_{S_{k1}}^{S_{k2}} p(S_k) dS_k \quad (7)$$

However, this description can only be used when  $S_k$  is in the positive distribution state, in order to resolve the negative  $S_k$  exists, we introduce:

$$RI_s = -\frac{1}{2} \log_2 \left\{ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W_{S_k}(t, \omega) dt d\omega \right\} \quad (8)$$

where  $W_{S_k}$  is the signal  $S_k$  the Wigner-View distribution, that is:

$$W_{S_k}(t) = \frac{1}{2\pi} \int S_k^* \left( t - \frac{1}{2} t \right) S_k \left( t + \frac{1}{2} t \right) e^{-j\omega} dt \quad (9)$$

Our definition of heart sounds certainty is:

$$HSSCD = \frac{(RI_s)_{\text{Compare signal}}}{(RI_s)_{\text{standard heart sound signal}}} \quad (10)$$



HSSCD has quantitatively described the overall performance of closeness of a signal with standard heart sound signal in the amount of information and complexity. Figure 3 is time-domain waveform diagram of the heart sound signal and a variety of automotive background noise; Table 1 is a signal corresponding HSSCD value.

We can have the following conclusions by the HSSCD analysis.

Judgment 1: when  $HSSCD = 1$ , that is exactly the same two heart sound signals;

Judgment 2: when HSSCD is approaching 1, indicating that the detection signal can be classified as a heart sound signal;

Judgment 3: when the HSSCD value is smaller, it signals, does not belong to the heart sound signal.

Experiments show that HSSCD have little relationship with the signal change in sampling frequency and length, even if there is a difference in length and the sampling frequency of the two signals, HSSCD remains essentially unchanged.

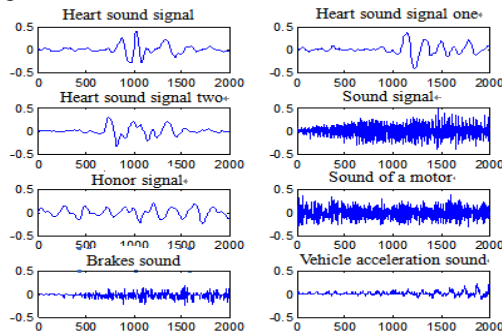


Figure 3. Waveform diagram of the part of the heart sound signals and car background noise

TABLE I. VARIOUS SIGNALS CORRESPONDING HSSCD VALUE IN FIGURE 1

Signal Type	HSSCD	Signal Type	HSSCD
Heart sound signal	1	Horn sound	0.8678
Heart sound signal 1	1.0991	Sound of a motor	0.3585
Heart sound signal 2	0.9941	Brakes sound	0.4536
Sound signal	0.4807	Acceleration sound	0.7323

V. HEART SOUND SIGNAL EXTRACTION METHOD A KIND OF STATISTICAL CHARACTERISTIC PARAMETERS: THE HEART SOUND INDEPENDENT WAVELET FUNCTION

Feature Extraction of heart sound is to find a kind of transformation. This transformation can convert the original heart sound signal into a certain state of feature space, but also be able to save all the original information, so that to

lay the foundation for the classification and identification of heart sounds.

Feature Extraction of heart sound has many forms, such as a time-domain characteristics parameters extracted cardiac cycle, heart rate, and adjacent interval of the first heart sound and second heart sounds, Extracting the spectral characteristics of the heart sounds, linear predictive spectral coefficients, Mel frequency spectral coefficients [11] as the characteristic parameters of frequency. These characteristic values can be used for classification and identification of the heart sounds. According to the characteristics of the heart sounds, this article uses the heart sound independent wavelet function as a new kind of statistical characteristic.

A. Description of independent sub-band function

For one channel of the heart sound signal  $s(t)$ , if we can break it into  $q$  layers data segment  $Z_q$ , with equal length ( $q=1,2, \dots, Q$ ), independent component analysis and signal divided into layers, by finding a full rank separation matrix, thereby defining the output signal  $[b_1, b_2, \dots, b_Q]$ , including the information of heart sound signal  $s(t)$  as independent as possible, this group statistically independent function with each other in the time domain, which is what we call cluster of independent sub-wave function [5].

The Independent sub-wave function is derived in the following way: (a) layers of the signal. This process can be understood as hierarchical linear transformation of the signal; (b) Analysis of the independent components of the signal. The result of this process is hierarchical signal matrix solutions mixing matrix and the hierarchical signals are converted into mutually independent vectors matrix. This process can also be considered to be the projection of the vector direction independent of each of the sub-wave function. From more experiments, we draw more conclusions, as explained below:

The independent sub-wave function has strong local characteristics of the time domain, since the independent sub-wave function obtained by the different signals are varied, DFT (Discrete Fourier Transformation), DCT (Discrete Cosine Transformation) transform guide [9] to the base function to the various signals are very similar, showing obvious global characteristics in the time domain.

The independent sub-wave function has strong local characteristics of the frequency domain because the frequency characteristic curves of the different signals acquired independent sub-wave function are different.

These characteristics indicate that the independent sub-wave functions can characterize on the signal not overlapped as a characteristic parameter, the identification and classification of the heart sounds may be used as an effective feature parameter.

B. Algorithm realization of heart sound independent sub-wave function

Specific steps to obtain heart sounds independent sub-wave function are as follows:

1) Stratify heart sound signal  $s(t)$ . Signal layering is equivalent to a set of orthogonal base projection, and the hierarchical signal had orthogonally, using independent component analysis of the signal, while the length of each of the hierarchical signal should remain the same. Hierarchical method can meet these two conditions, such as wavelet hierarchical method, EMD (Empirical mode decomposition) hierarchical method;

2) Remove the mean, and preliminary bleach the stratified signal;

3) Independent component analysis of the hierarchical signal after pretreatment, eventually to get heart sound independent sub-wave function cluster.

Figure 4 (a) is a single cycle of heart sound signal  $s$ , first use empirical mode of EMD decomposition  $t$  that broke  $s$  into a series of approximate simple component signal combination, that is:

$$s_k(t) = \sum_{l=1}^L \lambda_l Z_l + r \quad (11)$$

$Z_l$  is the number one intrinsic mode function IMF, which is self-adaptive decomposition by the nature of the signal;  $r$  is the residual function of the average trend of the representative signal,  $\lambda_l$  is a coefficient, Figure 4 (b) is the result of empirical mode decomposition.  $S_k(t)$  will be merged into two groups hierarchical signals  $s_{11}(t)$  and  $s_{12}(t)$  with equal amplitude, as shown in Figure 4 (c); then we use fastica (fixed point iterative algorithm), Infomax (information maximization algorithm), and dwt\_ica, three algorithms to process  $s_{11}(t)$  and  $s_{12}(t)$  statistically and independently, to obtain the heart sound independent sub-wave, as shown in Figure 4 (d), (e), (f) below. We can see from the figure that the same sample points of independent heart sounds sub-wave as the source signal, but these three ways to get the heart sounds independent sub-waves are very different in magnitude. When using the dwt\_ica algorithm, the independent heart sound sub-wave obtained is basically the same, however when obtained with fastica and infomax method, there is slightly different in the heart sound independent sub-wave  $b_{1f}$ ,  $b_{2f}$ ,  $b_{1i}$ ,  $b_{2i}$ , especially, the infomax method has a significant uncertainty; its best result shows the effect can be better than dwt\_ica, FastICA algorithm, but sometimes appears to be worse than those two methods. Cluster of independent sub-wave function shown in the picture is the average of the three test results. (Wherein the x-axis is the sample point, the y-axis is the amplitude.)

Coefficient description of reconfiguration similarity: two heart sounds independent sub-wave similarity coefficients between the heart sound signals reconstructed with the original heart sound signal is the bigger the better; Irrelevancy describes with correlation coefficient  $r$ , which is the metric of independent heart sounds sub-wave of irrelevance degree, relevance between  $b1$  and  $b2$  is the smaller the better. Compare the data results shown in Table

2; the value in the table is the average results of the three experiments.

Using the data shown in the table, we can make a decision from the signal remodeling, the effect of the three algorithms are both good. Judging from the relevance of independent heart sounds wavelet, effects using fastica [9] or dwt\_ica [9] algorithm are good, but the wavelet relevance obtained using infomax algorithm is significantly bigger. Analyzing from the time spent in computation, dwt\_ica algorithm is the most efficient.

In summary, dwt\_ica algorithm shows obvious advantages in all aspects, so we will use dwt\_ica algorithm to get the heart sound independent sub-wave function. A heart sound signal stratified results and three BSP algorithms of obtaining independent sub-wave function as shown in Figure 4.

TABLE II. THREE BSP ALGORITHMS EFFECTS OF OBTAINING INDEPENDENT SUB-WAVE FUNCTION

Three BSP algorithms	Reconfigurability	Correlation	Computing time(s)	Comprehensive Evaluation
fastica	0.9979	0.0000	0.4836	Good
dwt_ica	0.9983	0.0000	0.0624	Very good
infomax	0.9919	0.4352	0.4212	Not bad

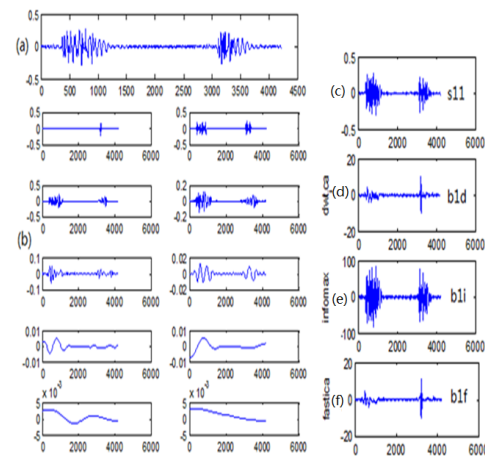


Figure 4. A heart sound signal stratified results and three BSP algorithms of obtaining independent sub-wave function

## VI. HEART SOUND SIGNAL EXTRACTION METHOD: HEART SOUND CLASSIFICATION AND RECOGNITION

The process of heart sound classification and identification is essentially an automated pattern matching process, as shown in Figure 5. There are many heart sound classification and recognition algorithms, such as statistical identification method, neural network identification method [11]. In this paper, as a new statistical characteristic parameter, the heart sounds independent sub-wave function, need a pattern matching method; so, we define a similar

distance to identify the distance of the independent wave function of two heart sound signals.

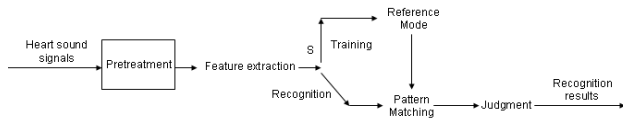


Figure 5. Heart sounds to identify general method

Suppose that the independent sub-wave function digital of the standard heart sound signal as signal of the independent sub-wave function digital to be identified as, then, the similar distance will be:

$$d_k = 1 - \frac{\left| \sum_{n=1}^N b_i^c(t) b_j^s(t) \right|}{\sqrt{\sum_{n=1}^N (b_i^c(t))^2 \sum_{n=1}^N (b_j^s(t))^2}} \quad (12)$$

The smaller the similar distance  $d_k$  is,  $b_i^c(t)$  and  $b_j^s(t)$  will be more similar; when  $d_k = 0$ , then  $b_j^s(t) = b_i^c(t)$ , which means they are exactly the same.

Therefore, after the signal preprocessing, to use the independent sub-wave function of the heart sounds to do pattern matching, find their similar distance, and classify accordingly identification.

## VII. THE EXPERIMENTAL RESULTS

### A. The acquisition and separation of heart sounds

In the vehicle initiative safety of heart sound acquisition device needs to set sampling frequency, channel parameters, etc. Frequency of the heart sound concentrates on the range of frequency from 10 to 400 Hz while heart murmur is below 1500 Hz. According to sampling law, we can avoid distortion when sampling rate is at least more than 3000 Hz. In the environment where vehicle noise is relatively large, we use 4 channels and 11025 Hz sampling rate. A group of acquisition results are shown in Figure 6(a). Then, according to the extraction method of heart sound signal stated in the section 3 and blind separation of four aliasing signal based on ICA [9], we can get four separation signal shown in Figure 6(b).

In order to judge the heart sound and the vehicle background noise from the four separated signals, we analyze the sound signal with heart sound signal certainty (HSSCD), and discover that the HSSCD of separated signal 1 is 0.9869, which is heart sound signal; the HSSCD of separated signal 2 is 0.4765, which is speech signal; the HSSCD of separated signal 3 is 0.3275, which is motor sound signal; and the HSSCD of separated signal 4 is 0.4532, that is brake sound signal. Through sound transmitter, we can verify the conclusion. In fact, we just need to extract heart sound signal, and do not need to spend time on the other signals.

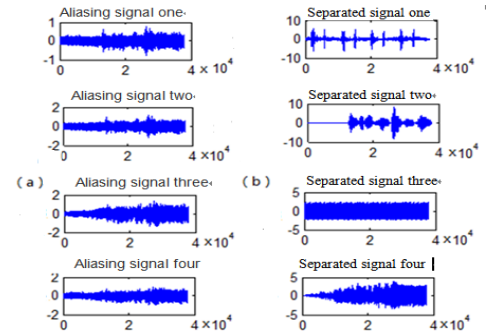


Figure 6. The extraction and separation of signal in heart sound acquisition device

### B. The heart sound signal pretreatment

The main task of the heart sound signal pretreatment is as follows [11]: (1) determine the beginning and the end point of the first heart sound and the second heart sound of each of the signal in a section of heart sound signal; (2) find out the beginning and end point in a periodic heart sound signal, and calculate the heart rate for displaying the characteristics of heart sounds better, highlighting the main component of heart sounds which are quite useful for the heart sounds identifying. Firstly, calculate the energy spectrum of the heart sound signal, namely separated signal 1 shown in Figure 7(a). According to the envelope extraction method based on Hilbert transformation [5], we can extract heart sounds envelope, and the result is shown in Figure 7(b). Then, we can get normalization energy envelope shown in Figure 7(c) with the threshold taken from the average of the envelope. The optimized envelope can reflect the section of heart sound signal more accurately than the envelope directly obtained. Among them, every broader pulse respectively represents the first heart sound, and the narrower respectively stands for the second heart sound. Thus, we can easily calculate heart sounds interval and heartbeat. Lastly, we section heart sounds according to the period, and the result is shown in the Figure 7(d).

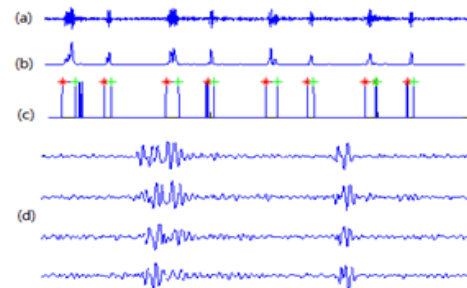


Figure 7. The heart sound signal pretreatment

### C. The classification identification of heart sound signals

In the identification mode, the heart sound information in the database is a standard group. The database includes normal heart sounds (like the compare heart sound signal 1), and abnormal heart sounds (like the compare heart sound signal 2 which is premature beat heart sound, and the

compare heart sound signal 3 which is fibrillation heart sound). Take discretionarily a periodic heart sound signal from the signals in Figure 7(d) as the test group. They are shown in Figure 8(a).

The standard group and the test group are required to use the same signal equipment and own the same magnification. They should not have saturated distortion (the best control is about the 70 ~ 80% of the maximum distortion amplitude). Then, we can extract heart sound wavelet function which is taken as a kind of statistical characteristic parameters from these data. Set the independent wavelet function of heart sound signal of the standard group as  $b_i^c(t)$  and the function of the identified heart sound signal as. Then, calculate their average similarity distance by the formula (13), which is respectively 0.0790, 0.6114, 0.8942. Choose directly the one which has the smallest similarity distance as the results of identification result, and this tested signal is the normal heart sound one. The similar phase diagram is shown in Figure 8(b). We can find out intuitively that the testing signal of classification recognition is most similar to the normal heart sound signal and the similar phase diagram is a thin oblique line of  $45^\circ$ . Several hundred classification identification experiments show that the same person's recognition rate at the same time is 100%, and the recognition rate at different time from the same person is up to 97%. For some specific person, the recognition rate is up to 99%, if we constantly update the normal database with the new acquisition of heart sounds. If we only classify heart sounds as normal heart sounds and abnormal, we would have a better classification recognition efficiency and effect. It basically reaches the level of practical application.

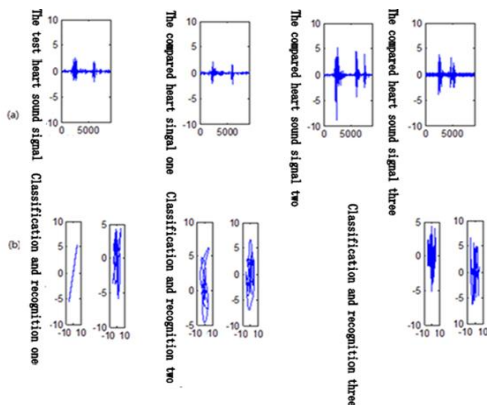


Figure 8. The effect of classification and identification experiments.

### VIII. CONCLUSION AND FUTURE WORK

This paper presented a method based on classification and recognition of heart sound detection for vehicle active safety. Through the analysis of the main characteristics between heart sound signals and car background noise, designed in the car on the heart sound acquisition, separation,

classification of soft and hardware scheme, compared fastica, dwt\_ica, infomax three algorithms to obtain the effect of heart sound independent wave function, the dwt\_ica algorithm the results of comprehensive evaluation is best, and discussed the heart sound signal model, noisy sound mixing signal separation model and heart sound recognition model. It is the foundation of designing simple information recognition system. In addition, the article also raised effective classification technology, such as heart sounds determined (HSSCD). Similar distance mode matching method for heart sound independent characteristic parameters of the wave function and normalized envelope heart sound cycle segmentation technology, it reduced the calculation and increased the recognition rate, and it opens a new way for the practical application of heart sound identification.

In the future, we will further improve the calculation speed and practical application.

### ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (Grant No.61271334).

### REFERENCE

- [1] D. Rus, R. Gray, and D. Kotz, Transportable Information Agents [J], 1997, pp. 215-238.
- [2] Xiao X. Q. and Wang Q. D., Study on automotive active safety technology based on driving behavior and intention, China Mechanical Engineering, OCT. 2010, pp. 2390-2393.
- [3] A. K. Jain, A. Ross, and S. Prabhakar, An Introduction to Biometric Recognition [J], IEEE Transactions on Circuits and Systems for Video Technology, vol.14. 2004, pp. 4-20.
- [4] Zhao L., Li Q., Shao Q. Y., and Zhu X. L., Studies on the comparison of normal and abnormal heart sound signals[J], Chinese Journal of Medical Physics, vol. 3. 2000, pp. 26-28.
- [5] Cheng X. F., Ma Y., and Tao Y. W., Three-stage heart sound identification technology based on data fusion [J], Chinese Journal of Scientific Instrument, August. 2010, pp. 1712-1720.
- [6] Maglogiannis I, Loukis E., and Zafiroopoulos E, Support Vectors Machine-based identification of heart valve diseases using heart sounds. Comput Methods Programs Biomed (Ireland), Jul. 2009, 95, pp. 47-61.
- [7] Cheng X. F. and Tao Y. W., Heart Sound Recognition--A Prospective Candidate for Biometric Identification [J], Advanced Materials Research, vol. 225. 2011, pp. 433-436.
- [8] Kumar D., Carvalho P., and Antunes M., Heart murmur recognition and segmentation by complexity signatures. Conf Proc IEEE Eng Med Biol Soc (United States), 2008, pp. 2128-32.
- [9] Cheng X. F. and Ma Y., A Single-Channel Mixed Signal BSS New Method of Does Not Use the Prior-Knowledge [J], ACTA ELECTRONICA SINICA, 2011, pp. 2317-2321.
- [10] Gil. J. J., Single-Channel Singnal Separation Using Time-Domai Basis Functions [J], IEEE Signal Processing Letters, vol. 10. 2003, pp.168-171.
- [11] Cheng X. F., Ma Y., and Liu C., Research on heart sound identification technology [J], Sci China Inf Sci, vol. 55. 2012, pp. 281-292.

# Global Illumination-Invariant Fast Sub-Pixel Image Registration

Andrew Gilman

Institute of Natural and Mathematical Sciences  
Massey University  
Auckland, New Zealand  
Email: a.gilman@massey.ac.nz

Arno Leist

School of Engineering and Advanced Technology  
Massey University  
Auckland, New Zealand  
Email: a.leist@massey.ac.nz

**Abstract**—Sub-pixel image registration is an important part of many image processing and computer vision applications. We propose a computationally simple direct (i.e., non-iterative) method for sub-pixel registration of translated images. To register two images, first a global space-invariant resampling filter is computed that is a least-squares optimal predictor of one image from the pixel values of the other image. Then, the coefficients of this filter are linearly combined to compute the offset between the two images. The computational cost of this algorithm is linear in the number of pixels. The accuracy and efficiency of the proposed algorithm is demonstrated to be better than a range of existing methods for images with various levels of high-frequency detail and at various noise levels.

**Keywords**—image registration; sub-pixel; direct; least-squares optimal; linear computational complexity.

## I. INTRODUCTION

Image processing often requires processing of data from multiple images of the same scene captured from different viewpoints with one or multiple cameras or maybe even with completely different types of sensors. Before this can be achieved, the pixel data in the images must be somehow 'synchronised' and this is achieved using registration. Image registration is the problem of finding a geometric transformation that maps the coordinate plane of one image to another using the image data itself. It is an important problem in image processing and computer vision, required as part of many applications in areas such as medical imaging, remote sensing and consumer electronics [1].

What makes this problem challenging is the fact that many applications require fast, yet accurate, registration. For some applications, such as image super-resolution or image fusion (see Figure 1 for an example), it is critical to register input images with accuracy down to sub-pixel level. But, unfortunately, speed or computational complexity and accuracy are generally trade-offs and can be hard to achieve simultaneously. Sub-pixel registration can also be hampered by the fact that the images may be contaminated by noise and aliasing, both of which may have a considerable effect, as well as possible lack of conformity of the image data to the proposed transformation model.

Even for applications that do not require precise sub-pixel registration, it is still very important, because registration of images with larger integer offsets is often achieved using a multi-resolution image pyramid [2] for efficiency reasons (for

examples see [3]–[5]). Ignoring the high-resolution detail at higher levels of the pyramid is not only more computationally efficient, but also necessary, because of aliasing of high spatial frequency components undergoing large motion. The use of a multi-resolution framework helps to alleviate this problem [6].

The typical approach to solving the registration problem is to place it into an optimisation framework [7], with the objective function defined by some metric that measures similarity (or alternatively distance) between one image and a transformed version of the other. The objective function is then optimised with respect to the transformation parameters. This approach was popularised by Lucas and Kanade [8] in 1981 and has seen many variants proposed since, such as using different optimisation methods and similarity metrics. A summary of some common similarity metrics and search strategies can be found in [1] and [9]. While many of these approaches are theoretically sound, they are generally computationally expensive and require a number of iterations to achieve convergence, due to the non-linear nature of the problem. This is highly inadequate for some applications that require the results to be computed in a timely manner. An example of such applications, which is becoming increasingly popular, is image processing (High Dynamic Range (HDR) [10], panorama stitching, etc.) on smartphone and tablet devices that possess only limited computing resources.

We are interested in a special case of image registration, where the geometric transformation is constrained to pure translation. The desire for fast registration of translated images with sub-pixel accuracy has led to the development of a number of direct approaches that attain lower complexity



Fig. 1: High Dynamic Range image produced from registered images with varying exposures.

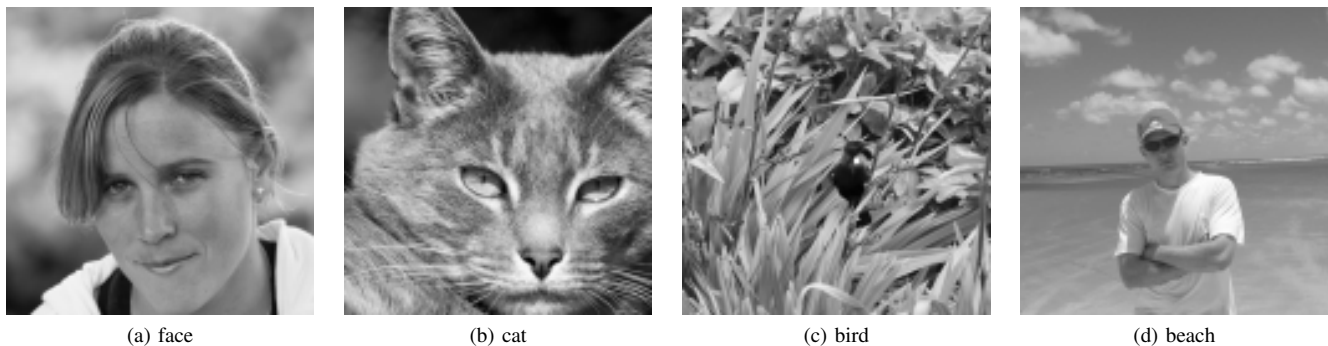


Fig. 2: Test images

by making various simplifying assumptions. Fourier phase correlation [11]–[14], for example, assumes that the images obey the Nyquist criterion [15], while interpolation of the cross-correlation function [16] makes assumptions about the shape of the scene's autocorrelation function. Unfortunately, when the real data deviates from the assumptions made in the model the accuracy can suffer.

We present a new direct technique for sub-pixel registration of translated images. Even though this method does not require iteration, its formulation has more similarities with the iterative methods than the direct, although it does not require the choice of an appropriate interpolation function for the resampling filter. Instead, it computes the optimal resampling filter from the image data itself using linear least-squares. The relative shift between the images is then computed in a novel way from the filter coefficients.

In Section II, we introduce the theory of image registration in the optimisation framework and present the new direct approach. In Section III, we describe the experimental method for sub-pixel accuracy evaluation. Section IV presents the experimental results and their discussion, followed by conclusions in Section V.

## II. THEORY

In this section, we provide a mathematical definition for the image registration problem and show how it is formulated in the optimisation framework, which requires a computationally intensive iterative solution. We then consider an alternative formulation, which results in a direct solution.

Let us consider the problem of registering a discrete reference image  $f(kT, lT)$  and an offset target image  $g(kT, lT) = f(kT + u_x, lT + u_y)$ , where  $k$  and  $l$  are integer pixel indices,  $T$  is the pixel pitch (which we will assume to be equal to one) and  $u_x, u_y$  are some unknown and possibly non-integer offsets between the two images. The goal of sub-pixel image registration is to estimate  $u_x$  and  $u_y$  with as much accuracy as possible. A conventional way of doing this is to iteratively minimise the difference (usually sum of squared differences) between  $f(k + \hat{u}_x, l + \hat{u}_y)$  and  $g(k, l)$ :

$$\arg \min_{\hat{u}_x, \hat{u}_y} \sum_{k,l} (g(k, l) - f(k + \hat{u}_x, l + \hat{u}_y))^2 \quad (1)$$

In practice, since the image  $f(k, l)$  is only known on an integer grid, some sort of interpolation must be applied to  $f(k, l)$  to get the shifted image  $f(k + \hat{u}_x, l + \hat{u}_y)$ . For computational reasons, only the values located on the shifted grid are interpolated and this procedure is generally performed using a resampling filter [17]. The filtering operation can be expressed as

$$\sum_{m=M^-}^{M^+} \sum_{n=N^-}^{N^+} f(k + m, l + n) h_{\hat{u}_x, \hat{u}_y}(m, n) \quad (2)$$

where  $m, n \in \mathbb{Z}$  and  $h_{\hat{u}_x, \hat{u}_y}(m, n)$  is a resampling filter, based on some predetermined interpolation basis function (see [18], [19] for some common examples), with a window span between  $M^-$  and  $M^+$  in  $x$  direction and between  $N^-$  and  $N^+$  in  $y$  direction. The resampling filter coefficients are samples of the interpolation basis function offset by  $(u_x, u_y)$  and are generally non-linear in  $u_x$  and  $u_y$ , making expression (1) a non-linear least-squares problem. Choosing a different interpolation basis function would result in a slightly different interpolated image, in turn affecting the results of minimisation.

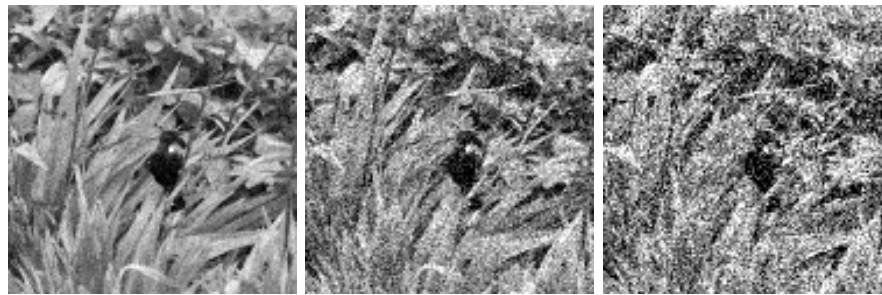
### A. Registration through Optimal Interpolation

Now, let us consider an alternative formulation of this problem. First, reformulate (1) to minimise with respect to the filter coefficients  $h(m, n)$  instead of the offsets  $u_x$  and  $u_y$ . This removes any dependence on an interpolation basis function and results in filter coefficients being optimised in a least-squares fashion. The resulting filter is optimal – no other resampling filter can do a better job of predicting image  $g(kT, lT)$  from  $f(kT, lT)$ .

$$\arg \min_{\mathbf{h}} \sum_{k,l} \left( g(k, l) - \sum_m \sum_n f(k + m, l + n) h(m, n) \right)^2 \quad (3)$$

The next step is to determine the relative offsets  $u_x$  and  $u_y$ . Given a set of filter coefficients, computed from an interpolation basis function, we could compute the offset by which this particular filter would shift the image through a linear combination of the filter coefficients. Consider the Keys' cubic convolution interpolation kernel:

$$CC(x) = \begin{cases} \frac{3}{2}|x|^3 - \frac{5}{2}|x|^2 + 1 & 0 < |x| < 1 \\ -\frac{1}{2}|x|^3 + \frac{5}{2}|x|^2 - 4|x| + 2 & 1 < |x| < 2 \\ 0 & 2 < |x| \end{cases} \quad (4)$$



(a) Image 'bird'. Noise std: 10.8, 28.8, 46.8



(b) Image 'face'. Noise std: 4.6, 12.3, 20.0

Fig. 3: Effect of noise strength on test images. Noise std stated in grayscale levels.

and a one-dimensional resampling filter  $h_u(m)$ , based on this kernel, defined as

$$h_u(m) = CC(-u + m), \quad m \in \mathbb{Z}. \quad (5)$$

It can be shown algebraically (see Appendix) that the offset  $u$  can be recovered from the filter coefficients through a weighted sum combination thereof:

$$u = \sum m h_u(m) \quad (6)$$

This is also true for the optimal filter coefficients, which do not depend on any explicit interpolation function, but are still related to the offset between the images through the image data itself. The offset estimate between the images can be computed in the same way as:

$$\hat{u}_x = \sum_{m=M^-}^{M^+} \sum_{n=N^-}^{N^+} m h(m, n) \quad (7)$$

$$\hat{u}_y = \sum_{m=M^-}^{M^+} \sum_{n=N^-}^{N^+} n h(m, n) \quad (8)$$

and we show that this works experimentally, in Section IV.

### B. Global Illumination Invariance

Cameras with automatic exposure setting will meter the scene prior to taking a picture. Capturing multiple images of the scene, even in quick succession, can result in global illumination variations due to the metering algorithm calculating slightly different exposure settings. This difference can be fairly large in some cases – for example when capturing different parts of the scene, such as when panning the camera to capture a panorama. Using different exposure settings may

also be intended, for example, when capturing images to produce a high-dynamic range reproduction of a scene. The image processing algorithms, including registration, must be able to cope well with these situations.

Global illumination difference between the two images under registration can be accounted for in the model by including gain  $A$  and offset  $S$ :  $g(kT, lT) = Af(kT + u_x, lT + u_y) + S$ . Because the filter coefficients in optimisation expression (1) are not constrained to sum to one, this optimisation would automatically account for gain  $A$ , resulting in

$$\sum_m \sum_n h(m, n) = 1/A \quad (9)$$

The relative offsets  $u_x$  and  $u_y$  can still be recovered from  $h(m, n)$  by normalising expressions (7) and (8) by the sum of filter coefficients:

$$\hat{u}_x = \frac{\sum_m \sum_n m h(m, n)}{\sum_m \sum_n h(m, n)} \quad (10)$$

$$\hat{u}_y = \frac{\sum_m \sum_n n h(m, n)}{\sum_m \sum_n h(m, n)} \quad (11)$$

The illumination offset  $S$  can be accounted for by including an additional variable  $\hat{S}$  into the optimisation procedure (1), subtracted from  $g(k, l)$ . This does not change the equations (10) and (11), by which the offset is estimated from the filter coefficients.

## III. METHOD

To validate the proposed method and assess its sub-pixel accuracy, we have performed an experiment using synthetically

generated images with known sub-pixel offsets.

#### A. Test Image Generation

A high-resolution image (10 times higher resolution than the test images) was taken as the source scene. A number of test scenes (shown in Figure 2) that contain different combinations of texture, sharp detail and smooth areas at a range of different scales were selected for this performance evaluation. These scenes vary in high-frequency content, resulting in different amounts of aliasing in the test images, from little aliasing in image 'beach' to a lot of aliasing in image 'bird'.

A simple model of the imaging process was used to simulate the capture of lower resolution test images. The imaging model consisted of area-sampling (any blurring due to the optics' Point-Spread Function (PSF) was assumed to be insignificant in comparison to the degradation from the sensor PSF – this assumption is not unrealistic for many consumer cameras) and additive white Gaussian noise [20]. Creation of test images in this way allowed for aliasing to occur providing the opportunity to test the algorithms on aliased data, rather than just with band-limited signals.

The  $1300 \times 1300$  pixel grayscale source images with 8-bit precision were filtered using a  $10 \times 10$  box average to simulate blurring during area-sampling, and then down-sampled by a factor of 10 in each dimension to result in  $124 \times 124$  pixel test images (low-resolution images were cropped to avoid edge effects from the filtering operation). Shifting the high-resolution image by an integer number of pixels prior to downsampling allowed for generation of low-resolution test images shifted by fractional (in steps of 0.1) amounts of a pixel.

Here, we test the ability of the method to register offsets smaller than one pixel by registering a set of one hundred image pairs with relative offsets ranging from (0, 0) to (0.9, 0.9).

#### B. Measurements

The uncertainty in estimating the relative offsets is not constant across the sub-pixel range. Registering image pairs that relate by a small sub-pixel translation (close to zero) or a translation of almost one whole pixel results in smaller MSE than registering a pair of images offset by about half a pixel (see Figure 4 for a typical error pin plot). Because there is no prior knowledge about the distribution of the translations, we estimate the performance of sub-pixel registration on average by assuming that all sub-pixel translations are equally likely to occur. Average performance can then be calculated simply by averaging the squared error over all possible sub-pixel translations – 100 translations, ranging from (0,0) to (0.9, 0.9) for this experimental set-up.

Each pair of images with a unique offset was registered 100 times, each time with a different instance of white Gaussian noise added to the images. A different pair of illumination gain  $A$  and offset  $S$  constants were also used each time to adjust the exposure of the target image. The gain constant was drawn from a normal distribution with a mean of one and a standard deviation of 0.1, and the offset constant was drawn from a normal distribution with a mean of 0 and a standard deviation of 25. The target image was clipped to only contain values between 0 and 255 after the transformation.

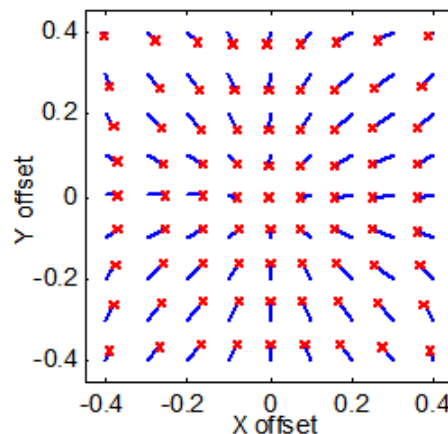


Fig. 4: Sub-pixel registration errors depicted as a pin plot for image 'bird' with no noise, using normalised cross-correlation method.

To make the results easier to interpret, registration error magnitude (12) was computed for each registration, and these were averaged over the 100 repeated results for each possible sub-pixel offset, and over all 100 sub-pixel offsets to give expected sub-pixel accuracy when registering two images.

$$\|e\|_2^2 = (u_x - \hat{u}_x)^2 + (u_y - \hat{u}_y)^2 \quad (12)$$

The measurements were repeated at different noise levels. We found the sub-pixel accuracy of registration to be very sensitive to the noise level. To be able to compare the results between the test scenes, different amounts of noise were added in each case. This was achieved empirically by matching the visual effect of noise on each scene; see Figure 3 for an example – image 'bird' required almost twice the noise standard deviation in comparison to image 'face' to achieve the same level of contamination.

#### C. Methods for Comparison

To put the experimental results in perspective, we compare them to three other popular methods: one standard iterative method and two direct methods. We employ Lucas-Kanade iterative registration [8] with bilinear interpolation for resampling. A detailed description of implementation of this method can be found in [21].

The two direct techniques that are used are the normalised cross-correlation and a Fourier phase based method by Stone et al. [22]. The Fourier phase method was chosen because it works well in the presence of aliasing. Sub-pixel accuracy of normalised cross-correlation was achieved by interpolating a  $3 \times 3$  window around the function peak using a 2D second order polynomial.

## IV. RESULTS AND DISCUSSION

We have performed the experiment with the proposed method using resampling filters of various sizes. We were able to successfully recover the relative shift between the images from the optimal filter coefficients according to equations 10 and 11, confirming our observations at the end of Section II-A.



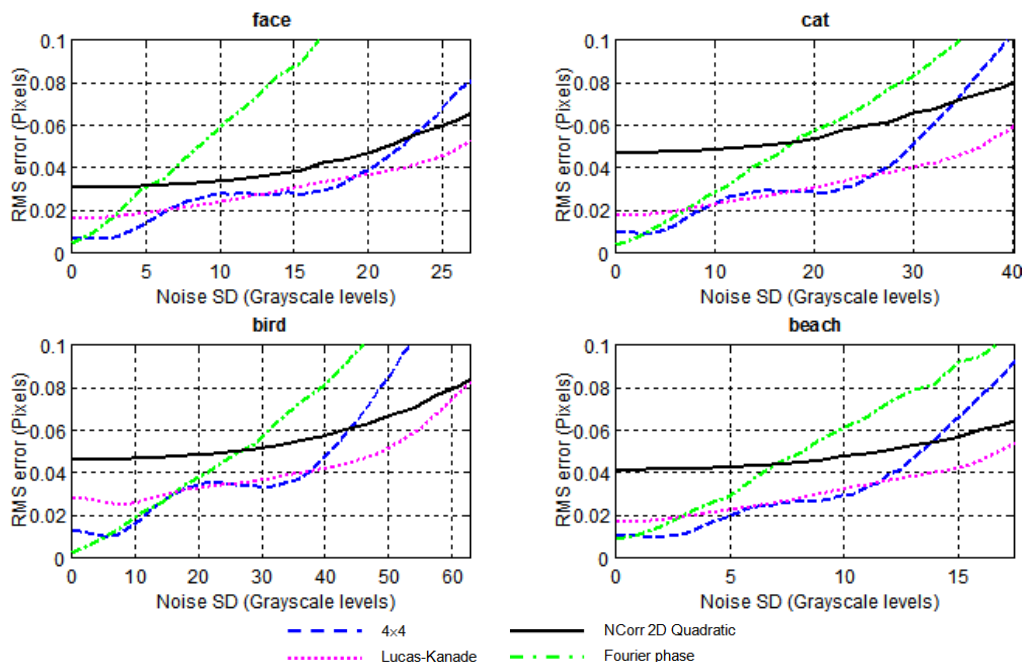


Fig. 5: Effect of noise strength on test images. Noise std stated in grayscale values.

Using filters of different sizes resulted in slightly different degrees of accuracy.

We report the results for the  $4 \times 4$  filter, as it had the best performance relative to its computational complexity – larger filters obviously require more computation. A plot of root-mean-square registration error versus the noise standard deviation can be seen in Figure 5. The registration error is expressed as a fraction of the spatial pixel size (same as the pixel pitch), whereas the noise standard deviation is expressed in grayscale levels.

It can be observed that the proposed method can achieve an accuracy of as high as 1% of a pixel in a low-noise situation (left part of the graph). As the noise level increases, the accuracy drops slightly to around 4% of a pixel (middle part of the graph). At even higher noise levels, the accuracy drops off more quickly, but at these noise levels the images are quite heavily contaminated by noise, as can be seen from Figure 3.

Comparing the results of the proposed method to that of the other methods one can observe that it performs significantly better than the cross-correlation method, better than Lucas-Kanade method and better than the Fourier phase-based method, apart from images 'cat' and 'bird' with no added noise. These are favourable results for the proposed method, taking into consideration that it was twice as fast as normalised cross-correlation and Fourier phase-based methods and eight times faster than the iterative Lucas-Kanade method. All of these methods, except Fourier phase-based method, see a linear increase in execution time with an increase in image size. Fourier phase-based method is in theory  $O(N \log N)$ ; however, because the transform is separable and image dimensions are small, it also sees almost linear increase in execution time with image size.

## V. CONCLUSION

Image registration is traditionally an optimisation problem that makes use of interpolation as a crucial link between the discrete and continuous domains. The use of interpolation functions that are common in image processing, such as piece-wise cubic polynomials for example, leads to a non-linear objective function, the optimisation of which requires an iterative solution. In the case of sub-pixel registration of translated images, some direct methods also exist, but these make certain assumptions that can limit the accuracy if the assumptions do not hold. The optimal filter based image registration proposed in this paper is an alternative two step method for sub-pixel registration of translated images that does not require iteration. The first step computes a resampling filter that is a least-squares optimal predictor of the target image using the pixel values of the reference image. The second step estimates the translation between the target and the reference from the filter coefficients. The method has a linear-time computational complexity.

Experimental evaluation of the proposed method, using synthetically generated images, demonstrated that its performance is better than that of a number of other image registration methods for a range of scenes. It was shown that the proposed method is capable of sub-pixel accuracy approaching 1% of a pixel in low-noise situations and 4% of a pixel for moderate noise. The proposed method has been shown to cope well with global illumination variations and is also the fastest, by at least a factor of two, out of the methods tested here. This method would be useful for a range of applications that require fast alignment of shifted images, such as image super-resolution, high-dynamic range reconstruction, panorama stitching.

## REFERENCES

- [1] A. Goshtasby, *2-D and 3-D image registration for medical, remote sensing, and industrial applications*. Hoboken, NJ: J. Wiley & Sons, 2005.
- [2] P. J. Burt and E. H. Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. 31, no. 4, pp. 532–540, 1983.
- [3] A. Rosenfeld and G. J. Vanderbrug, "Coarse-fine template matching," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 7, no. 2, pp. 104–107, 1977.
- [4] G. J. Vanderbrug and A. Rosenfeld, "Two-stage template matching," *IEEE Transactions on Computers*, vol. C-26, no. 4, pp. 384–393, 1977.
- [5] P. Thévenaz, U. E. Ruttimann, and M. Unser, "A pyramid approach to subpixel registration based on intensity," *IEEE Transactions on Image Processing*, vol. 7, no. 1, pp. 27–41, 1998.
- [6] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation," *Computer Vision - Eccv 92*, vol. 588, pp. 237–252, 1992.
- [7] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge, UK ; New York: Cambridge University Press, 2004.
- [8] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. 7th Int. Joint Conf. Artificial Intelligence*, vol. 2, 1981, pp. 674–679.
- [9] L. G. Brown, "A survey of image registration techniques," *Computing Surveys*, vol. 24, no. 4, pp. 325–376, 1992.
- [10] S. Mann, "Compositing multiple pictures of the same scene," in *Proceedings of the 46th Annual IS&T Conference*, vol. 2, 1993.
- [11] H. Shekarforoush, M. Berthod, and J. Zerubia, "Subpixel image registration by estimating the polyphase decomposition of cross power spectrum," in *Proc. CVPR '96*, Inria,F-06902 Sophia Antipolis,France, 1996, pp. 532–537.
- [12] C. D. Kuglin and D. C. Hines, "The phase correlation image alignment method," in *IEEE 1975 Conf. Cybernetics and Society*. IEEE, 1975, pp. 163–165.
- [13] M. Guizar-Sicarios, S. T. Thurman, and J. R. Fienup, "Efficient subpixel image registration algorithm," *Optics Letters*, vol. 33, no. 2, pp. 156–158, January 2008.
- [14] E. Vera and S. Torres, "Subpixel Accuracy Analysis of Phase Correlation Registration Methods Applied to Aliased Image," in *16th European Signal Processing Conference (EUSIPCO 2008)*, Lausanne, Switzerland, August 25-29 2008.
- [15] H. Nyquist, "Certain Topics in Telegraph Transmission Theory," *Transactions of the American Institute of Electrical Engineers*, vol. 47, no. 2, pp. 617–644, April 1928.
- [16] V. N. Dvornychenko, "Bounds on (deterministic) correlation functions with application to registration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 5, no. 2, pp. 206–213, 1983.
- [17] K. Turkowski, "Filters for common resampling tasks," in *Graphics gems*, A. S. Glassner, Ed. San Diego, CA, USA: Academic Press Professional, Inc., 1990, pp. 147–165.
- [18] T. M. Lehmann, C. Gnner, and K. Spitzer, "Survey: Interpolation methods in medical image processing," *IEEE Transactions on Medical Imaging*, vol. 18, pp. 1049–1075, 1999.
- [19] P. Thévenaz, T. Blu, and M. Unser, "Image interpolation and resampling," in *Handbook of Medical Imaging, Processing and Analysis*. Academic Press, 2000, ch. 25, pp. 393–420.
- [20] C. F. Gauss, "Theoria motus corporum coelestium in sectionibus conicis solem ambientium," 1809, Translation: Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections.
- [21] S. Baker and I. Matthews, "Lucas-kanade 20 years on: A unifying framework," *International Journal of Computer Vision*, vol. 56, no. 3, pp. 221–255, 2004.
- [22] H. S. Stone, M. Orchard, E. chien Chang, and S. Martucci, "A fast direct fourier-based algorithm for subpixel registration of images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, pp. 2235–2243, 2001.

## APPENDIX

Equation 5 describes a resampling filter  $h_u$  based on Keys' cubic convolution interpolation kernel. The filter can be used to "shift" a discrete sequence by resampling it on a uniform grid which is offset by  $u$ . Because the interpolation kernel has finite support (see Equation 4), the resampling filter has only 4 non-zero coefficients. Assuming  $0 \leq u < 1$ , these four coefficients can be calculated by substituting  $m = \{-1, 0, 1, 2\}$  into Equation 5:

$$h_u = [CC(-u - 1), CC(-u), CC(1 - u), CC(2 - u)] \quad (13)$$

To show that Equation 6 is correct, we can substitute the above filter coefficients into the right hand side of this equation and simplify:

$$\begin{aligned}
 u &= \sum mh_u(m) \\
 &= -CC(-u - 1) + CC(1 - u) + 2CC(2 - u) \\
 &= -\left(-\frac{1}{2}|-u - 1|^3 + \frac{5}{2}|-u - 1|^2 - 4|-u - 1| + 2\right) \\
 &\quad + \left(\frac{3}{2}|1 - u|^3 - \frac{5}{2}|1 - u|^2 + 1\right) \\
 &\quad + 2\left(-\frac{1}{2}|2 - u|^3 + \frac{5}{2}|2 - u|^2 - 4|2 - u| + 2\right) \\
 &= \frac{1}{2}(u + 1)^3 - \frac{5}{2}(u + 1)^2 + 4(u + 1) - 2 \\
 &\quad + \frac{3}{2}(1 - u)^3 - \frac{5}{2}(1 - u)^2 + 1 \\
 &\quad - (2 - u)^3 + 5(2 - u)^2 - 8(2 - u) + 4 \\
 &= u.
 \end{aligned}$$

# Filtering of Large Signal Sets: An Almost Blind Case

Anatoli Torokhti, Phil Howlett  
 CIAM, School of Inf. Techn. & Math. Sci.  
 University of South Australia,  
 Adelaide, SA 5095, Australia

Email: anatoli.torokhti@unisa.edu.au, phil.howlett@unisa.edu.au

Hamid Laga  
 PBRC, University of South Australia  
 ACPFG,  
 Adelaide, SA 5095, Australia  
 Email: hamid.laga@unisa.edu.au

**Abstract**—We propose a new technique which allows to estimate any random signal from a large set of noisy observed data on the basis of information on only a few reference signals. The conceptual device behind the proposed estimator is a linear interpolation of an optimal incremental estimation applied to random signal pairs interpreted an extension of the Least Squares Linear (LSL) estimator. We consider the case of observations corrupted by an arbitrary noise (not by an additive noise only) and design the estimator in terms of the Moore-Penrose pseudo-inverse matrix. Therefore, it is always well defined. The proposed estimator is justified by establishing an upper bound for the associated error and by showing that this upper bound is directly related to the expected error for an incremental application of the optimal LSL estimator. It is shown that such an estimator is possible under quite unrestrictive assumptions.

**Keywords**—large signal sets; filtering; least squares linear estimator.

## I. INTRODUCTION

### A. Motivation

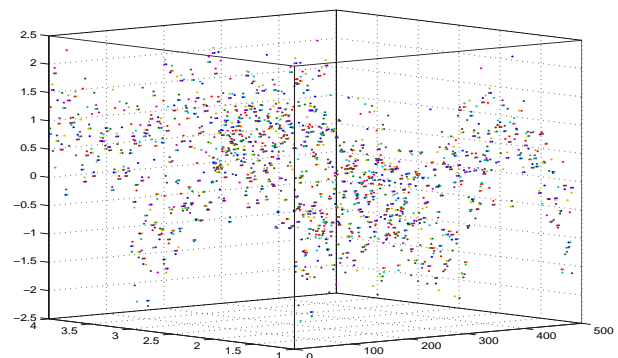
We write  $\mathbf{x}_\omega = \mathbf{x}_\omega(t)$  for a stochastic vector  $\mathbf{x}_\omega(t)$  associated with a random outcome  $\omega$  and time  $t \in T = [a, b] \subset \mathbb{R}$ . A rigorous notation is given in the section that follows.

In many applications associated with a difficult environment, *a priori* information on a large set of signals of interest,  $\mathcal{K}_x = \{\mathbf{x}_\omega(t)\}$ , can only be obtained for a few signals  $\{\mathbf{x}_\omega(t_j)\}_1^p \subset \mathcal{K}_x$  where  $p$  is a small number. Typical examples are devices and equipment exploited in the stratosphere, underground and underwater such as those in defence and the mining industry. Signals  $\mathbf{x}_\omega(t_1), \dots, \mathbf{x}_\omega(t_p)$ , are associated with given times,  $t_1, t_2, \dots, t_p$ , respectively, such that

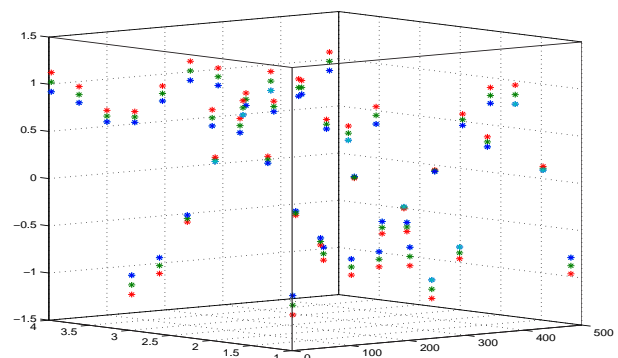
$$a = t_1 < t_2 < \dots < t_{p-1} < t_p = b. \quad (1)$$

A choice of signals  $\mathbf{x}_\omega(t_1), \dots, \mathbf{x}_\omega(t_p)$  might be beyond our control (in geophysics and defence tasks, for instance). At the same time, it is required to estimate each reference signal in the set  $\mathcal{K}_x$  from the corresponding set of noisy observations. Thus, all we can exploit to develop an associated filter is observed noisy data and a sparse information on reference signals.

*Example 1:* Suppose we need to process a set  $\mathcal{K}_y$  of  $N = 121$  random signals over set  $T = [\tau_1, \tau_2, \dots, \tau_N]$  so that each input signal from this set,  $\mathbf{y}(t, \cdot)$ , enters a filter at time  $t = \tau_k$



(a) Observed signals from the set  $\mathcal{K}_y$ .



(b) Samples of reference signals  $\bar{X}^{(j)}$  at times  $t_j$  for  $j = 1, \dots, 11$ .

Fig. 1. Signals and samples considered in Example 1.

where  $\tau_1 = 0$  and  $\tau_{k+1} = \tau_k + 0.05$ , for  $k = 1, \dots, 120$ . At time  $\tau_k$ , for  $k = 1, \dots, N$ , the observed signal  $\mathbf{y}(\tau_k, \cdot)$ , is represented by its realizations as a  $4 \times 4$  matrix

$$Y^{(k)} = \{y_{\ell,r}^{(k)}\}_{\ell,r=1}^4 = [\mathbf{y}(\tau_k, \omega_1), \dots, \mathbf{y}(\tau_k, \omega_4)]. \quad (2)$$

A column of matrix  $Y^{(k)}$ ,  $\mathbf{y}(t_k, \omega_i) \in \mathbb{R}^4$ , represents the realization of the signal  $\mathbf{y}(t, \omega_i)$  at time  $t = \tau_k$  generated by the random event  $\omega_i$ , for each  $i = 1, 2, \dots, 4$ . Thus, all observed signals are given by the  $4 \times 484$  matrix  $Y = [Y^{(1)}, \dots, Y^{(121)}]$  represented in Fig. 1 (a).

Suppose that, for  $j = 1, \dots, p$ , information on the reference signals can only be obtained at some times  $t_1 = \tau_1$ ,

$t_{j+1} = \tau_{12j+1}$  where  $j = 1, \dots, 10$  (see (1)) in the form of samples given by  $4 \times 4$  matrices

$$\tilde{X}^{(j)} = [\tilde{\mathbf{x}}(t_j, \omega_1), \dots, \tilde{\mathbf{x}}(t_j, \omega_4)] = \{\tilde{x}_{\ell,r}^{(k)}\}_{\ell,r=1}^4. \quad (3)$$

Fig. 1 demonstrates a typical situation with noisy observed signals and sparse information on the reference signals. In Example 2 below we show that, under certain conditions, the proposed technique allows us to estimate the signals of interest with an acceptable accuracy.

### B. Formalization of the problem

To formalize the problem, we write  $\{\Omega, \Sigma, \mu\}$  for a probability space where  $\Omega$  is the set of all experimental outcomes,  $\Sigma \subset \Omega$  is a sigma-algebra of measurable sets known as the event space and  $\mu$  is a non-negative probability measure with  $\mu(\Omega) = 1$ . We denote by  $\mathcal{K}_x = \{\mathbf{x}_\omega \mid \omega \in \Omega\}$  a set of reference stochastic signals and by  $\mathcal{K}_y = \{\mathbf{y}_\omega \mid \omega \in \Omega\}$  a set of observed signals.

In an intuitive way,  $\mathbf{y}$  can be regarded as a noise-corrupted version of  $\mathbf{x}$ . For example,  $\mathbf{y}$  can be interpreted as  $\mathbf{y} = \mathbf{x} + \mathbf{n}$  where  $\mathbf{n}$  is white noise. We do not restrict ourselves to this simplest version of  $\mathbf{y}$  and assume that the dependence of  $\mathbf{y}$  on  $\mathbf{x}$  and  $\mathbf{n}$  is arbitrary. Note that, theoretically,  $\mathcal{K}_x$  and  $\mathcal{K}_y$  are infinite signal sets. In practice, however, sets  $\mathcal{K}_x$  and  $\mathcal{K}_y$  are finite and large, each with, say,  $N$  signals.

To each random outcome  $\omega \in \Omega$  we associate a unique signal pair  $(\mathbf{x}_\omega, \mathbf{y}_\omega)$  where  $\mathbf{x}_\omega : T \rightarrow \mathcal{C}^{0,1}(T, \mathbb{R}^m)$  and  $\mathbf{y}_\omega : T \rightarrow \mathcal{C}^{0,1}(T, \mathbb{R}^n)$ . The space  $\mathcal{C}^{0,1}(T, \mathbb{R}^p)$  is the set of vector-valued Hölder continuous functions  $\mathbf{f}$  of order 1 with  $\mathbf{f}(t) \in \mathbb{R}^p$  and  $\|\mathbf{f}(s) - \mathbf{f}(t)\| \leq K|s - t|$  (see [1], p. 96.) Write

$$\mathcal{P} = \mathcal{K}_x \times \mathcal{K}_y = \{(\mathbf{x}_\omega, \mathbf{y}_\omega) \mid \omega \in \Omega\} \quad (4)$$

for the set of all such signal pairs. For each  $\omega \in \Omega$ , the components  $\mathbf{x}_\omega = \mathbf{x}_\omega(t), \mathbf{y}_\omega = \mathbf{y}_\omega(t)$  are Lipschitz continuous vector-valued functions on  $T$  [1].

We wish to construct an estimator  $F^{(p-1)}$  that estimates each reference signal  $\mathbf{x}_\omega(t)$  in  $\mathcal{P}$  from related observed input  $\mathbf{y}_\omega(t)$  under the restriction that *a priori* information on only a few reference signals,  $\mathbf{x}_\omega(t_1), \dots, \mathbf{x}_\omega(t_p)$ , is available where  $p \ll N$ .

In more detail, this restriction implies the following. Let us denote by  $\mathcal{K}_x^{(p)}$  a set of  $p$  signals  $\mathbf{x}_\omega(t_1), \dots, \mathbf{x}_\omega(t_p)$  for which *a priori* information is available. A set of associated observed signals  $\mathbf{y}_\omega(t_1), \dots, \mathbf{y}_\omega(t_p)$  is denoted by  $\mathcal{K}_y^{(p)}$ . Then for all  $\mathbf{y}_\omega(t)$  that do not belong to  $\mathcal{K}_y^{(p)}$ ,  $\mathbf{y}_\omega(t) \notin \mathcal{K}_y^{(p)}$ , estimator  $F^{(p-1)}$  is said to be the *blind* estimator [2], [3], [4], [5] since no information on  $\mathbf{x}_\omega(t) \notin \mathcal{K}_x^{(p)}$  is available. If  $\mathbf{y}_\omega(t) \in \mathcal{K}_y^{(p)}$  then  $F^{(p-1)}$  becomes a *nonblind* estimator since information on  $\mathbf{x}_\omega(t) \in \mathcal{K}_x^{(p)}$  is available. Thus, depending on  $\mathbf{y}_\omega(t)$ , estimator  $F^{(p-1)}$  is classified differently. Therefore, such a procedure of estimating reference signals in  $\mathcal{K}_x$  is here called the *almost blind* estimation.

### C. Differences from known techniques

We would like to note that the *almost blind* estimation is different from known methods such as nonblind [6]–[18],

semiblind and blind techniques [2]–[5], [19]–[22]. Indeed, at each particular time  $t \in T$ , the input of the *almost blind* estimator  $F^{(p-1)}$  developed below in this paper, is a random vector  $\mathbf{y}_\omega(t)$ . Thus, for different  $t \in T$ , the input is a different random vector  $\mathbf{y}_\omega(t)$  but we wish to keep *the same estimator*  $F^{(p-1)}$  for any  $t \in T$ , i.e., for any observed signal  $\mathbf{y}_\omega(t)$  in the set  $\mathcal{K}_y$ . The literature on these subjects is very abundant. Here, we listed only some related references.

By known techniques in [2]–[16] and [19]–[22], an estimator (here, we choose the united term ‘estimator’ to denote an equalizer or a system) is specifically designed for *each* particular input–output pair represented by random vectors. That is, for different inputs (observed signals)  $\mathbf{y}_\omega(t)$ , known techniques require different specified estimators and the number of estimators should be equal to a number of processed signals. In the case of *large signal sets*, such approaches become inconvenient because the number of signals  $N$  can be very large as it is supposed in this paper. For example, in problems related to DNA analysis,  $N = \mathcal{O}(10^4)$ . Therefore, the inconvenient restriction of using *a priori* information on only  $p$  reference signals, with  $p \ll N$ , is quite significant. At the same time, beside difficulties that this restriction imposes on the estimation procedure, we use it in a way that allows us to avoid the hard work associated with known techniques applied to large signal sets. To the best of our knowledge, the exception is the methodology in [17], [18], where, for estimation of a set of signals, the single estimator is constructed. The estimation techniques in [17], [18] exploit information in the form of a vector obtained, in particular, from averaging over signals in  $\mathcal{K}_x^{(p)}$ .

Further, the semiblind techniques are not applicable to the considered problem because they require a knowledge of some ‘parts’ of each reference signal in  $\mathcal{K}_x$  (e.g., see [3], [5], [19]) but it is not the case here. Although the blind techniques allow us to avoid this restriction, it is known that they have difficulties related to accuracy and computational load. In the problem under consideration, the advantage is a knowledge of some (small) part of the set of reference signals. It is natural to use this advantage in the estimator structure and we will do it in Section II.

Nonblind estimators [6]–[16] are not applicable here because they require *a priori* information on each reference signal in  $\mathcal{K}_x$  (e.g., a knowledge of covariance matrix  $E[\mathbf{x}_\omega \mathbf{y}_\omega^T]$  where  $E$  is the expectation operator). In particular, it is known that there are significant advantages in adaptive or recursive estimators (e.g., associated with Kalman filtering approach [23]) and it may well be possible to embed our estimator into such an environment—but that is not our particular concern here. Further, we note that much of the literature on piecewise linear estimators [24]–[28] seems to be *not directly relevant* to the estimator proposed here. In the first instance papers such as [24]–[28] are mostly concerned with the theoretical problems of approximation by piecewise linear functions on multi-dimensional domains which is *not the case here*.

Also, unlike many known techniques, we consider the case of observations corrupted by an arbitrary noise (not by an

additive noise only) and design the estimator in terms of the Moore-Penrose pseudo-inverse matrix [29]. Therefore it is always well defined.

## II. THE MAIN RESULTS

In this section, we outline the rationale for the proposed estimator and state the main results.

### A. Some preliminaries

The proposed estimator  $F^{(p-1)}$  is adaptive to a sparse set  $\mathcal{K}_x^{(p)}$ .

The conceptual device behind the proposed estimator is a linear interpolation of an optimal incremental estimation applied to random signal pairs  $(\mathbf{x}_\omega(t_j), \mathbf{y}_\omega(t_j))$  and  $(\mathbf{x}_\omega(t_{j+1}), \mathbf{y}_\omega(t_{j+1}))$ , for  $j = 1, \dots, p-1$ , interpreted an extension of the Least Squares Linear (LSL) estimator (see, for example, [6], [11], [16]). Although this idea may seem reasonable, the detailed justification of the new estimator is not straightforward and requires careful analysis. We shall do this by establishing an upper bound for the associated error and by showing that this upper bound is directly related to the expected error for an incremental application of the optimal LSL estimator. In Section II-B below, we will show that such an estimator is possible under quite unrestrictive assumptions.

Since the estimator proposed below is based on an extension of the LSL estimator it is convenient to sketch known related results here. Consider a *single* random signal pair  $(\mathbf{x}(\omega), \mathbf{y}(\omega))$  where  $\mathbf{x} \in L^2(\Omega, \mathbb{R}^m)$  and  $\mathbf{y} \in L^2(\Omega, \mathbb{R}^n)$  with zero mean  $(E[\mathbf{x}], E[\mathbf{y}]) = (\mathbf{0}, \mathbf{0})$ , where  $\mathbf{0}$  is the zero vector. Note that here,  $\mathbf{x}$  and  $\mathbf{y}$  do not depend on  $t$  as above. The estimate  $\hat{\mathbf{x}}$  of the reference vector  $\mathbf{x}$  by the optimal least squares linear estimator is given by

$$\hat{\mathbf{x}}(\omega) = E_{\mathbf{x}\mathbf{y}} E_{\mathbf{y}\mathbf{y}}^\dagger \mathbf{y}(\omega) \quad (5)$$

where  $E_{\mathbf{x}\mathbf{y}} = E[\mathbf{x}\mathbf{y}^T]$  and  $E_{\mathbf{y}\mathbf{y}} = E[\mathbf{y}\mathbf{y}^T]$  are known covariance matrices and  $E_{\mathbf{y}\mathbf{y}}^\dagger$  is the Moore-Penrose pseudo-inverse of  $E_{\mathbf{y}\mathbf{y}}$ . By the LSL estimator, matrices  $E_{\mathbf{x}\mathbf{y}}$  and  $E_{\mathbf{y}\mathbf{y}}^\dagger$  should be specified for each signal pair  $(\mathbf{x}(\omega), \mathbf{y}(\omega))$ .

Further, for a justification of our estimator, we need some more notation as follows. It is convenient to denote  $\mathbf{x}(t, \omega) = \mathbf{x}_\omega(t)$  and  $\mathbf{y}(t, \omega) = \mathbf{y}_\omega(t)$  so that  $\mathbf{x}(t, \omega) \in \mathbb{R}^m$  and  $\mathbf{y}(t, \omega) \in \mathbb{R}^n$ .

### B. The piecewise LSL interpolation estimator

For each signal pair (or vector function pair) in the set  $\mathcal{P}$ ,  $(\mathbf{x}(t, \omega), \mathbf{y}(t, \omega))$ , we assume that  $(E[\mathbf{x}(t, \cdot)], E[\mathbf{y}(t, \cdot)]) = (\mathbf{0}, \mathbf{0})$ . To begin the estimation process we need to find an initial estimate  $\hat{\mathbf{x}}(t_1, \omega)$ . It is assumed this can be found by some known method. Further, let us consider a signal estimation procedure at  $t_2, \dots, t_p$ . We use an inductive argument to define an incremental estimation procedure. Consider a typical interval  $[t_j, t_{j+1}]$  and define incremental random vectors

$$\mathbf{v}_j(\omega) = \mathbf{x}(t_{j+1}, \omega) - \mathbf{x}(t_j, \omega) \in \mathbb{R}^m, \quad (6)$$

$$\mathbf{w}_j(\omega) = \mathbf{y}(t_{j+1}, \omega) - \mathbf{y}(t_j, \omega) \in \mathbb{R}^n \quad (7)$$

and construct the optimal linear estimate

$$\hat{\mathbf{v}}_j(\omega) = E_{\mathbf{v}_j \mathbf{w}_j} E_{\mathbf{w}_j \mathbf{w}_j}^\dagger \mathbf{w}_j(\omega) \quad (8)$$

of the increment  $\mathbf{v}_j(\omega)$  for each  $j = 1, \dots, p-1$ . We will write

$$B_j = E_{\mathbf{v}_j \mathbf{w}_j} E_{\mathbf{w}_j \mathbf{w}_j}^\dagger \in \mathbb{R}^{m \times n}. \quad (9)$$

Define the estimate at point  $t_{j+1}$  by setting  $\hat{\mathbf{x}}(t_{j+1}, \omega) = \hat{\mathbf{x}}(t_j, \omega) + \hat{\mathbf{v}}_j(\omega)$ . Thus we have

$$\begin{aligned} \hat{\mathbf{x}}(t_{j+1}, \omega) &= \hat{\mathbf{x}}(t_j, \omega) + B_j[\mathbf{y}(t_{j+1}, \omega) - \mathbf{y}(t_j, \omega)] \\ &= \epsilon_j(\omega) + B_j \mathbf{y}(t_{j+1}, \omega) \end{aligned} \quad (10)$$

where we write

$$\epsilon_j(\omega) = \hat{\mathbf{x}}(t_j, \omega) - B_j \mathbf{y}(t_j, \omega). \quad (11)$$

Note that this definition can be rewritten more suggestively as

$$\hat{\mathbf{x}}(t_j, \omega) = \epsilon_j(\omega) + B_j \mathbf{y}(t_j, \omega) \quad (12)$$

for each  $j = 1, \dots, p-1$ .

The formula (10) shows that on each subinterval  $[t_j, t_{j+1}]$  the estimate of the reference signal at  $t_{j+1}$  is obtained from the initial estimate at  $t_j$  by adding the optimal LSL estimate of the increment.

Now, consider a signal estimation at any  $t \in [a, b]$ . The first step is simply to extend the formulæ (10) and (12) to all  $t \in [t_j, t_{j+1}]$  by defining

$$\hat{\mathbf{x}}(t, \omega) = \epsilon_j(\omega) + B_j \mathbf{y}(t, \omega). \quad (13)$$

Thus, the incremental estimation across each subinterval is extended to every point within the subinterval. Because of determining estimate  $\hat{\mathbf{x}}(t_{j+1}, \omega)$  in the form (8)–(10) we interpret this procedure as the *LSL piecewise interpolation*.

The incremental estimations are collected together in the following way. For each  $j = 1, 2, \dots, p-1$ , write

$$F_j[\mathbf{y}(t, \omega)] = \epsilon_j(\omega) + B_j \mathbf{y}(t, \omega) \quad (14)$$

for all  $t \in [t_j, t_{j+1}]$  and hence define the *piecewise LSL interpolation estimator* by setting

$$F^{(p-1)}[\mathbf{y}(t, \omega)] = \sum_{j=1}^{p-1} F_j[\mathbf{y}(t, \omega)][u(t-t_j) - u(t-t_{j+1})] \quad (15)$$

for all  $t \in [a, b]$  where  $u(t) = \begin{cases} 1 & \text{for } t > 0 \\ 0 & \text{otherwise.} \end{cases}$  is the unit step function. Thus we can now use the estimate

$$\hat{\mathbf{x}}(t, \omega) = F^{(p-1)}[\mathbf{y}(t, \omega)] \quad (16)$$

for all  $(t, \omega) \in T \times \Omega$ . The idea of a piecewise LSL interpolation estimator on  $T$  seems intuitively reasonable for signals with a well defined gradient over  $T$ .

We note that by (9)–(16), the estimator  $F^{(p-1)}$  is adaptive to a variation of signals in  $\mathcal{K}_x^{(p)}$ . A change of signals in  $\mathcal{K}_x^{(p)}$  implies a change of determinations of sub-estimators  $B_j$  in (9) and keep the same structure of the  $F^{(p-1)}$ .

### C. Justification of the LSL interpolation estimator

We wish to justify the proposed estimator by establishing an upper bound for the associated error.

To explain the technical details we introduce some further terminology.

Let us denote  $\|\mathbf{x}(t, \cdot)\|_{\Omega}^2 = \int_{\Omega} \|\mathbf{x}(t, \omega)\|^2 d\mu(\omega)$ . Assume that for all  $t \in T$ , we have

$$\|\mathbf{x}(t, \cdot)\|_{\Omega}^2 < \infty \quad \text{and} \quad \|\mathbf{y}(t, \cdot)\|_{\Omega}^2 < \infty, \quad (17)$$

where  $\|\mathbf{x}(t, \omega)\|$  and  $\|\mathbf{y}(t, \omega)\|$  are the Euclidean norms for  $\mathbf{x}(t, \omega)$  and  $\mathbf{y}(t, \omega)$  for each  $(t, \omega) \in T \times \Omega$ , respectively. Thus we will say that the signals are square integrable in  $\omega$  and write  $\mathbf{x}(t, \cdot) \in L^2(\Omega)$  and  $\mathbf{y}(t, \cdot) \in L^2(\Omega)$  for each fixed  $t \in T$ .

For each  $t \in T$ , let  $\mathcal{F} = \{\mathbf{f} : T \times \Omega \rightarrow \mathbb{R}^m \mid \mathbf{f}(t, \cdot) \in L^2(\Omega, \mathbb{R}^m)\}$  and define

$$\begin{aligned} \|\mathbf{f}\|_{T, \Omega} &= \frac{1}{b-a} \int_{T \times \Omega} \|\mathbf{f}(t, \omega)\| dt d\mu(\omega) \\ &= \frac{1}{b-a} \int_T E[\|\mathbf{f}(t, \cdot)\|] dt \end{aligned}$$

for each  $\mathbf{f} \in \mathcal{F}$  where  $\|\mathbf{f}(t, \omega)\|$  is the Euclidean norm of  $\mathbf{f}(t, \omega)$  on  $\mathbb{R}^m$  for all  $(t, \omega) \in \mathbb{R}^m$ . Suppose that for all  $(\mathbf{x}, \mathbf{y}) \in \mathcal{P}$  there exist constants  $\gamma_j, \delta_j > 0$  such that

$$\|\mathbf{x}(s, \omega) - \mathbf{x}(t, \omega)\| \leq \gamma_j |s - t|, \quad (18)$$

$$\|\mathbf{y}(s, \omega) - \mathbf{y}(t, \omega)\| \leq \delta_j |s - t| \quad (19)$$

for all  $(s, \omega), (t, \omega) \in [t_j, t_{j+1}] \times \Omega$ , i.e. we suppose that the Lipschitz constants in (18) are independent of  $\omega$ .

The error bound for the piecewise LSL interpolation estimator is established in Theorem 1 below.

*Theorem 1:* If condition (18) is satisfied then the error  $\epsilon_p = \|\mathbf{x} - F^{(p-1)}[\mathbf{y}]\|_{T, \Omega}$  associated with the piecewise LSL interpolation estimator satisfies the inequality

$$\epsilon_p \leq \max_{j=1, \dots, p-1} \{(\gamma_j + \|B_j\|_2 \delta_j) |t_{j+1} - t_j| \quad (20)$$

$$+ \left[ \|E_{\mathbf{v}_j}^{1/2} \mathbf{v}_j\|_F^2 - \|E_{\mathbf{v}_j} \mathbf{w}_j (E_{\mathbf{w}_j}^{1/2} \mathbf{w}_j)^\dagger\|_F^2 \right]^{1/2} \} \quad (21)$$

where  $\|B_j\|_2$  denotes the 2-norm given by the square root of the largest eigenvalue of  $B_j^T B_j$  and  $\|\cdot\|$  denotes the Frobenius norm.

*Example 2:* The time interval  $T$  is the same as in Example 1. At each time  $\tau_k$ , for  $k = 1, \dots, N$ , the training reference signal  $\mathbf{x}(\tau_k, \cdot)$  is represented by its realizations as a  $4 \times 4$  matrix

$$X^{(k)} = [\mathbf{x}(\tau_k, \omega_1), \dots, \mathbf{x}(\tau_k, \omega_4)] = \{\tilde{x}_{\ell, r}^{(k)}\}_{\ell, r=1}^4. \quad (22)$$

where  $x_{1,1}^{(k)} = 0.918\eta_1^{(k)}$ ,  $x_{1,2}^{(k)} = 1.02\eta_2^{(k)}$ ,  $x_{1,3}^{(k)} = 1.122\eta_3^{(k)}$ ,  $x_{1,4}^{(k)} = 0.918\eta_4^{(k)}$ ,  $\eta_1^{(k)} = -\cos(2k)$ ,  $\eta_2^{(k)} = \sin(\cos(k))$ ,  $\eta_3^{(k)} = -\cos(k)$ ,  $\eta_4^{(k)} = \cos(\sin(k))$ . All training reference signals are simulated as a  $4 \times 484$  matrix  $X = [X^{(1)}, \dots, X^{(121)}]$  shown in Fig. 2 (a). Note that in (3), due to measurement errors the values of  $\tilde{x}_{\ell, r}^{(k)}$  are different from

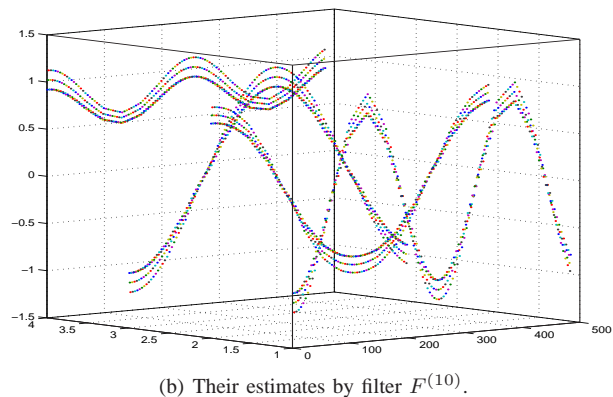
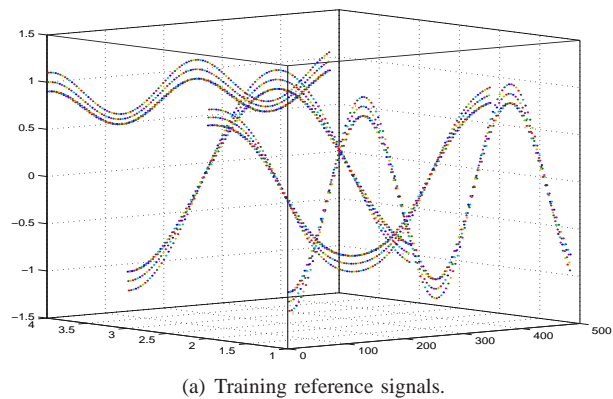


Fig. 2. Training signals and their estimates considered in Example 2.

the values of  $x_{\ell, r}^{(k)}$ . The observed signals in Example 1 were simulated from  $X$  by adding random noise.

The estimates of the reference signals by filter  $F^{(p-1)}$ , for  $p = 11$ , obtained on the basis of the information represented in Fig. 1 (b) are given in Fig. 2 (b). The covariance matrices are estimated from samples  $Y^{(j)}$  and  $\tilde{X}^{(j)}$  taken at times  $t_j$ , for  $j = 1, \dots, 11$  (see Example 1). The averaging polynomial filter [16] gives much worse accuracy.

### III. CONCLUSION

The piecewise least squares linear (LSL) interpolation estimator was developed to estimate a large set of random signals of interest from the set of observed data. The distinctive feature is that *a priori* information can be obtained on only a few reference signals in the form of samples. Since no information of the major part of the set of reference signals is known, such a procedure is called *almost blind* estimation.

The proposed estimator mitigates to some extent the difficulties associated with existing estimation approaches such as the necessity to know information (in the form of a sample, for instance) on *each* random reference signal; invertibility of the matrices used to define the estimators; and demanding computational work.

### REFERENCES

- [1] E. Zeidler, *Applied Functional Analysis, Applications to Mathematical Physics*, Applied Mathematical Sciences 108, Springer, 1997.

- [2] Y. Hua, Fast maximum likelihood for blind identification of multiple FIR channels, *IEEE Trans. on Signal Processing*, 44, no. 3, pp. 661-672, 1996.
- [3] K. Georgoulakis and S. Theodoridis, Blind and semi-blind equalization using hidden Markov models and clustering techniques, *Signal Processing*, 80, Issue 9, pp. 1795-1805, 2000.
- [4] V. Zarzoso and P. Comon, Blind and Semi-Blind Equalization Based on the Constant Power Criterion, *IEEE Trans. on Signal Processing*, 53, 11, pp. 4363-4375, 2005.
- [5] C.-Y. Chi, C.-H. Chen, C.-C. Feng, C.-Y. Chen, *Blind Equalization and System Identification*, Springer, 2006.
- [6] Y. Hua, M. Nikpour, and P. Stoica, Optimal Reduced-Rank estimation and filtering, *IEEE Trans. on Signal Processing*, vol. 49, pp. 457-469, 2001.
- [7] Y. Hua and W. Q. Liu, Generalized Karhunen-Loève transform, *IEEE Signal Processing Letters*, vol. 5, pp. 141-143, 1998.
- [8] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, ISBN 0-13-048434-2, 2002.
- [9] J. Chen, J. Benesty, Y. Huang, and S. Doclo, New Insights Into the Noise Reduction Wiener Filter, *IEEE Trans. on Audio, Speech, and Language Processing*, 14, no. 4, 1218 - 1234, 2006.
- [10] M. Spurbek and P. Schreier, Causal Wiener filter banks for periodically correlated time series, *Signal Processing*, 87, 6, pp. 1179-1187, 2007.
- [11] J. S. Goldstein, I. Reed, and L. L. Scharf, A Multistage Representation of the Wiener Filter Based on Orthogonal Projections, *IEEE Trans. on Information Theory*, vol. 44, pp. 2943-2959, 1998.
- [12] V. J. Mathews and G. L. Sicuranza, *Polynomial Signal Processing*, J. Wiley & Sons, 2001.
- [13] A. P. Torokhti and P. G. Howlett, An Optimal Filter of the Second Order, *IEEE Trans. on Signal Processing*, 49, No 5, 1044-1048, 2001.
- [14] A. Torokhti and P. Howlett, Method of recurrent best estimators of second degree for optimal filtering of random signals, *Signal Processing*, 83, 5, 1013-1024, 2003.
- [15] A. Torokhti and P. Howlett, Optimal Transform Formed by a Combination of Nonlinear Operators: The Case of Data Dimensionality Reduction, *IEEE Trans. on Signal Processing*, 54, no. 4, 1431-1444, 2006.
- [16] A. Torokhti and P. Howlett, *Computational Methods for Modelling of Nonlinear Systems*, Mathematics in Science and Engineering, 212, Series editor C. K. Chui, Elsevier, 2007.
- [17] A. Torokhti and P. Howlett, Filtering and Compression for Infinite Sets of Stochastic Signals, *Signal Processing*, 89, pp. 291-304, 2009.
- [18] A. Torokhti and J. Manton, Generic Weighted Filtering of Stochastic Signals, *IEEE Trans. on Signal Processing*, 57, issue 12, pp. 4675-4685, 2009.
- [19] H. A. Cirpan and M. K. Tsatsanis, Stochastic Maximum Likelihood Methods for Semi-Blind Channel Estimation, *IEEE Signal Processing Letters*, 5, no. 1, pp. 21-24, 1998.
- [20] G. Kutz and D. Raphaeli, Maximum-Likelihood Semiblind Equalization of Doubly Selective Channels Using the EM Algorithm, *EURASIP Journal on Advances in Signal Processing*, Springer Open J., 2010.
- [21] D. He and H. Leung, Semi-Blind Identification of ARMA Systems Using a Dynamic-Based Approach, *IEEE Trans. on Circuits and Systems-I*, 52, 1, pp. 179-190, 2005.
- [22] J. Even and K. Sugimoto, An ICA approach to semi-blind identification of strictly proper systems based on interactor polynomial matrix, *Int. J. Robust Nonlinear Control*, 17, pp. 752768, 2007.
- [23] R.E. Kalman, A New Approach to Linear Filtering and Prediction Problems, *Transactions of the ASME Journal of Basic Engineering*, 82 (Series D), 35-45 1960.
- [24] S. Kang and L. Chua, A global representation of multidimensional piecewise-linear functions with linear partitions, *IEEE Trans. on Circuits and Systems*, 25 Issue:11, pp. 938 - 940, 1978.
- [25] L.O. Chua and A.-C. Deng, Canonical piecewise-linear representation, *IEEE Trans. on Circuits and Systems*, 35 Issue:1, pp. 101 - 111, 1988.
- [26] J.-N. Lin and R. Unbehauen, Canonical piecewise-linear approximations, *IEEE Trans. on Circuits and Systems I: Fundamental Theory and Applications*, 39 Issue:8, pp. 697 - 699, 1992.
- [27] P. Julian, A. Desages, B. D'Amico, Orthonormal high-level canonical PWL functions with applications to model reduction, *IEEE Trans. on Circuits and Systems I: Fundamental Theory and Applications*, 47 Issue:5, pp. 702 - 712, 2000.
- [28] J.E. Cousseau, J.L. Figueroa, S. Werner, T.I. Laakso, Efficient Nonlinear Wiener Model Identification Using a Complex-Valued Simplicial Canonical Piecewise Linear Filter, *IEEE Trans. on Signal Processing*, 55 5, pp. 1780 - 1792, 2007.
- [29] A. Ben-Israel and T. N. E. Greville, *Generalized Inverses: Theory and Applications*, John Wiley & Sons, New York, 1974.
- [30] P. G. Howlett, P. Pudney, and X. Vu, Local energy minimization in optimal train control, *Automatica*, 45(11), 2692-2698, 2009. DOI: 10.1016/j.automatica.2009.07.028.

# A P300-Based Word Typing Brain Computer Interface System Using a Smart Dictionary and Random Forest Classifier

Faraz Akram, Hee-Sok Han, and Tae-Seong Kim

Department of Biomedical Engineering

Kyung Hee University,

Yongin-si, Republic of Korea

farsun@gmail.com, hshan@khu.ac.kr, tskim@khu.ac.kr

**Abstract**— The conventional P300 brain computer interface (BCI) system for character spelling is typically composed of a paradigm that displays flashing rows or columns of characters and a P300 classifier from which a target character gets recognized. One significant drawback of this system in practice is its typing speed which could take a few minutes to type each character of a target word. In this work, we propose a novel BCI system through which a whole word can be typed with much higher word typing speed and accuracy. In our presented system, we have integrated a custom-built smart dictionary to give word suggestions upon few key characters initially typed by the user. Upon the suggested words, the user can select one out of the given suggestions to complete word typing. Our novel paradigm significantly reduces the word typing time and makes words typing more convenient. In the classification part, we have also adopted a new classifier, Random Forest (RF) instead of a commonly used Support Vector Machine (SVM). Our results with four subjects using the presented word typing system demonstrate an average typing time of 1.66 minutes per word, whereas the conventional took 2.9 minutes, improving the typing time by 42.75%. Also RF improves the P300 classification accuracy significantly, outperforming SVM. Our presented system could be useful for practical human computer interaction applications.

**Keywords**-P300; Brain Computer Interface; Word Typing; Human Computer Interaction

## I. INTRODUCTION

Brain Computer Interface (BCI) is a system that provides a way of communication between a computer or device and the brain. The primary goal of a BCI is to enable severely disabled peoples to communicate and control their external environment without relying on peripheral nerves and muscles. Yet it is not restricted to disabled peoples only, many applications for normal peoples are emerging. A BCI system that utilizes P300 event related potentials of the brain is known as P300 BCI. One application of P300 BCI was first demonstrated for character spelling by Farwell and Donchin [1]. Since then Farwell and Donchin (FD) paradigm has been most widely used for P300-based character spelling [2,3]. Most of the later research followed the same scheme and was focused on improving the classification accuracy and speed.

To improve the classification accuracy, many techniques have been suggested for P300 classification such as support vector machine (SVM) [4], linear discriminant analysis (LDA) [5], and neural network [6] and lately for P300

extraction such as independent component analysis (ICA) [3] and constrained ICA [7]. Recently, there is a growing interest in designing efficient paradigms. Various attempts have been made to modify the FD paradigm. For instance, in [8] Salvaris and Sepulveda made various changes to the visual aspects of the FD paradigm. In [9], Allison and Pineda tried three different matrix sizes such as 4×4, 8×8, and 12×12 to investigate the effect of matrix size on the amplitude of P300. In [10], Guan et al. used a single character flipping instead of row and column intensifications to improve classification accuracy. Single character flipping reduced the target probability, hence increasing the amplitude of P300. In [11], Townsend et al. used an 8×9 checkerboard paradigm to eliminate this double flash problem, improving accuracy.

A common drawback of all these conventional P300 spelling systems is that to type a word a user has to spell each character of a word one at a time. This spelling process is slow and can take several minutes to finish typing a single word. A new paradigm is in need by which a user can type a whole word in less typing time and effort. Some attempts have been made in this regards. In [12], Ahi et al. used a dictionary driven P300 speller. They integrated a custom-built dictionary of 942 four-lettered words into the classification system of P300 speller for automatic correction of misspellings. However, the dictionary was used only for word correction and the user had to spell all the characters of a target word. In [13], Ryan et al. proposed a predictive spelling scheme with words suggestions, but in their work, classification accuracy was decreased due to a higher workload in their displays. In [14], Kaufmann et al. used a German language predictive speller with some commonly used German words. They obtained the similar results due to the required high attentions.

In this paper, we propose an improved BCI system in the words suggestions and selection paradigms by reducing the workload to the user and by employing a better classifier. In this work, we have implemented a smart dictionary to give words suggestions to the user in order to reduce the typing time. In our proposed scheme, a user only needs to spell few initial characters of a word. Then possible words matching the typed initials will be suggested to the user. Finally, the user only needs to select one of the suggestions. This scheme reduces the spelling time and thereby increases information transfer rate. Our proposed scheme also reduces the visual fatigue by reducing the number of characters required to type a word, hence a user can better concentrate on typing in longer trials. We have tested our P300 words typing BCI and



have increased the typing speed by 42.75% in comparison to the conventional character typing under the same conditions. Also, we have adopted a new RF classifier for P300 classification to improve the accuracy of typing. Our results, in comparison to a commonly used classifier in BCI, show that RF outperforms SVM in the typing accuracy. Our presented system should be useful as a practical P300 BCI spellers. Also it could be used in the application areas of human computer interaction for word typing.

The paper is organized as follows: Section II describes the conventional speller and our proposed word typing methodology with classification. The results are presented in Section III, followed by a conclusion in Section IV.

## II. METHODOLOGY

### A. The Conventional P300 Speller

The conventional P300 speller consists of a 6x6 matrix of characters and numbers (much similar to Fig. 1 (a)) in which each row and column flashes randomly. A user is asked to focus on a target character. P300 is elicited when a row or column containing the target character flashes. The target character is detected by identifying the row and column containing P300s. The user completes word typing by spelling every character of the target word.

### B. Overview of Our Proposed P300-based Word Typing System

In our proposed system, we have added a words suggestion mechanism with a smart dictionary to the conventional character spelling paradigm. Our system consists of two paradigms: the first paradigm is a matrix of 6x5 with characters, as shown in Fig. 1(a). When a session starts, the first paradigm is shown to a user and the user spells few initial characters of a desired word. The typed initial characters get in to the dictionary module and the dictionary searches for the words matching the initial characters. When the number of suggested words is less than a pre-specified threshold (in our case of nine), the searched words are displayed as suggestions. Finally, the user is asked to select one out of those suggestions as shown in Fig. 1(b). The user selects the target word using the second paradigm.

### C. Initial Character Spelling Paradigm

In the implementation of the initial character spelling paradigm as shown in Fig. 1 (a), we have used a 6x5 matrix of characters similar to the FD paradigm in which each row and column gets intensified randomly. Intensifications are block randomized and in each block of eleven intensifications, each row or column gets intensified exactly once in a random order. For one character epoch, this block of intensifications is repeated fifteen times. After each character epoch, there is a 2.5s blank time before starting for the next character. This blank time indicates a user that one character is completed and the user must spell the next character of a word to be typed. Intensification time is 100ms with a 75ms blank time between the intensifications according to the standard P300 BCI data of BCI competition III [2].

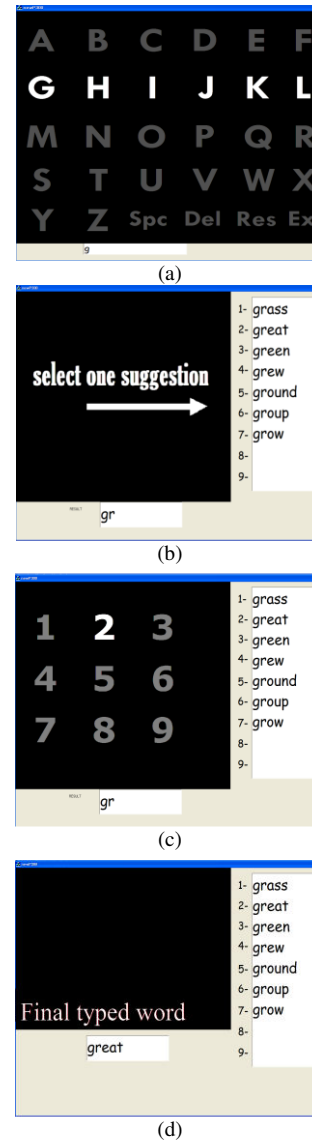


Figure 1. Our word typing paradigms. (a) The first paradigm for spelling initial characters, (b) a words suggestion screen, (c) the second word selection paradigm, and (d) a display showing the final typed word

A randomly chosen target word is displayed before the start of intensifications. The user focuses on a target character and silently counts the number of times of each row or column containing the target character intensified. P300 potentials are elicited when the row or column containing each target character gets flashed.

### D. Dictionary Module

The dictionary module is implemented in the form of a Ternary Search Tree (TST). TST is a special prefix tree ('Trie') data structure that can find all key words having a given prefix. Partial matches can easily be searched. The advantage of using the prefix tree is its fast searching, but it has a disadvantage of its high storage requirements [15]. TST handles the storage requirement by combining prefix tree with the binary search tree [16]. For an online system, a

method is needed to search the dictionary with less access time. TST can perform this very efficiently with less storage requirements. Implementation of TST was done based on the work of Bentley and Sedgewick [16]. Our dictionary consists of 2,000 most commonly used English words.

E. The Second Word Selection paradigm

The second paradigm is a matrix of 3x3 used to select a word out of the suggested words given by the dictionary as shown in Fig. 1(c). Since it is known that the row or column-wise intensifications with the 3x3 matrix size decrease P300 amplitude and the P300 amplitude has an inverse relationship with a priori probability of target stimulus [9], in the second paradigm, we used a single number intensification, instead of intensifying rows and columns, to decrease the priori probability of the target stimulus. After the word selection, the final word gets typed as shown in Fig. 1(d). Table I shows the timing specification for both paradigms.

TABLE I. TIMING INFORMATION OF BOTH PARADIGMS

	First Paradigm	Second paradigm
Intensification time	100ms	100ms
Blank time between intensifications	75ms	75ms
Total stimuli	11	9
Character repeat	15	15
Blank time between characters	2.5s	2.5s

Timings of both the paradigms are same except the total number of stimuli: the first paradigm has 11 stimuli (i.e., 6x5 matrix) whereas the second paradigm only 9 (i.e., 3x3 matrix).

F. Classification

In this study, we adopted the RF classifier, an ensemble learning technique introduced by Breiman [17]. It is a powerful classifier that has been used efficiently in other areas of classification but is relatively unknown in the field of BCI. The main idea of RF is to combine multiple independent decision trees and let them vote for the popular class. We have compared the performance of RF against SVM. Each subject participated in training and testing sessions. In the training session, each subject was instructed to spell ten randomly selected characters. We applied a band-pass filter with the cutoff frequencies of 0.1Hz and 25Hz. Then epochs of 600ms after the stimulus onset were extracted. One character data contains two targets and 9 non-target epochs (i.e., two out of eleven rows and columns contain P300s). To balance the training data, only two randomly chosen non-targets and two targets are used. Segments of data are concatenated over the channels to create a single feature vector. A total of six channels was used for classification. Both classifiers were trained using the same features.

III. RESULTS

We conducted the word typing experiments with four healthy subjects. The consent forms were obtained from each subject. The EEG data was acquired through a 32-channel

BrainAmp MR amplifier [18] with a sampling frequency 250Hz. Electrodes were placed according to the 10-20 international standards. Six channels out of 32: namely Cz, Pz, P3, P4, O1, and O2, were used in typing words with the proposed system. There were ten test sessions and in each session, each subject typed one word, totaling ten words per subject. Table II shows the accuracy of four subjects in terms of words using SVM and RF.

TABLE II. ONLINE ACCURACY

Subject	Accuracy (%)	
	SVM	RF
S1	60	70
S2	80	90
S3	80	80
S4	90	100
Mean	77.5	85

Based on the timing information given in Table I, we have computed the time required to spell one character using both the paradigms. For the conventional and our first paradigm, the total number of intensification is 165 (i.e., 11 stimuli x 15 repetitions) and 175 ms (i.e., intensification time + blank time between intensifications) is the time for one intensification. Therefore the time required to spell a single character using the conventional paradigm comes out to be 31.375s (i.e., [165 x 175] ms + 2.5s blank time between characters). In the same way, this comes out to be 26.125s for the second paradigm.

With these numbers, we computed the time required by the conventional paradigm to spell the target words and compared it with the actual time taken by the proposed paradigm for the same words. Table III shows the time required to spell these words using the conventional paradigm with 100% accuracy.

TABLE III. WORD TYPING TIME USING THE CONVENTIONAL SCHEME

Word Number	Word Typing Time Using the Conventional Scheme (minutes)			
	S1	S2	S3	S4
1	2.61	2.09	2.09	4.18
2	2.61	2.61	3.13	2.09
3	3.13	2.61	2.09	3.66
4	2.09	2.61	2.61	4.18
5	4.71	4.18	4.18	2.09
6	3.13	3.66	3.13	2.09
7	3.13	2.61	3.66	3.13
8	2.61	2.61	2.61	2.61
9	2.09	3.13	2.61	2.09
10	2.61	4.18	2.61	2.09
Mean	2.87	3.03	2.88	2.82
Grand mean	2.9			

Table IV shows the time taken by the proposed scheme to spell the same words. The conventional spelling required an average time of 2.9 minutes per word while with the proposed took an average typing time of 1.66 minutes per word, decreasing the typing time by 42.75 %. Note that in this study, we have used the standard timings as used in the conventional spellers to compare the performance of the proposed against the conventional paradigm. Typing speed

can be improved by reducing the number of stimulus repetitions and stimulation rate.

TABLE IV. WORD TYPING TIME USING THE PROPOSED SCHEME

Word Number	Word Typing Time Using the Proposed Scheme (minutes)			
	S1	S2	S3	S4
1	1.48	1.48	2.00	1.48
2	1.48	2.00	0.96	1.48
3	2.00	1.48	1.48	2.00
4	0.96	2.00	2.00	2.00
5	1.48	2.00	2.00	1.48
6	2.00	2.00	2.00	1.48
7	1.48	2.00	1.48	2.00
8	1.48	1.48	1.48	1.48
9	1.48	2.00	2.00	1.48
10	1.48	0.96	2.00	1.48
<b>Mean</b>	<b>1.53</b>	<b>1.74</b>	<b>1.74</b>	<b>1.64</b>
<b>Grand mean</b>	<b>1.66</b>			

We also compared the offline classification results of RF against SVM for different number of stimulus repetitions. Fig. 2 shows the grand mean classification accuracy vs. the number of stimulus repetitions for both SVM and RF over all subjects.

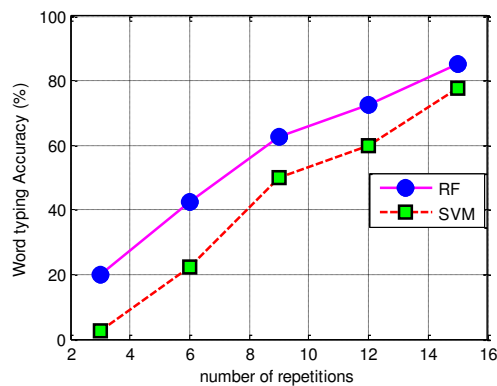


Figure 2. Grand averaged classification results over four subjects for RF vs. SVM with different number of repetitions

Our results show that the RF classifier gives a better performance on P300 classification even under lesser repetitions.

IV. CONCLUSION

This study is aimed at developing an efficient and convenient P300-based word typing BCI. The word typing will be enhanced in terms of speed and accuracy through the proposed word typing paradigm, a smart dictionary, and RF classifier. Our results show that the proposed method of P300 word typing system not only increases the speed of typing but also makes the spelling task easier for users with less fatigue. This could be one important advantage to become a practical BCI.

ACKNOWLEDGMENT

This research was supported by MKE (The Ministry of Knowledge Economy), Korea under the ITRC (Information Technology Research Center) support program supervised

by the NIPA (National IT Industry Promotion Agency) (NIPA-2013-(H0301-13-2001)).

REFERENCES

- [1] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and Clinical Neurophysiology*, vol. 70, issue 6, Dec. 1988, pp. 510-528.
- [2] B. Blankertz, et al. "The BCI competition III: Validating alternative approaches to actual BCI problems," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, June 2006, pp. 153-159.
- [3] H. Serby, E. Yom-Tov, and G. F. Inbar, "An improved P300-based brain-computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 13, no. 1, Mar. 2005, pp. 89-98.
- [4] M. Kaper, P. Meinicke, U. Grossekhoefer, T. Lingner, and H. Ritter, "BCI competition 2003 - Data set IIb: Support vector machines for the P300 speller paradigm," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, June. 2004, pp. 1073-1076.
- [5] U. Hoffmann, J. M. Vesin, T. Ebrahimi, and K. Diserens, "An efficient P300-based brain-computer interface for disabled subjects," *Journal of Neuroscience Methods*, vol. 167, no. 1, Jan. 2008, pp. 115-125.
- [6] H. Cecotti and A. Graser, "Convolutional neural networks for P300 detection with application to brain-computer interfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, Mar. 2011, pp. 433-445.
- [7] O. I. Khan, et al. "Robust extraction of P300 using constrained ICA for BCI applications," *Medical & Biological Engineering & Computing*, vol. 50, no. 3, Mar. 2012, pp. 231-241.
- [8] M. Salvaris and F. Sepulveda, "Visual modifications on the P300 speller BCI paradigm," *Journal of Neural Engineering*, vol. 6, no. 4 Aug. 2009, p. 046011.
- [9] B. Z. Allison and J. A. Pineda, "ERPs evoked by different matrix sizes: Implications for a brain computer interface (BCI) system," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, June. 2003, pp. 110-113.
- [10] C. Guan, M. Thulasidas, and J. Wu, "High performance P300 speller for brain-computer interface," *Proc. IEEE Int. Workshop Biomed. Circuits Syst.*, Dec. 2004, pp. S3-5/INV-S3/13-16.
- [11] G. Townsend, et al. "A novel P300-based brain-computer interface stimulus presentation paradigm: Moving beyond rows and columns," *Clinical Neurophysiology*, vol. 121, no. 7, July. 2010, pp. 1109-1120.
- [12] S. T. Ahi, H. Kambara, and Y. Koike, "A dictionary-driven P300 speller with a modified interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 19, no. 1, Feb. 2011, pp. 6-14.
- [13] D. B. Ryan, et al. "Predictive spelling with a P300-based brain-computer interface: Increasing the rate of communication," *International Journal of Human-Computer Interaction*, vol. 27, no. 1, Dec. 2010, pp. 69-84.
- [14] T. Kaufmann, S. Völker, L. Gunesch, and A. Kübler, "Spelling is just a click away—a user-centered brain-computer interface including auto-calibration and predictive text entry," *Frontiers in Neuroscience*, vol. 6, no.72, May. 2012, doi: 10.3389/fnins.2012.00072.
- [15] E. Fredkin, "Trie memory," *Communications of the ACM*, vol. 3, no 9, Sept. 1960, pp. 490-499, doi:10.1145/367390.367400.
- [16] J. L. Bentley and R. Sedgewick "Fast algorithms for sorting and searching strings," *SODA '97 Proceedings of the eighth annual ACM-SIAM symposium on Discrete algorithms*, Society for Industrial and Applied Mathematics Philadelphia, USA, Jan. 1997, pp. 360-369.
- [17] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, Oct.2001.
- [18] BrainProducts, Munich, Germany, BrainAmp MR, <http://www.brainproducts.com>

# Research and Application of the Radar Intelligent Fault Diagnosis System Based on Dual-Mode Fusion

Cheng Xie-feng

College of Electronic Science & Engineering  
Nanjing University of Posts and Telecommunications,  
Nanjing, China  
chengxf@njupt.edu.cn

Cheng Hui-zhong

Atmospheric Observation Technology  
Center of Yunnan protection  
Kunming, China  
ynchenghuizhon@163.com

**Abstract**—The emergence of a new generation of weather radar increases the complexity of radar control system and the difficulty of troubleshooting. In order to ensure the reliability and maintenance performance of weather radar systems, this paper describes a dual-mode weather radar intelligent fault diagnostic expert system based on the author's many years of experience in radar maintenance. The radar fault diagnosis principle of wavelet back propagation neural network with improved case-based reasoning is firstly introduced, and the dual-mode fusion algorithm based on the improved diagnosis system is given later. At the end, the paper used 3830 weather radar system as an example, focusing on analyzing the failure phenomenon of system drive and the failure types, combined with the practical application to show that dual mode fusion algorithm can effectively detect the fault of weather radar in device level, plate level and chip level, three levels of diagnosis. This application example of radar fault diagnosis demonstrated that the method described in this article is effective and practical.

**Keywords**-Radar Fault Diagnosis; wavelet BP Neural Network; Case-Based Reasoning; Dual-Mode Fusion.

## I. INTRODUCTION

In recent years, many scholars study the Fault Diagnosis Expert Systems, the use of human experts to provide expertise, simulated the thinking process of human experts, like a human expert to solve some technical problems in the field of intelligent systems. Among them, Khomfoi and Tolbert [1] proposed the Fault Diagnosis System for a multilevel makes using a Neural Network. Zhihong and Le [2] studied about radar test and repair support system. Chien [3] offered the Design and Implementation of a CBR (Case-based Reasoning) system for Marketing Plans. Cheng and Ma [4] proposed the data of multi-feature extraction-based on the data layer fusion and improved D-S information fusion method. New expert systems are developed for simple monotype and single subject to modalities and comprehensive, with many experts coordination, a variety of knowledge representation, artificial neural networks, the latest artificial intelligence technology such as knowledge acquisition and learning mechanism has become the important features of the new expert system.

The emergence of a new generation of weather radar

improves the radar function and accuracy, but also increases the complexity of radar control systems and the difficulty of troubleshooting them. In order to ensure the radar system reliability and fast, accurate diagnosis, various fault diagnosis technologies combined with artificial intelligence method is the research hot spot, radar fault diagnosis technology has developed to the intelligent diagnosis stage. Diagnostic expert system based-on our many years of experience in radar maintenance, this paper designs a BP (Back Propagation)-CBR-based weather radar intelligent fault diagnosis expert system.

The article first introduces the principle of radar fault diagnosis based on BP neural network and CBR, and then discusses intelligent fault diagnosis method, finally gives applications in weather radar. As the long-term mechanical wear, radar servo system motor drive and motor are a high failure probability, we focus on CINRAD / CC radar antenna servo system hardware [6] characteristics detailed introduction of radar servo system motor drives and motor failure of reasoning method to judge, as well as hardware replacement parameter detection set expert guidance program. Three CINRAD/CC radar fault intelligent diagnosis cases show that this proposed method is effective and practical.

## II. PRINCIPLES OF RADAR FAULT DIAGNOSIS BASED ON DUAL-MODE FUSION

### A. Wavelet BP method

BP (Back Propagation) [1] is a widely used feed-forward neural network model, which is superior to other network forms for the systematicness, completeness and ease application of its algorithm; it is particularly suitable for fast pattern recognition and classification applications.

In order to improve the operating speed, the improved BP neural network model of Morlet wavelet function [4] is adopted in this paper. It has three layer structures: input layer X, middle H layer and output layer Y, each layer of network is composed by multiple neurons, as shown in Figure 1. The input layer of the network takes M nodes according to the type number of failure, output layer for fault diagnosis conclusion, suppose that there is N conclusion and implicit layer was washed with K nodes reasoning judgment. BP

generally chooses S function as K node role function  $f(x)=1/(1+e^{-x})$ . The system selects the Morlet wavelet activation function as hidden layer neurons according to the characteristics of the radar system circuit failure [5, 6].

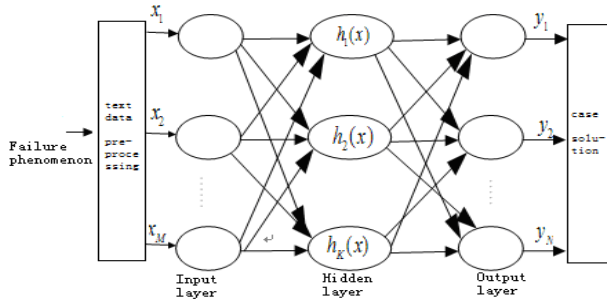


Figure 1. Wavelet BP neural network process

$$h\left(\frac{x-b}{a}\right) = \cos\left(1.75\frac{x-b}{a}\right) \exp\left(-0.5\left(\frac{x-b}{a}\right)^2\right) \quad (1)$$

The input and output samples is the given  $P(p=1,2,\dots,P)$  group, learning rate is  $\eta(\eta > 0)$ , momentum factor is  $\lambda(0 < \lambda < 1)$ , the target error function is

$$E = \sum_{p=1}^P E^p = \frac{1}{2P} \sum_{p=1}^P \sum_{n=1}^N (d_n^p - y_n^p) \quad (2)$$

$d_n^p$  is the n node expected output of the output layer,  $y_n^p$  is the network actual output.

The goal of the algorithm is to make the error function reach a minimum by constantly adjust the parameters of the network.

Hidden layer output is [5]:

$$O_k^p = h\left(\frac{I_k^p - b_k}{a_k}\right), \quad I_k^p = \sum_{m=1}^M w_{km} x_m^p \quad (3)$$

$x_m^p$  is the input of the input layer,  $O_k^p$  is the output of the hidden layer,  $w_{km}$  is the weight between input layer node m and hidden layer k node,  $h()$  is the Morlet wavelet function.

The output of the output layer is [5]:

$$y_n^p = h(I_n^p), \quad I_n^p = \sum_{k=1}^N w_{nk} O_k^p \quad (4)$$

$I_n^p$  is the input of the output layer,  $w_{nk}$  is the weight between hidden layer nodes.

k and output layer node n,  $y_n^p$  is the fault diagnosis result.

The weight adjustment formula between the nodes of the input layer and the hidden layer nodes is:

$$w_{km}^{new} = w_{km}^{old} + \eta \sum_{p=1}^P \delta_{km} + \lambda \Delta w_{km}^{old} \quad (5)$$

$w_{km}^{old}$ 、 $w_{km}^{new}$  is respectively the weight value before the adjustment and the weight value after the adjustment

between the input layer node m and the hidden layer node k,  $w_{km}^{old}$  is the momentum item.

In the training, the momentum is added in the weights and threshold correction algorithm and the former correction value is used to smooth learning path step, to avoid falling into the local minimum value and accelerate learning speed. In order to avoid causing the oscillation of the weights and threshold correction in the sample-by-sample training, we adopt batch training methods. The output of the network is not a simple weighted sum, but the network hidden layer wavelet node is weighted summation firstly, and then converted by the Sigmoid function [7], to obtain the final output of the network, this is conducive to the handling classification problems and reduce the possibility of divergence in the training process at the same time.

The direct fusion of wavelet and BP neural network, namely wavelet yuan [1] instead of neurons, is that the weights of the input layer to the hidden layer and the hidden layer threshold are respectively instead by the scaling of wavelet function and the translation parameters. On the same learning task, wavelet BP network has the characteristics of self-learning, adaptive, and fault tolerance, and it could avoid blindness on the BP neural network structure design, the structure is more simple, the speed of convergence is faster.

### B. Improved CBR method

CBR puts forward solutions for the current fault by summing up the expert fault diagnosis experiences, using experts have successfully solved similar problems, comparing background and conditions of the old and new fault analysis, adjustment, modifying the existing fault diagnosis experience; so, it is also known as case-based reasoning method [2].

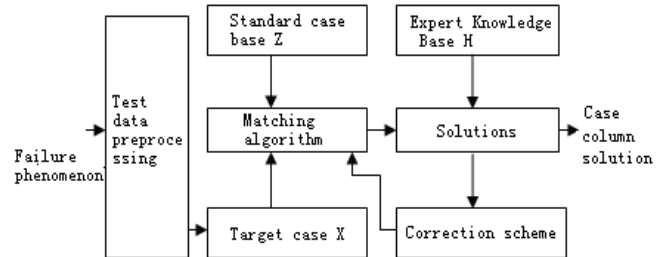


Figure 2. Improved CBR process diagram

Case Representation: Disposed case model,  $G = (A, B, C, S, R)$ , which is based on the case model of the feature vector, each case G should contain the following: (1) Case No. A; (2) Case information B (fault phenomenon description, fault source); (3) Case description C behavior characteristics 1 (index 1, value 1), ..., behavior characteristics N (index N, value N); (4) Solution S is the solution method for the failure; (5) Case retrieve information R (retrieval sequence number, frequency of use).

Standard case base Z stores series of case G in an interconnected database, Standard case base is divided into three layers: The first layer is the system layer, which is divided according to the main components in the weather

radar system, such as antenna device system layer, power plant system layer, operating platform system layer, display device system layer, etc. The second layer is the representative case layer, corresponding to the typical representative cases that are under the system layer. The third layer is the sub-case layer which is a collection of a certain class of cases, each case of the same sub-case layer should be as similar as possible, each case of the different sub-case layer should be as far as possible not similar, and the same type of sub-case number should be at least 3.

Case Retrieval: The fault system uses the “hierarchical retrieval, the distance matching” method, focusing on the fault eigenvectors of target case, it makes sure that the failure occurred in what parts, and then finds the corresponding representative case in the system layer, the last fine in the sub-case select the cases corresponding to the minimum similarity distance, finally selects the corresponding minimum similarity distance case in the sub-cases. In a predetermined threshold range, there may be more than one sub-case to meet the requirements, or cannot find a matching case, at the time it could call the expertise libraries knowledge, for those who probability of occurrence is high, or have occurred many times of failure as a fixed case for feedback matching, regard the final solution as a case for a new solution and add to the standard case base after new solutions normalized. If there are hardware replacement requirements in solution, then the hardware replace adjusting parameter database will update the parameter testing procedures and specific parameter value debugging method. An improved CBR process is shown in Figure 2.

Similar distance function matching method: Case matching algorithm uses the similarity distance function matching method, which describes the credits of characteristic data group between target case X and the standard case Z. the similar distance function is:

$$S_i = 1 - \frac{\left| \sum_{j=1}^N X_j \cdot W_{ij} \cdot Z_{ij} \right|}{\sqrt{\sum_{j=1}^N X_j^2 \cdot \sum_{j=1}^N Z_{ij}^2}} \quad (6)$$

where  $X_j$  is the j-dimensional vector-valued of target case,  $Z_{ij}$  is the j-th vector-value of i-th standard sub case,  $S_i$  is the similarity between the target case and i- th standard sub case,  $W_{ij}$  is the weight coefficient, which is usually determined by expert experience data.

The definition of the similar distance function between the target case X and i- th standard case  $Z_i$  is:

$$d_i = 1 - S_i < \theta \in (0,1) \quad (7)$$

where  $\theta \in (0, 1)$  is the determination threshold, when d is less than the threshold, we can see that the fault exists similarity between standard case and target case.

### C. Dual mode fusion algorithm

Multi-modal fusion fault diagnosis is the information processing technology that developed in the last decade, it males a variety of diagnostic information intelligent process as a whole, results more accurate and complete judgment than a single diagnosis and ultimately gives higher reliable

fault diagnosis results [6,7]. It involves a multifaceted theories and techniques, such as signal processing, estimation theory, uncertainty theory, pattern recognition, and optimization techniques, neural networks, artificial intelligence, and so on. Multi-modal fusion fault diagnosis usually belongs to the decision-making level fusion, which associates the diagnostic results of different algorithms, and then gets the joint inference results through the correlation processing, decision-making fusion decision. The common methods are Bayesian inference [7], Dempster-Shafer evidence theory [8] and fuzzy set theory [8], the expert system [3], etc.

The overall diagnostic result is actually the decision that is made after integrate the individual diagnostic information. In the decision-level fusion of dual-mode radar fault diagnosis results, We actually obtained the number of diagnostic information is at least two or more than two, but it does not say that they have completely and accurately represent the overall diagnostic conclusions; so, the decision-making level fusion is needed. In the view of Bayesian statistical point, their diagnostic information can be viewed as random variables, these random variables is not necessarily the true value of the overall diagnosis, but only a random performance of the overall truth value, but they are all fraught with some information of the overall true value. Therefore, we could make re-integration of this information, eventually infer the overall diagnostic information, and give the best fault diagnosis results.

According to the Dempster-Shafer theory of evidence [8,9], setting up  $h_i$  ( $1 \leq i \leq m$ ) as the failure basic probability distribution function exported by m independent diagnostic, the probability distribution function  $h$  jointly acted by m independent diagnosis is:

$$h = h_1 + h_2 + \dots + h_m = \begin{cases} h(\gamma) = 0 & (D = 1) \\ h(K) = \frac{\sum_{(k_1 \cap k_2 \cap \dots \cap k_m = K)} \prod h_i(k_i)}{1 - D} & (D \neq 1) \end{cases} \quad (8)$$

The  $h(K)$  is the basic credibility of k; D represents the measure between different fault, The D is the greater, the greater the conflict between each fault, with:

$$D = \sum_{k_1 \cap k_2 \cap \dots \cap k_m \neq \gamma} \prod h_i(k_i) \quad (1 \leq i \leq m) \quad (9)$$

When there are M different diagnostic conclusions for the target, the sample library has J fault samples, the similar distance between target and sample is d, then the rate membership matrix of m times diagnosis is:

$$\beta_{m \times J} = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1J} \\ d_{21} & d_{22} & \dots & d_{2J} \\ \vdots & \vdots & \dots & \vdots \\ d_{m1} & d_{m2} & \dots & d_{mJ} \end{bmatrix} \quad (10)$$

The basic probability function value can be determined by the following formula [7, 8]:

$$h_m(k_i) = \delta_m \frac{1}{d_{mi}} \quad (i=1,2,\dots,J) \quad (11a)$$

$$h_m(\Theta) = 1 - \sum_i^J \frac{1}{d_{mi}} \delta_m \quad (11b)$$

$h_m(k_i)$  indicates the confidence level of events "for diagnosis fault is the  $i$  fault of the standard fault library" confidence level,  $\delta_i$  is the weight. 2 factors should be mainly considered in setting the weight  $\delta_i$ : (a) Performance weight factors: Different fault feature extraction methods have different characteristics, so the fault diagnosis rate and stability are different too. After a large number of experiments, statistical diagnostic results in the actual environment, this paper uses average correct diagnosis rate of failure as performance weights. (b) Correlation weight factors: Making use of the improved DS combination rule requires the independence of each other between the various types of fault cases. However, there is a inevitably correlation between the information provided by each failure characterization in the practical application, so we should give larger weights for the separate fault cases, on the contrary, give smaller weights.

Thus, when  $k_1, k_2 \subset \Theta$ ,  $\Theta$  is the fault diagnostic framework, we take J sample of fault library. We could get:

$$h(k_1) = \max\{h(k_i)\}, (k_i \subset \Theta, i = 1, \dots, J),$$

$$h(k_2) = \max\{h(k_i)\}, (k_i \subset \Theta, \text{且 } k_i \neq k_1, i = 1, \dots, J) \text{ and:}$$

$$\begin{cases} h(k_1) - h(k_2) > \alpha_1 \\ h(\Theta) < \alpha_2 \\ h(k_1) > h(\Theta) \end{cases} \quad (12)$$

The final diagnosis result is  $h(k_i) \cdot h(\Theta)$  represents uncertainty,  $\alpha_1, \alpha_2$  is the set threshold. A dual-mode fusion radar intelligent fault diagnosis system based on improved DS is shown in Figure 3. Failure phenomena were respectively sent into wavelet BP neural network module and the improved case-based reasoning module for fault diagnosis of the first phase, the resulting case solution  $h_i, h_j$  and then fed into the improved the DS dual mode fusion module for fusion of the second order to obtain the highest reliable fault diagnosis conclusion and failure solutions, if the fault solutions exist hardware replacement requirements, the hardware replacement adjustment parameter database will give the corresponding parameter test steps after replace the components and specific debugging method of parameter values, which is a very practical expert knowledge base designed based on the author's many years of work experience [6, 10].

In the maintenance of the electronic systems, judging which one overall device damage to replace the whole device is often called a device-level maintenance, judging which one Printed Circuit Board (PCB) damage of device and replace the circuit board is called plate level maintenance, judging which one element, chip burnt out of the circuit board and replace the components and chip is called the maintenance of chip level. Therefore, we divide fault

diagnosis level of the weather radar system into corresponding three levels: device level fault diagnosis, plate level fault diagnosis, chip level fault diagnosis, and give the corresponding failure maintenance solutions of the device-level, plate level and chip-level. The experimental data show that the wavelet BP method is particularly suitable in the chip-level fault diagnosis, improved CBR method is particularly suitable for device-level fault diagnosis; the plate level fault diagnosis method can use both.

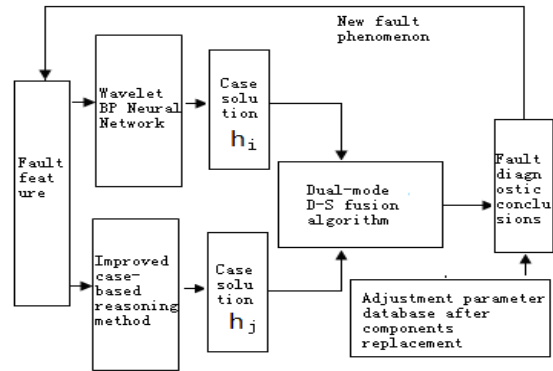


Figure 3. The block diagram of radar intelligent fault diagnosis system based on the improved d-s dual mode fusion

### III. CASE ANALYSIS

Based on 3830 weather radar system [6], as an example, the analysis of the system of drive fault phenomena, fault type is the key. The drive function is receiving control instruction sent by servo control panel and R/D conversion board driver interface circuit, including antenna speed steering instructions, positioning position command, control mode selection instruction, etc., receiving motor with rotary encoder (code disc) delivery of the antenna at present speed, steering, etc. state information, and eventually generating a driving antenna rotational drive signal from the internal operation treatment sent to drive antenna scanning bearing motor and pitching machine. The 3830 weather radar system drive function diagram is shown in Figure 4. The drive maintenance of 3830 weather radar system is defined as a device-level fault diagnosis, the maintenance of drive plate and chip, is defined as the drive plate and chip plate level and chip-level fault diagnosis.

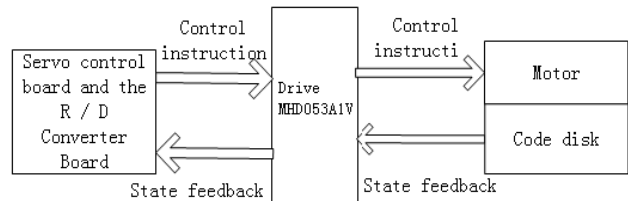


Figure 4. Functional diagram of 3830 weather radar system drive

#### A. Application of wavelet BP cases

For the maintenance characteristics of the drive plate and chip we set eight test points on the drive plate, only consider

the 11 main consumable components transistors and integrated circuits, define the type of fault as 11, wavelet BP output nodes as 11, the input nodes as 8, the number the hidden layer wavelet function nodes can be adjusted according to the actual training effect, in this case is taken as 16. BP wavelet fault diagnosis algorithm specific implementation steps are as follows:

1) Initialization of the network parameters. The random wavelet dilation factor, translation factor, the initial value of the network connection weights, Learning rate = 0.8, momentum factor = 0.1, learning number = 2000, counter = 1, system error = 0.001, is shown in Figure 5.

2) Entering the learning samples and network training. The training samples are usually obtained by a number of experimental data, or by entering the value after the normalization of Electronic Design Automation (EDA) software simulation data [6] into the network as training sample, less than 1000 training, and network convergence.

3) Circuit Fault Diagnosis. The voltage value of the test fault circuit and the test data after pretreatment input to the wavelet BP network have been trained well for the fault diagnosis of the circuit. For example, let the servo drive MHD053A1V plate in the circuit shown in Figure 6 VDC = 0V, namely the circuit's power supply is disconnected; the 7-th test point voltage value is 0V. Input all eight test point voltage value after normality to the wavelet BP network and the trained wavelet BP network can give the circuit fault diagnosis results for "VDC = 0V" number. Diagnostic results are the same with the actual situation, showing that the present method is effective for circuit fault diagnosis.

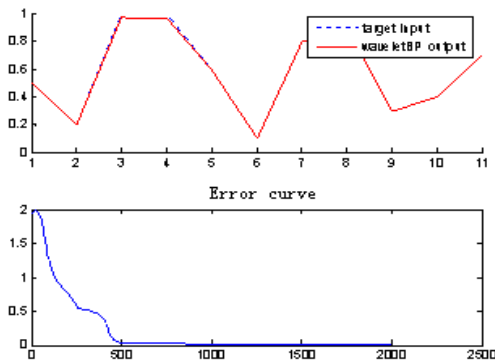


Figure 5. Wavelet BP neural network learning Effect and error curve

**B. CBR application cases**

3830 weather radar drive a fault case is shown in Table 1. Failure case Numbers is TQ3 and the T of TQ3 represents system layer of antenna device, Q represents case layer drive, 1 represents drive of the cases is 1 case, take  $W_{ij}=1$ .

**C. The application of the dual-mode Fusion**

According to the previous dual-mode fusion algorithm and fault determination rules, wavelet BP usually takes larger value in credibility of the chip-level fault diagnosis, So their weights in the chip-level fault diagnosis can be set at 95%, 75%;

However, the improved CBR method take a larger value in the credibility of device-level fault diagnosis, their weights were set at 75%, 95%; Both methods have their own strengths in plate level fault diagnosis, the weights can be set at 50%, 50%. Then, follow the improved DS's dual mode fusion algorithm described in Section 2.3 to fusion, the  $\alpha_1, \alpha_2$  were respectively 0.6, and 0.1.

**D. The adjustment parameter database after components changing**

Because after components replacement, usually, there is a need for adjustment and settings of some parameters to ensure that the radar system works in the best state; so, the fault diagnosis system especially adds an adjustment parameter database. For example, the CINRAD / CC radar servo system drive replacement parameter settings shown in Table 2. The calibration setting after the radar servo system antenna replacement is shown in Table 3.

**E. Failure phenomenon: Antenna azimuth fixed**

Feature 1: Without electricity, antenna could be more easily rotated, no abnormal noise; Feature 2: It has overload alarm; Feature 3: The electromagnetic brake act correctly; Feature 4.1: Motor connecting wire normal; Feature 4.2: The resistance value of the 1, 2, 3 line and ground resistance on XS03 is infinite; Feature 4.3: The mutual resistance of 5, 6, 7 line on XS03 is not basic equal.

If there is an input failure characteristics into the dual-mode weather radar intelligent fault diagnosis system, the diagnostic conclusion of these characteristics in wavelet BP is 0, the fault diagnostic conclusion in improved CBR is: The drive is damage, similar distance function  $d_{12} = 0.0012$ . The fault diagnostic conclusion in the improved DS's dual-mode is: the drive is damage, the credibility is 95%, exclusion program; solution is to replace the azimuth drive, and process according to the requirements of components replacement adjustment parameters database after components replacement.

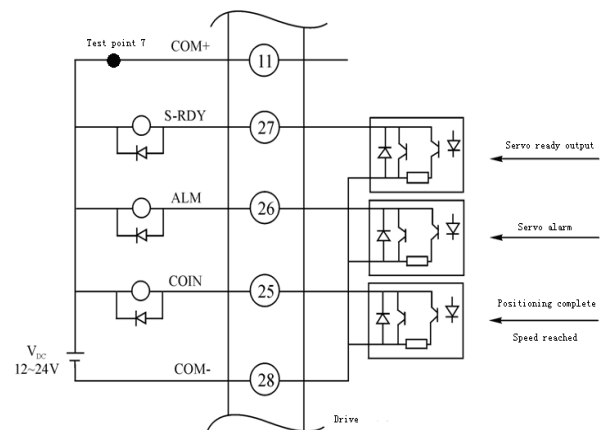


Figure 6. Feedback signal connecting circuit of the drive



TABLE I. THE CASE OF A FAILURE OF THE DRIVE

Fault case Number A	Fault phenomenon on B	Feature Description C	Troubleshooting solutions S	Retrieval information R
TQ1	Antenna pitching fixed	1 Without electricity, manual antenna can turn normal, and presence of abnormal noise; 2 Whether for a long time more than the rated load and torque run, overload alarm; 3 Whether electromagnetic brake malfunction; 4 Whether motor fault; 4.1 Check motor connecting line is normal or not; 4.2 Whether the resistance value of the 1,2,3 line and ground resistance on XS03 is infinite; 4.3 Whether the mutual resistance of 4,5,6 line on XS03 is the basic equal;	1 Eliminate antenna running met blocking or drive mechanism to be blocked; 2 Changing the pitch drive; 3 Remove electromagnetic brake in wrong action; 4 Repair or replace motor 4.1 Reconnect; 4.2 Motor on short circuit, the motor burnout, replace the motor; 4.3 Brush, collecting ring bad contact; Clean or replace carbon brush and collecting ring;	Has been used five times

TABLE II. CINRAD/CC RADAR SERVO SYSTEM DRIVE REPLACEMENT PARAMETER SETTINGS

Parameter code	Function	The set up in 3830	Setting Description
PRA 02	Control mode selection	This system is set to 2	Drive work in position control mode.
PRA 42	Instruction pulse input mode	This system is set to 3	Input command signal for "direction" and "pulse" way, changing input "direction" (high/low level) signal can change the motor rotation direction, changing the input "pulse" the signal frequency can change the speed of the motor speed.
PRA 4B	Instruction pulse frequency division denominator	This system is set to 500	Divide the frequency of the signal input "pulse": $1 / (10000/500)$ , this parameter can be fine-tuning; when the radar antenna speed partial small (or Angle positioning "less than"), make the parameter a small adjustment; When radar antenna speed partial big (or Angle positioning "overshoot"), make the parameter a big adjustment.
PRA 5E	The first torque limit	This system is set to 12	We set the maximum motor output torque is 120% of the rated torque to prevent winter power the radar antenna instantaneous starting torque overload shutdown phenomenon.

TABLE III. CINRAD/CC RADAR SERVO SYSTEM DRIVE REPLACEMENT PARAMETER SETTINGS

Parameter code	Function	The set up in 3830	Setting Description
T1. 1	Antenna Angle calibration	The system uses solar tracking	(a) Using the control software to make the antenna beam aim at the sun gradually (the receiver output noise maximum just at the sun); (b) Record the azimuth and elevation angle indicated by servo system as the noise level reach the maximum, and compare the angle with the sun actual angle; (c) Correct the antenna directive angle according to the comparison result.
T1. 2	Receiver noise factor test	The system is set to "noise coefficient calibration"	In the noise power on and off two kinds of state, the signal processor continuously samples receiver output noise level, shows the noise coefficient test value after calculation in the terminal display. The noise coefficient is generally $2 \sim 3$ db.
T1. 3	The systems phase noise test	Doppler processing of the output I, Q signals	Make the antenna at a fixture, receive the echo signal. Doppler process the I and Q signals outputted by linear channel, give the phase $\varphi$ , then find out phase noise $\sigma_{\varphi}$ , Phase noise should be less than $0.3^{\circ}$ .

The fault diagnosis conclusion could be regarded as a new phenomenon of the failure sent to the input of the system and then have a chip-level diagnosis of the drive failure. Collect eight test points voltage of the drive firstly, and then set the eight voltage values into the dual-mode weather radar intelligent fault diagnosis system as the new fault feature for secondary diagnosis. For example, the seven test point voltage = 0V, the voltage of the other test points is normal, then the diagnostic conclusion of these characteristics in wavelet BP is 10: the feedback signal connection circuit of drive power is abnormal; the fault diagnostic conclusion in CBR is 0; the fault diagnostic conclusion in the improved DS's dual-mode fusion is that the feedback signal connection circuit of drive power is abnormal, the credibility is 95%, exclusion program; remedy is to check the DC power supply of the feedback signal connection circuit or the circuit is partly open.

#### IV. CONCLUSION AND FUTURE WORK

The dual-mode weather radar intelligent fault diagnosis system is a practical intelligent fault diagnosis system. It is result of combined with theory and practice. The wavelet BP method is particularly suitable in the chip-level fault diagnosis, improved CBR method is particularly suitable for device-level fault diagnosis, and the plate level fault diagnosis method can use both. The fusion of the two kinds of fault diagnosis model increased the "adaptability" of radar circuit fault diagnosis system, which is effective for the radar fault diagnosis of device level, plate level, chip level, and give correspond the fault repair solutions at the same time, the improved d-s fusion algorithm of dual mode actually improved the credibility of the radar fault diagnosis, increased the radar equipment analytical, the radar fault components positioning accuracy can be improved. The parameter settings database after components replacement and the feedback processing functions of fault diagnosis conclusion are newly added in the system, these features are fully demonstrated the practicality and effectiveness of the system.

But, this system is relatively complex, calculations require long times. On the basis of the guarantee the accuracy of fault diagnosis, still need to improve the efficiency; this is the future research direction in this paper.

#### ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (Grant Nos. 61271334).

#### REFERENCES

- [1] S. Khomfoi and M. Tolbert, Fault Diagnosis System for a Multilevel Inverter Using a Neural Network [J]. IEEE Industrial Electronics Conference, June. 2005, pp. 1458-1460.
- [2] X. Zhihong and C. Le, Study for radar test and repair support system [J]. Modem Radar, vol. 28. 2006, pp. 122-125.
- [3] C. Chien, Design and Implementation of a Case based Reasoning System for Marketing Plans [J]. Expert Systems with Applications, 2005, 28, pp. 43-53.
- [4] Cheng X. F. and Ma Y., A Single-Channel Mixed Signal BSS New Method of Does Not Use the Prior-Knowledge [J]. ACTA ELECTRONICA SINICA, 2011, pp. 2317-2321.
- [5] Varaprasad B., Patnaik L. M., Jamadagni H. S. A new ATPG technique (Multi Detect) for testing of analog macros in mixed-signal circuits [J]. Computer – Aided Design of Integrated Circuits and Systems, vol. 23. 2004, pp. 273- 287.
- [6] Cheng H. Z. and Chai X. F., Failure analysis and maintenance of CINRAD/CC radar servo motor drive [J]. Meteorological hydrological Marine instrument, Feb. 2011, pp.9-11.
- [7] Cheng X.F., Ma Y., and Tao Y. W., Three-stage heart sound identification technology based on data fusion [J]. Chinese Journal of Scientific Instrument, August. 2010, pp. 1712-1720.
- [8] Zhu D. and Yu S. I., Based on d-s evidence theory data fusion algorithm and its application in the circuit fault diagnosis [J]. Chinese Journal of Electronics, vol. 30. 2002, pp. 221- 224.
- [9] Cheng X. F., Ma Y., and Liu C., Research on heart sound identification technology [J]. Sci China Inf Sci, vol. 55. 2012, pp. 281-292.
- [10] Cheng X. F. and Tao Y. W., Heart Sound Recognition--A Prospective Candidate for Biometric Identification [J], Advanced Materials Research, vol. 225. 2011, pp. 433-436.

## Automated Analysis of CT Slices for Detection of Ideal Midline from Brain CT Scans

Xuguang Qi, Ashwin Belle, Sharad Shandilya,  
Kayvan Najarian

Department of Computer Science, School of Engineering,  
Virginia Commonwealth University,  
Richmond, VA  
E-mail: qix2@vcu.edu, bellea@vcu.edu,  
shandilyas2@vcu.edu, knajarian@vcu.edu

Charles Cockrell, Yang Tang

Department of Radiology, School of Medicine,  
Virginia Commonwealth University,  
Richmond, VA  
E-mail: chcockrell@vcu.edu, ytang2@vcu.edu

Rosalyn S. Hobson Hargraves

Department of Electrical and Computer Engineering,  
Virginia Commonwealth University,  
Richmond, VA  
E-mail: rhobson@vcu.edu

Kevin R. Ward

Department of Emergency Medicine,  
Michigan Center for Integrative Research in Critical Care  
University of Michigan,  
Ann Arbor, MI  
E-mail: keward@med.umich.edu

**Abstract**—Midline shift detection with high accuracy is crucial in quantitatively analyzing the severity of a brain injury in clinical environments. Accuracy of the estimated ideal midline (IML) significantly affects the accuracy of the computed midline shift. In this work, a two-step process, which consists of computed tomography (CT) Slice Selection Algorithm (SSA) and IML detection, is proposed to automatically estimate the IML in brain CT images. SSA is designed for automatic slice selection. Skull fracture level and intracranial area are used as vital features in the selection. Using skull symmetry and anatomical features, IML detection accurately estimates the position and rotation angle of the IML before calibrating. Experimental results of the multi-stage algorithm were assessed on 1762 CT slices of 40 patients. The accuracy of the proposed system is 91.6%, which makes it viable for use under clinical settings.

**Keywords**—ideal midline; IML; midline shift; MLS; CT slice; SSA; mid-sagittal plane

### I. INTRODUCTION

In the United States alone, nearly 1.7 million cases of Traumatic Brain Injury (TBI) are recorded annually [1]. Midline Shift (MLS), which is the shift in the brain's midline, is a common aftermath due to the head injury. It is an important index for clinicians to assess the severity of TBI. MLS greater than 5 mm can lead to subfalcine herniation and possibly death [2]. Ideal midline is the symmetric midline of the brain without injury or illness. Estimating Ideal Midline (IML) [2] is a vital step in MLS calculation.

Skull symmetry and anatomic features have been widely used to detect the IML in last two decades [3, 4, 5]. Ruppert

et al. extracted the mid-sagittal plane (MSP) based on bilateral symmetry maximization [6]. Chen et al. used a combination of bone symmetry and anatomical features in CT images for detection of IML [7]. This method works effectively and accurately on a single CT slice, but does not consider the connection among CT slices. Furthermore, all the methods above mentioned [3, 4, 5, 6, 7] cannot automatically select the proper slices before analysis. In practice, dozens of CT images can be acquired in one patient's brain scan. It is crucial to choose a few appropriate slices that contain clear anatomical features and limited noise, to be used for MLS quantification. Prior to this work, no automated method to perform this task existed.

In this work, we propose a two-step algorithm for automated detection of the IML. As the first step, a CT Slice Selection Algorithm (SSA) is proposed to select appropriate slices from a large number of raw CT images. SSA proposed in this work realizes the real automated slice selection which is the initial step for automated IML detection. We did not find any existing automated method to perform this task. The second step focuses on the IML detection through anatomical features extracted from the selected slices and the consideration of the connection among CT slices. A database of 1762 CT slices of 40 patients with TBI cases were used for this study. As shown later, the proposed algorithm yields highly desirable accuracy and efficiency when tested against this dataset.

The rest of this paper is organized as follows: the methodology is introduced in Section II. The results are presented and discussed in Section III. The work is summarized in Section IV.

II. METHODOLOGY

2.1 CT Slice Selection Algorithm (SSA)

Among the dozens of raw CT images acquired from a single patient, only a few are useful for the physician in a diagnosis that includes midline estimation. Some images taken from the lower section of the head contain too much interference/noise from other organs, such as the eye and nose in the top image in Figure 1-a. Some images capture a small intracranial area because the scan position is too close to the calvaria, as seen in Figure 1-b. From the viewpoint of anatomical features, the ideal CT slices usually contain integrated skull bone and larger intracranial area, such as Figure 1-c. Therefore, CT slice selection should ideally be based on the above mentioned features.

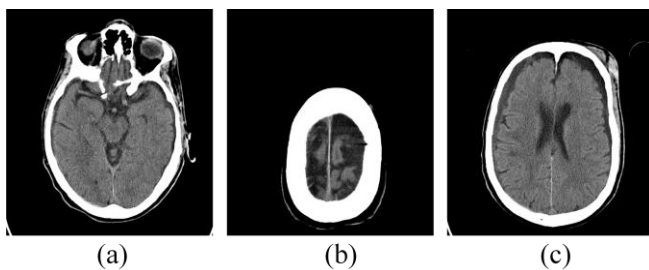


Figure 1. Three raw CT slices from one patient's head CT scan.

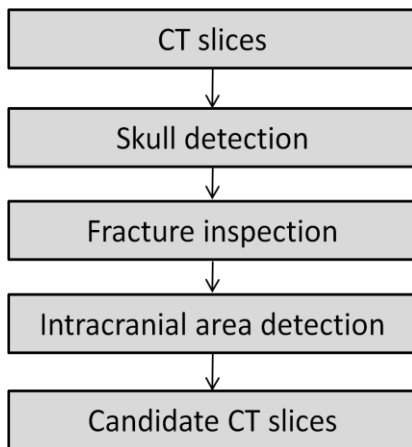


Figure 2. Flowchart of CT Slice Selection Algorithm (SSA).

The CT Slice Selection Algorithm (SSA) was designed to effectively select a few appropriate CT slices from a large number of images. As the flowchart shows in Figure 2, this algorithm analyzes every slice by examining multiple anatomic features.

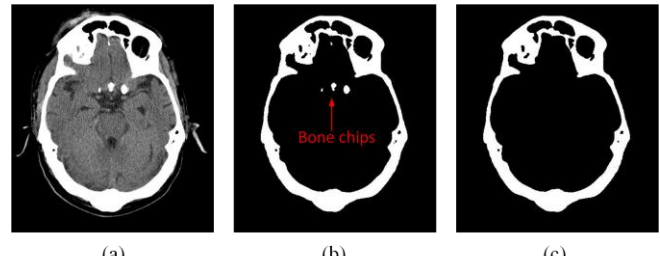


Figure 3. Skull detection process. (a) Raw CT slice (b) Detected bones B with little bone chips (c) Detected skull.

As the first step in SSA algorithm, skull detection is implemented on every raw CT slice. Using a threshold method, potential bone pixels can be extracted from the raw image. In this study, based on experimentation, the value for the threshold is set to 250 (out of 255), which lies within the common range for bone intensity within CT images. Using the connected component algorithm (CCA) [8], the discrete bone chips can be removed (Figure 3-b). Bone pixels form a certain number of connected regions. We choose the one containing the largest number of elements as the candidate skull (Figure 3-c).

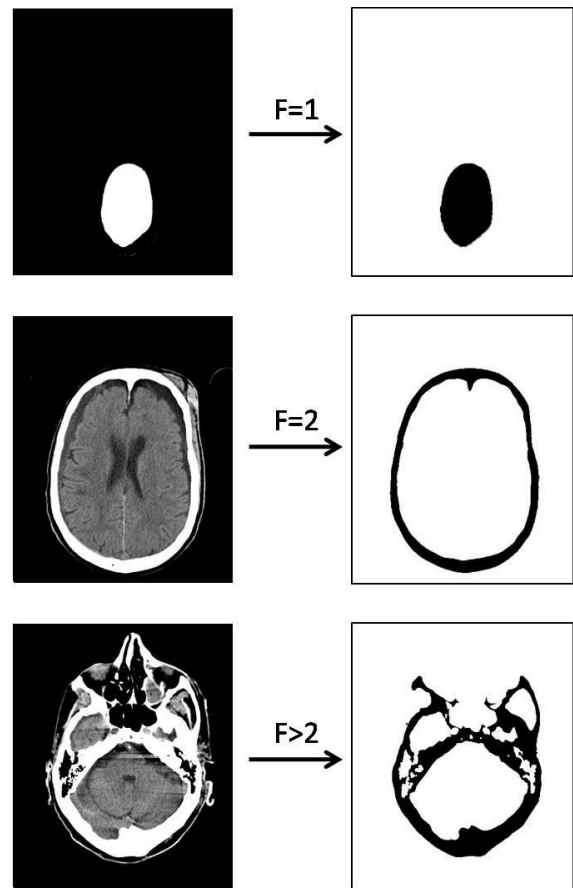


Figure 4. Skull fracture inspection. The left three images are the raw CT images while the right three images show the detected skull.

The second step in the SSA algorithm is skull fracture inspection to remove slices with either skull fractures or partial skull. Such "non-integrated" skull affects IML identification since symmetry value calculation, through the exhaustive symmetric position search, is sensitive to skull contour. We define a new measure, called skull fracture level  $F$ , to estimate the integrity of the skull. Skull fracture level  $F$  is defined by the number of isolated regions separated by the skull. To prevent any small holes in the skull from affecting the calculation, a minimum threshold (of 200 pixels in this work) is set for the area of those isolated regions. If the computed skull fracture level  $F$  is equal to 2, it implies that the skull is integrated and ideal for the following steps of detection. An example is the middle slice in Figure 4. If the skull fracture level is not equal to 2, the image cannot be used in detection of IML due to either an inappropriate scan position or a serious fracture in the skull. Examples are the top and bottom slices in Figure 4. After skull fracture inspection, all images with  $F \neq 2$  are removed from the slice subset.

Based on clinical experience, CT slices with larger intracranial area generally contain more information for IML detection. Hence, in the third step of the SSA algorithm, the intracranial area is calculated and sorted for all remaining slices. After skull fracture inspection, every CT image should contain only two dark regions, which are separated by the detected skull. An example is the middle-right image in Figure 4. In order to calculate intracranial area, the intracranial region has to be distinguished from the region outside of the skull. This can be achieved using the coordinate of the skull's mass-center. The image moment  $m_{pq}$  of the order  $p+q$  can be defined as below,

$$m_{pq} = \sum_{j=1}^n \sum_{i=1}^m i^p \cdot j^q \cdot \Gamma_{ij}, \quad (p=0,1; q=0,1) \quad (1)$$

where  $\Gamma_{ij}$  with the value of either 1 or 0 represents the intensity of the element at the  $i$ th row and  $j$ th column in the detected skull matrix  $I$ . The coordinate of the mass center  $(x, y)$  of the identified skull can be obtained by

$$\begin{cases} x = \frac{m_{10}}{m_{00}} \\ y = \frac{m_{01}}{m_{00}} \end{cases} \quad (2)$$

Thus the region containing the coordinate of the skull mass center  $(x, y)$  is the intracranial region. The intracranial area of all remaining slices is calculated and sorted in descending order. The first  $\lambda$  slices with larger intracranial areas are selected as the candidate slices for IML detection. This number of  $\lambda$  is a variable that depends on the number of slices for one patient or on physician's requirement. In this work, we choose  $\lambda=3$  candidate slices for the detection that follows.

## 2.2 Ideal midline detection

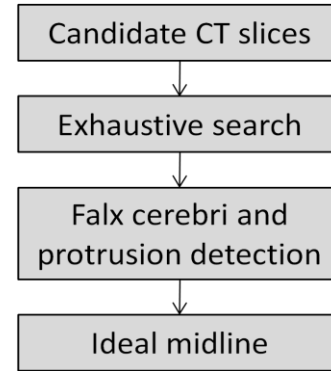


Figure 5. Flow chart of ideal midline detection

After slice selection is performed using SSA algorithm, all candidate slices are appropriate for IML detection. Ideal midline detection consists of an exhaustive search as well as falx cerebri and protrusion detection, as shown in Figure 5.

To find the approximate IML, we use the exhaustive symmetric position search algorithm, which was developed in a prior work by our research group [9]. The row symmetry is defined as the difference in distance between each side of the skull edge and the current approximate midline. The CT image is rotated around the mass center of the skull, which is calculated by (2). The symmetry cost  $S_\theta$  of the image at each rotation angle  $\theta$  is calculated as the sum of all row symmetry in the resulting image as follows.

$$S_\theta = \sum_{i=1}^m |l_i - r_i| \quad (3)$$

where  $m$  is the number of rows in the image with the rotation angle  $\theta$  (in this study ;  $-45 \leq \theta < +45$  ) and measures  $l_i$  and  $r_i$  are the distance between the edge of the skull on the left or right side, respectively, and the current approximate midline at the  $i$ th row. More details can be found in [9]. Finally, the rotation angle  $\theta$  with the minimum symmetry cost  $S_\theta$  determines the rotation direction of the midline of the brain on each particular CT slice.

$$\theta_k = \underset{\theta_{kj}}{\operatorname{arg\,min}} [ S_{\theta_{k1}}, S_{\theta_{k2}}, \dots, S_{\theta_{k\lambda}} ], \quad (1 \leq k \leq \lambda) \quad (4)$$

where  $\theta_k$  is the rotation angle of the midline on the  $k$ th slice and  $S_{\theta_{kj}}$  is the symmetry cost of the  $k$ th slice at the rotation angle  $\theta_{kj}$ . Then, the  $k$ th candidate slice is calibrated to the vertical direction by rotating the skull by  $-\theta_k$  angle.

In addition, the accuracy of approximate IML can be improved by utilizing other features of the skull and the brain. Thus, following the approximate IML estimation using exhaustive search, brain anatomical features, such as the position of the falx cerebri and protrusion of skull bone, are used to refine the position of the IML. Here, we use the algorithm proposed in our previous work [9, 10]. The falx cerebri is a strong arched fold of dura mater that descends vertically in the longitudinal fissure between the left and right cerebral hemispheres. In this work, we use edge

detection method and Hough transform to execute the detection. Additionally, a bone protrusion located in the anterior section of the skull is used in the refinement. To locate the lowest point of the protrusion curve, the derivative of the curve is calculated in a limited neighborhood area (10-15 pixels in this work). The local minimum point  $a$  is determined by

$$x_a = \arg \max_x [\mathfrak{R}(x+w) + \mathfrak{R}(x-w) - 2 \cdot \mathfrak{R}(x)] \quad (5)$$

where the function  $\mathfrak{R}(x)$  is the extracted curve of the interior bone edge and  $w$  is the neighborhood width. Using the detected falx cerebri and the bone protrusion, we can obtain the refined rotation angle  $\varphi_k$  of the midline in the  $k$ th slice. Therefore, the rotation angle for the  $k$ th slice should be  $\phi_k = \theta_k + \varphi_k$ .

It is worth notice that those obtained rotation angles may be different for different slices. However, since the patient usually keeps the same gesture during CT scan, the slices in one CT scan should have the same rotation angle. In order to fully consider the connection among slices, we use a global rotation angle  $\phi$  given by (6).

$$\phi = \text{median}[\phi_1, \phi_2, \dots, \phi_\lambda], \quad (6)$$

As shown in (6), it is the median value of the rotation angles of all  $\lambda$  slices after CT slice selection.

Lastly, each slice is calibrated to the vertical direction. Therefore, during the IML detection process, the IML is centered by the mass center of the skull and rotated by an angle of  $-\phi$  from the original position in the slice.

### III. RESULTS AND DISCUSSION

The SSA algorithm is primarily based on the anatomical characteristics of the skull and closely simulates the process of manual CT slice selection and decision making in IML by physicians. In our database, all CT slices selected by the SSA algorithm have been found to be acceptable for IML detection by physician's approval. Result of the IML detection is displayed in Figure 6. It can be noticed that the detected IML is accurately located in the middle of the skull. Additional, the direction of the skull is calibrated by moving the IML to vertical direction.

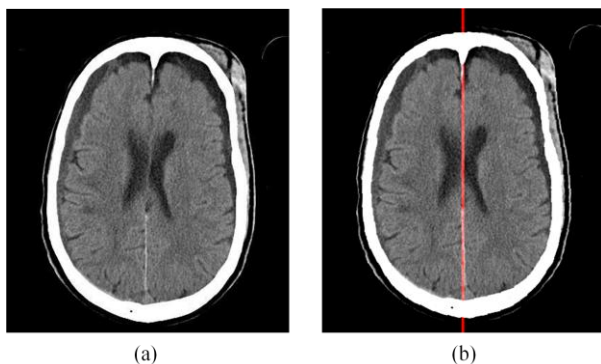


Figure 6. The ideal midline detection on a candidate slice selected by the SSA algorithm. (a) Original CT slice, (b) ideal midline is detected and the skull direction is calibrated.

This database contains original 1762 axial CT scan slices acquired across 40 patients with cases of both mild and severe Traumatic Brain Injuries (TBI). Collaborating physicians manually labeled the IML. With a strict definition of accuracy, which is an allowed error of three pixels in the horizontal direction and 2 degrees of the rotation angle, accuracy of our algorithm is 91.6% and the mean value of the error  $\delta$  in horizontal direction is only 2.4 pixels as shown in TABLE I.

We have evaluated our method using a previous work designed by some authors of this paper [9] as baselines. With the same criteria, the accuracy of this work is 6% higher and the mean value of horizontal error is 17% smaller than the method in [9].

TABLE I. COMPARISON ON THE ACCURACY OF IML ESTIMATION

Method	Our method	Method in [9]
Number of patients	40	40
Accuracy	91.6%	85.7%
Mean value of error $\delta$	2.4	2.9

The improvement on accuracy shows that the implementation of global rotation angle after IML detection greatly enhances the accuracy on skull rotation calibration by fully considering the connection among slices.

### IV. SUMMARY AND FUTURE WORK

In this work, we developed a system with the combination of the SSA algorithm and the ideal midline (IML) detection process to identify the IML using CT scans of patients with head injuries. The proposed SSA algorithm is used to closely simulate the process of manual selection of CT slice by physicians. Fully considering the symmetry of the skull and anatomical features, IML detection algorithm with the adjustment of global rotation can accurately identify the IML on the candidate CT slices selected by the SSA algorithm. The obtained results show high accuracy (91.6%) and a potential for the system to be implemented in clinical settings. In the future, more work can focus on the actual midline detection which is the shifted midline after brain injury or illness. Then IML detection can be used the midline shift estimation which is one of the key index in TBI assessment in clinical practice.

### REFERENCES

- [1] M. Faul, L. Xu, M. M. Wald, and V. G. Coronado, "Traumatic Brain Injury in the United States: Emergency Department Visits, Hospitalizations and Deaths, 2002-2006," Centers for Disease Control and Prevention, National Center for Injury Prevention and Control, Atlanta, GA, USA, 2010.
- [2] J. S. Broder, "Head computed tomography interpretation in trauma: a primer," *The Psychiatric Clinics of North America*, vol. 33, 2010, pp. 821-854.
- [3] Q. Hu and W. L. Nowinski, "A rapid algorithm for robust and automatic extraction of the midsagittal plane of the human

- cerebrum from neuroimages based on local symmetry and outlier removal,” *NeuroImage*, vol. 20, no. 4, 2003, pp. 2153–2165.
- [4] G. Ruppert, L. Teverovskiy, C. Yu, A Falcao, and Y. Liu. “A new symmetry-based method for mid-sagittal plane extraction in neuroimages,” *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2011, pp. 285-288.
- [5] C.C. Liao, F. Xiao; J. Wong, and I. Chiang, “A simple genetic algorithm for tracing the deformed midline on a single slice of brain CT using quadratic Bezier curves,” *Sixth IEEE International Conference on Data Mining Workshops*, 2006, pp. 463-467.
- [6] R. Guillemaud, P. Marais, A. Zisserman, B. McDonald, T. J. Crow, and M. Brady, “A three dimensional mid sagittal plane for brain asymmetry measurement,” *Schizophrenia Research*, vol. 18, no. 2, 1996, pp. 183–184.
- [7] W. Chen, R. Smith, S. Y. Ji, K. Ward, and K. Najarian, “Automated ventricular systems segmentation in brain CT images by combining low-level segmentation and high-level template matching,” *BMC Medical Informatics and Decision Making*, vol. 9, 2009, pp. S4.
- [8] R. M. Haralick and L. G. Shapiro, “*Computer and Robot Vision*,” Volume I, Addison-Wesley, 1992, pp. 28-48.
- [9] W. Chen, R. Smith, S.Y. Ji, and K. Najarian, “Automated segmentation of lateral ventricles in brain CT images” *Bioinformatics and Biomedicine Workshops, IEEE International Conference*, 2008, pp. 48-55.
- [10] W. Chen, A. Belle, C. Cockrell, K. Ward, and K. Najarian, “Automated Midline Shift and Intracranial Pressure Estimation based on Brain CT Images,” *J. Vis. Exp.*, vol. 74, 2013, pp. 3791-3871

# Actual Brain Midline Detection using Level Set Segmentation and Window Selection

Xuguang Qi, Ashwin Belle, Sharad Shandilya,

Kayvan Najarian

Department of Computer Science, School of Engineering,  
Virginia Commonwealth University,  
Richmond, VA

E-mail: qix2@vcu.edu, bellea@vcu.edu,  
shandilyas2@vcu.edu, knajarian@vcu.edu

Charles Cockrell, Yang Tang

Department of Radiology, School of Medicine,  
Virginia Commonwealth University,  
Richmond, VA

E-mail: chcockrell@vcu.edu, ytang2@vcu.edu

Rosalyn S. Hobson Hargraves

Department of Electrical and Computer Engineering,  
Virginia Commonwealth University,  
Richmond, VA

E-mail: rhobson@vcu.edu

Kevin R. Ward

Department of Emergency Medicine,  
Michigan Center for Integrative Research in Critical Care  
University of Michigan,  
Ann Arbor, MI

E-mail: keward@med.umich.edu

**Abstract**—Detection of actual brain midline is essential for accurate estimation of midline shift due to traumatic brain injury. An effective method to estimate the actual midline is to use the positions of the identified ventricles. In this work, a level set algorithm combined with Ventricle Window Selection Algorithm and Ventricle Identification Algorithm is proposed to detect the ventricular system in computed tomography (CT) images. The system automatically selects appropriate slices from numerous raw slices and confines the focus to the region of interest prior to segmentation. The variational level set method performs ventricle segmentation without any requirement of re-initialization or intensity-homogeneity of CT images. Combined with ventricle identification, the level set segmentation successfully extracts ventricle contours and estimates actual midline. Experimental results assessed on 391 CT slices of 40 patients support that the proposed system is accurate (90%) and useful for clinical practice.

**Keywords**—actual midline; ventricle; CT slice; window selection; level set segmentation

## I. INTRODUCTION

Automated actual midline detection precedes accurate midline shift estimation, which is a key clinical index to assess the severity of Traumatic Brain Injury. Midline Shift is known to be highly correlated with elevated Intracranial Pressure levels (ICP) [1]. Computer-aided estimation of the actual brain midline using medical images has attracted a great deal of attention [2-4]. Using quadratic Bezier curve [5] in CT image segmentation, the model introduced by Liao et al., is simple and effective, but it suffers from low accuracy when CT slices show spontaneous intracranial hemorrhage [5]. Chen et al. proposed a method based on Gaussian Mixture Model and template matching to detect the ventricle system on brain CT image. This method has high-accuracy

but is time-consuming during the computation of brain midline estimation [6, 7]. Variational level set method overcomes the intrinsic limitation of re-initialization and high sensitivity on intensity-homogeneity [8]. It has been widely used in medical image processing [9, 10].

In this work, a system based on variational level set segmentation combined with Ventricle Window Selection Algorithm (VWSA) and Ventricle Identification Algorithm (VIA) has been proposed, for ventricle detection and midline estimation. The system achieves higher accuracy with less time consumption on midline estimation as compared to other proposed methods.

The rest of this paper is organized as follows: the methodology is introduced in Section II, results are presented and discussed in Section III, and the conclusion of this work is given in Section IV.

## II. METHODOLOGY

A flowchart of the four-step algorithm for actual midline estimation is shown in Figure 1.

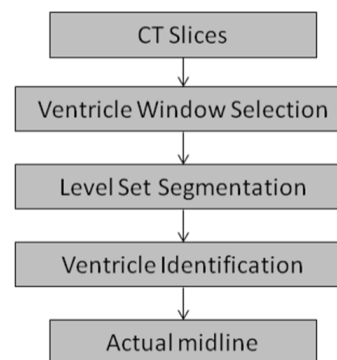


Figure 1. Flowchart of actual midline estimation



First, Ventricle Window Selection Algorithm is run with all raw CT slices to select appropriate ones. Second, utilization of the level set segmentation effectively extracts candidate contours. Subsequently, the right and left lateral ventricles are determined in ventricle identification. Lastly, the brain midline is estimated using the position of the ventricles.

### 2.1 Ventricle Window Selection Algorithm (VWSA)

In clinical practice, even though dozens of CT images can be acquired from the head CT scans of each patient, only a few of these slices contain clear ventricle information owing to appropriate scan position. Hence, Ventricle Window Selection Algorithm (VWSA) is designed to narrow down the selection of relevant CT slices and confine the region of focus.

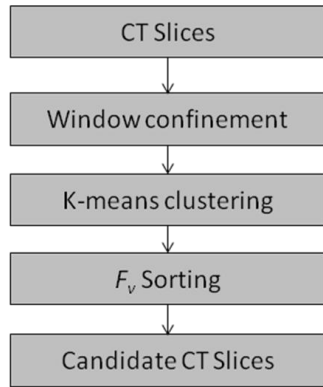


Figure 2. Flowchart of Ventricle Window Selection Algorithm (VWSA)

Raw CT slices may contain too much noise and other irrelevant tissue information, which affects the visibility of ventricles. Moreover, since the speed of contour evolution in subsequent level set segmentation largely depends on size of image, narrowing down the region of interest can accelerate the algorithms throughput. Therefore, as the flowchart shows in Figure 2, the first step of VWSA is confining the focus of operation to a window within a fixed area in the upper center of the brain. We set a window with size of 140\*200 pixels. Because of anatomical characteristics and observations during the algorithm development process, this region contains ventricles and is hardly affected by skull or by the presence of other tissues. Using K-means clustering on the gray scale of every pixel, the pixels belonging to the ventricle class are distinguished from all other pixels within the window. In order to evaluate the visibility of ventricles in slices, a new measure, called Ventricle Fidelity Measure ( $F_v$ ), has been introduced in this study.  $F_v$  is a proportion defined by the number of pixels labeled as the ventricle class over the total number of pixels in the window. The larger the Ventricle Fidelity Measure is, the more visible the ventricle should be within the image. Three slices with the largest  $F_v$  values are selected as the candidate slices for the following detection.

### 2.2 Ventricle segmentation based on level set method

Region-based geometric active contour model implemented by variation level set method [9] is used in ventricle contour detection.

Here, we consider a two-dimensional gray level image  $I: \Omega \rightarrow \mathbb{R}$ , where  $\Omega \subset \mathbb{R}^2$  is the image domain and  $I$  is the image intensity [9].  $C$  represents closed contour in the 2-D image domain ( $C \subset \Omega$ ) and segments the image into two regions:  $\Omega_1 = \text{outside}(C)$  and  $\Omega_2 = \text{inside}(C)$ . Contour  $C$  can be expressed by the zero level set of a Lipschitz function  $\varphi: \Omega \rightarrow \mathbb{R}$ , which is a level set function. The energy function  $F(\varphi, f_1, f_2)$ , which is subject to minimization is defined as;

$$F(\varphi, f_1(x), f_2(x)) = \varepsilon_f(\varphi, f_1(x), f_2(x)) + \nu |C(\varphi)| + \mu P(\varphi) \quad (1)$$

where the fitting energy  $\varepsilon_f$  minimizes the gray value variance in separated phases, the contour length  $|C(\varphi)|$  smooths the curve, and the internal energy  $\mu P(\varphi)$  helps stabilize curve evolution.  $f_1(x)$  and  $f_2(x)$  reflect the intensity in the region with center  $x$ . With a fixed  $f_1$  and  $f_2$ , the energy function  $F(\varphi, f_1, f_2)$  in (1) can be minimized by solving the gradient flow equation as follows:

$$\begin{aligned} \frac{\partial \varphi}{\partial t} = & -\delta_\varepsilon(\varphi) \left( \lambda_1 \int K_\sigma(x-y) |I(x) - f_1(y)|^2 dy \right. \\ & \left. - \lambda_2 \int K_\sigma(x-y) |I(x) - f_2(y)|^2 dy \right) \\ & + \nu \delta_\varepsilon(\varphi) \text{div} \left( \frac{\nabla \varphi}{|\nabla \varphi|} \right) + \mu (\nabla^2 \varphi - \text{div} \left( \frac{\nabla \varphi}{|\nabla \varphi|} \right)) \end{aligned} \quad (2)$$

with

- $\delta_\varepsilon(\varphi) \left( \lambda_1 \int K_\sigma(x-y) |I(x) - f_1(y)|^2 dy - \lambda_2 \int K_\sigma(x-y) |I(x) - f_2(y)|^2 dy \right)$  being the data fitting term;
- $\nu \delta_\varepsilon(\varphi) \text{div} \left( \frac{\nabla \varphi}{|\nabla \varphi|} \right)$  represents the arc length term; and
- $\mu (\nabla^2 \varphi - \text{div} \left( \frac{\nabla \varphi}{|\nabla \varphi|} \right))$  is the level set regularization term.

The segmentation used in this work is based on the level set evolution equation (2). Further details can be found in [9]. Level set regularization term effectively eliminates time consumption due to re-initialization, which is one of the main purposes of using this method in segmentation.

Using the variational level set method in segmentation, candidate contours are extracted from the candidate CT slice. Generally, two ventricle contours is expected to be found corresponding to the right and left lateral ventricles. However, due to the complexity of CT images, in practice, more (or less) than two contours are possibly found. Therefore, we need the following ventricle identification process to identify the actual ventricles.

### 2.3 Ventricle identification and actual midline estimation

As the flowchart shows in Figure 3, the first step of the VIA is to identify the right and left lateral ventricle contours from a set of contours obtained from level set segmentation. We remove the unclosed contours using the Connected Component Algorithm (CCA) [11]. Assume that n-candidate

contours are retained in the VWSA window. If  $n$  equals 0 or 1, it means that either no ventricle or a mangled ventricle image is displayed in the CT slice. The algorithm automatically skips this slice and moves back to VWSA to compute on the next candidate slice. In the case where two or more candidate contours are found, the two contours with the largest length are chosen as the right and left lateral ventricles.

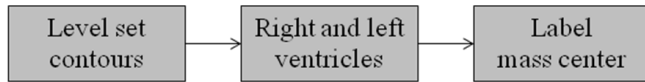


Figure 3. Flowchart of ventricle identification

In order to determine the mass center of a contour, the moment  $m_{pq}$  [12] with the order of  $p+q$  of the digital image  $F$  is defined as below,

$$m_{pq} = \sum_{j=1}^n \sum_{i=1}^m i^p \cdot j^q \cdot \Gamma_{ij}, \quad (p=0,1; q=0,1) \quad (3)$$

where  $\Gamma_{ij}$  is the image intensity of the element at the  $i$ th row and the  $j$ th column in the matrix  $F$ . This means that  $\Gamma_{ij}$  equals 1 when the element is on the contour, else equals 0. Then, the coordinate of the mass center  $(x, y)$  [12] of the object contour is given by

$$\begin{cases} x = \frac{m_{10}}{m_{00}} \\ y = \frac{m_{01}}{m_{00}} \end{cases} \quad (4)$$

Finally, the step of estimating the actual midline can be performed. Assuming that the right and left lateral ventricles have the mass center positions of  $(x_1, y_1)$  and  $(x_2, y_2)$ , respectively, the actual midline is believed to be at the middle of the two ventricles and have the slope  $K$  as shown below;

$$K = -\tan \left( \frac{x_2 - x_1}{y_2 - y_1} \right) \quad (5)$$

### III. RESULTS AND DISCUSSION

#### 3.1 Data

The dataset in this study contains 391 axial CT scan slices acquired across 40 patients with cases of both mild and severe Traumatic Brain Injuries.

#### 3.2. Results of Ventricle Window Selection Algorithm

Ventricle Window Selection Algorithm (VWSA) is aimed at selecting the most appropriate slices and confining the window of focus. Figure 4 shows that three CT slices with the largest  $F_v$  values are selected as candidate slices from a patient's CT scans. As can be seen, the first candidate slice Figure 4-(a) with the largest  $F_v$ , shows ventricles more clearly than the other two. In our experiment, 30 out of 36 cases, in which ventricle boundaries were correctly extracted with segmentation, obtained their actual midline using the VWSA slice with the top  $F_v$ .

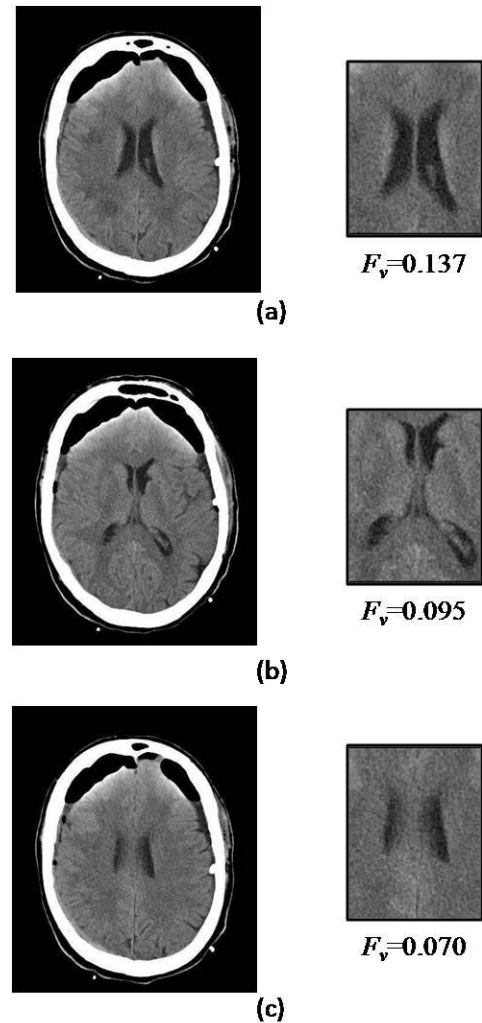


Figure 4. Results for Ventricle Window Selection Algorithm: three candidate slices are selected from one patient's CT images.

With the focus region confined from whole image to only the VWSA window, the time-cost associated with level set segmentation is greatly reduced. Additionally, the contour evolution becomes much easier due to the avoidance of the artifacts present outside the windowed region.

#### 3.3 Level Set Segmentation

The algorithm has been implemented in Matlab on a PC with Intel core I7 3.40 GHz processor with 8GB RAM. The initial level set function  $\phi$  assigns the value 2 for all the pixels within the VWSA window. The parameters of the level set function are set as follows,  $\lambda_1=1$ ,  $\lambda_2=1$ ,  $\sigma=3.0$ , time step  $\Delta t=0.1$ ,  $\mu=1$ , and  $\nu=0.001 \times 255 \times 255$ .

Figure 5 depicts the results of the level set method at various iterations. It demonstrates how the algorithm evolves from the initial stage 0 iteration to the final stable contour at iteration 200. The images from the left to right shows 0, 50, 100, and 200 iterations in level set evolution, respectively. The evolution becomes stable after 200 iterations. From the above results, it can be seen that the level set segmentation

successfully extracts the final contours that match the ventricles boundary well. It provides a good foundation for the subsequent ventricle identification and midline estimation.

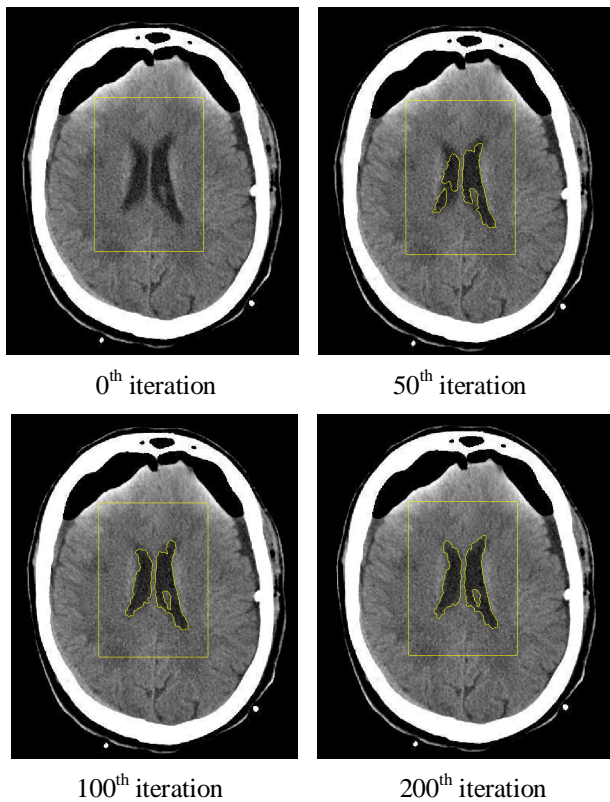


Figure 5. Contour evolution using the level set algorithm. From left to right images show different levels of iterations.

In this work, the average CPU time used on the level set evolution on one candidate CT slice is 129.2 seconds. In order to test the computational cost of the algorithm, we compare the experimental results of the level set segmentation and the Gaussian Mixture Model (GMM) method [7] on the same database under the same hardware configuration. It is found that the level set segmentation takes 17% lesser computational time as compared to the GMM method.

### 3.4 Ventricle identification and actual midline estimation in post-processing

The ventricle identification and midline estimation process is shown in Figure 6. Three candidate contours are extracted by level set segmentation (Figure 6-a). VIA selects the largest two to represent the left and right lateral ventricles. Then, their mass centers, calculated by (11), are labeled in Figure 6-b. According to the positions of the ventricle mass centers, the actual midline is estimated and shown in Figure 6-c.

The collaborating physicians manually labeled the actual midline for every patient in the database. With a strict

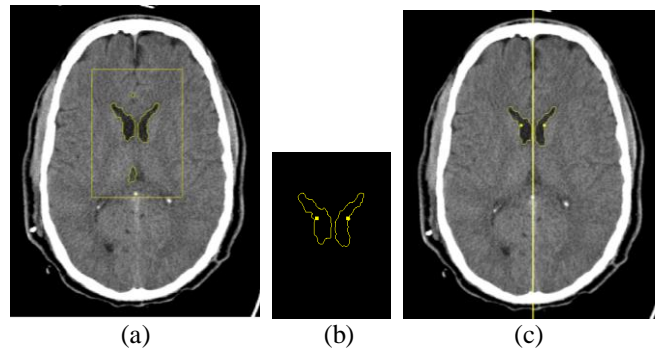


Figure 6. Ventricle identification and midline estimation. (a) The candidate contours extracted by level set segmentation, (b) ventricle contour identification and mass centers determination, (c) midline estimation.

definition of accuracy, which is an allowed error of three pixels in the horizontal direction and two degrees of the rotation angle, the accuracy of the midline estimation of our system is 90%, which is much higher than 87.5% using GMM method in [7].

## IV. CONCLUSION

In this work, an actual midline detection system based on the level set segmentation and window selection has been proposed. The Ventricle Window Selection Algorithm not only selects appropriate CT slices containing clear ventricle information but also confines the focus of operation. It greatly enhances the efficiency of the whole system by reducing processing time for the level set segmentation. The variational level set segmentation model, combined with the ventricle identification process, successfully extracts ventricle contours. Actual midline is estimated by the positions of the ventricle contours. With physician's validation, the results show a high accuracy of 90% for midline estimation. This makes the new system viable in clinical settings.

## REFERENCES

- [1] J. S. Broder, "Head computed tomography interpretation in trauma: a primer," *The Psychiatric Clinics of North America*, vol. 33, 2010, pp. 821–854.
- [2] S. Prima, S. Ourselin, and N. Ayache, "Computation of the Mid-Sagittal Plane in 3-D Brain Images," *IEEE Transactions on medical imaging*, vol. 21, no. 2, 2002, pp. 122-138.
- [3] Q. Hu and W. L. Nowinski, "A rapid algorithm for robust and automatic extraction of the midsagittal plane of the human cerebrum from neuroimages based on local symmetry and outlier removal", *NeuroImage*, vol. 20, no. 4, 2003, pp. 2154-2166.
- [4] A. V. Tuzikov, O. Colliot, and I. Bloch, "Brain symmetry plane computation in MR images using inertia axes and optimization", *Proceedings of the International Conference on Pattern Recognition on 1 (ICPR1)*, 2012, pp. 516-519.
- [5] C. C. Liao, I. J. Chiang, F. Xiao, and J. M. Wong, "Tracing the deformed midline on brain CT," *Biomedical Engineering Application, Basis and Communications*, vol. 18, December 2006, pp. 305-311.
- [6] W. Chen, R. Smith, S.Y. Ji, K. Ward, and K. Najarian, "Automated ventricular systems segmentation in brain CT images by combining low-level segmentation and high-level

- template matching,” *BMC Medical Informatics and Decision Making*, vol. 9, no. Suppl 1, 2009, pp. S4.
- [7] W. Chen, K. Najarian, “Segmentation of ventricles in brain CT images using Gaussian mixture model method,” *Proceedings of IEEE International Conference on Complex Medical Engineering (CME)*, April 2009, pp. 1-6.
- [8] C. Li, C. Xu, C. Gui, and M. D. Fox, “Level Set Evolution Without Re-initialization: A New Variational Formulation”, *Computer Vision and Pattern Recognition*, vol. 1, June 2005, pp. 430–436
- [9] C. Li, C. Kao, J. C. Gore, and Z. Ding, “Minimization of Region-Scalable Fitting Energy for Image Segmentation”, *IEEE Transactions on Image Processing*, vol. 17, no. 10, October 2008, pp. 1940-1949.
- [10] C. Li, R. Huang, Z. Ding, C. Gatenby, D. N. Metaxas, and J. C. Gore, “A Level Set Method for Image Segmentation in the Presence of Intensity Inhomogeneities with Application to MRI”, *IEEE Transactions on Image Processing*, vol. 20, no. 7, July 2011, pp. 2007-2016.
- [11] R. M. Haralick and L. G. Shapiro, “Computer and Robot Vision,” Volume I, Addison-Wesley, 1992, pp. 28-48.
- [12] R. Mukundan and K. R. Ramakrishnan, “Moment Functions in Image Analysis: Theory and Applications,” Chapter 2, World Scientific, 1998, pp. 9-38.

# Detection of Brain Tumor Using Zernike Moments on Magnetic Resonance Images

Kiran Thapaliya and Goo-Rak Kwon

Department of Information and Communication Engineering

Chosun University

Gwangju, South Korea

{[kiranthapaliya9@gmail.com](mailto:kiranthapaliya9@gmail.com), [grkwon@chosun.ac.kr](mailto:grkwon@chosun.ac.kr)}

**Abstract-** In this study, a new method has been developed for the detection of brain tumor in magnetic resonance images. Magnetic resonance images are analyzed using the Zernike moments and different orders of Zernike moments are calculated. The image is divided into two parts from the center of the image. The average value of the pixels located at the central line is calculated. The new vectors of the pixel are formed based on the calculated average value. The value obtained from the low and high order of Zernike moments are used to calculate the proper threshold value which can extract the tumor efficiently. The proposed method was tested with the different magnetic resonance images containing tumor and the algorithm was successful to analyze the tumor part from the brain image.

**Keywords-** Zernike moments; Zernike polynomials; Mean; MRI; Segmentation.

## I. INTRODUCTION

According to the data of World Health Organization (WHO), more than 400000 persons take the treatment of brain tumor every year [29]. Treatment of brain tumor is a difficult task. It is believed that a careful diagnosis might save patients life. The segmentation of brain tumor from Magnetic Resonance Images (MRI) is a difficult task that involves several disciplines such as pathology, MRI physics, radiologist's perception, and image analysis based on intensity, shape and size. There are several issues and challenges in proper segmentation of brain tumor. Tumors differ with shape, size and location and varies with their intensity. The accurate segmentation of brain tumor is of great interest. Most patients in hospital undergo Computerised Tomography (CT) scan and MRI for the identification of tumor part in the brain. Manual segmentation is risky most of the time and may create problem during the diagnosis by naked eye. Therefore, it is always necessary to use an automatic method to extract the tumor part of the brain; this reduces the human error.

Different methods and approaches have been proposed for the extraction of brain tumor. Some methods are manual, others are semi-automatic and others are automatic. Fuzzy, clustering, edge detection, region growing, and level set methods are introduced in the field of segmentation. Clustering-based methods, such as K-means clustering, have a fast speed even on large datasets but they doesn't provide the same result in each run due to their dependency on initial random assignments [1-2]. Hierarchical self-organizing map- based multiscale image segmentation [3]

and 3D variational segmentation-based methods [4] were also used for the image segmentation [5,6].

These methods used artificial intelligence techniques for automated tumor segmentation. Statistical pattern recognition based methods [7,8-10] fall short, partly because large deformations occur in the intracranial tissues due to the growth of the tumor and edema. However, these techniques need to significantly modify the brain atlas to accommodate the tumors, which specially lead to poor results. The method presented by Singh and Dubey method [27] used the marker based watershed approach to segment the tumor, but it is not an efficient method because it loses tumor information and it is time consuming. Most researchers are using Markov random fields (MRFs) [11,12], which involve estimating the parameters for a parametric model that has one set of parameters to express the probability that each specific voxel is a tumor, and another set to express the distribution over the labels for a pair of adjacent voxels. In Gray Level Co-occurrence Matrix (GLCM) method [14], there is an inherent problem to choose the optimal inter-pixel distance in a given situation. Zernike moments are the mapping of an image onto a set of complex Zernike polynomials [15]. J.K. Udupa et al. [16] combined morphological process and the region growing methods in order to determine tumor volume. Based on the fuzzy logic, Khotanzad and Hong[17] used the fuzzy clustering or the fuzzy connectedness for addressing the problem of abnormal tissue segmentation and classification. Some authors [18-21] used Zernike moments and implemented in the area of image analysis. Zernike polynomials are orthogonal to each other. Zernike moments can represent the properties of an image with no redundancy or overlap information between the moments [22]. Due to these properties, Zernike moments have been widely used in different types of applications [15]. Zernike moments have been used in shape based image retrieval [23], feature set [24] and edge detection in pattern recognition [25]. Even though Zernike moments are used in various fields, they have drawbacks of computational complexity that make unsuitable for real time application.

In this paper, a new method has been developed for the diagnosis of the MR images containing tumor. The main aim of the paper is to extract the tumor part efficiently from MR images and reduce human manual interaction. The detail of Zernike moments has been described in section II, the detail algorithm for the extraction of brain tumor has

been described in section III, followed by the experimental results and conclusion in section IV and V respectively.

## II. ZERNIKE MOMENTS

Zernike moments are based on complex polynomials that form a complete orthogonal set on a unit circle. The Zernike moments over the unit circle  $x^2 + y^2 = 1$  is described by

$$v_{nm} = v_{nm}(\rho, \theta) = R_{nm}(\rho) e^{im\theta} \quad (1)$$

where,  $n$  is a non-negative integer and  $m$  is a non-zero integer, under the condition  $n - |m|$  is even and  $|m| \leq n$ , and  $\rho$  is a vector from the origin of the disc to a point on it.  $\theta$  is the angle that the vector  $\rho$  makes with the positive direction of x-axis in the counter clockwise direction.  $R_{nm}(\rho)$  are the Zernike radial polynomials in polar coordinates and are defined by

$$R_{nm}(\rho) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \frac{(n-s)!}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!} \rho^{n-2s} \quad (2)$$

Polynomials in above equation are orthogonal and according to orthogonality condition

$$\int \int_{x^2+y^2 \leq 1} v_{nm}^*(x, y) v_{pq}(x, y) dx dy = \frac{\pi}{n+1} \delta_{np} \delta_{mq} \quad (3)$$

where,  $\delta_{np} = 1$  when  $n = p$  and zero otherwise.  $\delta_{np}$  and  $\delta_{mq}$  is the Kronecker delta.

Zernike functions corresponding to continuous function  $f(x, y)$ . Zernike moment for order  $n$  and repetition  $m$  is given by

$$A_{nm} = \frac{n+1}{\pi} \int \int_{x^2+y^2 \leq 1} f(x, y) v_{nm}^*(\rho, \theta) dx dy \quad (4)$$

If  $f(x, y)$  is a digital image, we replace the integral by summations to get Zernike moments for the image. Then,

$A_{nm}$ , in this case, reduces to

$$\begin{aligned} A_{nm} &= \frac{n+1}{\pi} \sum_x \sum_y f(x, y) v_{nm}^*(\rho, \theta) \\ &= \frac{n+1}{\pi} \sum_{\rho=0}^1 \sum_{\theta=0}^{2\pi} f(\rho, \theta) R_{nm}(\rho) e^{-im\theta} \\ &= \frac{n+1}{\pi} \sum_{\rho=0}^1 R_{nm}(\rho) \sum_{\theta=0}^{2\pi} f(\rho, \theta) e^{-im\theta} \end{aligned} \quad (5)$$

## III. PROPOSED METHOD

In this paper, the value of low and high order Zernike moments are used to segment the tumor from the MR images. The input image is divided into two parts with different pixels value. The division of pixel is performed

vertically from the center of the image. The low order and high order Zernike moments are calculated using Equ. (5). The values of low and high order Zernike polynomials are used to calculate low and high order of Zernike moments respectively. In the proposed method, different values of Zernike moments are calculated at different order. Thus, obtained value is utilized to calculate the mean value, which separates the tumor from the image. The detail procedure for extraction of tumor image is discussed in the following section.

### A. Feature extraction and Selection

The image is divided into two left and right hemispheres. These two hemispheres of the image contain complex Zernike moment value. The division of pixel is based on the average value of the pixels located at the center boundary of image.

Let us assume that  $f(x, y)$  is the image formed by the complex Zernike moments value.  $b(x, y)$  is the image pixels located at the center boundary of image. We calculate the average value of the pixel  $b(x, y)$  located at the center boundary of the image as shown in the fig. (1). Average pixel value is calculated using following equation.

$$Avgvalue = \frac{1}{N} \sum_{x=0}^{x=x-1} \sum_{y=0}^{y=y-1} b(x, y) \quad (6)$$

where,  $N$  is the total number of pixels located at the center boundary of the image.

The image is divided into left and right hemisphere; choose the minimum pixel values (that is not zero) from the left hemisphere of the divided image. Thus, form vector of all the pixels that lies between minimum values of the left hemisphere of the image to the average value *Avgvalue* of the boundary. Similarly, choose the minimum value of pixel from the right hemisphere of the image to the average value *Avgvalue* of the center boundary of the image. Thus, vector form is the range of minimum pixel value from the left and right hemisphere to the average value of central boundary pixel. These pixels vector are combined together and treated as a single image  $k(x, y)$ .  $I_1(x_n, y_n)$  is the range of pixel values from left hemisphere to the average pixel value *Avgvalue* of the pixel located at the center boundary. Similarly,  $I_2(x_n, y_n)$  is the range of pixel values from right hemisphere to the average pixel value *Avgvalue* of the pixel located at the center boundary.  $k(x, y)$  is the vector formed from the pixel  $I_1(x_n, y_n)$  and  $I_2(x_n, y_n)$

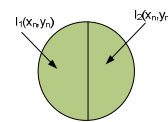


Fig.1. Division of image into two parts from the center of the image

The low order and high order Zernike moments are calculated for the different order  $N$  and with repetition of  $M$  of the image. The table below shows the low and high order Zernike moments for the image.

TABLE I. SAMPLE ORDER REPETITION COMBINATION

Group	N	M	Number of moments
Low order Zernike moments	2	0,2	18
	3	1,3	
	4	0,2,4	
	5	1,3,5	
	6	0,2,4,6	
High order Zernike moments	7	1,3,5,7	18
	7	3,7	
	8	0,4,8	
	9	1,5,9	
	10	2,6,10	
	11	3,7,11	
	12	0,4,8,12	

In, the table above, 18 Zernike moments are selected for the low and high order Zernike moments that satisfy the following conditions

$$LZM = \{Z_{n,m}\} \forall \begin{cases} 2 \leq n \leq 7 \\ |m| \leq n \\ n - |m| = 2k \\ k \in N \end{cases} \quad (7)$$

$$HZM = \{Z_{n,m}\} \forall \begin{cases} 7 \leq n \leq 12 \\ |m| \leq n \\ n - |m| = 4k \\ k \in N \end{cases}$$

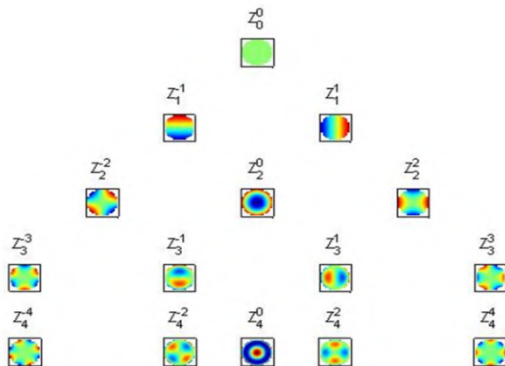


Fig.2. Plots of the magnitude of low order Zernike basis function in the unit disk.

### B. Segmentation

The final step of the proposed method is the extraction of tumor from the image. The vectors contain the complex Zernike moment value; we calculate the appropriate thresholding value, which can separate the tumor from the image using these complex Zernike moments value. Complex Zernike moments value contain low order Zernike moments and high order Zernike moments. Here, we calculate the mean value for each real and imaginary part of low order Zernike moments separately. Similarly, the calculation of mean value for real and imaginary part for high order complex Zernike moments are carried out separately.

$$\mu_1 = \left\{ \max \left( \frac{1}{x \times y} \sum_{x=0}^{x-1} \sum_{y=0}^{y-1} k_1(x, y) \right) \right\}^{\frac{1}{2}}$$

$$\mu_2 = \left\{ \max \left( \frac{1}{x \times y} \sum_{x=0}^{x-1} \sum_{y=0}^{y-1} k_2(x, y) \right) \right\}^{\frac{1}{2}} \quad (8)$$

$$\mu_3 = \left\{ \max \left( \frac{1}{x \times y} \sum_{x=0}^{x-1} \sum_{y=0}^{y-1} k_3(x, y) \right) \right\}^{\frac{1}{2}}$$

$$\mu_4 = \left\{ \max \left( \frac{1}{x \times y} \sum_{x=0}^{x-1} \sum_{y=0}^{y-1} k_4(x, y) \right) \right\}^{\frac{1}{2}}$$

where,  $\mu_1$  and  $\mu_2$  are the mean value of real and imaginary part of low order Zernike moment respectively.  $k_1(x, y)$  and  $k_2(x, y)$  are pixel values that contains only real and imaginary pixel values of low order Zernike moments respectively.  $\mu_3$  and  $\mu_4$  are the mean value of real and imaginary part of high order Zernike moments respectively.  $k_3(x, y)$  and  $k_4(x, y)$  are the pixel values that contains only real and imaginary pixel values of high order Zernike moments respectively.

Now, we subtract the mean value of real part of low order Zernike moments from real part of high order Zernike moments

$$\mu_5 = \mu_3 - \mu_1 \quad (9)$$

where,  $\mu_5$  is the mean value after subtracting two mean values of real part of low and high Zernike moments.

Similarly, we subtract the mean value of imaginary part of low order Zernike moments from imaginary part of high order Zernike moments which is given as

$$\mu_6 = \mu_4 - \mu_2 \quad (10)$$

where,  $\mu_6$  is the mean value after subtracting mean values of imaginary part of low and high Zernike moments

Now, we calculate the final thresholding value, which can separate the tumor from other objects in the image.

$$T = \frac{\mu_5 + \mu_6}{2} \tag{11}$$

$k(x,y)$  is the image that contains the low and high order Zernike moments value. Now using this thresholding value, we obtained only the tumor part from the image  $k(x,y)$ .

$$final(x,y) = \begin{cases} 1 & \text{if } k(x,y) \leq T \\ \text{else} & \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

$final(x,y)$  is the final tumor part extracted from the image.

Finally, the replacement of the pixel is done to get the desired output tumor segmented image. The pixel replacement is performed as

$$\begin{aligned} \text{if } final(x,y) = 1 \text{ then } final(x,y) = 0 \\ \text{else otherwise } final(x,y) = k(x,y) \end{aligned} \tag{13}$$

#### IV. EXPERIMENTAL RESULTS

The proposed method was tested with different MR images. The proposed method was tested with the brain having different intensity, shape and size. The low order Zernike moments were calculated from different kinds of brain tumor images and proposed method was successful to efficiently extract the tumor part from the brain tumor images. The method was tested using MATLAB 2012. The images size of 240X240 for Fig. 3 and 200X200 for Fig. 4-7 was taken for the experimental purpose. Experimental results for the different kinds of images are shown below.

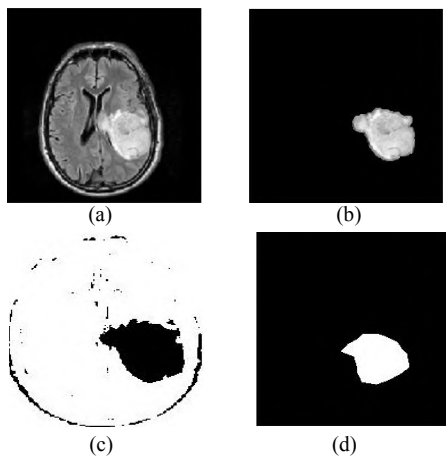


Fig. 3. Extraction of huge mass of brain tumor. (a) Original image, (b) extracted tumor from proposed method, (c) Region grows method, and (d) Singh and Dubey method.

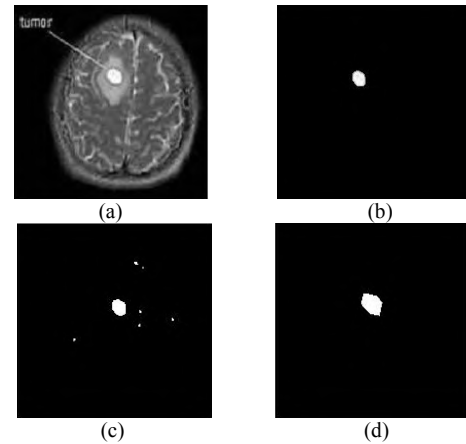


Fig. 4. Segmentation of tumor using proposed method. (a) Original image, (b) Segmented tumor from proposed method, (c) Region grows method, and (d) Singh and Dubey method.

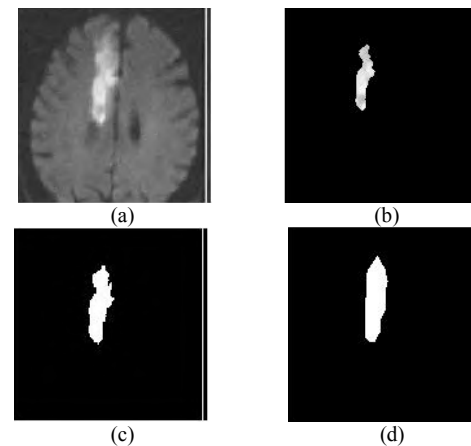


Fig. 5. Segmentation of brain tumor having cylindrical shape (a) Original image, (b) Proposed method, (c) Region grows method, and (d) Singh and Dubey method.

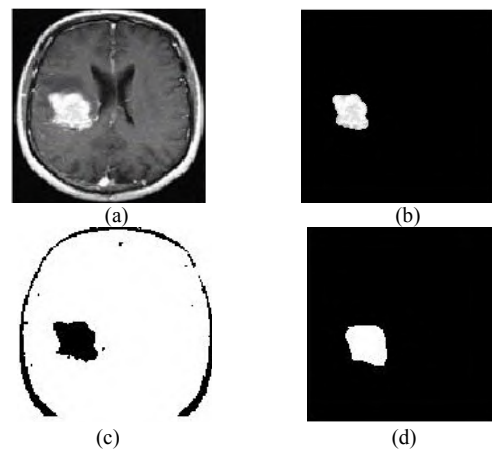


Fig. 6. Extracted star shaped tumor using our proposed method. (a) Original image, (b) Proposed method, (c) Region grows method, and (d) Singh and Dubey method



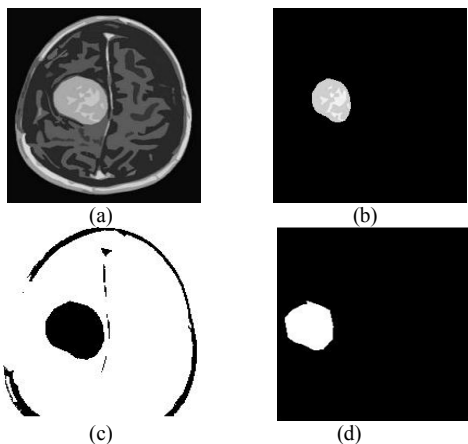


Fig. 7. Extracted oval shaped brain tumor. (a) Original image, (b) Proposed method, (c) Region grows method, and (d) Singh and Dubey method.

Fig. 3-7 show the different images obtained from the proposed method and other two methods, region grow and Singh and Dubey method. In Fig. 3 (a) is the original image with the huge mass of tumor. Fig. 3(b), (c) and (d) are the tumor images obtained by the proposed, region grow and Singh and Dubey method respectively. It is clear from the experimental results that the proposed method effectively extracts the tumor part accurately rather than other two methods. Similarly, from Figs. 4-7, the proposed method gives better results in comparison to other methods. Therefore, by subjective analysis it is clear that the proposed method outperforms the other two methods for the extraction of tumor part from the input brain MRI images.

To prove the proposed method is better than the region grows and Singh and Dubey method quantitatively, we calculate the Root Mean Square Error (RMSE) and time complexity.

Root mean square error (RMSE): it is a quadratic scoring rule which measures the average magnitude of the error. The RMSE value is calculated as

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Q_m - Q_A)^2}{n}} \quad (14)$$

where,  $Q_m$  and  $Q_A$  are the tumor voxels segmented manually and the different methods, respectively, and  $n$  is the size of the image.

RMSE value calculated for different images using different method is shown in Table II.

TABLE II. RMSE OF PROPOSED AND OTHER TWO METHODS

Figure	Proposed	Region grow	Singh and Dubey
Fig. 3	0.09	0.98	0.96
Fig. 4	0.05	1.25	0.99
Fig. 5	0.22	1.01	0.87
Fig. 6	0.20	1.31	0.97
Fig. 7	0.24	1.76	0.93

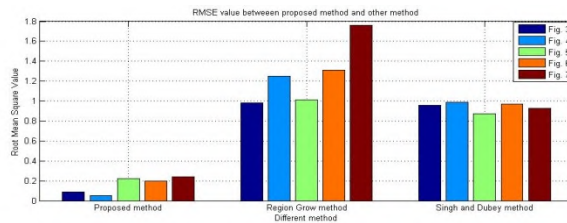


Fig. 8. RMSE value of proposed method and other two methods

The RMSE value of Table II is shown in graphical representation. From the above table and graph , it clear that RMSE value of proposed method is less than the region grow and Singh and Dubey method. The data above clearly mention that proposed method is better than other two methods to segment the tumor part.

We analyze the complexity of the proposed method with region grow and Singh and Dubey method. Table III below shows the calculation of time complexity in second performed on MATLAB 10, Intel 2.40GHz, with memory 2.0GB. From the table below, it clear that time taken to run the proposed method is less than the other two methods. Therefore, we can say that the proposed method has less time complexity in comparison to region grow and Singh and Dubey method and proposed method outperformed other two methods.

TABLE III. TIME COMPLEXITY

Figure	Proposed	Region grow	Singh and Dubey
Fig. 3	2.99	4.3	4.83
Fig. 4	2.39	4.23	5.01
Fig. 5	2.23	4.75	5.05
Fig. 6	2.13	5.10	4.34
Fig. 7	3.04	4.55	5.19

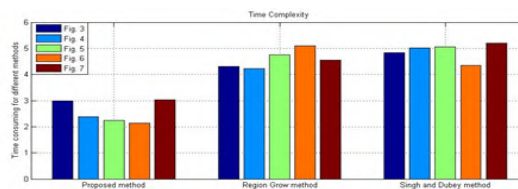


Fig. 9. Time complexity calculated for proposed and other two methods.

## V. CONCLUSION AND FUTURE WORK

In this paper, an efficient method for the extraction of brain tumor has been introduced. The proposed method is based on the Zernike moments. The high order Zernike moments not only have very high computational complexity but they are even sensitive to noise. Therefore, the proper value should be chosen that is shown in above table. Some values from the low order Zernike moments are chosen, so that high order value does not have much effect during the computation value of mean value. The combination of mean values from the low and high order Zernike moments are able to extract the tumor part from the different kinds of

brain tumor images invariant of its shape, size and intensity. The method proposed here is simpler and easy to understand. Even the proposed method is able to extract the tumor efficiently from the MR images; it also paves the way for the expert to decide whether the extracted brain tumor is benign and malignant due to many pathological features. This will be the subject of future research.

#### ACKNOWLEDGEMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No.2010-0008974)

#### REFERENCES

- [1] V. Rajamani and S. Murugavalli, "A High Speed Parallel Fuzzy C-means Algorithm for Brain Tumor Segmentation," *ICGIST International Journal on BIME*, vol. 6, Dec. 2006, pp. 29-34
- [2] D. Vignion et al., "A New Method based on both Fuzzy set and Possibility Theories for Tumor Volume Segmentation on PET images," 30th International Conference of IEEE EMBS, Aug. 20-25, 2008, pp. 3122-3125.
- [3] S.M. Bhandarkar and P. Nammalwar, "Segmentation of Multispectral MR images Using a Hierarchical Self-Organizing Map Computer-Based Medical System CBMS 2001," *Proc. 14th IEEE Symposium*, vol 26, no. 27, 2001, pp. 294-299.
- [4] J. Chunyan, X. Zhang, W. Huang, and C. Meinel., "Segmentation and Quantification of Brain Tumor," *IEEE international Conference on Virtual Environment, Human-Computer Interfaces and Measurement Systems*, 2004, pp. 61-66.
- [5] Clark et al., "Automatic Tumor Segmentation using Knowledge based Techniques," *IEEE Transactions on Medical Imaging*, vol. 117, 1998, pp. 187-201.
- [6] Fletcher-Heath et al., "Automatic Segmentation of Non-enhancing Brain Tumor in MR Images," *Artificial Intelligence in Medicine*, vol. 21, 2001, pp. 43-63.
- [7] Kaus MR et al., "Segmentation of Meningiomas and Low Grade Gliomas in MRI," *MICCAI*, Cambridge, UK, 1999, pp. 1-10.
- [8] N. Moon, E. Bullitt, K. Leemput, and G. Greig, "Model-based Brain and Tumor Segmentation," *Proc. of ICPR*, Aug. 2002, pp. 528-531.
- [9] N. Otsu, "A Threshold Selection Method from Greylevel Histogram," *IEEE Transactions on System, Man and Cybernetic*, vol. 9, 1979, pp. 62-66.
- [10] M. Prastawa, E. Bullitt, S. Ho, and G. Gerig, "A Brain Tumor Segmentation Framework based on Outlier Detection," *Medical Image Analysis*, vol. 8, 2004, pp. 275-283.
- [11] AS. Capelle, O. Alata, C. Fernandez, S. Lefevre, and J.C. Ferrie, "Unsupervised Segmentation for Automatic Detection of Brain Tumors in MRI," *Proc. of Int. Conf. on Image Processing*, 2000, pp. 613-616.
- [12] MB. Caudra et al., "Atlas based Segmentation of Pathological Brains Using a Model of Tumor Growth," *MICCAI*, 2002, pp. 380-387.
- [13] F. Lefebvre, G. Berger, and P. Laugier, "Automatic Detection of the Boundary of the Calcaeus from Ultrasound Parametric Images Using an Active Contour Model-Clinical Assessment," *IEEE Transactions on Medical Imaging*, vol. 17(1), 1998, pp. 45-52.
- [14] W. Wang, J.E. Mottershead, and C. Mares, "Mode-shape Recognition and Finite Element Model Updating Using the Zernike Moment Descriptor," *J. Mechanical Systems and Signal Processing*, vol. 23, 2009, pp. 2088-2112.
- [15] P. Gibbs, D. Buckley, S. Blackband, and A. Horsman, "Tumour Volume Determination from MR Images by Morphological Segmentation," *Physics in Medicine and Biology*, vol. 13, 1996, pp. 2437-2446.
- [16] J.K. Udupa, P.K. Saha, and R.A. Lotufo, "Relative Fuzzy Connectedness and Object Definition: Theory, Algorithms, and Applications in Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24(11), 2002 pp. 1485-1500.
- [17] A. Khotanzad and Y.H. Hong, "Rotation Invariant Image Recognition Using Features Selected via a Systematic Method," *Pattern Recognition* vol. 23(10) , 1990, pp. 1089-1101.
- [18] S. Ghosal and R. Mehrotra, "Robust Optical Flow Estimation Using Semiinvariant Local Features," *Pattern Recognition*, vol. 30(2), 1997, pp. 229-237.
- [19] S.X. Liao and M. Pawlak, "On the Accuracy of Zernike Moments for Image Analysis," *IEEE Trans. Pattern anal. Mach. Intell.*, vol. 20, 1998, pp. 1358-1364.
- [20] R.R. Bailey and M. Srinath, "Orthogonal Moment Features for use with Parametric and Non-parametric Classifiers," *IEEE Trans. Pattern anal. Mach. Intell.*, vol. 18, 1996, pp. 389-400.
- [21] S.K. Hwang and W.Y. Kim, "A Novel Approach to the Fast Computation of Zernike Moments," *Pattern Recognition*, vol. 39, 2006, pp. 2065-2076.
- [22] Sh. Li, M.Ch. Lee, and Ch.M. Pun, "Complex Zernike Moments Features for Shape- based Image Retrieval," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 1 (39), 2009, pp. 227-237.
- [23] X. Li and A. Song, "A New Edge Detection Method Using Gaussian-Zernike Moment Operator," In *Proceedings of the IEEE, 2nd International Asia Conference on Informatics in Control*, vol. 1, March, 2010, pp. 276-279.
- [24] J. Haddadnia, M. Ahmadi, and K. Faez, "An Efficient Feature Extraction Method with Pseudo-Zernike Moment in RBF Neural Network-based Human Face Recognition System," *Journal of Applied Signal Processing*, vol. 9, 2003, pp. 890-901.
- [25] Z. Iscan, Z. Dokur, and T. Olmez, "Tumor Detection by Using Zernike Moments on Segmented Magnetic Resonance Brain Images," *Expert Systems with Application*, vol. 37, 2010, pp. 2540-2549.
- [26] A. Tahmasbi, F. Saki, and S. B. Shokouhi, "Classification of Benign and Malignant Masses based on Zernike Moments," *Computers in Biology and Medicine*, vol. 41, 2011, pp. 726-73
- [27] L. Singh, R.B. Dubey, and Z.A. Jaffery, " Segmentation and Characterization of Brain Tumor from MR Images," *Proc. Int. on Advances in Recent Technologies in Communication and Computing*, 2009, pp. 815-819.
- [28] K. Thapaliya and G.R. Kwon, "Extraction of Brain Tumor Based on Morphological Operations," *Proc. Int. ICCM*, vol. 1, 2012, pp. 515-520.
- [29] P.Narendran, V.K Narendira Kumar, K. Somasundaram, "3D Brain Tumor and Internal Brain Structures Segmentation in MR Images," *IJ Image, Graphics and Signal Processing*, vol. 1, 2012, pp-35-43

# A Meta Model-Based Web Framework for Domain Independent Data Acquisition

Dominic Girardi and Johannes Dirnberger  
 Research Unit Medical Informatics  
 RISC Software GmbH  
 Hagenberg, Austria  
 Email: dominic.girardi@risc.uni-linz.ac.at  
 Email: johannes.dirnberger@risc.uni-linz.ac.at

Johannes Trenkler  
 Institute for Radiology  
 Landesnervenklinik Wagner Jauregg  
 Linz, Austria  
 Email: johannes.trenkler@gespag.at

**Abstract**—We present a generic, web-based data acquisition system that is based on a domain-independent meta data model and which is able to collect, store and manage data of almost arbitrary structure. Due to the use of abstract meta data models, completely generic applications can be built. The additional level of abstraction guarantees the independence of database structure and source code from the actual domain of application and allows to create software systems that can be customized for a certain application without changing any internals of the system. So, domain experts and researchers are able to create, run, and adapt their own web interface for data acquisition without depending on external IT experts. We demonstrate our approach on a registry for intracranial aneurysms.

**Keywords**—*Meta-Modelling; Web-based Data Acquisition; Generic Data Acquisition Systems.*

## I. INTRODUCTION

Data analysis and data mining are well established and indispensable tools in medical research. But they cannot be seen as isolated steps. They are included into a process of knowledge discovery (KD). Cios et al. [1] define this process as the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data. Besides Cios et al., several other definitions and descriptions of the knowledge discovery process (KDP) can be found in [2], [3], and [4]. They all describe KD as a process usually beginning with data selection, data cleaning and transformation, followed by data mining, finalized by interpretation and presentation.

However, one particular aspect is often neglected: data acquisition. The reason for this neglect is the fact that, usually, data, which is produced anyway in business processes, genetic experiments, etc. is analyzed. But, especially in scientific medical research, it is often necessary to acquire the data of need. Data stored in hospital information systems (HIS) is hardly suitable for scientific research because it is often semi-structured, textual data [5] or contains data mostly for billing and documentation purposes [6]. Although data mining has already been performed directly on HIS, those results are less scientifically applicable than for management purposes [7], [8].

Since medical data structures are usually non-trivial, a professional data acquisition system is needed to acquire and store the data. Since every field of study requires its own individual data structure, these data acquisition systems are hardly reusable and need to be developed individually for each

new domain. This is elaborate and expensive and often makes study authors prefer suboptimal data storage solutions (i.e., MS Excel sheets). In order to overcome these drawbacks, we have developed a generic, web-based data acquisition system which is based upon an abstract meta data-model. Several existing approaches try tackling this goal. Frameworks like *Hibernate* [9] or new programming languages like *Ruby* [10] *on Rails* [11] have their focus on supporting the programmer rather than the end user. Our approach is focused on providing a system that can be easily adjusted by the end-user directly without information technology (IT) knowledge.

The application of a meta data-model allows the creation of an abstract domain-independent system that is then domain-customized by the domain expert himself. Intuitive user interfaces allow the medical domain expert (who is probably not an IT expert) to define his data structures of interest. The rest of the system (UI forms, overview tables, search masks, etc.) is automatically created based on the stored meta-data. So, the user is now able to set up and maintain its own medical study without dependence on an external IT contractor.

In Section II, we provide an overview over related publications. Section III contains a detailed description of the generic meta model and the application itself. In Section IV, we present our results on the example of a disease register, which was realized using the system in the Landesnervenklinik Wagner-Jauregg Linz. Section V contains our conclusion and an outlook for further research.

## II. RELATED RESEARCH

Our literature review in scientific sources did not yield any scientifically published system that is directly comparable to our approach. However, in the field of medical data recording there are a number of systems which offer electronic case report forms (eCRF). Franklin et al. [12] compare three of the most popular electronic data capture (EDC) systems Catalyst Web Tools, OpenClinica and REDCap [13]. These systems are very complex and offer additional features for study management and planning. But still, they are limited to clinical data acquisition and not as generic as our system. A more generic approach can be found in Zavaliy et al. [14]. The one page position paper describes the basic concept of an ontology-based data acquisition system for electronic medical record data. A very simple ontology is used, which contains four

concepts: Person, Hospital, Diagnosis and Medication. They point out, that the main reason for using an ontology-based approach is the need for adaptive data structures. This work is closely related to our work. However, their paper contains no information on how data can be entered into the system, neither is there information about the system architecture or semantic data checking. There is also no information given on how adaptable their ontology is and if those four main concepts can be replaced or not. Despite the fact that they follow a very similar basic idea, our system seems to be more extensive and matured.

Apart from these, at least strongly customizable but not absolutely, generic systems some domain specific web-based data acquisition systems can be found, such as Kodama et al. [15], for collecting wind data.

### III. METHODS

#### A. Meta Model

1) *Motivation:* Conventional data acquisition systems are based on a descriptive data model, which directly reflects the domain of application. Since the whole application is based on the data model, there is a strong semantic dependency from the domain of application through the data model to all other aspects of the application (logical layer, user interfaces, data exchange interfaces). Due to this dependency, changing requirements (additional attributes, changing tables, adding relations, etc.) cause changes throughout the whole application, which usually needs to be performed by IT experts. Furthermore, this dependency strongly limits the reusability, since different domains of applications cause completely different data models.

The usage of meta data-models leverages this dependency. The domain-specific data structures do not define the data structure of the application but are stored into the meta data-structure. So, the generic meta data-structure always remains the same, regardless of the actual field of application. The stored meta data allows the automatic creation of GUI elements, so changes to the domain-specific data structures are automatically propagated throughout the application.

2) *Meta ER Model:* Our meta model is based on Peter Chen's entity-relationship (ER) model [16]. It is basically an ER model which is able to store another ER model and its data. Figure 1 shows the core elements of our model. The elements on the left hand side are primarily used to define the data structure, while the elements on the right contain the actual data. For further technical details the reader is referred to [17].

3) *Consequences:* Due to the abstraction of the whole system from the domain of application, a lot of advantages arise for both the user and the developer. The user is now able to set up and maintain the system without dependence on an IT contractor. The changes he makes to the domain data model have immediate effects on the user interfaces. By using meta models, we were able to create a system that is highly adaptable and reusable. However, developing generic systems also means dealing with an additional level of abstraction. Especially, when it comes to database queries, meta data-models can be very challenging. Data, which can

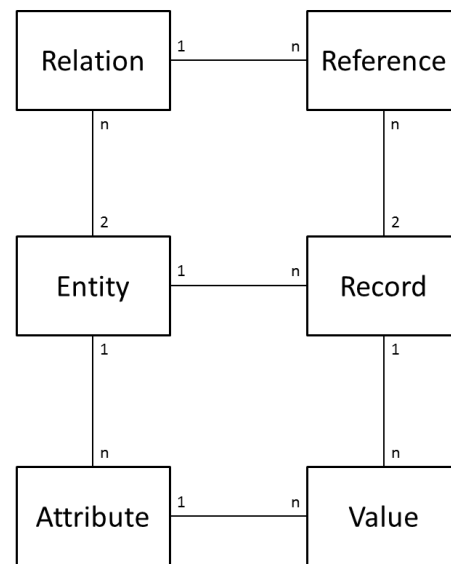


Fig. 1. ER diagram of the meta data-model [17]

be easily queried from a conventional database, needs to be joined from multiple tables in the meta model. This is not only a performance issue; it also complicates the query structure.

4) *Relation to OWL Ontologies:* Ontologies are content theories about the sorts of objects, properties of objects, and relations between objects that are possible in a specified domain of knowledge [18]. So, the presented meta data-model can be seen as an ontology storage and the system itself as a ontology-based data acquisition system. Consequently, the question arises, why the ontology definition languages OWL or RFD were not used for this system. The main reason is the paradigm mismatch between the Open World Assumption of (OWL) ontologies and the Closed World Assumption of relational data models. In OWL anything, that is not forbidden is allowed. So, for example, a record (individual) can be instance of more than one class, which does not make sense for medical data acquisition. For our purpose, the Closed World Assumption is more accurate and more comprehensible to the user. Furthermore, the whole system is a database-based solution, while OWL ontologies are file based. However, there are ways to store OWL ontologies in relation databases [19], but this solution includes transformations from OWL files to SQL statements. This would require to user to define his domain of application in an OWL editor, which is considered tricky, even for experienced users, and then, transfer the whole ontology into the database. So, this approach is not applicable for our purpose, either.

For a more detailed argumentation on OWL ontologies and the described system the reader is referred to paper Girardi et al. [17].

#### B. System Architecture

The system is written in object-oriented PHP [20] and uses a MySQL [21] database for data storage. The meta model entities are implemented as PHP classes that allow the creation of structures (entity, attribute, relation, type) and data (record,

reference, value, user, usergroup). A central main class controls the collaboration of the classes with the user interface. The standard user interface is controlled by the PHP class *GUI*, which encapsulates most of the code for the generation of the HTML output. Furthermore, numerous system functions such as record searching, record tree, calendar, charts, data import, reporting, etc., are provided.

Basically, the system architecture consists of three main parts

1) *Meta Model*: Every entity of the meta model is represented as a separate class in our system.

- **User**: A user represents a physical user that accesses the data and meta data of the system. A user must have a unique name and password with which he is authenticated to the system. Every data manipulation is tracked based on the user that has performed the action. Every user is member of a user group that can define special access permissions. A user owns the records he as created.
- **Entity**: An entity represents a class of objects with equal properties. It has a unique name and it has multiple attributes of different types and names. In contrast to the common relational database paradigm we have implemented the object-oriented inheritance paradigm, thus entities can inherit attributes from other entities. Moreover, we distinguish between tree entities and lookup entities. Tree entities are affected by the end user data acquisition, whereas lookup entities are used for selection boxes of attributes in tree entities. Each entity defines an access level that can either be User, Group or Everyone. Dependent on this access level different users see or do not see records of this entity. If the level is set to User, every user only sees records he has created. If access level is Group, a user has access to all records that have been created by users of his group.
- **Attribute**: An attribute represents one field of data inside an entity. It corresponds to a table field in a relational database but has more properties. An attribute has name and a label. It can also have a group name for grouping multiple attributes in the GUI. Attributes also define their order of appearance in record captions, record tables and selection dropdown lists. Attributes can also have different colors on the GUI and an attribute defines the valid range for numerical values. A regular expression for data validation can also be defined. Attributes can have a tooltip that is display in the GUI and they can have a link with further information that will also be displayed as info button in the GUI (see Figure 3). Relations are treated as attributes of type *Foreign Key*.
- **Type**: The system provides several different types for attributes of entities. The common types are ID, integer, float, string, text, date, time, file. For consistency reasons, we do not allow user-defined types. An attribute as selection of a pre-defined set of choices can be created by adding a relation to a lookup entity.

- **Relation**: A relation represents a 1:n connection between entities just as in a relational database schema. Our relation can also represent an m:n connection resulting in a multi-select list element in the GUI (see attribute 'Others' in Figure 3).
- **Crosslink**: A crosslink is a weak reference between records that are not directly part of a relation. Thus, our system allows to create links between records on different levels of the relational hierarchy. This helps linking information across the hierarchy.

2) *Auxiliary System Functions*: The system provides extra data analysis and visualization features that are integrated in the user interface.

- **Search**: The class RecordSearch provides vast search functionality for records in the system. It is also capable of displaying a user interface for defining and storing search queries.
- **Charts**: Every list of records can be used to fill a 2D chart that can then be displayed anywhere in the user interface. There are several chart types available like line chart, bar chart, pie chart, self organizing maps (SOM), etc.
- **Reports**: Often used queries can be stored as reports for quick reuse in multiple places in the user interface. Reports can be generated quickly and present their results in table format where every cell can be a customized query field.
- **Dependency**: Our system allows to define dependencies between records, which can either result in hiding selection choices that are not plausible in certain conditions or even in hiding whole attributes under certain conditions. Dependencies can be defined over multiple levels in the entity hierarchy providing a powerful tool for increasing data acquisition quality. Dependencies can also be used to check plausibility of entered or imported data, which allows easy determination of data quality for studies.
- **Import/Export**: Our system provides an I/O interface that currently supports XML import and export, PDF export, CSV import and export.

3) *Web Interface*: The class GUI encapsulates most of the code for generating the HTML user interface of our system. There are functions for generating a header, a footer, a Windows Explorer-like record tree with full navigation functionality, dropdown areas, etc. All data dependent code is generated by the meta model classes themselves. A record can display itself either in a table row or as an HTML form. The representation can be customized by parameters. An entity can display all its records as a table. Clicking on a record row automatically displays the record's HTML form without further implementation.

The system also provides a comprehensive administrator interface for meta model manipulations. Users with administrator access can add/delete entities, manipulate attributes and relations, or manage users.

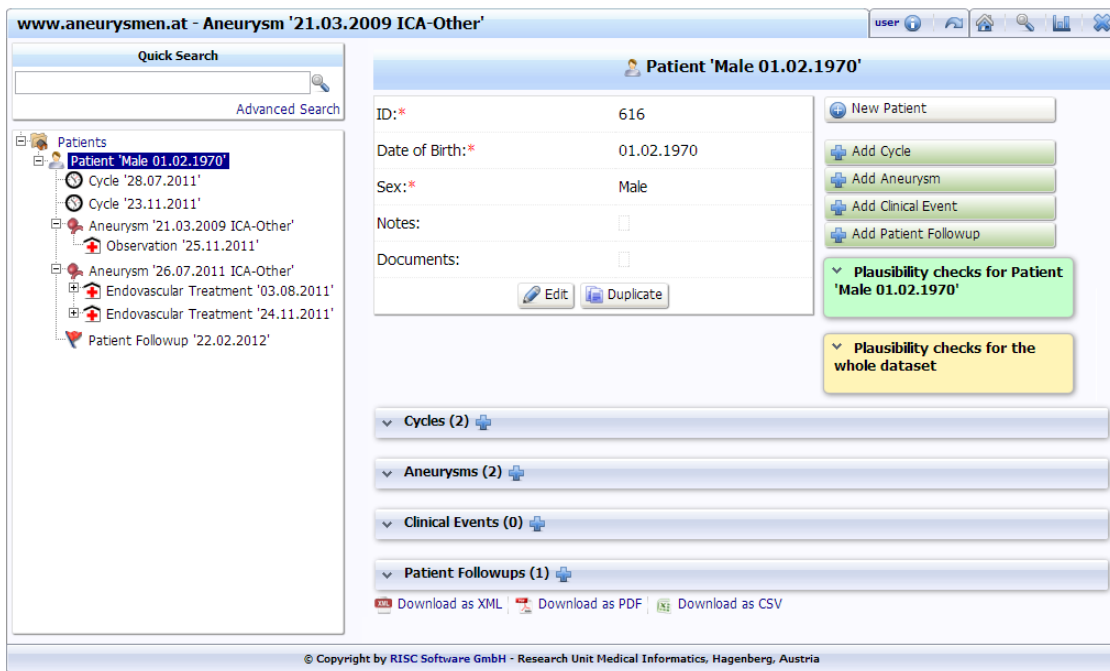


Fig. 2. Standard user interface in cerebral aneurysm registry for a record of entity *Patient* with *Add* buttons for all relations and showing the results of plausibility checks of this record, as well as the record hierarchy navigation tree on left side

The appearance of the user interface (as depicted in Figure 2) can also be customized by the concept of multiple startpages. Every instance of our system can define its own startpage providing almost infinite possibilities. Every instance has a unique configuration file, which defines the database connection and the startpage of the system. Thus, multiple instances of our system can be hosted on one server.

### C. Dynamic GUI Generation

Due to the usage of a meta model as basis for the generation of the user interface, no source code changes are necessary, if the data structure of the system is changed. Administrators can arbitrarily add, delete or modify entities, attributes and relations. The user interface changes instantaneously according to the changes. Every record can display itself on the user interface by showing a table of all attributes of its entity, a collection of 'Add' buttons for every relation and automatic plausibility checks based on the defined dependencies (see Figure 2).

Every record automatically creates an HTML interface for editing all its attributes according to their types. An HTML form is created as table with one row for every attribute of the record's entity. Textual and numeric attributes are displayed as text fields (see attributes 'Admission Date', 'Presentation Date' and 'NIH at Presentation' in Figure 2, whereas relations create a selection box filled with the records of the referenced entity (see attributes 'mRS before Presentation', 'Cycle Presentation', and 'Hunt & Hess' in Figure 2). Our meta model also allows relations of one record with multiple other records, which is displayed as a box of checkboxes with the referenced records in the HTML form of the record (see attribute 'Others' in Figure 2).

Attributes can contain various additional properties according to our meta model such as minimum value, maximum value, regular expression, labels, font colors, tooltip text, information link, ordering, and various others. If entered values mismatch the properties minimum value, maximum value or regular expression, a warning hint is displayed in the HTML user interface.

Below, the records attribute fields and *Add* buttons all existing related records are shown as tables in dropdown areas entitled by the related entity name and the number of related records. Thus, it is possible to quickly navigate through the whole data tree of a root object, which is in our case the *Patient* record. The plus button next to every related entity caption allows the quick creation of new related records of that particular entity. For example, by clicking *plus* next to *Cycles (2)*, the system shows the interface for creating a new record of entity *Cycle* that has automatically set the prior patient as parent record.

## IV. RESULTS - A CEREBRAL ANEURYSM REGISTRY

### A. Definition of a Cerebral Aneurysm

A cerebral aneurysm is the dilation, bulging, or ballooning-out of part of the wall of an artery in the brain. Cerebral aneurysms can occur at any age, although they are more probable in adults than in children and are slightly more common in women than in men. The signs and symptoms of an unruptured cerebral aneurysm will partly depend on its size and rate of growth. For example, a small, unchanging aneurysm will generally produce no symptoms, whereas a larger aneurysm that is steadily growing may produce symptoms such as loss of feeling in the face or problems with the eyes. Immediately after an aneurysm ruptures, an individual may experience such

Fig. 3. User interface for new Cycle record after clicking plus on related records caption

symptoms as a sudden and unusually severe headache, nausea, vision impairment, vomiting, and loss of consciousness [22].

### B. Epidemiological Aspects

Intracranial aneurysms occur in 1 to 5 percent in adult population, which results in about 12 million patients in the US. Most of these aneurysms (50 to 80 percent) are rather small and do not rupture during a patient's life time. The estimated incidence of subarachnoid hemorrhage (SAH) from a ruptured intracranial aneurysm is 1 case per 10,000 persons (in the US). SAH is more common in women than in men (2:1) and the peak incidence is in persons 55 to 60 years old. Although the causes of intracranial aneurysms are not yet known, smoking and hypertension are supposed to have big influence on the development of aneurysms [23].

### C. Purpose of the Registry

The aneurysm registry [24] was established in 2008 by the Institute for Radiology, Landesnervenklinik Wagner Jauregg in Linz, Austria. Its aim is to properly gather and store all relevant information of a patient with cerebral aneurysms. As opposed to hospital information systems, the aneurysm database contains solely medically relevant data in a structured way (no free text or semi-structured information). This allows further automatic processing and analysis of the data, which is used for medical research and internal quality benchmarking.

### D. Data Structure

The data structure is organized in a single rooted acyclic entity graph (figure 4). The top-most entity, in case of the aneurysm data set, is the entity *Patient*. It contains the basic demographic key data (sex, age) and it has four relations to sub-entities. The first kind of sub-entity is *Cycle*. It encapsulates all relevant information about a single hospital stay of the Patient such as admission date and discharge date,

and several medical scores at admission and discharge. In the sub-entity *Clinical Event* all complications are recorded. Each complication is linked to a hospital stay by a crosslink relation. The next sub-entity is *Aneurysm*. It contains all information about the aneurysm's morphology and pathology. This entity is the container for all *Treatment* entities where all treatments (observations, surgical or endovascular treatments) are recorded. Each treatment is linked to a cycle by a crosslink relation as well. The last sub-entity of Patient is the *Follow up*, which represents follow-up results over an interval of several years.

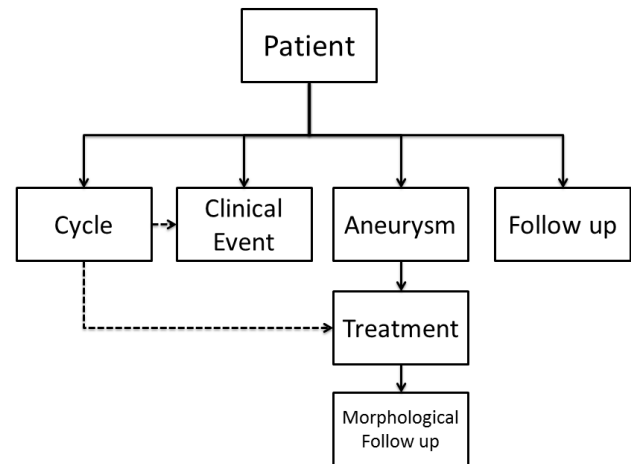


Fig. 4. Data structure of the aneurysm registry (boxes depict Entities, solid arrows depict Relations, dotted arrows depict Crosslinks)

### E. Results

The aneurysm registry was started in 2009 together with Dr. Johannes Trenkler of the Landesnervenklinik Wagner-Jauregg Linz. Since then, over 570 cerebral aneurysms have been recorded, including detailed data on treatments, clinical events, anamnesis, and clinical findings. Based on this data set, several medical dissertations have already been written.

## V. CONCLUSION AND FUTURE WORK

### A. Conclusion

We have shown that our system is capable of storing arbitrary data structures and corresponding data. Once installed, the back-end web interface allows complete customization of the data structure with more possibilities than common relational database systems. The key benefit of our system is the automatically generated front-end based on the data structure stored in the meta model. Therefore, changes to data structure do not affect the underlying source code of the system and can be done after the data acquisition has already started. Acquired data is transformed automatically regarding to the structural changes. Another benefit of our system is the automatic check for plausibility for both entered and imported data. A simple but powerful search interface allows immediate data exploration for users, who can then export their search results as XML, CSV or PDF.

The cerebral aneurysm registry was the first application of our system. Meanwhile, we have successfully instantiated

the system as a neuro-surgical complication registry, a measurement acquisition for child proportions, a tracking system for eye surgeries, and a benchmarking system for clinical performance. Experience from these applications show that the meta-model based approach allows very low lead times for new data acquisition projects. Given that the data structure is already known - e.g. from a former data collection via excel - the system can be set up and online within hours. Furthermore, in the first phases of a new project, where, as experience teaches, a lot of little changes are to be made, the ability for the user to perform these changes on their own was a valuable feature. Long-term experience with large data acquisition systems showed that the additional layer of abstraction does not cause any performance bottlenecks, although the whole transactional data is actually stored in only three tables (record, value, reference).

### B. Further Research

The system is still subject of further development and research. Currently, we are working on a generic expression engine which allows the user to define arbitrary expressions consisting of existing attributes and operators. These expressions can then be used for complex search queries or for defining logical rules for checking the semantic plausibility. For example, if the size and the weight of a patient are part of the domain specific data structure, then an expression can be created that calculates the body mass index (BMI) for this patient. The BMI can then be used to find patients with a BMI in a certain range or to check if the BMI is in a plausible interval. Furthermore, the BMI can also be used as a new variable for subsequent data analysis, which is the second big subject of research.

We are about to extend the data acquisition system in a way that allows the user to explore his data. The stored meta data helps to decrease the effort the user has to put into data preparation and analysis. Furthermore, the meta data allows to automatize many pre-processing steps for data mining and enables the automatic transformation in external file formats, like the ARFF data format [25] of the Weka machine learning framework. So, like the meta model based approach reduces the user's effort for creating a data acquisition infrastructure it is going to actively support the user with data analysis and mining. For more detailed explanation the reader is referred to [26] and [27].

## VI. ACKNOWLEDGEMENTS

We would like to thank the Landesnervenklinik Wagner-Jauregg of the GESPAG and the federal state of Upper Austria for their financial support. Special thanks go to Dr. Johannes Trenkler and Dr. Raimund Kleiser for their effort in creating and deploying the aneurysm registry presented in this paper.

## REFERENCES

- [1] K. J. Cios, W. Pedrycz, R. W. Swiniarski, and L. A. Kurgan, *Data Mining: A Knowledge Discovery Approach*. Springer, 1 ed., Feb. 2007.
- [2] J. He, "Advances in data mining: History and future," *Intelligent Information Technology Applications, 2007 Workshop on*, vol. 1, pp. 634–636, 2009.
- [3] U. Fayyad, G. Piatetsky-shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI Magazine*, vol. 17, pp. 37–54, 1996.
- [4] T. Runkler, *Data-Mining : Methoden und Algorithmen intelligenter Datenanalyse ; mit 7 Tabellen*. Wiesbaden: Vieweg + Teubner, 2010.
- [5] M. Bursa, L. Lhotska, V. Chudacek, J. Spilka, P. Janku, and M. Huser, "Practical problems and solutions in hospital information system data mining," in *Information Technology in Bio- and Medical Informatics* (C. Bhm, S. Khuri, L. Lhotska, and M. Renda, eds.), vol. 7451 of *Lecture Notes in Computer Science*, pp. 31–39, Vienna: Springer Berlin Heidelberg, 2012.
- [6] F. Leiner, W. Gaus, R. Haux, and P. Knaup-Gregori, *Medical Data Management - A Practical Guide*. Springer, 2003.
- [7] S. Tsumoto and S. Hirano, "Data mining in hospital information system for hospital management," in *Complex Medical Engineering, 2009. CME. ICME International Conference on*, (Tempe, AZ), pp. 1–5, april 2009.
- [8] S. Tsumoto, S. Hirano, and Y. Tsumoto, "Information reuse in hospital information systems: A data mining approach," in *Information Reuse and Integration (IRI), 2011 IEEE International Conference on*, pp. 172–176, aug. 2011.
- [9] Hibernate, "http://www.hibernate.org/," May 2013.
- [10] Ruby, "http://www.ruby-lang.org/en/," May 2013.
- [11] R. on Rails, "http://rubyonrails.org/," May 2013.
- [12] J. D. Franklin, A. Guidry, and J. F. Brinkley, "A partnership approach for electronic data capture in small-scale clinical trials," *Journal of Biomedical Informatics*, vol. 44, Supplement 1, no. 0, pp. S103–S108, 2011. AMIA Joint Summits on Translational Science 2011.
- [13] P. A. Harris, R. Taylor, R. Thielke, J. Payne, N. Gonzalez, and J. G. Conde, "Research electronic data capture (redcap) a metadata-driven methodology and workflow process for providing translational research informatics support," *Journal of Biomedical Informatics*, vol. 42, no. 2, pp. 377–381, 2009.
- [14] T. Zavalij and I. Nikolski, "Ontology-based information system for collecting electronic medical records data," in *Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET), 2010 International Conference on*, p. 125, feb. 2010.
- [15] N. Kodama and T. Matsuzaka, "Web-based data acquisition system of wind conditions and its application to power output variation analysis for wind turbine generation," in *SICE-ICASE, 2006. International Joint Conference*, pp. 3747–3750, oct. 2006.
- [16] P. P. S. Chen, "The entity relationship model - toward a unified view of data," *ACM Transactions on Database Systems*, vol. 1, pp. 9–36, March 1976.
- [17] D. Girardi, K. Arthofer, and M. Giretzlehner, "An ontology-based data acquisition infrastructure," in *Proceedings of 4th International Conference on Knowledge Engineering and Ontology Development*, (Barcelona), pp. 155–160, October 2012.
- [18] B. Chandrasekaran, J. Josephson, and V. Benjamins, "What are ontologies, and why do we need them?," *Intelligent Systems and their Applications, IEEE*, vol. 14, no. 1, pp. 20–26, 1999.
- [19] N. K. Irina Astrova and A. Kalja, "Storing owl ontologies in sql relational databases," *Engineering and Technology*, vol. 29, 2007.
- [20] PHP, "http://php.net/," May 2013.
- [21] MySQL, "http://www.mysql.com/," May 2013.
- [22] NIH, "Cerebral aneurysm information page," April 2010.
- [23] J. Brisman, K. K. Song, and D. Newwell, "Cerebral aneurysms," *The New England Journal of Medicine*, vol. 355, pp. 929–939, 2006.
- [24] OEGNR, "www.aneurysmen.at," May 2013.
- [25] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: An update," *SIGKDD Explorations*, vol. 11, 2009.
- [26] D. Girardi, K. Arthofer, and Giretzlehner, "Ontology-guided data acquisition and analysis," in *Proceedings of 1st International Conference on Data Analytics DATA ANALYTICS*, (Barcelona), 2012.
- [27] D. Girardi, M. Giretzlehner, and J. Küng, "Using generic meta-data-models for clustering medical data," in *ITBAM*, (Vienna), pp. 40–53, 2012.



## Cancer and Deadly Infection in Institutions: Developing Use Cases for an MBE Application to Prevent another Enron or Barings

Thang N. Nguyen  
 Department of Information Systems  
 California State University Long Beach  
 CA, USA  
[Thang.Nguyen@csulb.edu](mailto:Thang.Nguyen@csulb.edu)

**Abstract**— One of the deadly diseases in human is cancer. A human cell becomes abnormal when its DNA is mutated and/or its genes are damaged for one reason or another. The abnormal cell produces other cells by division or mitosis. The uncontrollable growing collection of the abnormal cells, called tumor, when invading nearby tissues, is classified as malignant. Malignant tumors eventually proliferate to other organs throughout the human body via the blood and/or lymph circulation, a process called metastasis. Tumors cause serious threats to human health and potentially death. If we think of an institution as a human body, then its employees can be analogously considered as the body's cells. In that sense, the group of "abnormal" or "special" employees led by Jeff Skilling, Andrew Fastow and others in Enron can be considered as a malignant "institution tumor". The group has influenced other organizational units and brought collapse to Enron. Human can also die due to infections caused by a single cell hosting a virus. It is analogous to the case whereas Nicholas Leeson single-handedly brought Barings bank to bankruptcy. We investigate these extreme deadly cases in humans, namely cancer and deadly infection, for insights into the construction of six use cases towards the development of an enterprise-wide MBE-based (management by exceptions) application for the prevention of another Enron or Barings bank.

**Keywords**—*Management by Exceptions; biologically-inspired system; bankruptcy prevention*

### I. COLLAPSES OF BUSINESS INSTITUTIONS DURING THE LAST TWO DECADES

In February 26, 1995, Barings bank in the UK bankrupted because its asset was in the \$635 million while its liability was around \$27 billion [1]. The bank collapsed. It was sold to ING for £1. It started with the last margin call on SIMEX on February 23, 1995. It requested an amount of \$835 million, higher than Barings asset, for transfer from Barings London to Barings Singapore due to futures and options positions in Nikkei 225 and SIMEX. It was found that the collapse was due to a single employee of Barings bank, Nicholas Leeson [2].

In October 2001, Enron announced \$618 million loss for the 2001 third quarter. During the preceding 9 months, Enron revenues were estimated to be roughly \$138 billion [3]. Followed the restatement of 2001 third quarter were the 1997-2000 financial statement reconsolidations. Enron stock went down quickly to 28 cents from roughly \$90. Enron filed for protection was December 2, 2001. It was

determined that the wrongdoings were done by a group of executives in Enron, led by primarily Jeff Skilling and Andrew Fastow [4].

In March 2002, Adelphia stock dropped 18% from \$20+ per share of the day before, after it was disclosed that Adelphia has cosigned a \$2.3M loan to the Rigas family business but it did not report in Adelphia books [5]. A few days later, Adelphia acknowledged a possible debt of \$500M. During May, Adelphia attempted to sell its assets to reduce debts while reworked its 1999, 2000 and 2001 statements to correct off-the-book debts. Stock price went down to 75 cents in early June. More troubling news surfaced: change of external auditor and more debts revealed. The company filed for bankruptcy on June 25, 2002. It was determined by SEC that the Rigas family committed an elaborate and extensive accounting fraud [5].

In June 2002 WorldCom announced restatement because its expenses were posted as capital expenditures in prior financial statements. The company further admitted irregularities in its reserve accounts in August. WorldCom accounting fraud was found to amount to \$11 billion [6]. On July 21, 2002 WorldCom filed for Chapter 11 Bankruptcy Protection. The fraud was committed by a group of executives headed by Bernie Ebbers and Scott Sullivan.

In February 2003, CFO Fuasto Tonna announced a €500M bond issue. He was fired by CEO Calisto Tanzi and was replaced by Alberto Ferraris. Ferraris resigned in November. He revealed he was unable to get access to some of the corporate books. Del Soldato replaced him and only a month later also resigned. In the same month, Parmalat forged paperwork to show it had €3.95B in its Cayman Islands subsidiary, claiming it was back by Bank of America. Bank of America denied. Subsequently, Parmalat defaulted €150M bond. This led to Parmalat bankruptcy on December 24, 2003 [7].

In 2008, Lehman Brothers incurred a substantial loss due to subprime market crisis. Between August and September 2009, Lehman failed to sell its assets to Korea Development Bank, Bank of America and Barclays. The talk with New York Federal Reserve Bank for the possibility of an emergency liquidation of Lehman's assets also failed. In September 15, 2008, Lehman filed for Chapter 11 [8].

After Enron, there were investigations after investigations. They resulted in multiple articles, reports, books, lessons learned as well as a number of proposed

solutions and recommendations based on *who did what, why, when, where and how*. One of the main objectives was to arrive at the prevention of future Enron-like as well as Baring-like institutions from bankruptcy.

Among the solutions and recommendations proposed and implemented, there were Sarbanes-Oxley Act in 2002 [9] for stricter accounting standards, Higgs report [10] on non-executive directors, and resolutions from AICPA [11]. At the US House Energy and Commerce Committee Hearing, B. Dharan suggested several recommendations to SEC and FASB (Financial Accounting Standard Board). The latter involved pro-forma earnings reporting, a complete set of financial statements and modifications to the mark-to-market (MTM) accounting methodology, among others. Also others investigators have proposed changes to the roles, processes, practices, controls, reporting, and the like.

One would think that it would be very hard for another Barings or Enron to occur. The reality is that (1) 13 years later, Jerome Kerviel single-handedly over three days of trading drove Societe Generale to a loss of €3.9 billion, and (2) the list of wrongdoings grew longer for two decades since the Barings case. The latter included some important scandals: Madoff, Tyco International, HealthSouth, HIH Insurance in Australia, Daiwa Bank in Japan, Liu Qibing in China, and others [12]. A few of them escaped bankruptcy, however.

Nevertheless, one cannot help wonder whether all the proposed solutions and recommendations have been collectively effective and/or enough. It appears that they are all good but obviously not enough since the collapses of institutions continued to persist. One can speculate that either the problems were rooted too deep that the prevention measures became ineffective or regardless of any new rules, regulations etc. somebody was able to break it.

Our work approaches the solution from a perspective different than most. The idea is to look at the collapses in institutions in a fashion analogous to the death caused by a cancer in humans. The main thrust is to address prevention from collapse by detecting early symptoms much like in cancer prevention. This in turn suggests a known scheme of enterprise-wide management by exceptions (MBE) [13] to be revisited for the detection of significant symptoms of wrongdoings in time to take appropriate decisions. To that end, we propose in Section 2 a partially biologically-inspired MBE solution for prevention. Section 3 is reserved for the development of use cases, the topic of this paper, towards such a solution. We include a discussion and present our concluding remarks and future work in Section 4.

## II. APPROACH TO PREVENTION AND SYNOPSIS OF A PARTIALLY BIOLOGICALLY-INSPIRED MBE SOLUTION

We start with the assumption that symptoms of wrongdoings in an institution are always there, just as cancerous symptoms in a human body. They are either

ignored, or went undetected. When the symptoms surface the situation is that it will be commonly too late. In effect when cancer is detected, it is already in later phases, and therefore quite often death is practically unavoidable.

Cancer is one of the deadly diseases in humans. Cancer is generally described as follows. The human body is made up of many types of cells. These cells grow and divide in a controlled way to produce more cells as the body needs to keep it working and being healthy. When cells become old, they die and are replaced with new cells, except maybe brain cells [14]. A human cell becomes abnormal when its DNA is mutated and/or its genes are damaged for one reason or another (internal or external).

The abnormal cell produces other cells by division or mitosis. The uncontrollable growing collection of the abnormal cells, called tumor, when invading nearby tissues, is classified as malignant. Malignant tumors eventually proliferate to other organs throughout the human body via the blood and/or lymph circulation, a process called metastasis. Tumors cause serious threats to human health and potentially death [15].

When an institution is considered as analogous to a human then the institution's employees can be analogously considered as similar to biological cells. "Institution tumors" in turn can be viewed analogously as groups of employees in the organization which grow uncontrollably and start some wrongdoings. When funded to do their way, they can influence other organizational units and may become "malignant". They could become seriously harmful to the institution health, and potentially lead it to bankruptcy. The analogy to cancer is elaborated for Enron case as follows.

We equate Jeffrey Skilling and Andrew Fastow and other Enron executives as "special" or "abnormal cells" in the institution body. Skilling used an accounting practice which he convinced SEC to approve. In a sense, Enron's MTM practice to accounting is analogous to a mutation to general accounting practices much like a mutated DNA/gene of a cell; therefore it is affecting its growth. The 360-degree review [16] is analogous to a mutation to human resource hiring/firing policy/process. Special purpose entities (SPE) and accounting schemes can be viewed as mutations (changes or deviations) to commonly practiced SPEs and GAAP.

When Arthur Anderson was influenced in the mishandling accounting audits [17], due to its dual role of external consultant and internal auditor, the institution "tumor" in fact has proliferated to the other institution's "organs and organ systems" (i.e. finance division, accounting, legal counsel, etc.). Cancerous symptoms started to surface: Skilling's replacement of Lay as CEO, followed by Skilling's resignation after only 6 months, followed by the cancellation of deals with Blockbusters and later with Dynegy [18]. SEC began its inquiries and the revision of financial statements from 1997-2000 was initiated [19-21].

Discoveries of wrongdoings in SPEs required Enron statements to include Chewco consolidations as well as those of other projects. Enron stock prices slipped. The consolidations showed Enron debts and liabilities were previously off balance sheets. This caused additional stock price slips and cash shortfalls. Within just 60 days, the company’s stock went down to 28 cents per share. Enron filed for bankruptcy protection [11].

	Human body		Institution
Guiding principles (top level)	“Milieu interieur” (Claude Bernard)	↔	Information environment
	Cybernetics (Norbert Weiner)	↔	Managerial cybernetics (Stafford Beer)
	Homeostasis (Walter Cannon)	↔	Stability
Organization (mid-level)	Structural	↔	Employees, Professionals, Dept.
	Functional	↔	Division
	Behavior	↔	Business processes
Supporting entities (low-level)	Cells		Employees
	Proteins	↔	Tasks
	Macromolecules	↔	Projects
	Cellular exchange	↔	Transactions
	Chemical products	↔	Accounts
	DNA (genes)	↔	Policy (regulations)

Figure 1. An analogy between human body and institution

The analogy to cancer motivates us to look into Enron structural, functional and behavioral organization within its business model and its underlying operations, in a fashion analogous to the human body for insights into cancer prevention. At the top-level of Figure 1 (top box), the first guiding principle in human is derived from the concept of “milieu interieur” (or internal environment) of the body in which all cells bath as stated by Claude Bernard [22]. Analogously, there exists an *information environment* in which all employees of an institution live in and act upon.

Next is the principle of *cybernetics* dealing with feedback and control, a concept owned by Norbert Weiner [23]. This concept is further exploited and applied to business management discipline, termed *managerial cybernetics* by Stafford Beer [24]. Cybernetics is to maintain *homeostasis* (equilibrium) in the human body. Homeostasis is a principle by Walter Cannon [25] originated from Ernest Starling [26] and expanded by Sherwin Nuland [27]. We can equate *business stability* as analogous concept to homeostasis.

The human tissues, organs, organ systems made up the structure, functionality and behavior of a human (mid-level, middle box) are determined by the constituent cells. Likewise, the employees are grouped into professionals, departments and divisions, across which run the business processes.

At the lower level (bottom box), it appears that the biological processes involve the basic constructs created by the cell’s organelles: the protein synthesis. Analogously, the tasks performed by the employees in the institution are much like the proteins created in the cells. The transactions

created between employees are like the cellular exchanges at the cell membranes. Similarly the tasks to the projects and the transactions to the accounts in an institution are like, respectively, the proteins to macro molecules and cellular exchanges to the chemical products.

The organizational analogy in Figure 1 does not suggest how we can address prevention, however. Therefore, we rearrange the guiding principles linking them to the supporting entities in terms of *activities, events* and *control mechanisms* (shown in the right side of Figure 2). The green dotted box shows business entities analogous to those entities in the human body. The red dotted lines and red entities exhibit the basic elements involved anomalies or wrongdoings.

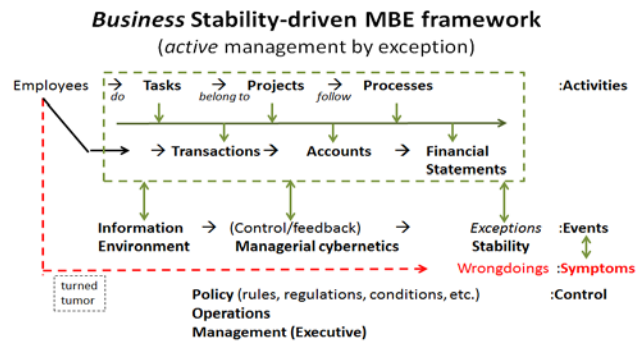


Figure 2. Institution model

More specifically, the operational, tactical and strategic activities of an institution are realized in the tasks (biologically analogous to proteins produced by cells) of projects (analogously carried out by macromolecules in tissues, organs, organ systems) to be executed following predefined business processes (biological processes) by the employees (cells). Transactions (amounts or volume of exchange or release of chemicals in cellular exchange) of various accounts (chemical products) are related to tasks (proteins produced by cells) within the projects (macromolecules) and among them.

Lastly, policies are statements that govern all entities and guiding principles. They can be specific, detailed, and measurable at the operational level of the organization. In fact, there is an operational process/procedure defined on anything and everything such as evaluating a loan application, startup a server, make a reservation for meeting room, etc. These processes grew from simple procedures defined at the start-up of the institution and become mature as the institution grows.

The policy can become sketchy, unstructured and robust at higher level (strategic level). The high-level plan and process (we call it strategy) for defining and achieving company’s business objectives can be considered as policy since it is about how to compete within a predefined scope.

We present above how a set of concepts are identified towards the formulation and realization of an approach to a partially biologically-inspired MBE solution. In the

following, we focus on the development of use cases for such application,

### III. DEVELOPING USE CASES FOR DETECTION OF CANCEROUS SYMPTOMS: ENRON CASE

In this paper, we show only the Enron case in our example implementation: We review the chronological wrongdoings (as we now known) in Enron for the capturing and detection of symptoms indicating wrongdoings which could lead to bankruptcy. For inputs, we classify them under the business entities defined in the MBE model: project, policy, account, task and transaction. Outputs are the exceptions detected by the said MBE.

#### A. Strategy (business plans) and projects

The first strategy was Jeff Skilling's MTM strategy with the concept of gas bank. MTM as a trading model was new in the sense that it was an energy derivative where Enron would act as an intermediary "bank". As such, Enron assumed the risks to buy gas from suppliers and to sell it to consumers at contractual fixed prices and service fees [3]. Skilling and Fastow used MTM to achieve two successful Projects: Cactus 3 with GE Credit Corp and other banks as partners, and JEDI with CalPERS partnership [28].

There were other strategies, e.g., the shift from Asset-heavy to Asset-light [29] and those nicknamed Death Star, Load Shift (creating appearance of electric power congestion), Get Shorty (buy low, sell big, buy back low, etc.), Fat Boy and Ricochet (California's energy market) [30] [31], and a diversification strategy. The latter was to extend Enron to businesses other than gas and electricity (water, broadband, etc.). The said strategies were realized in many subsequent projects: Dabhol, Bolivia, Azurix, Blockbuster, etc. Funds from Chewco, JEDI II, LJM Cayman, LJM 2 and LJM 3, Braveheart, Raptors, etc. in SPEs partnerships were used to off-balance its liabilities. Many of these projects were managed mainly by Fastow with SPEs created to avoid consolidation in financial statements.

#### B. Policy and regulations

To support the MTM strategy and others further, Skilling set out to hire the best and the brightest traders. Skilling devised a new policy on performance review called "360-degree review" and a new mantra (RICE) for Respect, Integrity, Communication and Excellence. Skilling successfully convinced the internal counsel, Andersen accounting audit and SEC to approve the MTM change to accounting practices. The Enron Board of directors also approved Fastow to have a dual role: Enron CFO and SPE manager in 1999. *The above are considered as mutations to policy and regulations on common practices.*

#### C. Finance and Accounting

All together, the MTM trading model, the group of top traders, the expansion to all other businesses, the bull

market during 1990's facilitated investment opportunities. Also included was the use of the Enron Online (EOL) developed by Louise Kitchen and promoted by Michael McConnell since the end of the 1999. The EOL was an electronic, real-time trading site. At its peak in 2000, EOL was handling \$335 billion in online commodity [11]. Until then, there were all success stories, at least on the surface.

Since 1997, however, Enron profits were squeezed due to new entrants and other smaller competitors: Dynegy, Duke Energy, El Paso and Williams, etc. [3]. Enron began to lose the competitive advantage. To be financially able and to maintain high credit ratings, Enron started to devise the use of SPEs to access capital and hedge funds as they entered into new mergers and acquisitions. The company has become more a hedge fund than a trading company [11].

SPEs were the shell partnerships sponsored by Enron, supposedly funded by independent financing. Two conditions must be satisfied in order to keep the SPEs separate from Enron: at least 3% equity and 50% or more control of financial interest given to the independent investors. SPEs were used to purchase forward contracts with producers and to sell under long-term contracts to consumers [18]. However, in Chewco for example 3% was owned and actually controlled by Enron executives. *Fastow began to hide losses in SPEs, thousands of which were created for that purpose.*

#### D. Deals and Transactions

The introduction of SPEs model and practice had brought successful results to the two first projects: Cactus 3 and JEDI [28]. SPEs became problematic in subsequent projects [32].

As we know now, Fastow, the master mind behind all SPEs, used them as a way to transfer losses off Enron's books. Thus, he was able to book profits, and maintained good credit ratings by reducing debt-to-total-assets ratio. He used thousands of SPEs in subsequent projects, e.g. Chewco, LJM Clayman, LJM 2 [11]. Using SPEs, Enron was able to obscure disclosures on Enron financial statements by taking advantages of the non-consolidation requirements.

It was reported [32] that in a deposition, Fastow confirmed that by transferring assets and loss off balance sheet, risks of the transactions were transferred to investors, therefore investors earned a return from the risks. As such, Enron rating would not be hurt. Fastow assured that he worked very closely with outside accountants to not violating any rules.

As later discovered and reported in various sources [18], starting from October 2001, \$287M for Azurix (acquired in 1998), \$180M for broadband with Blockbuster, and \$544 for others, a total of 22% of Enron expenditures from 1998-2000 were write-offs. *Portland General Corp (acquired in 1997) was sold for \$1.9B at \$1.1B loss.*

### E. *Wrongdoings and symptoms*

Some symptoms leading to wrongdoings can be accounted as follows, extracted from happenings of (A-D):

1) The initial success with Cactus 3 and JEDI with hedge fund from CalpPERS, as shown in financial statements of 1992 did not go unquestioned. *Toni Mack, in her Forbes' article "Hidden Risks" in 1993 has pointed out some risk issues related to MTM strategy.*

2) The next symptom was *the lawsuit by Bernard Glatzer in 1997 that his business model was stolen by Enron in moving assets of Enron into partnerships via SPEs.*

3) The symptom which followed and should be noticed was the *Enron's Board's approval of Fastow's dual role in 1999, with the condition that the task of monitoring the deals was given to Skilling, COO, Causey (CAO – accounting, and CRO - risk). There were also concerns raised by David Duncan and Thomas Bauer of Andersen's Houston office in February 2001* however these were not disseminated outside of Anderson.

4) Another (now known) symptom was a *confidential memo from Jordan Mintz in May 2001 to Skilling, but Mintz got no response. The next major symptom was the resignation of Skilling citing personal reasons 6 months after he became Enron CEO. Ken Lay resumed the CEO position on August 2001.*

5) Followed was the *anonymous letter to Ken Lay in late August (later known as being written by Sharon Watkins, an Enron VP) in August 2001, and subsequently a meeting was held between Lay and Watkins on accounting irregularities. Following the anonymous letter, Vinson and Elkins, Enron legal counsel, who helped draft the documents of some partnerships, was advised not to look into accounting by Arthur Andersen when the heart of the problem was really there [32]. It was known that an Andersen lawyer advised the Houston office not to retain documents which were no longer needed. Andersen auditors shredded documents subsequently by order of David Duncan of Andersen.*

6) Meanwhile there were other symptoms during the last part of 1990's:

- *Top management's large amounts of compensation, Excessive executive compensation as a result of financial successes as stock options reached almost 100M shares by end of 2000,*

- *Huge amount of management fees for Enron executives (Fastow alone was paid \$30M)*

7) *Arthur Andersen's conflict of interest acted both as external auditors and consultants. Corporate Audit Committee failed to recognize those symptoms.*

8) *Two days after Lay stepped down as CEO, Clifford Baxter, Vice Chairman committed suicide.*

The first official wrongdoing which brought everybody's attention was the *press release of Q3 2001 (October 16, 2001) without the balance sheet disclosure until after the markets closed. It revealed a \$1.2B charge against equity*

[33]. Breakdowns of footnotes were detailed in Chatterjee [29]. Roughly one week later, SEC inquired about the SPEs. Also, faced the pressure from Wall Street, in November 2001, *Enron admitted buried hidden losses in the SPEs, and posted the re-statements for 4 years (1997-2000).* They were accounting fraud.

The discovery of wrongdoings, re-statements of financial report in the third quarter of 2001, accounting frauds, and financial consolidations resulted in the Enron stocks from \$90+ to the 28 cents within a couple of months [34]. It brought the company to collapse in November 2001.

The classification of the above entities (A-D) was to summarize and to suggest that if attention was paid to these entities as entries to the MBE, the symptoms in (E) would have yielded important indicators of wrongdoings. The summary we cite here is not be a complete and totally accurate account of what really happened and how they happened but we thought it is sufficient for us to illustrate the possible discovery of wrongdoings based on observed symptoms.

We show a typical output produced by our initial prototypical MBE application in Figure 3 (*at the end of the paper after references*). If such exception report is produced and made transparent to all responsible parties, some attention would be given, and some decisions be reached.

### IV. DISCUSSIONS AND CONCLUDING REMARKS

Recall that our proposed model in Section 2 is *employee-centered* much as an organism is *cell-centered*. Everything in the institution is considered as product of the employee's *tasks* much as everything in the human body is the product of the cell's organelles as *proteins*. The products are described not only in terms of tasks but also projects, transactions or accounts, and are not limited to other entities the institution might use.

Secondly, the system is *policy-driven* across all level of organizations much like *gene-driven* in a human body, to measure results against "faulty strategy, managerial mishandlings, intended wrongdoings, diverted tasks within projects, out-of-the-ordinary transactions hidden in financial accounts and underreporting statements". These are either supported by existing policy and/or regulations or questionably violate them as we mentioned earlier in Enron case.

Thirdly, the system is geared towards evaluating and labeling exceptions in terms of severity level as a set of relevant symptoms for diagnostics of wrongdoings as shown in Figure 3. All records have drill-down and roll-up capability, and sideway links. Thus, the MBE system is capable of displaying the set of information on a particular issue as complete as possible including all analyses (as background calculations or evaluations, not addressed here) substantiating it, and all actions/decision for or against it (overriding decisions). Such transparent MBE scheme, with highly vertical and horizontal integration and correlation among the business entities can help detect symptoms

towards wrongdoings. Implementation of this MBE enterprise-wide is difficult, time consuming and complex because it will involve many existing applications. However, it can be done in increment.

The six collapses we consider for the development as use cases are complex. They all had sophisticated strategies and processes, huge projects, complicated transactions in numerous areas of business from trading to CDO's, from SPE to Repo 105, and the like. The Enron illustration shown here is highly simplified to convey the idea. There are a lot more in each and in future cases. We believe however it does suggest an improved internal control toward prevention. Future work is on the mechanisms, partially biologically-inspired for the detection and reporting of symptoms leading to collapses, for managerial decisions.

REFERENCES

[1] H. Edwards, (2004). "Barings – A Case Study in Risk Management and internal Control. The Risk" Regulatory Forum. 2004

[2] S. J. Brown, and O. W. SteenBeek (2001). "Doubling: Nick Leeson's trading strategy". Pacific-Basin Finance Journal. 2001.

[3] B. G. Dharan and W. R. Bufkins (2002). "Red Flags in Enron's Reporting of Revenues and Key Financial Measures". www.ruf.rice.edu/~bala/files/dharan-bufkins\_enron\_red\_flags.pdf,

[4] B. G. Dharan, and N. Rapoport, Eds., (2009). "Enron and Other Corporate Fiascos", Foundation Press 2009

[5] K. V. Peurseem, M. Zhou, T. Flood and J. Buttimore (2007). "Three Cases of Corporate Fraud: An Audit Perspective", Number 94, University Of Waikato, June 2007.

[6] W. Lyke and M. Jicking. (2002). "WorldCom: The Accounting Scandal". CRS Report for Congress, RS21253, August 2002

[7] C. Celani. (2004). "The Story behind Parmalat's Bankruptcy". Executive Intelligence Review. January 2004.

[8] A. Azadinamin. (2012). "The Bankruptcy of Lehman Brothers: Causes of Failures & Recommendations Going Forwards", Swiss Management Center, 2012. http://ssrn.com/abstract=2016892.

[9] Sarbanes-Oxley Act. (2002). http://www.soxlaw.com, 2002

[10] D. Higgs. (2003). "Review of the role and effectiveness of non-executive directors". Department for Business, Enterprise and Regulatory Reform. http://www.berr.gov.uk/files/file23012.pdf.

[11] C. W. Thomas, (2002). "The Rise and Fall of Enron", Journal of Accountancy, April 2002.

[12] L. Davies (2013) "Classic Financial and Corporate Scandals". http://projects.exeter.ac.uk/RDavies/arian/scandals/classic.html 2013

[13] J. A. R. and R. R. Nelson (1987). "Management-by-exception reporting: An empirical investigation", Information & Management, Volume 12, Issue 5, May 1987

[14] B. Alberts, D. Bray, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter. (1998). Essential Cell Biology, Garland Publishing, 1998.

[15] R. J.B. King, (1996). Cancer Biology, Longman, 1996.

[16] K. Eichenwald (2005). Conspiracy of Fools: A True Story, Broadway, 2005

[17] C. N. Smith and M. Quirk. (2004). "From Grace to Disgrace: The rise and fall of Arthur Andersen", Journal of Business Ethics Education, 1(1), 2004

[18] P. M. Heally and K. G. Palepu, (2003). "The Fall of Enron", Journal of Economic Perspectives, Vol. 17, Number 2, Spring 2003.

[19] EnronAnnualReport1998. (1998). picker.uchicago.edu/Enron/EnronAnnualReport1998.pdf

[20] EnronAnnualReport1999. (1999). picker.uchicago.edu/Enron/EnronAnnualReport1999.pdf

[21] EnronAnnualReport2000. (2000). picker.uchicago.edu/Enron/EnronAnnualReport2000.pdf

[22] C. G. Gross. (1998) "Claude Bernard and the constancy of the internal environment". Neuroscientist 4 (1)

[23] N. Weiner (1948). Cybernetics or Control and Communication in the animal and the machine, The Technology Press, 1948

[24] A. S. Beer (1972). Brain of the firm: Managerial Cybernetics of Organization,

[25] W. Canno, (1963). The wisdom of the body, The Norton Library, Norton & Company, 1963

[26] E. Starling (1923). The wisdom of the body, 1923.

[27] S. Nuland (1997). The wisdom of the body, Alfred A Knoff, 1997

[28] K. Eichenwald and M. Brick, "Enron Collapse: The Strategy", NY Times, 2002

[29] S. Chatterjee and B. Fellow. (2002). "Enron's Asset-Light Strategy: Why It went Astray", 2002

[30] A. Haglund (2011). "Failed Strategies of Enron", http://mg312.wordpress.com/2011/09/02/failed-strategies-of-enron/, 2011

[31] R. A. Oppel Jr., (2002). "Enron's Many Strands: The Strategies". NY Times May 8, 2002

[32] K. Eichenwald and M. Brick, "Enron Collapse: The Strategy", NY Times, 2002

[33] B. G. Dharan (2002). "Enron's Accounting Issues – What We Can Learn to Prevent Future". Prepared Testimony. US House Energy and Commerce Committee's Hearing on Enron Accounting. 2002.

[34] S. Deakin, and S. J. Konzelmann. (2003). "Learning from Enron", ESRC Centre for Business Research, University of Cambridge, 2003

ID	Related entity	Entity type	Date	Severity	Link to	Action by	Override		
							Date	Reason	Severity
P-003	SPE creation	Policy	1991	●	P-003	Fastow	1991	Partnership	●
E-191	Rhythms Net	Project	1999	●	PR-99	Skilling	1999	Investment	●
E-192	Financial analysis	Task	1999	●	PR-99	Kaminski	1999	Risk	●
E-193	LJM Cayman	Account	1999	●	A-099	Fastow	1999	Hedge	●
E-194	Cayman Islands	Transaction	1999	●	TR-99	Fastow	1999	Enron stock	●
E-195	FS-1999	Others	1999	●	O-199	Skilling	1999	Statement	●
...	...	...	...	...	...	...	...	...	...
E-211	FS-2001 Q3	Others	2001	●	O201	Skilling	2001	Consolidation	●
E-212	SEC Inquiry	Others	2001	●	T-991	Fastow	2001	Consolidation	●

Figure 3. Partial exception list in Enron Rhythms NetConnections project

# Temporal Data Management

## Uni-temporal Table Modelling Update

Michal Kvet

Department of Informatics  
Faculty of Management Science, University of Zilina  
Zilina, Slovakia  
Michal.Kvet@fri.uniza.sk

Anton Lieskovský

Department of Informatics  
Faculty of Management Science, University of Zilina  
Zilina, Slovakia  
Anton.Lieskovsky@fri.uniza.sk

**Abstract**— Today, it is necessary to store not only actual data – data that are valid at this moment, but also historical data, by which the progress and frequency of changes can be monitored. Managing historical data offers creating future prognoses and analyses. Temporal tables (mostly modelled using uni-temporal and bi-temporal tables) in comparison with conventional tables can process and retain the information of the validity in the past. Using procedures, functions, triggers, cursors – snapshot of the database object or whole database at any time point can be reconstructed easily. Thus, each database object is not represented by only one row, but by the set of rows representing the same object during a different period of time. This paper deals with the principles of temporal data modelling and offers the solution based on uni-temporal table model. It contains the implementation methods based on changes monitoring and deals with the problem of managing undefined states of the objects. Operations implemented in temporal system is compared with the conventional model.

**Keywords**-conventional table; temporal table; uni-temporal model; valid time;

### I. INTRODUCTION

Massive development of data processing requires access to extensive data using procedures and functions to provide easy and fast manipulation. The basis is the database technology.

Database systems are the root of any information system and are the most important parts of the information technology. They can be found in standard applications, but also in critical applications such as information systems for energetics, industry, transport or medicine.

Most of the data in the database represent actual states. However, properties of objects and states are changed over time - customer changes its status, address, products are modified and updated. If the object state is going to be changed, data in the database are updated and the database will still contain only currently applicable object states. But everything has time evolution, thus, history and future that can be useful to store. History management is very important in systems processing very important or sensitive data; incorrect change would cause a great harm or in the systems requiring the possibility of restoring the previous states of the database. Therefore, it is necessary to store not only the

current state, but also the previous states and progress. It can also help us to optimize processes or make future decisions.

Historical data are now possible to be saved using log files and archives. Thus, the historical data can be obtained, but it is complicated process. In addition, management requires quick and reliable access to data defined by any time point, but also getting information about the attributes changes in the future without significant time delays [10] [11].

Temporal data processing is not a new problem. Early in the development of the databases transaction, log files were created and database was regularly backed up. Log files were usually deleted after the backup, because all transactions were recorded in the production database. Thus, it was possible to obtain an image database at any time point. However, these data were in the raw form (raw material) and handling them was difficult, lasted too much time. Nowadays, historical data management is easier than in the past, requires less processing time, but there is still need to make significant progress in temporal database processing research to create a complex module allowing to run existing applications without modifying source code and settings.

Fundamental paradigm of database systems used since the beginning of the data processing focuses on the actual data processing [10] [11] [12].

The situation in the computer field changed significantly in the early 80-ies of the 20th century- price of the disk storage space decreased allowing greater and easier way to save backups. So, there was an opportunity to compare multiple images (backup) from different time periods.

Each backup is a snapshot of the database table or the whole database. However, if the values do not change (or it is not necessary to store historical values of them), too much duplicities are stored. Later, the first concept of the data warehouse based on the database table level was created by Barry Devlin and Paul Murphy [6].

Recent history shows the potential opportunities of historical data processing, which could be faster and more efficient than managing backups or log files. The main disadvantage of the above-mentioned ways is the need of the administrator intervention (operation manager). An administrator must manage not only the running applications but also requirements for accessing historical backups. Decisions are based on historical data and the progress, so it was necessary to load historical backup to get the database

snapshot in the historical time point. Operational decisions could not be based on the historical data because of the time consumption (sometimes even days to load all needed snapshots). In addition, the granularity of the data is still growing, so number of backup was above the acceptable level. An important task for administrators was to define the time frame between two backup of the database. If the interval is too large (assuming backups without using their own log files - they are too large for more images), not all operations are stored there, e.g. the insert and also delete of the record could be between two images, so user has no information about the existence of the object in the database. Another example is multiple updates of the same record; see Figure 1. The opposite way, if the interval was too small, large images containing a lot of the same attribute values were created, which include high demands on disc storage. Interval between two backups can also be defined dynamically, but the problem with uncaught of some changes remains unresolved. Another solution is to delete old images, which is unacceptable because of the management requirements [8].

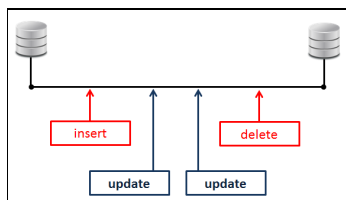


Figure 1. Backup problem - possible loss of data

It is not enough to find a faster solution for historical data processing. What does the term “faster” mean? To order shorten the processing time from days to hours, from tens of hours to a couple of hours? The usage of that solution is unacceptable and inapplicable. The aim is to create support for the temporal data, that the difference between the processing in the currently valid data and historical data is minimized. Thus, it is necessary to create a system with easy data. However, historical data management is not a complete temporal system, because it must be allowed to process values, the validity of which begins in the future. If the begin time of the validity of the object occurred, the database system must be able to update the data without user intervention.

System requirements can be divided into two parts with special aspects [6] [13] [14]:

**1. aspect of usability** (easy methods) – the aim is to provide access to outdated information as easily and quickly as to the actual values. Transactions for managing temporal data must be as simple as for current data processing. Moreover, it is necessary to define a way to combine the past and present.

**2. aspect of performance** (speed and correctness of results) requires results in the same form as when accessing the actual data with adequate processing time. The difference in accessing the object at any point in time should be minimal.

Following the history of computer systems for data management, we come to the conclusion that the databases have been developed and deployed for the current values administration - currently valid data processing. None of the above-mentioned solution does not have the structure to represent objects and their states during the time, support for managing attribute value changes in the temporal dimension. However, developers require those functionality and data access.

Temporal databases define a new paradigm for selecting one or more rows based on the specified criteria, for projecting of one or more columns to the output sets and for joining the tables by specifying relationship criteria. Rows with the different values of the primary key (PK) can represent one object at different times. Transactions for inserting, updating and deleting the rows must therefore specify not only the object itself, but also processed period. If the valid time of the object is defined by time interval, the transaction must include time period - 2 time point values - begin and end timestamps (dates). This means that the update query does not cause only update of existing data, but also insert of the new row based on the validity intervals [2] [6] [8].

This paper consists of five sections. The second deals with the structure of conventional and temporal table, the third describes the implementation techniques and methods. The section named “Experiments” deals with the improvements of the model and the current performance. The last one is the conclusion.

II. CONVENTIONAL AND TEMPORAL TABLE

Row in a relational database table can be defined in three different ways using time (Figure 2). "ID" is a unique identifier, "PK" refers to a primary key. "BD1" and "ED1" are a pair of columns defining the beginning and end value of the period, "BD2" and "ED2" define the second time interval. The primary key can be defined composite without changing the functionality of the methods used. The first model does not use time for definition. This is a standard model used today, called the conventional model. The primary key is defined by the attribute "ID". All not key attributes in the table, regardless of their number, are merged into a common block called "data" [1] [3] [4] [5].

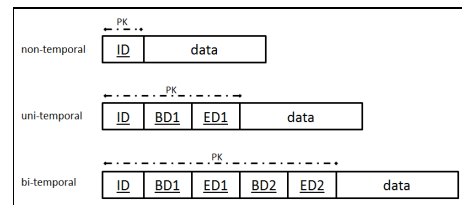


Figure 2. Conventional and temporal table

In this paper, we use the standards used in scientific articles, the names of the attributes forming the primary key are underlined. The following table shows the object identified by an attribute value ID = 1 in the time period <September 2012 - December 2013>. We can see that the attribute "data" has the value “123” at the time <September



2012 - June 2013>, in the second part of the interval, attribute "data" has the value of the "234". The representation of the time interval is the closed-closed type.

TABLE I. UNI-TEMPORAL

ID	BD1	ED1	data
1	September 2012	June 2013	123
2	January 2013	November 2014	555
1	July 2013	December 2013	234

The problem occurs, if the end date of the time period is unknown. There are two opportunities, in general. The first is to deal with undefined values - NULL. In this case, if there is a request to obtain the current state of the object, respectively overall database, we cannot use standard methods to manipulate time and we have to add more conditions. The second solution is to replace an undefined time with the maximal value that can be used - "December 9999", or other variants depending on time granularity. If we need to update the state of the object, the attribute value defining the end of the validity (attribute "ED") is replaced by the actual time value.

The last - third basic model – bi-temporal model - is based on the concept of uni-temporal tables, but uses two time intervals. Thus, it allows not only defining several lines for one object, but also multiple rows for a particular object's state at the time. The reason for this model is the need to record an updated status in the past.

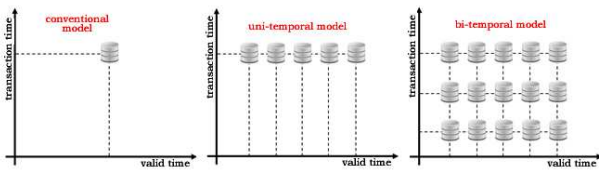


Figure 3. Time representation

The method for transformation uni-temporal model to bi-temporal is based on the same principle as the conversion of the conventional model (conventional model) to uni-temporal model - two additional attributes defining a timestamp (begin and end time of the validity) are added to the primary key. The values of the new attributes after the transformation can be identical to the first time interval defined by attributes "BD1" and "ED1" (Figure 4).

The transformation from bi-temporal model to uni-temporal model is also possible, but there is possibility of the loss of the updated data [6] [7] [8].

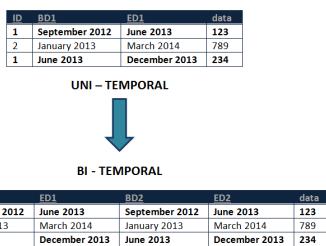


Figure 4. Transformation of the uni-temporal table to bi-temporal table

Figure 4 shows the principle of bi-temporal modelling. Primary key of this model consists of three logical units:

- **The object identifier (ID).**
- **Interval (BD1, ED1)** - the time during which the object has been describing the characteristics of the row, e.g. the period during which a customer has the characteristics - name, address, status, etc.
- The last component of the logical primary key is a **pair of dates** (or timestamps according to the representation of a time granularity of data). These dates limit the period during which we believe the value of the row is correct. This component limits the time interval defined by the second component (BD1, ED1).

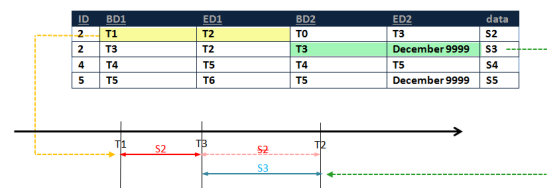


Figure 5. Bi-temporal table

Figure 5 shows the principle of bi-temporal table modelling.

### III. UNI-TEMPORAL TABLE IMPLEMENTATION

Transformation of the conventional table to temporal model is not trivial problem and needs special structures and resources. In addition, the requirement of the users is to provide compatibility, easy manipulation and proper time consumption. Therefore, the triggers, procedures and functions must be declared, the original tables can be transformed to views (if necessary). Each database record is defined by the primary key. In most cases, it is the unique identifier (ID), sometimes, we use composite primary key. Conventional database tables transformation recommends single attribute primary key in every table to create a common temporal table for all tables containing temporal attributes. Thus, ID is suitable; each record can be clearly defined and referenced. Moreover, the ID does not have special denotation and the need for its change is irrelevant, e.g. personal identification number contains the birthdate and if the mistake in time of insert occurs, the record must be updated. Thus, the best way is to create sequence for ID; trigger before insert sets the correct value.

Information of any change of temporal column is recorded in the table managing changes - temporal\_table. However, it contains information only about table containing temporal column. This table consists of these attributes (see also Figure 5) [7]:

- **ID change**
- **ID previous change** – references the last change of an object identified by ID. This attribute can also have NULL value that means, the data have not been updated yet, so the data were inserted for the first time in past and are still actual.

- **ID\_tab** – references the table, record of which has been processed by DML (Data modelling language) statement (INSERT, DELETE, UPDATE).
- **ID\_orig** - carries the information about the identifier of the row that has been changed.
- **ID\_column, ID\_row** – hold the referential information to the old value of attribute (if the DML statement was UPDATE). Only update statement of temporal column sets not null value.
- **BD** – validity of the new state of an object starting.

“ID\_column” also contain null value (example for TAB1) (see also Figure 5) [7]:

```
INSERT INTO temporal_table (ID_change, ID_previous_change,
ID_tab, ID_orig, ID_column, ID_row, BD)
values (ID_TEMPORAL_TABLE_sequence.nextval, NULL, 1,
ID_tab1_sequence.currval, NULL, NULL, sysdate);
```

Figure 8. Temporal table insert

```
CREATE OR REPLACE TRIGGER Upd_tab_2_trigger
BEFORE UPDATE ON TAB2
FOR EACH ROW
BEGIN
if (:old.street<>:new.street) then
--data processing
end if;
if (:old.post_code<>:new.post_code) then
--data processing
end if;
END;
```

Figure 6. Update trigger

The following figure shows the data representation and manipulation.

```
INSERT INTO TAB_Z_COLUMN (ID, Z) values
(ID_TAB_Z_COLUMN_sequence.nextval, :old.Z);
```

Figure 9. Updating column "z"

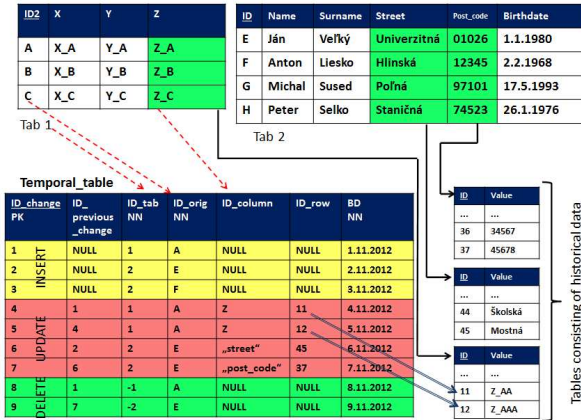


Figure 7. Temporal\_table model [7]

A. Insert trigger

New record containing information about the change of the temporal column is inserted into the temporal table after inserting into conventional table. These operations are provided by insert trigger. The new value for attribute “ID\_change” is set using the sequence. Value of “ID\_previous\_change” attribute is null, which means, the new data have been inserted (for the first time). There is no reference to old value of the attributes, so the “ID\_row” and

B. Update trigger

Updating existing data requires saving old data – not the whole row, but only changed temporal attribute values. The original table consists of the actual data, so the data manipulation – actual snapshot is easy to get. Historical data – the snapshot of the whole database, database table or only object – must be also accessible, but are obtained by passing historical conditions defined by insert, delete or update statement. Thus, the update trigger is started before update. First of all, the data that will be changed are stored in the table consisting only of the ID of the record and the value itself [7].

Then, the reference to the change is stored in the temporal table (Figure 5) [7]:

- **ID\_change** is set using the sequence and trigger.
- **ID\_previous\_change** is maximum of “ID\_change” “used for those ID original and ID table (*select max(ID\_change) INTO cislo from temporal\_table where ID\_orig=:old.ID AND ID\_tab=1;*).
- **ID\_column** references the temporal column, data of which is going to be changed.
- **ID\_row** associates the table with historical values.

C. Delete trigger

The task of the trigger starting before delete is to save old data to the table for deleted objects. The information about delete is also inserted to the temporal table; “ID\_tab” now has the negative value. The relevance of it will be described later [7].

```

INSERT INTO TAB1_DELETED(ID, X, Y, Z) values( old.ID, : old.X, :
old.Y, : old.Z);

INSERT INTO temporal_table (ID_change, ID_previous_change, ID_tab,
ID_orig, ID_column, ID_row, BD)
values (ID_TEMPORAL_TABLE_sequence.nextval, max_change,
-1, : old.ID, NULL, NULL, sysdate);
    
```

Figure 10. Body of the trigger - delete

D. RestoreData

Deleted data recovery is a further problem. If the object data in the main table are not valid or we do not know the correct value in this moment, the object must be deleted, respectively relocated to table of deleted data. However, if the object attributes are again valid, we need to restore data to the original table. The problem causes a trigger which sets the new value of ID before inserting, so the old value of ID cannot be used. However, the database structure should have information, that the ID has been changed, but the object is the same. In addition, if the user is still using the old identifier, the methods working with original (old) ID would not work correctly. There can be used more solutions, one of them is to disable temporary the trigger for ID, restore the data and enable it. This solution is absolutely incorrect.

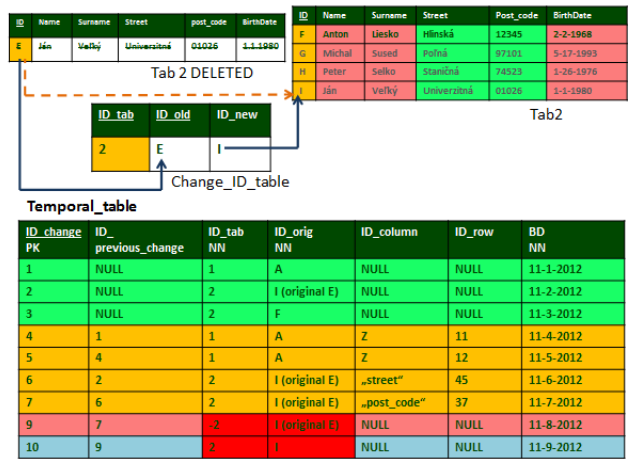


Figure 12. Data restoration

Our solution uses the new ID (Figure 11, Figure 12), old and new object is connected together. Thus, these steps must be done [7]:

- Data relocation from „deleted objects table“ to main table (insert + delete).
- Insert info about data restoration into temporal table.
- Update old „ID\_orig“ (temporal table) to actual.
- Insert into „change\_ID\_table“ new values – „ID\_tab“, „ID\_old“, „ID\_new“.
- Update „the change\_ID\_table“ - if necessary. The point is that the data could be relocated several times and therefore the old reference (but stored in the attribute „ID\_new“) is not actual. Figure 12 describes the problem.

IV. EXPERIMENTS

The overall adjustment and optimization of the procedures, functions, triggers and the model itself, we got very good values of the processing time; the worst result (update and delete statement) is a slowdown of 34% in comparison with conventional table model. It should be noted, that the traditional model does not retain any information about changes of the database. Our designed solution stores also old (historical) values - any snapshot of the database (database object) defined by time point can be got easily using programmed methods.

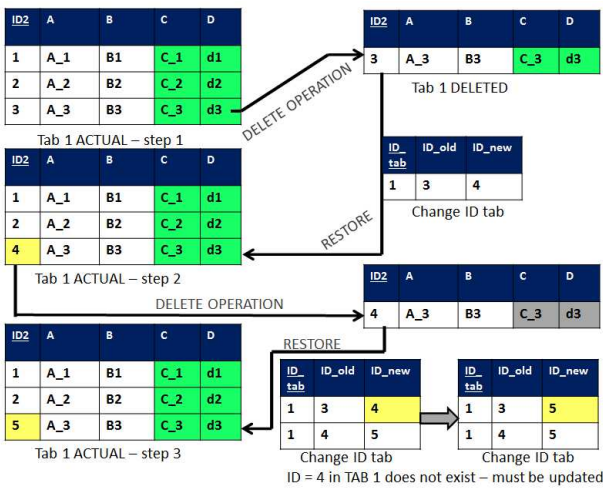


Figure 11. Principle of data restoration [7]

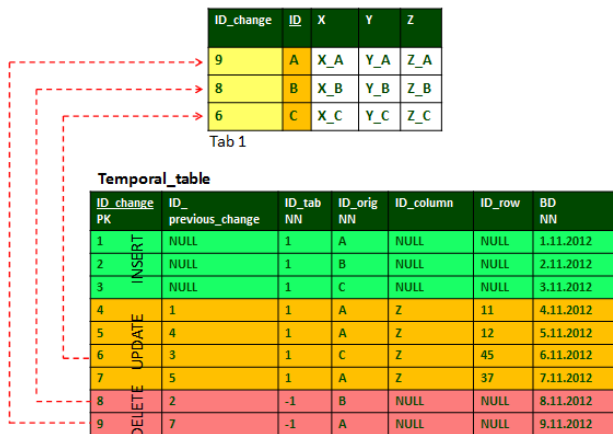


Figure 13. Principle of model 3

Figure 14 shows the total execution time of DML operations (insert, delete and update) for conventional and temporal tables. The total number of operations is 10 000. The experiments were provided using Oracle 11g instance.

TAB1		model 1		%		model 2		%	
INSERT	temp	0:26:28	0:26:449	105%	104%	0:26:345	0:28:538	104%	101%
	conv	0:25:00	0:25:475	100%	100%	0:25:112	0:28:173	100%	100%
UPDATE (10)	temp	7:36:267	6:53:765	178%	196%	7:01:754	7:36:109	158%	162%
	conv	4:12:192	3:33:79	100%	100%	4:44:265	4:55:91	100%	100%
DELETE	temp	0:4:099	0:4:045	173%	170%	0:3:825	0:4:226	140%	138%
	conv	0:2:364	0:2:38	100%	100%	0:2:732	0:3:063	100%	100%
TAB1		model 3		%					
INSERT	temp	0:26:123	0:25:994	104%	104%				
	conv	0:25:236	0:25:712	100%	100%				
UPDATE (10)	temp	6:32:273	6:12:109	138%	131%				
	conv	4:45:185	4:43:572	100%	100%				
DELETE	temp	0:3:212	0:4:412	131%	137%				
	conv	0:2:452	0:3:263	100%	100%				

Figure 14. Research results

Model 1 does not use the table with maximal values of changes, data are inserted into historical tables only if they have not been there yet. Model 2 uses table with maximal values of changes and historical data are always saved.

Model 3 stores the value of the identifier of the last change of the object directly in the main production table. This is due to a reduction in the number of records. Specifically, the table “Change\_ID\_tab” stores values for all temporal objects, and also contains data about deleted objects. If the number of deleted objects is great, the reduction of processing time is significant.

The overall slowdown of this model is:

- insert operations – 3%,
- delete operations – 34%,
- update operations – 34%.

As it was already mentioned, the results of the experiments can be considered satisfactory, because of the storing the entire history of temporal columns.

## V. CONCLUSION

Each instance in the conventional database is represented by one row. Temporal database concept offers new opportunities by adding additional time attributes limiting the validity of the object.

Temporal support brings new possibilities by extending the primary key with time interval. Two different time period defining the same object validity cannot be valid at the same time – cannot overlap. The aim of the temporal databases is to store information about all states of the objects during their life cycle, even after DELETE operation.

Comparison of the processing time of the conventional methods and temporal model is the root for the further research. Experiments, that have been made, show that the operation UPDATE is the critical factor. If we update the row, we need to save old value (insert it into historical table).

Proposed methods can be extended by the special attribute defining the end of the validity. This will create possibility of defining period during which the state of the object is not valid without the necessity of deleting it.

The begin time of the validity can be also in the future. In the future development, we are going to create methodology for future states management. Data valid in the future will be stored in special tables. New record will be automatically inserted into the database at the time of the beginning of the validity.

## VI. ACKNOWLEDGMENT

This contribution is the result of the project implementation: Centre of excellence for systems and services of intelligent transport II., ITMS 26220120050 supported by the Research & Development Operational Programme funded by the ERDF.

This work was supported by the project "Creating a new diagnostic algorithm for selected cancers," ITMS project code: 26220220022 co-financed by the EU and the European Regional Development Fund. The work is supported by the project VEGA 1/1116/11, Adaptive data distribution.

## REFERENCES

- [1] C. J. Date, "Date on Database". Apress, 2006.
- [2] C. J. Date, H. Darwen, and N. A. Lorentzos, "Temporal data and the relational model", Morgan Kaufmann, 2003.
- [3] Ch. S. Jensen, "Introduction to Temporal Database Research"
- [4] Ch. S. Jensen and R. T. Snodgrass, "Temporally Enhanced Database Design"
- [5] P. N. Hubler and N. Edelweiss, "Implementing a Temporal Database on Top of a Conventional Database"
- [6] T. Johnson and R. Weis, "Managing Time in Relational Databases", Morgan Kaufmann, 2010.
- [7] M. Kvet, A. Lieskovský, and K. Matiaško, "Temporal data modelling", 2013.(IEEE conference ICCSE 2013, 4.26. – 4.28.2013 ), pp. 452 - 459

- [8] M. Kvet and K. Matiaško, "Conventional and temporal table", EDIS, 2012. (virtual conference ARSA, 12.3. – 12.7.2012), pp. 1948-1952
- [9] N. Mahmood, K. Rizwan, and S. A. K. Bari, "Fuzzy-Temporal Database Ontology and Relational Database Model", 2012.
- [10] K. Matiaško, M. Vajsová, M. Zábovský, and M. Chochlík, "Database systems", EDIS, 2008.
- [11] K. Matiaško, M. Vajsová, M. Zábovský and M. Chochlík, "Database systems and technologies", STU, 2009.
- [12] G. Ozsoyoilu and R. T. Snodgrass, "Temporal and Real-Time Databases: A Survey", 1995.
- [13] R. T. Snodgrass, "Developing Time- Oriented Database Applications in SQL", Morgan Kaufmann, 1999.
- [14] R. Snodgrass, "Adding Valid Time to SQL/Temporal"

# Evaluation of Visual structure for Industrial size Software Product Line Architecture

Abeer Khalid, Salma Imtiaz  
 Department of Software engineering  
 International Islamic university  
 Islamabad Pakistan

abeer.msse234@iiu.edu.pk, salma.imtiaz@iiu.edu.pk

**Abstract**-Information visualization helps in facilitating comprehension of conceptual information. Information visualization plays a powerful role in software product line architecture. Much work has been done for variability representation, but little consideration has been given to scalability, visualization of traceability links, evolution of variance, etc. There is a greater need for a suitable visual structure that can illustrate “industrial sized” software product line architecture in managing these highlighted factors. There is also a need to perform an evaluation of the visual structures used for visualization of product line architecture, finding out their suitability for visualization of a large and complex software product lines. Our results show that hyperbolic trees are best suited for visualization of product line architecture due to their multiple attributes such as: Exponential growth, scalability, interaction without hindering the structure and clearer (un-complex) perception.

**Keywords**-software product line architecture; information visualization; visual structure

## I. INTRODUCTION

A software product line is a cluster of software intensive systems, “sharing a common”, administered set of features, resulting from the “core assets”, in an agreed upon way [2]. Software product line is hierarchal in nature, with intense complexity immersed in it, with increase in size, to industrial level; the intense complexity has increased exponentially.

The focus, on information visualization has increased, during the last decade; due to advanced information processing technology. Its essence is to augment “cognition” with the help of interacted visual illustrations [4], keeping in mind that “visual display provides the highest bandwidth channel from the computer to the human mind” [9].

The representation of software product line architecture in visualized form promotes understanding variation and reduces complexity of the data [1].

The incentive behind representation of software product line data in visual form is to provide an increased perception of data with highest interaction capability.

The benefit of mapping of software product line data to visual structure is best consummated if the visual structure preserves the specified data. There is a need for performing an evaluation of visual structure to identify their strengths and weaknesses. Information visualization has not been able to make much progress in software product line area.

The focus of this paper is to evaluate the existing visual structures for software product line architecture. The scope of the evaluation includes general visual structure along with ones specifically made for software product line

visualization. The evaluation criterion are Scope [5], Abstraction [32], Hierarchy [18], Traceability [14][15], Scalability [16], Evolution [17], General visual content [4][8] and Perception [4]. These criterion are carefully chosen from literature, keeping in mind the attributes of software product line architecture and the attributes necessary for good visual structure [4][5][8][14][15][16][17][18][32].

This paper is organized in seven sections: Section I gives a brief introduction to the concerned problem. Section II is concerned with the related work. Section III defines the significance of conducting this evaluation. Section IV defines the visual structures chosen for evaluation. Section V presents and defines the metrics used for evaluation purpose. Section VI presents the evaluation of the visual structure. Section VII presents the conclusion and direction for future work.

## II. RELATED WORK

Current visualization techniques in the literature are presented with a different focus: software visualization techniques in general, code visualization techniques, tools and techniques for software architecture visualizations, and visualization of static characteristic of software.

Price et al. [5] proposed a principled systematic approach for categorization of characteristics to be visualized, stating six main categories (scope, context, form, method, interaction, effectiveness) leading to further subcategories, which are further extended to other categories. These categorizations, helped in construction, as well as selection, of visualization techniques, but they lacked in specifying the domains for visualization. In short, visualization techniques, specific to some specialized area, were not considered, but they were reflected upon as an individual entity.

Bassil and Keller [10] evaluated qualitatively and quantitatively, software visualization tools using Price’s proposed framework. Maletic et al. [11] proposed a task oriented-approach wherein they incorporated Price’s [5] proposed framework, taking into consideration development and maintenance tasks of large scale software’s. Gallanger et al. [12] took Price’s [5] and Maletic’s [11] approaches and proposed their own, focusing on stakeholder perspectives for evaluation of SAVT (Software Architecture visualization tool).

Roman and Cox [6] and Storey et al. [7] emphasized the need for evaluation, and proposed their own taxonomies.

Caserta and Zendra [8] presented a literature review on the static features of visualization techniques and then assessed them on the basis of three points of evolution

criteria “changes over versions”, “relationship between components” and “evolving of metrics with releases” [8].

The focus and scope of all of the above mentioned papers is different from ours in the sense that, first, our focus is towards evaluating only visual structures. Second, evaluation is done exclusively for software product lines. The scope of this work only includes structures, which are common and most favorable for software product lines.

### III. SIGNIFICANCE OF EVALUATION OF VISUAL STRUCTURE FOR SPL ARCHITECTURE

Evaluation of visual structure for software product line architecture has major implications. Its importance is gained from the fact that if visual structure is not aligned with the needs of the concerned data, all the patterns may not be visible; a lot of data may be obscured, and context may not be correctly perceived. In short, departure towards a debacle takes place, if the right representation mechanism for data presentation is not selected [3].

Architecture of the development software is a cause of major apprehension for the stakeholders, and a proper depiction of architecture is raised in priority, therefore, software product line architecture has increased the stakes even more. “By its very nature, architecture is a statement about what we expect to remain constant and what we admit may vary” [13]. This, in short, is the motto of software product line.

The visual structure has a major hold in the visualization of software product line architecture, in the sense, that if visual structure is not up to mark then as many visual interaction techniques are placed; it would never confer a complete image. As stated by Tufte [3], we considered, that “There are right ways and wrong ways to show data; there are displays that reveal the truth and displays that do not” [3]. For this complication to be overcome; there is a major need for evaluation of software product line architecture. Our contribution is on evaluation of visual structures, for the sole purpose of finding a formfitting structure for software product line architecture, which would depict the essence of SPL architecture, with minimum of visual manipulation or interaction techniques needed for stakeholder task completion.

### IV. VISUAL STRUCTURES

We have chosen visual structures based on the fact that some of them are already being implemented in SPL and others are more suited for representing Software Product Line Data.

- A. Matrix Tables (MT)
- B. Cone Tree (CT)
- C. Tree Maps (TM)
- D. Conventional Trees (CNT)
- E. UML Notations (UMLN)
- F. Textual Form (TXF)
- G. Use Case Map notations (UCM)
- H. Hyperbolic Trees (HBT)

Matrix Tables (MT) are a form of representation mechanism used for illustration of SPL architecture data [20][21]. Advantage of this system of demonstration is that it plainly depicts the number of variables linked with an assortment of data [4]. MT are used in the context of hierarchal as well as network data. Additionally, “visualization of data tables is used for detection of data” [4].

Cone Trees (CT) are another form of representation of SPL data. They are favorable in nature because of the fact that they are in essence set in 3D surface space; also, they are a good form of representation for hierarchal data. Their visual structure has a significant effect on the perceptive of the viewer [4].

Tree Maps (TM) are favorable for software product line data representation simply for the fact that they are a space filling technique with a true visual format for representation of hierarchal data [31]. Its essence is to optimize utilization of full window space with a rectangular region mapping the hierarchal structure resulting in a “space filling manner” [31].

Conventional Trees (CNT) either being it be a vertical tree or horizontal tree [32][31] are another established form of representation of SPL data [32]. It is a popular form of depicting hierarchal data using link and containment, with the links being used to connect nodes (containment notation). Conventional tree structures are based on the fact that there are no cycles in it and only one axis is used for division of levels in the tree, making it easy to map and extrapolate data [4].

UML notations are probably the most utilized form of representation of architectural data [27][28][26][24][29][25]. Specifically, UML class diagrams are among the most cited in UML representation. Some used natural language with UML based notations for representation purpose. This is a relatively common form of representation for showing generalization, association, composition, inheritance and it is also used for providing platform independence.

Textual Form (TXF) is also used for representation of software product line data [23][22]. This form is useful as software product line data is extremely large in size, and folded axis is used to fit the data in the height of the window and Seesoft [30] form is used in stating file as columns and line of code as colored strip. These are some forms which focus on textual representation of SPL data.

Use case map path notations (UCM) are also used for representation of SPL. It is a concept which is used for capturing “requirements” at a reasonable level of detail. Its use in SPL is focused towards capturing of requirement for construction of architecture [19].

Hyperbolic Trees (HBT) are a unique way of representing structures. Its nature is that parallel lines deviate away from each other, “making the circumference of the circle to grow exponentially with its radius” [18]; thus, resulting in space being accessible with mounting distance.

## V. EVALUATION CRITERIA

The evaluation criteria are singled out from the literature on the merit that a visual structure would be most suited for use in illustrating architecture of a software product line if it has these features in its essence rather than having them incorporated externally. Based on the fact that a visual structure is made on the truth that its impact would amplify the “cognition” of the stakeholder, in short increase their perceiving power.

### A. Scope

Here, scope defines the intake (visual scope) of all the structure in one window and does not require the viewer to scroll or shift windows to see the whole structure. This means that the whole context is depicted in one window and a viewer does not have to rotate between different views [18].

### B. Abstraction

This metric is chosen for the fact that at an architecture level of SPL, the structure is at an abstract level it is not concerned with the minuscule details such as LOC. Its main concern is that it can see what “decisions” present are leading to which “features”, which further lead to “component”, whose code is used for implementation, is not its concern at that level. This is a major metric, in a sense that fine grain detail (are needed in implementation phase are not its concern) are not need of the time at this level.

### C. Hierarchy

This is one of the major metrics chosen based on the fact that hierarchy is embedded in the nature of software product line. If some structure does not pander towards the hierarchy then its use as part of software product line architecture is not required, also keeping in mind that those structures that do not show hierarchy of the SPL architecture data are of no importance, those structures can be good illustrations for some task, but are not good representation of the whole SPL architecture structure. Hierarchy is subdivided into complex hierarchy and clear hierarchy. Clear hierarchy is a structure that shows hierarchy of data clearly, whereas complex hierarchy makes, perception of data complex.

### D. Traceability

This is another major metric taken into consideration. Basically the concept behind it is the fact that the comprehension of “what” is affected by some “particular change” is known if “traces” are present [14, 15]. Not just the fact that they are present, but they should be visible as well. This fact is favorable in a sense that it gives viewer the power to not just mentally perceive but visually; see, e.g., forward and backward traceability [32]. This metrics is further subdivided into complex viewer traceability and easy viewer traceability. These sub factors state the fact that if a structure is supporting traceability but that it’s not shown clearly results in “complex factor” and a structure which shows trace links clearly falls in “easy factor”.

### E. Scalability

This metrics is drawn because of the simple fact that we are evaluating “industrial size” SPL architecture, which clearly states the fact that the structure is not going to be small or medium by any standard. This is a major issue and is highlighted by [16], but the literature suggests, that the present techniques might resolve other issues in SPL, but they are not scalable for “industrial size” SPL architecture [16].

### F. Evolution

This metric is a simple and well cited one in SPL architecture in a sense that evolution happens in “space” and “time” and the structure should be such that it can support evolution and the changes which occur because of it should be traceable back to its source. If not then even if you add changes to its architecture it would bear no consequence because SPL architecture evolution has “dependencies, mismatching of variance, high cognitive complexity etc” incarcerated in it [17].

### G. Perception

This metrics is extracted from the fact that visual structure is always considered to be “good” if what the data wants to convey is clearly presented, helping in perceiving patterns, association, and relationship and so on. In short, a structure is more effective if the data mapped to it is faster to understand, can express more peculiarity and tend to be less error prone than other mapping techniques [4]. Also, it is a well-known fact [9], that a visual image clearly depends upon the properties of human perception; so a structure is said to be considered good, if it conveys, only the mapped data and not something which is not required.

### H. General visual content

This general metric contains three sub headings known as overlapping components, overlapping links and data obscuring. These, being chosen on the basis that the obscuring of data is a critical flaw which can lead to disasters results, overlapping links, tend to be misleading in evolution handling matter, overlapping of components can lead to wrong assumption and complexity and shorten the power of perception considerably. These points combined together result in disaster if not taken into view when considering a visual structure for software product line.

## VI. EVALUATION OF VISUAL STRUCTURES

The evaluation of visual structures on the basis of criteria mentioned above is depicted in the below mentioned table. The (✓) mark states that the structure supports it, the (✗) mark states that the structure does not support it, and the (○) mark states that the concerned feature is not in contention. Evaluation criteria are a combination of SPL and visual structure attributes. As indicated before, this evaluation does not cover general visual structure (physical data); it only covers, those visual structures which are accounted for, by abstract data, plus they are already in use, in SPL architecture.



TABLE I. TABULAR EVALUATION OF VISUAL STRUCTURES

Evaluation criteria		Visual structure									
		MT	CNT		TM	CT	UMLN		TXF	UCM	HBT
			CNT-V	CNT-H			UMLN-CD	UMLN-NL			
Scope		x	x	x	✓	x	x	x	x	x	✓
Abstraction		✓	✓	✓	✓	✓	x	x	x	x	✓
Hierarchal	Clear hierarchy	x	✓	✓	x	x	○	○	○	○	✓
	Complex hierarchy	✓	x	x	✓	✓	○	○	○	○	x
Traceability	Complex viewing traceability	✓	x	x	✓	✓	○	○	○	○	x
	Easy viewing traceability	x	✓	✓	x	x	○	○	○	○	✓
Scalability		x	x	x	✓	x	x	x	x	x	✓
Evolution		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Perception		x	✓	✓	x	x	x	x	x	x	✓
General visual content	Overlapping components	○	x	x	x	✓	○	○	○	○	x
	Overlapping links	○	x	x	x	✓	○	○	○	○	x
	Data obscuring	○	x	x	x	✓	○	○	○	○	x

The evaluation suggests that the current techniques (UMLN, TXF UCM) which have been used, for representation of software product line architecture are not favorable. They, as a whole, are falling short in all areas, except in evolution handling; the point to be highlighted here is that even evolution handling is not feasible, because they do not support traceability. The only technique in use for SPL, which to some extent, is feasible, is MT. But, that too totally lacks in Scope, Scalability and Perception features. MT partially satisfies hierarchal and traceability features.

Conventional Trees (CNT) are then quite favorable as compared to other visual structures, but, they too also lack in area of scope and scalability.

Cone Tree (CT) falls short in the area of scope, scalability, and perception (affected because of falling short in general visual context).

Tree Maps (TM) are quite good visual structures but the area in which they fall short is hierarchal and traceability feature because, though they support it, they fall in the complex area of the above mentioned features. TM also falls short in the area of perception because of the fact that there structure is so crowded that overlapping of component, overlapping of links, and data obscuring due to component overlapping makes it hard to perceive.

Hyperbolic Trees (HBT) are the ideal structure, which falls, in all the right categories. It is a good structure, for scalability, perceiving, evolution, scope, abstraction, traceability, hierarchy. In short, they are “good” for representation of Software product line architectures.

## VII. CONCLUSION AND FUTURE WORK

This paper identified the best suited visual structure and its need, through evaluation of visual structures for software product line architecture. Initially, visual structures were extracted, and then evaluated on the basis of a defined criterion. The result showed that those techniques used by SPL architectures lack a lot of features, only one technique was identified HBT, which was found to be good for use, by software product line architecture.

There is a need for interactive techniques, to be incorporated with HBT for the construction of full information visualization technique for SPL architecture, which is seen as part of our future work.

## ACKNOWLEDGMENT

We would like to thank all our teachers and colleagues who helped. A.K, thanks MR. Mushtaq and MR. Abdullah for their endearing support.

## REFERENCES

- [1] K. Berg, J. Bishop, and D. Muthig, “Tracing Software Product Line Variability — From Problem to Solution Space,” In Proceedings of the 2005 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries (SAICSIT '05). South African Institute for Computer Scientists and Information Technologists, South Africa, pp. 182-191, 2005. [Retrieved: June, 2013].

- [2] D. John. McGregor, "Software product lines". Journal of Object Technology, vol. 3, no. 3, pp. 65-74, April/March, 2004. [Retrieved: June, 2013].
- [3] R. Edward. Tufte. "Visual explanation: Images and Quantities, Evidence and Narrative". Cheshire, CT: Graphics Press, 1997.
- [4] S. Card, J. Mackinlay, and B. Shneiderman. "Readings in Information Visualization - Using Vision to Think", Morgan Kaufmann, 1999.
- [5] A. Blaine. Price, R. Baecker, and S. Ian. Small, "A Principled Taxonomy of Software Visualization," J. Visual Languages and Computing, vol. 4, no. 3, pp. 211-266, September, 1993. [Retrieved: June, 2013]. doi: 10.1006/jvlc.1993.1015.
- [6] G. Catalin. Roman and C. Kenneth. Cox. "A Taxonomy of Program Visualization Systems," Computer, vol. 26, no. 12, pp. 11-24, December, 1993. [Retrieved: July, 2013]. doi: 10.1109/2.247643.
- [7] M. Storey, F. Fracchia, and H. Muller. "Cognitive Design Elements to Support the Construction of a Mental Model During Software Exploration," J. Systems and Software, vol. 44, pp. 171-185, January, 1999. [Retrieved: June, 2013]. doi: 10.1016/S0164-1212(98)10055-9.
- [8] P. Caserta and O. Zendra. "Visualization of the static aspects of software: A survey," IEEE Transactions on Visualization and Computer Graphics, vol. 99, no. RapidPosts, August, 2010. [Retrieved: July, 2013]. doi: 10.1109/TVCG.2010.110.
- [9] C. Ware. "Information visualization: Perception for design". Morgan Kaufman Publishers, 2nd ed, 2004.
- [10] S. Bassil and R. Keller. "A Qualitative and Quantitative Evaluation of Software Visualization Tools," Proc. 23rd IEEE Int'l Conf. Software Eng. Workshop Software Visualization, pp. 33-37, 2001. [Retrieved: July, 2013].
- [11] J. Maletic, A. Marcus, and M. Collard. "A Task Oriented View of Software Visualization," Proc. IEEE Workshop on Visualizing Software for Understanding and Analysis (VISSOFT 2002), Paris France, pp. 32-40, June, 2002. [Retrieved: July, 2013]. doi: 10.1109/VISSOFT.2002.1019792
- [12] K. Gallagher, A. Hatch, and M. Munro. "Software Architecture Visualization: An Evaluation Framework and Its Application," IEEE Trans. Visualization and Computer Graphics, vol. 34, no. 2, pp. 260-270, March/April, 2008. [Retrieved: June, 2013]. doi: 10.1109/TSE.2007.70757.
- [13] L. Bass, P. Clements, and R. Kazman. "Software Architecture in practice", Second Edition, Publisher: Addison Wesley, 2003.
- [14] A. Luis. Sequeira. Sousa. "Traceability Support in Software Product Lines". Thesis report, Lisboa, 2008.
- [15] A. Samuel. Ajila and B. Ali. Kaba. "Using Traceability Mechanisms to Support Software Product Line Evolution" Information Reuse and Integration, 2004. IRI 2004. Proceedings of the 2004 IEEE International Conference on , vol., no., pp.157-162, 8-10 November, 2004. [Retrieved: July, 2013]. doi: 10.1109/IRI.2004.1431453.
- [16] L. Chen, M. Ali. Babar, and N. Ali. "Variability Management in Software Product Lines: A Systematic Review". In Proceedings of the 13th International Software Product Line Conference, SPLC '09, pp. 81-90, Pittsburgh, PA, USA, August, 2009. Carnegie Mellon University. [Retrieved: June, 2013].
- [17] J. Bosch, G. Florijn, D. Greefhorst, J. Kuusela, J. Henk. Obbink, and K. Pohl. "Variability Issues in Software Product Lines". Springer. LNCS 2290, pp. 13-21, October, 2002, [Retrieved: June, 2013]. doi: 10.1007/3-540-47833-7\_3.
- [18] J. Lamping and R. Rao. "Hyperbolic Browser: A focus+context Techniques for visualizing large hierarchies". Journal of visual languages and computing, vol. 7, no. 1, pp. 33-55, March, 1996. [Retrieved: July, 2013]. doi: 10.1006/jvlc.1996.0003
- [19] T. John. Brown, R. Gawley, R. Bashroush, I. Spence, P. Kilpatrick, and C. Gillan. "Weaving behavior into feature models for embedded system families". Software Product Line Conf. pp. 52-61, August, 2006. Baltimore, Md [Retrieved: July, 2013]. doi: 10.1109/SPLINE.2006.1691577
- [20] H. Ye and H. Liu. "Approach to modelling feature variability and dependencies in software product lines". IEEE. vol. 152, pp. 101-109, June, 2005. [Retrieved: June, 2013]. doi: 10.1049/ip-sen:20045007.
- [21] S. Ferber, J. Haag, and J. Savolainen. "Feature Interaction and Dependencies: Modeling Features for Reengineering a Legacy Product Line". Software Product Lines (SPLC2): Springer. pp. 235-256, August, 2002. [Retrieved: July, 2013]. doi: 10.1007/3-540-45652-X\_15
- [22] K. Chul. Kang, S. Kim, J. Lee, K. Kim, E. Shin, and M. Huh. "FORM: A feature-oriented reuse method with domain specific reference architectures", Annals of Software Engineering, vol. 5, pp. 143-168, 1998. [Retrieved: July, 2013]. doi: 10.1023/A:1018980625587
- [23] A. van Deursen, M. de Jonge, and T. Kuipers. "Feature-Based Product Line Instantiation Using Source-Level Packages". Software Product Lines (SPLC2): Springer. pp. 217-234, August 2002. [Retrieved: June, 2013]. doi: 10.1.1.16.6376.
- [24] D. Muthig and C. Atkinson. "Model-Driven Product Line Architecture". Software Product Lines (SPLC2): Springer. pp. 110-129, August, 2002. [Retrieved: June, 2010]. doi: 10.1007/3-540-45652-X\_8
- [25] D. Fey, R. Fajta, and A. Boros. "Feature Modeling: A Meta- Model to Enhance Usability and Usefulness". Software Product Lines (SPLC2): Springer. pp. 198-216, August, 2002. [Retrieved: June, 2013]. doi: 10.1007/3-540-45652-X\_13
- [26] S. Salicki and N. Farcet. "Expression and Usage of the Variability in the Software Product Lines". Software Product-Family Eng (PFE-4): Springer. pp. 304-318, October, 2002. [Retrieved: July, 2013]. doi: 10.1007/3-540-47833-7\_27.
- [27] G. Halmans and K. Pohl. "Communicating the variability of a software-product family to customers". Software and Systems Modeling, vol. 2, pp. 15-36, March, 2003. [Retrieved: June, 2013]. doi: 10.1007/s10270-003-0019-9.
- [28] F. Bachmann, M. Goedicke, J. Leite, R. Nord, K. Pohl, B. Ramesh, and A. Vilbig. "A Meta-model for Representing Variability in Product Family Development". Software Product-Family Eng (PFE-5):

- Springer. pp. 66-80, November, 2004. [Retrieved: July, 2013]. doi: 10.1007/978-3-540-24667-1\_6.
- [29] D. Lyn. Webber and H. Goma. "Modeling variability in software product lines with the variation point model". *Sci. Comput. Program.*, vol. 53. pp. 305-331, December, 2004. [Retrieved: June, 2013]. doi: 10.1016/j.scico.2003.04.004.
- [30] S. G. Eick, J. L. Steffen, and E. E. Sumner, "SeeSoft –a Tool for Visualizing Line Oriented Software Statistics". *IEEE Transactions on Software Engineering*, vol. 18, pp. 957-968, November, 1992. [Retrieved: June, 2013]. doi: 10.1109/32.177365.
- [31] B. Johnson and B. Shnedierman. "Tree-maps: a space-filling approach to the visualization of hierarchical information structures," *Visualization*, 1991. *Visualization '91, Proceedings, IEEE Conference on*, vol., no., pp. 284-291, 22-25 October, 1991. [Retrieved: July, 2013]. doi: 10.1109/VISUAL.1991.175815.
- [32] D. Nestor, L. O'Malley, A. Quigley, E. Sikora, and S. Thiel, "Visualisation of Variability in Software Product Line Engineering," in *1st International Workshop on Variability Modelling of Software Intensive Systems (VaMoS-2007)*, Limerick, Ireland, 2007. [Retrieved: July, 2013]. doi: 10.1.1.136.9399.

# A Novel Multiple Attributes Decision Making Approach For Multimedia Session Selection

Tein-Yaw Chung, Ibrahim Mashal, Fong-Ching Yuan, Yuan-Hao Chiang, Osama Alsaryrah  
College of Informatics  
Yuan Ze University  
Taoyuan, Taiwan  
{csdchung, imyuan}@saturn.yzu.edu.tw, {ibrahimmashal, bpmania, osanaji}@netlab.cse.yzu.edu.tw

**Abstract**— The fourth generation (4G) network integrates various access technologies, such as UMTS, WiMAX, and WLAN. In the 4G environment, a handset with multiple interfaces can switch among access networks to achieve ubiquitous service. However, how to select the best access network is critical for users to obtain their preferred services. In the past, Always Best Network Connection (ABNC) has been presented to select an access network based on properties of communication session, which is composed of both source and destination access networks. Nevertheless, ABNC fails to consider current multimedia services, which require an end-to-end bandwidth much larger than that of voice sessions and media types, such as voice, video, and data. In this paper, an extension to ABNC is presented to support multimedia services by exploiting parallel paths between multimode handsets. We formulated the problem as a Multiple Attributes Decision Making (MADM) hierarchy with a utility-based media assignment and utilized a hybrid Analytic Hierarchy Process (AHP) and Simple Additive Weighting (SAW) scheme to solve the problem. Finally, an example is illustrated to show how the scheme works.

**Keywords**-4G; ABC; ABNC; MADM; AHP; SAW.

## I. INTRODUCTION

The main feature of 4G is in its ability to integrate different access technologies, such as universal mobile telecommunication system (UMTS) [1], Worldwide Interoperability for Microwave Access (WiMAX) [2], and wireless local area networks (WLAN) [2], leading us to all-IP networking era. The vision of a 4G mobile system is to integrate various radio access technologies (RATs) into a common network called the open wireless architecture (OWA) platform.

To enjoy the convenience of 4G, nowadays, handsets are equipped with multiple interfaces and can switch among various access networks to achieve ubiquitous service. However, 4G networks still face a number of challenges; one of them is to select the best access network based on user preferences.

An earlier solution was using Always Best Connected (ABC) schemes. ABC enables a user to choose the best available access network that best suits his/her needs based on receiving signal strength (RSS), cost, quality, security, etc. Later on, a new paradigm called Always Best Network Connection (ABNC) is presented. ABNC selects an access network based on properties of communication session, which is composed of both source and destination access

networks. It has been proved that ABNC indeed works better to meet user's requirements [3].

ABNC, however, has a number of deficiencies when it is applied for multimedia communications: first, ABNC considers only RSS as an index of quality; it does not consider other indices, e.g., bandwidth. Second, ABNC does not exploit path diversity between the source and the destination; it only chooses a path for end-to-end communication between two peers. Finally, ABNC only considers voice sessions, which only require small bandwidth. Thus, ABNC cannot meet the need of multimedia communications, which require exchange of various media data such as video that needs a large bandwidth. Thus, a more comprehensive model than ABNC is required for multimedia session services.

In multimedia communication, multiple connections need to be set up between the source and the destination for exchange of various media type of data, such as data, audio, and video. Besides, many paths may exist between handsets with multimode interfaces. Thus, each connection may be set up on different path. Therefore, to support multimedia services, the new model must allow us to select which path or set of paths for connection setup to best meet user's preference. In this paper, we extended ABNC and presented a new paradigm called Always Best Multiple Network Connections (ABMNC) for multimedia services.

ABMNC is formulated as a Multiple Attributes Decision Making (MADM) problem. The ABMNC decision hierarchy chooses a combination of network connections instead of each individual connection based on a number of attributes such as cost, quality, security, and power consumption. The ABMNC decision hierarchy also weights each media type in a multimedia service according to its significance in communication. To assign a media type on a network connection with best communication quality, the media assignment problem was formulated as a utility-based Multiple Knapsack Problem (MKP). Furthermore, to reduce the complexity of the MADM and have a reasonable number of alternatives, the original MADM hierarchy is decomposed into three iterated sub-MADM hierarchies. A novel hybrid AHP and SAW scheme was presented to efficiently solve the iterative strategy. Finally, an example is used to illustrate how ABMNC works.

In summary, this work has three contributions: A new paradigm called ABMNC is defined; a utility-based media assignment problem based on the significance of each media on multimedia communication is formulated as a MKP

problem; an efficient iterative problem solving scheme that integrate SAW, AHP and path selection is presented to solve ABMNC.

The rest of the paper is organized as follows. Section II summarizes related work. Section III describes and formulates the multimedia session selection problem. Section IV depicts the proposed methodology, and a hybrid scheme of SAW and AHP is presented. An example to illustrate ABMNC is given in Section V. Finally, Section VI concludes the paper.

## II. RELATED WORK

ABC has been extensively studied by a number of researchers. The Concept and architecture of ABC are described in [4]. V. Gazis et al. [1] formulates the ABC problem as a variation of the Knapsack problem with multiple knapsacks, and further proved it to be NP-Hard. Given a set of flows, [5] formulate ABC as a variant of bin packing problem and presents a series of approximation algorithms based on First Fit Decreasing (FFD) algorithm that selects a network to meet user's Quality of Service (QoS) requirements and with maximal admitted flows for a user. However, the approach used in [1] addresses the problem from a network perspective. What it is maximized is the admitted flow, not the user satisfaction. J. Jackson et al. [6] intends to achieve ABC over WLAN and WiMAX by using a new mechanism to detect QoS support of the underlie networks.

In [7], both user preferences and attributes of access networks were considered, and ABC was formulated as a problem of MCDM. A hybrid Analytic Network Process (ANP) [8] with RTOPSIS [9] model based on several decision making algorithms to select the best candidate networks from the user's perspective is proposed in [10]. It solves the rank irregularity problem and eliminates the interdependence between criteria. All previous researches in ABC focus on source access network selection, which is appropriate when a multimode handset is in a standby mode. However, when a service session is being set up, ABC is no longer efficient as the performance of a service session depends not only on the source access network, but also on the corresponding access network.

In [3], ABNC was proposed. ABNC enables users with multimode handsets to select the best network connection for a voice session, which consists of source and destination access network pair, to satisfy quality constraints and users' preferences. However, when a multimedia session with multiple media type and bandwidth requirement is being setup, ABNC cannot meet the need.

## III. MULTIMEDIA SESSION SELECTION

The characteristics of a multimedia session are determined by the source and the destination access networks. A multimedia session consists of three media types: video, voice, and data. The problem of selecting the best connection becomes more complicated since a configuration or alternative may consist of one or more connections and the significance of each media type in communication quality must be considered. In this section,

multimedia session attributes are described first. Then, the utility functions for media assignment are presented. Finally, the problem of multimedia session selection is formulated. Table I shows a summary of notations used in this section.

### A. Multimedia Session Attributes

Multimode handsets came with multiple interfaces, which mean that a combination of different access networks can be exploited during communications. Given source handset S and correspondent handset C,  $A_i^S$  and  $A_j^C$  denote access networks of source handset and access networks of corresponding handset, respectively.

TABLE I. SUMMARY OF NOTATION FOR OUR PROBLEM FORMULATION

Symbol	Meaning
$k$	Transmission configuration identifier
$C_k$	The cost for a multimedia session with configuration $k$
$P_k$	The power consumption for a multimedia session with configuration $k$
$SL_k$	The security level for a multimedia session with configuration $k$
$U_k$	The utility for a multimedia session with configuration $k$
$W_c, W_p, W_u, W_{sl}$	Weight for criteria cost, power, quality, and security level, respectively

Let  $S_{ij}$  denote a connection established from source handset interface  $i$  to corresponding handset's interface  $j$ , and be defined as  $S_{ij} = (A_i^S, A_j^C)$ . For example,  $S_{UMTS, WiFi}$  means the connection is established from UMTS interface of source to the WiFi interface of destination. Due to the nature of multimedia session ABMNC considers more than one connection. We define  $k$  as a set of configurations that indicates how caller and callee are connected through their access networks. For example, the configuration  $k = (S_{UMTS, WiFi}, S_{WiFi, WiFi})$  means there are two connections established from UMTS and WiFi interface of source to the WiFi interface of destination.

Connections can be characterized by cost, power, security level, and utility, which are specified in the user profile.

- Cost: The cost information can be acquired from ISP. Charging depends on both source and destination networks, the cost is presented as  $C(S_{ij}) = C(A_i^S, A_j^C)$ . The cost should be accumulated on the number of interfaces used by both sides during communication.
- Power: The energy consumed per unit of time. It depends not only on the type of  $A_i^S$  but also on the number of interfaces used during a call.
- Security Level: Access networks implement different security mechanism and have different security levels. The security level of access network

without considering the application-level encryption mechanism follows the order of  $SL(A_{UMTS}) > SL(A_{WiMAX}) > SL(A_{WiFi})$  [11]. The security level of a path is defined as the minimal security level of the access networks that consist of the path,  $SL_k(S_{ij}) = \min\{SL(A_i^S), SL(A_j^C)\}$ .

- Utility (quality): Although delay, packet loss rate of candidate connections and data rate allocation for each medium can be used to measure the quality of a multimedia session, we only consider the data rate as the quality index. The utility function  $U_k$  for configuration  $k$  will be defined in the next section.

The parameter cost, power, and security can be easily specified in the user profile but not the utility because the available bandwidth of each access network changes dynamically. In this study, it is assumed that each handset uses an IS service on Media Independent Handover (MIH) to report its available access networks and their respective available bandwidth [2]. Thus, with the assistance from IS, a handset can acquire all needed information for network selection.

### B. Utility-Based Media Assignment

In this section, a utility-based measurement scheme is presented to evaluate the quality of multimedia services, based on which, the media session assignment is formulated as a MKP.

The design of the utility functions is to strike a balance between system utilization and system QoS requirements. Let  $U_k^m$  be the utility function of a given configuration  $k$  for some medium  $m$ , where  $m \in \{video, voice, data\}$ , and is defined as follows:

$$U_k^m = w_m \log_2 \left( 1 + \frac{X_m}{b_m} \right) \quad (1)$$

where  $X_m$  and  $b_m$  are the allocated and requested BW for media  $m$ , respectively.  $w_m$  is the weight for media  $m$  which is obtained from AHP. Because humans are more sensitive to voice (speech) than video, and data is considered the least important, the weights for video, voice, and data are:

$$w_{voice} > w_{video} > w_{data}.$$

In the ABMNC, the network connection problem is modeled as a simplified MKP. In MKP, every path of different bandwidth between the caller and callee is considered as a bucket of different size. Each medium of different bandwidth requirement is considered an object of different size that to be included in one of many possible buckets. The objective is to assign objects to the available buckets such that the maximal utility value can be achieved without overflow. Thus, given  $N$  different paths, the media assignment problem of ABMNC is defined as follows.

$$\begin{aligned} U_k &= \max_{X_m} \sum_{i=1}^N \sum_m U_k^m x_{im} \\ &= \max_{X_m} \sum_{i=1}^N \sum_m w_m \log_2 \left( 1 + \frac{X_m}{b_m} \right) x_{im} \\ s.t. & \sum_m X_m x_{im} \leq W_i, \\ & \sum_{i=1}^N x_{im} = 1, \quad x_{im} \in \{0, 1\}, \\ & \sum_m w_m = 1, \\ & T_m \leq \frac{X_m}{b_m} \leq 1 \end{aligned} \quad (2)$$

where  $U_k$  is the utility or scores acquired by optimizing the allocated rate for a configuration  $k$  comprising a set of dedicated connection(s),  $x_{im}$  denotes the  $m^{th}$  type of media object assigned to the  $i^{th}$  path,  $U_k^m$  represents the utility value of media  $m$ , which depends on its weight  $w_m$  and allocated bandwidth  $X_m$  with respect to its required bandwidth  $b_m$ .

The first constraint ensures that the total allocated bandwidth to path  $i$  will not exceed its limit (knapsack weight)  $W_i$ , the available bandwidth of path  $i$ . The second one indicates each medium can only be assigned to one path.  $X_m/b_m$  is ranged from zero to one, so the maximal value of  $U_k$  is one.  $T_m$  limits  $X_m$  to have a minimal threshold value to ensure an acceptable quality level for media  $m$ . A connection is feasible only if it has bandwidth larger than the threshold value.

### C. Problem Formulation

Given a pair of communicating parties, each is equipped with  $n$  interfaces. Then, there exist  $n^2$  possible network connections between them. A multimedia session contains three medium sessions: video, voice, and data session, and each can take any one of the connections. Assume one medium can pass only on one connection and more than one medium can share one connection when the connection has sufficient bandwidth. With this assumption, there exist  $(n^2)^3 = n^6$  possible transmission configurations. So, the problem of ABMNC is to choose the best configuration from the large number of possible configurations.

Based on the previous analysis of session attributes, choosing the best configuration belongs to a MADM problem and can be defined as follows.

$$\begin{aligned} S^* &= \arg_k \max (w_c C_k + w_p P_k + w_u U_k + w_{sl} SL_k) \\ s.t. & CB > T_B \text{ and } CS < T_S \end{aligned} \quad (3)$$

where  $S^*$  is the best configuration obtained by comparison between score of  $k$ , which is the addition of the normalized contribution from each criterion multiplied by the weight factors. Let  $CB$  be the set whose elements are adequate when they are bigger than a threshold  $T_B$ , e.g., bandwidth. Let

$CS$  be the set whose elements are adequate when they are smaller than a threshold  $T_s$  such as delay.

The MADM decision hierarchy for ABMNC can be shown in Fig. 1. The goal of the hierarchy is to select the most suitable configuration  $k$  for transmission at the top level. Solution alternatives are located at the bottom nodes. Video, voice and data are represented as  $v$ ,  $s$ , and  $d$ , respectively.

#### IV. ABMNC

Applying a simple MADM solution to our problem is not feasible because we have to compare a large number of alternatives, which makes it impractical for real time applications. In the proposed scheme, the original hierarchy is decomposed into three iterated sub-MADM hierarchies and a hybrid scheme of SAW and AHP is used to efficiently solve the problem. AHP is used to decide the relative importance of criteria and media, which is then used by SAW for scoring. In the hybrid SAW and AHP scheme, each sub-MADM rates the path alternatives by considering all criteria except utility to produce a partial score first. Then, the overall score of a configuration is computed by merging the utility value after each media type is assigned.

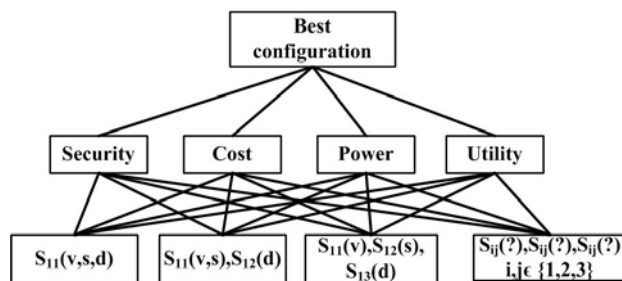


Figure 1. MADM hierarchy

The MADM hierarchy is divided into the single-path sub-MADM hierarchy that takes one path as input, the dual-path sub-MADM hierarchy that takes two paths, and the tri-path sub-MADM hierarchy that takes three paths. The single-path Sub-MADM hierarchy outputs its best configuration as an input to dual-path Sub-MADM. The dual-path Sub-MADM tries to find a better result by combining the input best single path with another path. Similarly, the tri-path Sub-MADM tries to find an optimal configuration that has three paths by integrating the best dual path from the dual-path Sub-MADM and another candidate path. Fig.2 shows single-path sub-MADM hierarchy.

##### A. AHP Weighting

In this study, AHP is used because it is easy to understand and does not need too much calculation. Furthermore, AHP is very convenient to establish a mathematical model [12]. AHP can be summarized as the process of constructing a problem as a hierarchy with a large number of attributes. The first step of AHP is to construct a decision hierarchy as shown in Fig.1. Then, pair-wise

comparison between alternatives for each attribute in each hierarchy is done.

Since our approach deals with multiple media sessions, AHP must not only weight criteria or attributes  $w_c, w_p, w_{sl}$ , and  $w_u$  of our MADM hierarchy, but also weight media  $w_m | m \in \{video, voice, data\}$  that is used during media assignment. Normally, attribute weights are decided according to user's preferences and AHP illustrates how important each attribute is by determining the relative weights of the compared attributes.

##### B. SAW

SAW is a simple and most often used multiple attribute decision technique. The method is based on the weighted average. In our research, the overall score of an alternative is computed as the weighted sum of all the attribute values. SAW classifies criteria into two categories, beneficial and non-beneficial. A criterion is beneficial if users are interested in having a higher value of it, e.g., security and utility; otherwise it is non-beneficial, e.g., cost and power.

In our study, the parameters used for different access technologies are derived as follows: the cost is referred by recent ISP charging mechanism, the power consumption, i.e., standby time is based on the specifications of some representative multimode handsets, and the relative values of security level are from some wireless technical reports as shown in Table III.

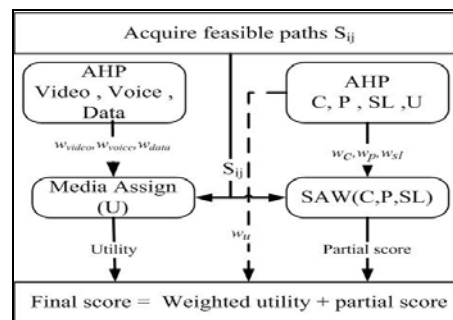


Figure 2. Single-path sub-MADM hierarchy

The scale values of 1-9 with calibration is used where 1 stands for extremely low while 9 represents extremely high. Table II shows the scale values, which are assigned to power and security.

#### V. ABMNC EXAMPLE

In this section, a simple example is shown to illustrate how ABMNC works.

##### A. Single-path

In the example, the source and the corresponding handset each have three interfaces, UMTS, WiMAX, and Wi-Fi. Thus, there exist 9 candidate network connections for the single-path sub-MADM hierarchy as shown in Table V.

First, AHP is used to elicit the weight for media ( $w_m$ ) as shown in Table IV. The voice row shows that voice is twice

as important as video, three times as important as data, and equal to itself. The video row shows that the video is twice as important as data. The matrix satisfies the transitive condition. The Consistency Ratio (CR) is applied to prove the matrix is consistency. [13] explains that when the CR of the comparison matrix is less than 0.1, a precise result can be achieved. Since our CR is equal to 0.0079, it is far below 0.1 so the matrix is consistent. A comparison matrix is also evaluated using Arithmetic Mean of the Normalized Column Vectors (AMNCV). AMNCV is defined as follows:

$$w_i = \frac{1}{|m|} \sum_{j=1}^{|m|} \frac{a_{ij}}{\sum_{i=1}^{|m|} a_{ij}} \quad i, j = 1, 2, \dots, |m| \quad (4)$$

TABLE II. SCALE VALUES OF 1-9

Value	Definition
1	Extremely low
3	low
5	average
7	high
9	Extremely high
2,4,6,8	Middle value between elements

TABLE III. PARAMETERS SETTING

		Caller	Callee
Power	UMTS	420mins	Irrelevant
	WiMAX	180mins	Irrelevant
	Wi-Fi	210mins	Irrelevant
Security	UMTS	Middle	
	WiMAX	High	
	Wi-Fi	Low	
Cost	UMTS	0.0003\$/s	0.0005\$/s
	WiMAX	0.0005\$/s	0.0003\$/s
	Wi-Fi	0.0001\$/s	0.0002\$/s

TABLE IV. COMPARISON MATRIX OF MEDIA WITH RESPECT TO THE CRITERION UTILITY

Utility	Video	Voice	Data
Video	1	1/2	2
Voice	2	1	3
Data	1/2	1/3	1

Given  $w_1 = w_{video}$ ,  $w_2 = w_{voice}$ , and  $w_3 = w_{data}$ ,  $a_{ij}$  denotes an element in the comparison matrix. The media weight  $w_m$  acquired by (4) is  $w_{video} = 0.297$ ,  $w_{voice} = 0.539$ , and  $w_{data} = 0.164$ . By a similar process, we can readily obtain weights for criteria  $w_c = 0.36$ ,  $w_p = 0.05$ ,  $w_{sl} = 0.05$ , and  $w_u = 0.54$ .

Second, SAW is used to construct a decision matrix based on parameters and scale values, where criteria considered are cost, power, and security level. In the un-normalized decision matrix,  $v_{ij}$  represents the original score of the  $j^{th}$  criterion of the  $i^{th}$  alternative. Table V shows the scores of all alternative paths in the un-normalized decision matrix.

Third,  $v_{ij}$  must be normalized to  $r_{ij}$ , the higher the  $r_{ij}$ , the better the alternative performs with respect to the  $j^{th}$  criterion. Beneficial criterion is normalized by using (5) while Non-beneficial criterion is normalized by using (6).

$$r_{ij} = \frac{v_{ij}}{\max_i (v_{ij})} \quad (5)$$

$$r_{ij} = \frac{\min_i (v_{ij})}{v_{ij}} \quad (6)$$

Finally, the best configuration is obtained by comparing the weighted average of all alternatives from normalized decision matrix by using (7).

$$MAX \sum_{j=1}^m w_j r_{ij} \quad (7)$$

where  $M$  denotes the number of criteria and  $w_j$  is the weight of the  $j^{th}$  criterion. We assume  $w_c = w_{j=1}$ ,  $w_p = w_{j=2}$ , and  $w_{sl} = w_{j=3}$ . By applying (7), we get Table VI, which shows the partial scores of alternatives without utility criterion.

It can be seen from Table VI that the best alternative is  $S_{33}$  (WiFi-WiFi). Intuitively,  $S_{33}$  is the right choice since the selection seems cost-preferred without taking utility into consideration. Assume the bandwidth of  $S_{33}$  is 90Kbps, the media assignment scheme will allocate 64 Kbps, 16 Kbps, and 10 Kbps to video, voice, and data, respectively. By applying (2), we get the utility of video  $U_k^{video} 0.067$ ,  $U_k^{voice} 0.161$  and  $U_k^{data} 0.083$ . Thus, the total utility score  $U_k$  is equal to 0.311 and the final score of  $S_{33}$  is equal to 0.588.

TABLE V. UN-NORMALIZED AND NORMALIZED DECISION MATRIX WITH RESPECT TO CRITERIA COST, POWER, AND SECURITY LEVEL

Decision matrix		UN-Normalized			Normalized		
Beneficial:1/Non-Beneficial:0		0	0	1	0	0	1
The weight of criteria		0.36	0.05	0.05	0.36	0.05	0.05
Criteria		Cost	Power	Security	Cost	Power	Security
UMTS-UMTS	$S_{11}$	0.0008	3	8	0.38	1.00	0.89
UMTS-WiMAX	$S_{12}$	0.0006	3	8	0.50	1.00	0.89
UMTS-WiFi	$S_{13}$	0.0005	3	7	0.60	1.00	0.78
WiMAX-UMTS	$S_{21}$	0.0010	7	8	0.30	0.43	0.89
WiMAX-WiMAX	$S_{22}$	0.0008	7	9	0.38	0.43	1.00
WiMAX-WiFi	$S_{23}$	0.0007	7	7	0.43	0.43	0.78
WiFi-UMTS	$S_{31}$	0.0006	6	7	0.50	0.50	0.78
WiFi-WiMAX	$S_{32}$	0.0004	6	7	0.75	0.50	0.78
WiFi-WiFi	$S_{33}$	0.0003	6	7	1.00	0.50	0.78



TABLE VI. PARTIAL SCORE FOR SINGLE-PATH

$S_{11}$	$S_{12}$	$S_{13}$	$S_{21}$	$S_{22}$	$S_{23}$	$S_{31}$	$S_{32}$	$S_{33}$
0.23	0.28	0.31	0.18	0.21	0.22	0.24	0.33	0.42

B. Dual-path

In dual-path, a large number of possible configurations is considered. The output of the single-path sub-MADM  $S_{33}$  is input to the dual-path sub-MADM, which tries to find the best result by combining  $S_{33}$  with another path in order to reduce the complexity.

When considering a dual-path alternative, the normalized scores of all  $v_{ij}$  must be reevaluated, because the computed results will be changed if  $\max_i(v_{ij})$  and  $\min_i(v_{ij})$  change. Fortunately, the  $v_{ij}$  with respect to all criteria except utility will become smaller for beneficial criteria and larger for non-beneficial criteria if more paths are involved. Thus, when we insert a new dual-path configuration by adding single-path to an existing single-path alternative,  $\max_i(v_{ij})$  and  $\min_i(v_{ij})$  remain the same when utility is not considered. Therefore, Table V remains unchanged.

When the same computations are applied on the dual-path configuration, the partial score of  $k_2 = \{S_{32}, S_{33}\}$  can be derived as 0.243, which is smaller than  $k_1 = \{S_{33}\}$ . The partial scores for dual-path are shown in Table VII.

TABLE VII. PARTIAL SCORE FOR DUAL-PATH

$S_{32}, S_{33}$	$S_{31}, S_{33}$	$S_{23}, S_{33}$	$S_{11}, S_{33}$
0.243	0.209	0.184	0.175

We assume the bandwidth of  $S_{32}$  is 90Kbps. Then, the media assignment scheme will assign 64 Kbps, 64 Kbps, and 20 Kbps to video, voice, and data, respectively to  $S_{33}$  and  $S_{32}$ . By applying (2), the utility of video  $U_k^{video}$ , voice  $U_k^{voice}$  and data  $U_k^{data}$  can be readily obtained as 0.067, 0.5, and 0.147, respectively. The total utility score  $U_k$  equals to 0.714. Finally, the final score of  $k_2 = \{S_{32}, S_{33}\}$  is derived to be equal to 0.629, which is larger than  $S_{33}$ . So the best configuration is  $S^* = (S_{32}, S_{33})$  by the dual-path sub-MADM. The triple-path sub-MADM, in this case, cannot find a better configuration and hence ABMNC returns configuration  $S^* = (S_{32}, S_{33})$  as the best connection.

VI. CONCLUSION

Multimedia communication has become part of daily live for many people. Therefore, how to choose an appropriate configuration for a multimedia session between multimode handsets is an important issue. In this paper, a new paradigm called ABMNC is presented to address this issue. Unlike traditional network selection problems, ABMNC is modeled as a MADM problem along with media assignment. To

reduce the size of candidate solutions, the ABMNC MADM hierarchy is divided to three Sub-MADMs and an efficient hybrid SAW and AHP scheme was proposed to solve the problem. An example has been illustrated to show how our presented approach works.

This paper only considered bandwidth as a quality index. In the future, we will consider more types of QoS parameters like packet loss rate, delay, etc., to make ABMNC more practical. Also, each criterion in connection evaluation may be correlated. For example, a larger bandwidth may also cost more. So, we will use different MADM model other than AHP in our future study.

REFERENCES

- [1] V. Gazis, N. Alonistioti, and L. Merakos, "Toward a generic "always best connected" capability in integrated WLAN/UMTS cellular mobile networks (and beyond)," *Ieee Wireless Communications*, vol. 12, Jun 2005, pp. 20-29.
- [2] A. A. Bathich, M. D. Baba, and M. Ibrahim, "IEEE 802.21 based vertical handover in WiFi and WiMAX networks," in *Computers & Informatics (ISCI), 2012 IEEE Symposium on*, 2012, pp. 140-144.
- [3] T. Y. Chung, F. C. Yuan, Y. M. Chen, and B. J. Liu, "S-3: Smart Session Selection for Voice Communications in Next Generation Wireless Network," *Ieice Transactions on Fundamentals of Electronics Communications and Computer Sciences*, vol. E91a, Oct 2008, pp. 2995-3002.
- [4] Y. P. Chen and Y. H. Yang, "A new 4G architecture providing multimode terminals always best connected services," *Ieee Wireless Communications*, vol. 14, Apr 2007, pp. 36-41.
- [5] E. Stevens-Navarro and V. W. S. Wong, "Comparison between Vertical Handoff Decision Algorithms for Heterogeneous Wireless Networks," in *Vehicular Technology Conference, 2006. VTC 2006-Spring. IEEE 63rd*, 2006, pp. 947-951.
- [6] J. Jackson Juliet Roy, V. Vaidehi, and S. Srikanth, "Always Best-Connected QoS integration model for the WLAN, WiMAX Heterogeneous Network," in *Industrial and Information Systems, First International Conference on*, 2006, pp. 361-366.
- [7] S. Qingyang and A. Jamalipour, "An adaptive quality-of-service network selection mechanism for heterogeneous mobile networks," *Wireless Communications & Mobile Computing*, vol. 5, Sep 2005, pp. 697-708.
- [8] J. W. Lee and S. H. Kim, "Using analytic network process and goal programming for interdependent information system project selection," *Computers & Operations Research*, vol. 27, APR 2000, pp. 367-382.
- [9] W. Chen, "On the Problem and Elimination of Rank Reversal in the Application of TOPSIS Method," *Operations Research and Management Science*, vol. 14, Oct 2005,
- [10] Y. Liu, "Access Network Selection in a 4G Networking Environment," master, University of Waterloo Canada, 2007.
- [11] Y. Zhang, J. Zheng, and M. Ma, *Handbook Of Research On Wireless Security ( 2 Vol Set )*: Information Science Reference, 2008.
- [12] A. Ishizaka and A. Labib, "Review of the main developments in the analytic hierarchy process," *Expert Systems with Applications*, vol. 38, 2011, pp. 14336-14345.
- [13] T. L. Saaty, *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*: McGraw-Hill, 1980.

## A Novel Algorithm for Selecting Multimedia Network Connections in Next Generation Networks

Tein-Yaw Chung Department of Computer Science and Engineering Yuan Ze University Taoyuan, Taiwan csdchung@saturn.yzu. edu.tw	Ibrahim Mashal Department of Computer Science and Engineering Yuan Ze University Taoyuan, Taiwan ibrahimmashal@netlab .cse.yzu.edu.tw	Fong-Ching Yuan Department of Information Management Yuan Ze University Taoyuan, Taiwan imyuan@saturn.yzu.ed u.tw	Yuan-Hao Chiang Department of Computer Science and Engineering Yuan Ze University Taoyuan, Taiwan bpmania@netlab.cse.y zu.edu.tw	Osama Alsaryrah Department of Computer Science and Engineering Yuan Ze University Taoyuan, Taiwan osanaji@netlab.cse.yzu .edu.tw
---	--	--	---	---

**Abstract**— In the past, Always Best Network Connection (ABNC) has been presented to selecting an access network based on the characteristics of communication session in heterogeneous wireless networks. The novelty of ABNC is to consider both source and destination access networks when a voice session is being setup. However, it is not suitable for multimedia service, which consists of video and data in addition to voice. This paper presents an Always Best Multimedia Network Connection (ABMNC) model to support multimedia services. ABMNC is an integration of a Multiple Attribute Decision Making (MADM) problem and a utility-based Multiple Knapsack Problem (MKP). We present a hybrid Analytic Hierarchy Process (AHP) and Simple Additive Weighting (SAW) scheme to solve the problem. Finally, a heuristic algorithm was proposed to reduce the complexity of access network selection. The simulation results show that the presented approach performs better than current approaches.

**Keywords**-4G; ABC; ABNC; MADM; AHP; SAW; Utility.

### I. INTRODUCTION

Nowadays, various access networks such as General Packet Radio Service (GPRS) [1], Long Term Evolution (LTE) [2], Microwave Access (WiMAX) [3], and Wireless Local Area Networks (WLAN) [3] have been deployed widely. Each access network provides different levels of Quality of Service (QoS), in terms of bandwidth, mobility, coverage area, and cost to the mobile users. Next-generation networks (NGN) commonly known as 4G integrate these different access technologies and support all IP-based traffic. Indeed, 4G users will be able to connect seamlessly to various access networks that offer the best possible quality at any time and any place.

Despite all potential advantages of 4G, selection of the best network connection remains one of the most important concerns and critical issues. Based on user preferences, the best access network must be selected to establish connection between source handset and corresponding handset. A widely accepted approach called Always Best Connected (ABC)[4] has been proposed to solve this problem. ABC is user-centered which means it makes users always choose the best available access networks at any place and any time. However, in many situations choosing a good access network is not good enough. For example, voice communication is charged not only based on source access

network but also on the chosen corresponding access network.

ABNC [5] extends ABC to let users select the best network connection, which consists of the source and the destination access network pair when handsets are equipped with multiple interfaces. It has been shown that ABNC indeed works better to meet user's requirements. However, ABNC is designed for voice sessions and is not suitable for multimedia services for the following reasons: Firstly, a multimedia session can be established on many paths and has more than one media type and thus, we may choose more than one path and have to assign each media type to the chosen paths, which is not considered in ABNC. Secondly, a multimedia session contains media type such as video that requires much larger end-to-end bandwidth than a voice session. This paper presents a more flexible model called Always Best Multimedia Network Connection (ABMNC). ABMNC extends ABNC to support multimedia services and deals with different multimedia sessions such as video, voice, and data. ABMNC constructs multiple end-to-end connections between users, and then chooses the best configuration that comprises one or more connections with respect to multiple media sessions.

In this study, ABMNC is formulated as a Multiple Attribute Decision Making (MADM) problem. The ABMNC decision hierarchy chooses a set of network connections based on attributes such as cost, quality, security, and power consumption. The ABMNC decision hierarchy weights all media types in a multimedia service and then assigns a media type to a network connection. In order to reduce the complexity of MADM and have a reasonable number of alternatives, the MADM hierarchy is decomposed into many iterated sub-MADM hierarchies. Also, the media assignment problem is formulated as a utility-based Multiple Knapsack Problem (MKP) and a heuristic approach is presented to solve the problem. Simulation results show that our heuristic approach can perform well as compared with that of an exhaustive searching scheme.

This paper adds contributions to [6] by introducing a heuristic path selection algorithm. Also, simulation was performed to study the performance of our presented algorithms.

The rest of the paper is organized as follows. Section II summarizes related work. Section III describes the multimedia session selection and formulates the problem.

Section IV depicts the proposed methodology, and illustrates a hybrid scheme of SAW and AHP. ABMNC performance is evaluated in Section V. Finally, Section VI concludes the paper.

## II. RELATED WORK

ABC has been extensively studied by a number of researchers. The concept and architecture of ABC are described in [3]. In [7], the ABC problem is formulated as a variation of the Knapsack problem with multiple knapsacks, and further proved it to be NP-Hard. Given a set of flows, in [8], authors formulate ABC as a variant of bin packing problem and present a series of approximation algorithms based on First Fit Decreasing (FFD) algorithm that select a network to meet user's QoS requirements and with maximal admitted flows for a user. However, the approach used in [7] addresses the problem from a network perspective. What it is maximized is the admitted flow, not the user satisfaction. Roy et al. [9] intends to achieve ABC over WLAN and WiMAX by using a new mechanism to detect QoS support of the underlie networks.

The work in [10, 11] take multiple criteria into consideration, however, still leaving some problems. For instance, weight elicitation has been ignored; and the rank irregularity problem is not addressed. [12] considers both user preference and attributes of access networks and formulate ABC as a problem of MCDM. In addition, in [13-15], an access link selection management is proposed to incorporate user preference profiles as a part of the selection policies.

Many hybrid approaches have been used in solving ABC. Liu [16] proposes a hybrid ANP [17] with RTOPSIS [18] based on several decision making algorithms to select the best candidate networks(s) from the user perspective. Mohamed et al. [19] presents selection strategy based on ANP and TOPSIS while [20] integrates AHP and TOPSIS in selecting a network. In [21], four MADM algorithms, namely SAW, MEW, TOPSIS and GRA, are simulated to show the impact of criterion's weight on the coefficients. They solve the rank irregularity problem and eliminate the interdependence between criteria. All previous researches in ABC focus on source access network selection, which appropriates when a multimode handset is in a standby mode. However, when a service session is being set up, ABC is no longer efficient as the performance of a service session depends not only on the source access network, but also on the corresponding destination access network.

In [5], ABNC is proposed to enable users with multimode handsets to select the best network connection for a voice session, which consists of source and destination access network pair, to satisfy quality constraints and users' preferences. However, when a multimedia session with more than one media type is being setup, ABNC cannot meet the need. Thus, a more comprehensive model than ABNC is required to meeting multimedia session services.

## III. MULTIMEDIA SESSION

### A. Problem Definition

Given a pair of communicating parties, each is equipped with  $n$  interfaces. Then there exist  $n^2$  possible network paths or connections between them. Media session contains video, voice, and data, and each can take any one of the connections. Assume one medium can pass only on one connection and more than one medium can share one path when the path has sufficient bandwidth. With this assumption, we have  $(n^2)^3 = n^6$  possible transmission configurations.

Given source handset S and corresponding handset C. Let  $A_i^S$  and  $A_j^C$  denote access networks of source handset and access networks of the corresponding handset, respectively. Let  $S_{ij}$  denote a connection established from source handsets interface  $i$  to corresponding handset's interface  $j$ , and be defined as  $S_{ij} = (A_i^S, A_j^C)$ . For example,  $S_{UMTS, WiFi}$  means the connection is established from UMTS interface of source to the WiFi interface of corresponding. Due to the nature of multimedia session, ABMNC considers more than one connection. Let  $k$  be a set of configurations that indicates how caller and callee are connected through their access networks. Then, the configuration  $k = (S_{UMTS, WiFi}, S_{WiFi, WiFi})$  means there are two connections established from UMTS and WiFi interface of source to the WiFi interface of destination. Characteristics of a multimedia session are determined by source and destination access networks. Connections can be characterized by cost, power, security level, and utility.

- Cost ( $C_k$ ): The cost information can be acquired from ISP. Charging depends on both source and destination networks, we represent cost as  $C(S_{ij}) = C(A_i^S, A_j^C)$ . The cost should be accumulated on the number of interfaces used by both sides during communication.
- Power ( $P_k$ ): The energy consumed per unit of time. It depends not only on the type of access networks of source handset but also on the number of interfaces used during a call.
- Security Level ( $SL_k$ ): The security level goes from UMTS, WiMAX, WiFi in a descending order. Since a path consists of a number of access networks, so, the access network with minimal security level will determine security level of the path  $SL_k(S_{ij}) = \min\{SL(A_i^S), SL(A_j^C)\}$ .
- Utility ( $U_k$ ): The quality of a multimedia session is represented by utility. Here, we use the data rate as the quality index. The utility function  $U_k$  for configuration  $k$  will be defined in the next section.

The best configuration  $S^*$  is selected by comparing scores of  $k$ , which is the addition of the normalized contribution from each criterion multiplied by the weight factors.  $S^*$  is defined as follows.

$$S^* = \arg_k \max(w_c C_k + w_p P_k + w_u U_k + w_{sl} SL_k), \quad (1)$$

$$\text{s. t } CB > T_B \text{ and } CS < T_s,$$

where  $w_c$ ,  $w_p$ ,  $w_{sl}$ , and  $w_u$  are weights for criteria cost, power, utility, and security level, respectively.  $CB$  is the set whose elements are adequate when they are bigger than a threshold  $T_B$ . Bandwidth is an example. Also,  $CS$  is the set whose elements, such as delay, are adequate when they are smaller than a threshold  $T_s$ . Based on (1), ABMNC is modeled as a MADM problem. The goal of the MADM hierarchy is to select the most suitable configuration  $k$  for transmission at the top level and solution alternatives are located at the bottom nodes. More information on mathematical model is presented in [6].

#### B. Utility-Based Media Assignment

In this section, a utility-based measurement scheme is presented to evaluate the quality of multimedia services, based on which, the media session assignment is formulated as a MKP, which is a generalization of the standard knapsack problem from a single knapsack to  $m$  knapsacks with different capacity. The objective is to assign each item to at most one of the knapsacks such that none of the capacity constraints are violated and the total profit of the items put into knapsacks is maximized.

The design of the utility functions is to strike a balance between system utilization and system QoS requirements. Let  $U_k^m$  be the utility function of a given configuration  $k$  for some medium  $m$ , where  $m \in \{\text{video}, \text{voice}, \text{data}\}$  and be defined as follows.

$$U_k^m = w_m \log_2 \left( 1 + \frac{x_m}{b_m} \right), \quad (2)$$

where  $X_m$  and  $b_m$  are the allocated and requested B.W. for media  $m$ , respectively.  $w_m$  is media weight which is obtained from AHP. Because humans are more sensitive to the voice (speech) than video, and data is considered to be the least important, the weight for video, voice, and data are  $w_{\text{voice}} > w_{\text{video}} > w_{\text{data}}$ .

In MKP, every path of different bandwidth between the caller and callee is considered as a bucket of different size. Each medium of different bandwidth requirement is considered an object of different size that to be included in one of many possible buckets. The objective is to assign objects to the available buckets such that we can get the maximal utility value while each bucket is not overflowed. Thus, given  $N$  different paths, the media assignment problem of the ABMNC is defined as follows.

$$U_k = \max \sum_{i=1}^N \sum_m U_k^m x_{im}$$

$$= \max \sum_{i=1}^N \sum_m w_m \log_2 \left( 1 + \frac{x_m}{b_m} \right) x_{im},$$

$$\text{s. t. } \sum_{i=1}^N X_m x_{im} \leq W_i, \quad (3)$$

$$\sum_{i=1}^N x_{im} = 1, x_{im} \in \{0,1\},$$

$$\sum_m w_m = 1, T_m \leq \frac{x_m}{b_m} \leq 1,$$

where  $U_k$  is the utility or scores acquired by optimizing the allocated rate for a configuration  $k$  comprising a set of dedicated connection(s),  $x_{im}$  denotes the  $m^{\text{th}}$  kind of media object assigned to the  $i^{\text{th}}$  path, and  $W_i$  denotes the available bandwidth for path  $i$ .

The first constraint ensures that the total allocated bandwidth to each path  $i$  will not exceed its limit (knapsack weight), and the second one indicates that each medium can only be assigned to one path.  $X_m/b_m$  is ranged from zero to one, so, the maximal value of  $U_k$  is one.  $T_m$  limits  $X_m$  to have a minimal threshold value to ensure an acceptable quality level for media  $m$ . A connection is feasible only if it has bandwidth larger than the threshold value.

#### IV. ABMNC

Selecting the best connection for multimedia session is a complicated process due to intensive comparison of a large number of alternatives. In this section, we propose a hybrid scheme of SAW and AHP in order to reduce the number of configurations (alternatives) and hence the computation complexity. AHP is used to decide relative importance of criteria and media, which is then used by SAW for scoring.

When the source handset want to start multimedia session it initiates a network discovery procedure to detect all available networks and candidate paths  $S_{ij}$ . These paths are input to our heuristic algorithm for path selection. To speed up the computation of MADM, we integrated the path selection scheme with our hybrid AHP and SAW approach. The hybrid AHP and SAW rates the path alternatives by considering all criteria, except utility, to produce a partial score first. We sort all candidate paths by their partial scores in a descending order. Then, we select paths based on their contribution on utility value or their bandwidth. After the heuristic algorithm selects a configuration, the overall score is computed by merging the utility value with media assignment. Fig.1 shows the process of multimedia session selection.

### A. SAW-AHP hybrid scheme

AHP illustrates the importance of each criterion by determining the relative weights of the compared criteria where weights are decided according to user's preferences. Since our approach deals with media sessions, AHP weights both criteria,  $w_c, w_p, w_{sl}$ , and  $w_u$ , and media which is used during media assignment.

Table I shows how AHP weights each media. Voice row shows that voice is twice as important as video, three times as important as data, and equal to itself. Video row shows the video is twice as important as data. We use Arithmetic Mean of the Normalized Column Vectors (AMNCV) and get  $w_{video} = 0.297$ ,  $w_{voice} = 0.539$ , and  $w_{data} = 0.164$ . By a similar process, we can readily obtain weight for  $w_c, w_p, w_{sl}$ , and  $w_u$  from AHP. AMNCV is defined as follows.

$$w_i = \frac{1}{|m|} \sum_{j=1}^{|m|} \frac{a_{ij}}{\sum_{i=1}^{|m|} a_{ij}} \quad i, j = 1, 2, \dots, |m|, \quad (4)$$

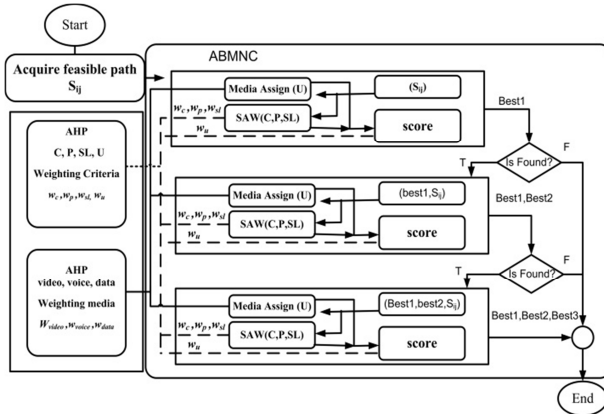


Figure 1. Process of multimedia session selection

TABLE I. COMPARISON MATRIX OF MEDIA WITH RESPECT TO THE CRITERION UTILITY

Utility	Video	Voice	Data
Video	1	1/2	2
Voice	2	1	3
Data	1/2	1/3	1

SAW is used to produce the overall score of an alternative and is computed as the weighted sum of all the attribute values. In our study, we use parameters for different access technologies and user preferences to construct a decision matrix with respect to cost, power, and security level. Then, we normalized the table based on whether the criteria is beneficial, e.g., security and utility or non-beneficial, e.g., cost and power. The best configuration is obtained by comparing the weighted average of all alternatives from the normalized decision matrix as follows.

$$MAX \sum_{j=1}^m w_j r_{ij}, \quad (5)$$

$M$  denotes the number of criteria and  $w_j$  is the weight of the  $j^{th}$  criterion.  $r_{ij}$  is the normalized score of the  $j^{th}$  criterion of the  $i^{th}$  alternative.

### B. Heuristic Algorithm.

Our heuristic algorithm classifies the sub-MADM hierarchies into three modes, single-path, dual-path, and triple-path. The first mode takes one path while the second takes two paths, and so on. Each mode outputs the best configuration as input to next mode if it exists. The next mode tries to find a better result by combining the input best configuration and one of the remaining  $S_{ij}$ . Note that each mode re-considers all remaining paths again for a better result. If a better configuration cannot be found, the result of the previous mode is the result. Also, the connection fails if nothing is found in the single-path mode. Table II shows notations used in the heuristic algorithm.

The heuristic algorithm is designed to choose the best configuration from a set of ranked paths for media transmission. The algorithm considers first the path with a larger score without utility value, and then a path with a larger bandwidth. It reduces the number of candidate paths by deleting a path with a lower bandwidth and ranking compared with the others.

Single-path, dual-path, and triple-path mode are shown in Algorithm 1. We specify the rules for the dual-path and the triple-path mode with some replacements to simplify the pseudo code because they're based on the same concept of the single-path mode. For example,  $SAW(f_i, f_{best1})$  at line 19 means scoring a new alternative  $(f_i, f_{best1})$  based on results at line 3.

#### ALGORITHM 1: Single-path heuristic path selection algorithm

```

1: BEGIN
2:    $S_k^* = \emptyset$  /*  $score_{all}(S_k^* = \emptyset)$  is zero*/
3:   SAW( $F$ )
4:    $\forall f_i \in F$  : Sort  $f_i$  in decreasing order by score.
5:    $\exists f_i \in F$  :  $B(f_i) \geq \sum b_m$  targets a  $f_i$  which has higher order
   and remove others lower than it from  $F$ .
6:    $Q = F$ 
7:   For  $f_i$  in  $Q$  remove  $f_{i+j}$  from  $Q$ 
   s.t.  $B(f_i) \geq B(f_{i+j}), 1 \leq j \leq |Q|$ . Endfor
8:   For  $f_i$  in  $Q$ 
9:     SPR( $f_i$ )
10:    Get the overall score of  $f_i$ 
11:    If ( $score_{all}(f_i) > score_{all}(S_k^*)$ )  $S_k^* = f_i$  Endif
12:   Endfor
    
```

**Step(3)**

13: **If** (IsUpdated(  $S_k^*$  ) is **TRUE**)  
 14:      $f_{best1} = S_k^*$   
 15:     **Jump** Dual-path  
 16: **Else Exit**  
 17: **Endif**

**Dual-path**

18: Repeat Step(1) but replace  $Q = F$  by  $Q = F - \{f_{best1}\}$  ,  
 19: Repeat Step(2) but insert **SAW** ( $f_i, f_{best1}$ ) after line 8,  
 20: Replace all  $f_i$  by ( $f_i, f_{best1}$ ) and **SPRA** by **DPRA**.  
 21: Repeat Step(3) but replace  $f_{best1}$  by ( $f_{best1}, f_{best2}$ ) ,  
 22: Replace Dual-path by Triple-path.

**Triple-path**

23: Repeat Step(1) but replace  $Q = F$  by  $Q = F - \{f_{best1}, f_{best2}\}$  .  
 24: Repeat Step(2) but insert **SAW** ( $f_i, f_{best1}, f_{best2}$ ) after line 10,  
 25: Replace all  $f_i$  by ( $f_i, f_{best1}, f_{best2}$ ) and **SPRA** by **TPRA**.  
 26: **END**

TABLE II. SUMMARY OF NOTATION FOR OUR HEURISTIC ALGORITHM

Symbol	Meaning	Symbol	Meaning
$S_k^*$	The final transmission configuration	$score$	Retrieving the saved partial score of an input configuration
$B$	Available bandwidth of some path	SPRA	Single-path rate allocation (media assignment)
$F$	A set of candidate paths.	DPRA	Dual-path rate allocation (media assignment)
$f_i$	Element in sorted $F$	TPRA	Triple-path rate allocation(media assignment)
$Q$	A set of good candidate paths	$f_{best1}$	The best result found in single-path mode
SAW	Score input configurations without utility	$(f_{best1}, f_{best2})$	The best result found in dual-path mode
$score_{all}$	Retrieve the saved overall score of an input configuration		

## V. PERFORMANCE EVALUATION

This section evaluates the performance of our proposed solution, ABMNC, through simulations. Fig.2 shows our simulation environment in which two handsets are willing to establish a multimedia session. Each handset is equipped with three interfaces. We assume there are nine feasible paths  $S_{ij}$  which satisfy the minimum bandwidth requirement, and  $S_{1j}, S_{2j}, S_{3j}$  denote the paths from source handset's interface of UMTS, WiMAX, and Wi-Fi to the  $j^{th}$  interface of destination, respectively.

Default simulation parameters used to analyse the performance are listed in the Table III. For different access technologies, the cost is referred by recent ISP charging mechanism, the power consumption, i.e., standby time is based on the specifications of some representative multimode handsets, and the relative values of security level are from some wireless technical reports. We assume monthly charged cost, and the cost is counted in both caller and callee no matter who initiates the call. Media weights are calculated from AHP.

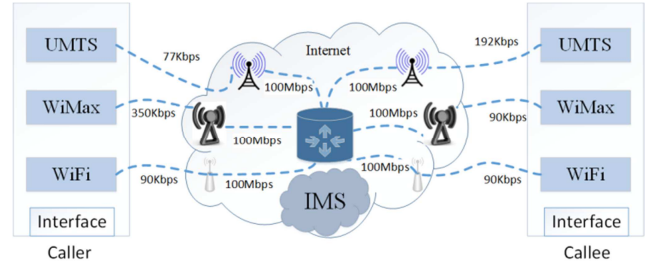


Figure 2. Environment for analysis

TABLE III. SIMULATION PARAMETERS

<b>Number of handsets</b>	2	<b>Cost for caller/UMTS</b>	0.0003\$/s
<b>Number of interfaces</b>	3	<b>Cost for caller/WiMax</b>	0.0005\$/s
<b>Number of feasible paths</b>	9	<b>Cost for callee/Wi-Fi</b>	0.0001\$/s
<b>Power/UMTS</b>	420mins	<b>Cost for callee/UMTS</b>	0.0005\$/s
<b>Power/WiMax</b>	180mins	<b>Cost for caller/WiMax</b>	0.0003\$/s
<b>Power/Wi-Fi</b>	210mins	<b>Cost for callee/Wi-Fi</b>	0.0002\$/s
<b>Security/UMTS</b>	Moderate	$w_{video}$	0.297
<b>Security/WiMax</b>	High	$w_{voice}$	0.539
<b>Security/Wi-Fi</b>	Low	$w_{data}$	0.164

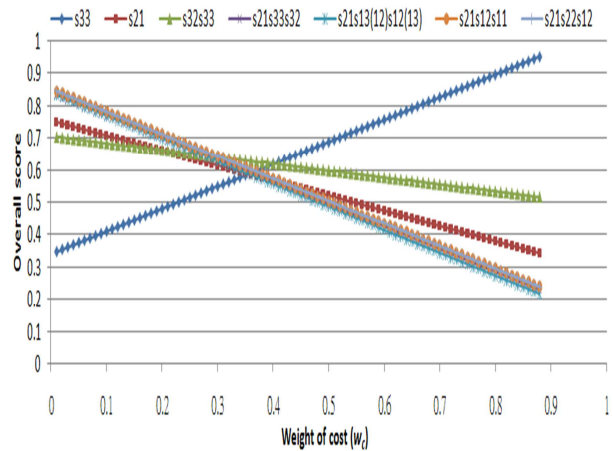


Figure 3. Scores of various alternatives

We compare ABMNC results with an exhaustive search (EX) method. The EX is a very general problem-solving technique that examines all configurations as alternatives. EX takes a large amount of time to complete a decision making when the number of candidates is large. In the simulation, the user is more interested in quality and cost, and cares less about security or power. Thus, based on user's preferences we set  $w_p = 0.05$ ,  $w_{sl} = 0.05$ , and  $w_u + w_c = 0.9$ . Fig. 3 illustrates the scores of all possible configurations at various weights of cost  $w_c$  and the respective weight of utility  $w_u$ .

Fig. 4 illustrates the decision making results by ABMNC and EX. We observed that the result converges to  $S_{33}$  when  $w_c$  increases because  $S_{33}$  has the smallest cost among all alternatives, and the cost dominates the overall scores when  $w_c$  is getting large. Thus,  $S_{33}$  is the best connection for general users that want to save money. On the contrary, configuration  $(S_{21}, S_{22}, S_{12})$  and  $(S_{21}, S_{12}, S_{11})$  are recommended for the users who strongly demand quality as  $w_u$  becomes large. One can observe that the overall score changes when weights change, which is a desirable feature for a scoring system. Note that the chosen configuration is not shown in this figure.

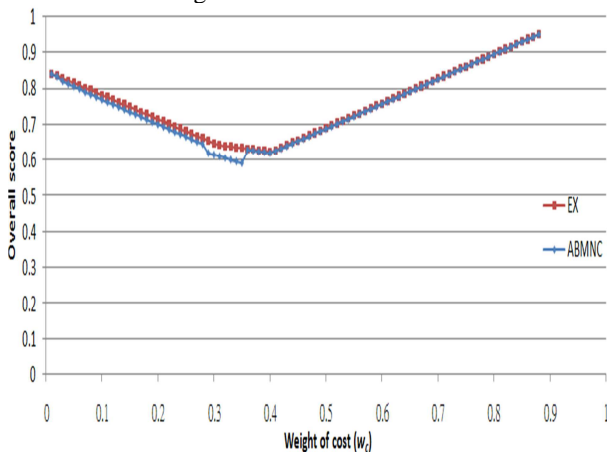


Figure 4. Comparison between ABMNC&EX

It can be observed from Fig. 4 that results are overlapped with one another except in the interval  $w_c = [0.25, 0.35]$ . This is due to the nature of the heuristic path selection algorithm. In this interval, ABMNC does not perform as well as the EX. EX chooses the dual-path or triple-path configurations rather than single-path  $S_{21}$  by ABMNC.

#### VI. CONCLUSION AND FUTURE WORK

In this paper, a new approach called ABMNC was presented to selecting best network connection for multimedia communication. Unlike all previous approaches, in ABMNC the problem was formulated as a MADM hierarchy with MKP media assignment. The complexity of the selection process was tackled by combining our hybrid SAW and AHP scheme with our novel heuristic algorithm, which all together work on reducing the number of alternative paths, and thus simplifying the computation

complexity. Simulation results show that ABMNC can choose network connections that match well with EX most of the time.

#### REFERENCES

- [1] C. Jain and D. Goodman, "General packet radio service in GSM," Communications Magazine, IEEE, vol. 35, 1997, pp. 122-131.
- [2] N. Man Hung, L. Shen-De, J. Li, and S. Tatesh, "Coexistence studies for 3GPP LTE with other mobile systems," Communications Magazine, IEEE, vol. 47, 2009, pp. 60-65.
- [3] Y. P. Chen and Y. H. Yang, "A new 4G architecture providing multimode terminals always best connected services," Ieee Wireless Communications, vol. 14, Apr 2007, pp. 36-41.
- [4] E. Gustafsson and A. Jonsson, "Always best connected," Ieee Wireless Communications, vol. 10, Feb 2003, pp. 49-55.
- [5] T. Y. Chung, F. C. Yuan, Y. M. Chen, and B. J. Liu, "S-3: Smart Session Selection for Voice Communications in Next Generation Wireless Network," Ieee Transactions on Fundamentals of Electronics Communications and Computer Sciences, vol. E91a, Oct 2008, pp. 2995-3002.
- [6] T.-Y. Chung, I. Mashal, F.-C. Yuan, Y.-H. Chiang, and O. Alsayrah, "A Novel Multiple Attributes Decision Making Approach For Multimedia Session Selection," to appear in ICCGI 2013.
- [7] V. Gazis, N. Alonistioti, and L. Merakos, "Toward a generic "always best connected" capability in integrated WLAN/UMTS cellular mobile networks (and beyond)," Ieee Wireless Communications, vol. 12, Jun 2005, pp. 20-29.
- [8] E. Stevens-Navarro and V. W. S. Wong, "Comparison between Vertical Handoff Decision Algorithms for Heterogeneous Wireless Networks," in Vehicular Technology Conference, 2006. VTC 2006-Spring. IEEE 63rd, 2006, pp. 947-951.
- [9] J. Jackson Juliet Roy, V. Vaidehi, and S. Srikanth, "Always Best-Connected QoS integration model for the WLAN, WiMAX Heterogeneous Network," in Industrial and Information Systems, First International Conference on, 2006, pp. 361-366.
- [10] F. Bari and V. C. M. Leung, "Automated network selection in a heterogeneous wireless network environment," Ieee Network, vol. 21, Jan-Feb 2007, pp. 34-40.
- [11] F. Bari and V. Leung, "Service delivery over heterogeneous wireless systems: networks selection aspects," presented at the Proceedings of the 2006 international conference on Wireless communications and mobile computing, Vancouver, British Columbia, Canada, 2006.
- [12] S. Qingyang and A. Jamalipour, "An adaptive quality-of-service network selection mechanism for heterogeneous mobile networks," Wireless Communications & Mobile Computing, vol. 5, Sep 2005, pp. 697-708.
- [13] H. Chaouchi and G. Pujolle, "A new handover control in the current and future wireless networks," Ieee Transactions on Communications, vol. E87b, Sep 2004, pp. 2537-2547.
- [14] E. Homayounvala and A. H. Aghvami, "User preference modelling for access selection in multiple radio access environments," Ieee Transactions on Communications, vol. E88b, Nov 2005, pp. 4186-4193.
- [15] H. Nguyen, H. Morikawa, and T. Aoyama, "Personal mesh: A design of flexible and seamless Internet access for personal area network," Ieee Transactions on Communications, vol. E89b, Apr 2006, pp. 1080-1090.
- [16] Y. Liu, "Access Network Selection in a 4G Networking Environment," master, University of Waterloo Canada, 2007.
- [17] J. W. Lee and S. H. Kim, "Using analytic network process and goal programming for interdependent information system project selection," Computers & Operations Research, vol. 27, Apr 2000, pp. 367-382.
- [18] W. Chen, "On the Problem and Elimination of Rank Reversal in the Application of TOPSIS Method," Operations Research and Management Science, vol. 14, Oct 2005.
- [19] L. Mohamed, C. Leghris, and A. Adib, "An Intelligent Network Selection Strategy Based on MADM Methods in Heterogeneous

Networks," International Journal of Wireless & Mobile Networks, vol. 4, Feb 2012, pp. 83-96.

- [20] L. Mohamed, C. Leghris, and A. Adib, "A Hybrid Approach for Network Selection in Heterogeneous Multi-Access Environments," in New Technologies, Mobility and Security (NTMS), 2011 4th IFIP International Conference on, 2011, pp. 1-5.
- [21] W. Lusheng and D. Binet, "MADM-based network selection in heterogeneous wireless networks: A simulation study," in Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology, 2009. Wireless VITAE 2009. 1st International Conference on, 2009, pp. 559-564.



# Reducing Power Consumption using Improved Wakelock on Android Platform

Joonkyo Kim and Jaehyun Park  
 School of Information and Communication Engineering  
 Inha University, Korea  
 Email: jkkim@emcl.org, jhyun@inha.ac.kr

**Abstract**—The power consumption is one of the most important issues on a mobile device because they usually use battery as a power source. The Android platform, a popular software platform for a hand-held device, also supports various power-saving schemes to reduce power consumption. This power-saving feature sometimes causes unwanted computational disruption. To avoid such disruption, the Android platform provides Wakelock to disable the power-saving mode temporarily. However, since Wakelock can be easily accessed through the user's API, improper use of Wakelock causes a huge extra power consumption. This paper proposes an improved Wakelock scheme that predicts the misuse of Wakelock. This improved Wakelock, called PR-Wakelock (Predict & self-Release Wakelock), does not only detect misuses of Wakelock, but also forcibly releases Wakelock for the system to go into a sleep mode. Several Android apps were implemented and an offline prediction software was also implemented on a linux platform for the simulating PR-Wakelock. The hit ratio of the proposed system obtained through an offline prediction software was close to 86.44%.

**Keywords**-Energy-aware systems; power management; Android; Wakelock

## I. INTRODUCTION

As the mobile technology evolves, the number of applicable areas of the hand-held devices including a smartphone has dramatically increased during the last decade. The power consumption of a mobile device has become an important issue because of the expanded application area and limited battery life. With an expansion of the application area, the power consumption of a mobile device has become an important issue because most of the hand-held devices use a battery as a power source. To reduce the power consumption and extend battery life, various dynamic power management technologies have been adopted in the hand-held devices.

Dynamic Voltage and Frequency Scaling (DVFS) technique is one of the key technologies used in modern mobile devices. It controls the clock frequency and CPU supply voltage in order to reduce the switching power consumption of the digital components [1], [2]. Since the power consumption ( $P$ ) is proportional to frequency ( $f$ ) and voltage ( $V$ ) as shown in (1), controlling the clock frequency of CPU and supplying voltage to CPU are effective ways to reduce power consumption at a hardware level [3].

$$P \propto f \cdot V^2 \quad (1)$$

Hence, the key idea of DVFS technique is to control both the supply-voltage and operating-frequency of CPU according to the software workload. To predict the workload of software, statistical analysis of the runtime distribution of software workload is usually used [4], [5].

Another power-saving technique at hardware level is Dynamic Power Management (DPM), which predicts the idle

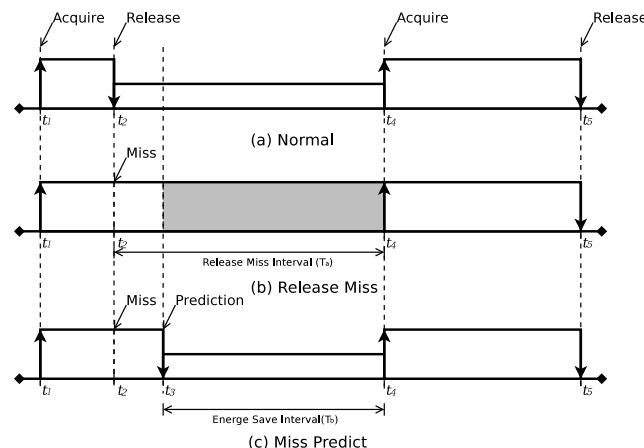


Figure 1. Behaviors of Wakelock

time of the system and controls the power supply to the unused components such as external peripherals as well as CPU's internal modules [6], [7]. To predict the idle time of the system, Program Counter-based Access Predictor (PCAP) algorithm that monitors the applications' function-calls based on the I/O system-calls, was proposed [8]. Previous work showed that DPM using a hardware-based PCAP algorithm can reduce the power consumption up to 29% [9].

Even though the hardware-based power management techniques described above are effective to reduce the overall power consumption of mobile devices, they may cause unnecessary side effects such as slowing down of an application program or frequently disabling peripherals that may lose connectivity. To compensate these side effects that DVFS/DPM may cause, these power management features are able to be controlled by software explicitly in most mobile platforms. To run a background process without disruption on an Android platform, which is one of the most widely used software platforms in the mobile device market, a special power management module called *Wakelock* was introduced. Wakelock prevents CPU from entering the sleep or power-saving mode even when a mobile device is in an idle state such as a screen-saving mode. Even though Wakelock is a useful feature to manage power-saving feature, it should be handled very carefully because misuse of Wakelock means CPU never goes into power-saving mode, and, in turn, a huge amount of power may be wasted [10]. Fig. 1 shows the timeline of normal and abnormal use-case of Wakelock. Fig. 1(a) shows the timeline of normal case of Wakelock. Since Wakelock is acquired at  $t_1$ , which means that the power-saving feature is disabled, and released at  $t_2$  normally, the system may enter into the power-saving mode until  $t_4$  at which Wakelock is acquired again. However, if Wakelock is not released adequately at  $t_2$  because

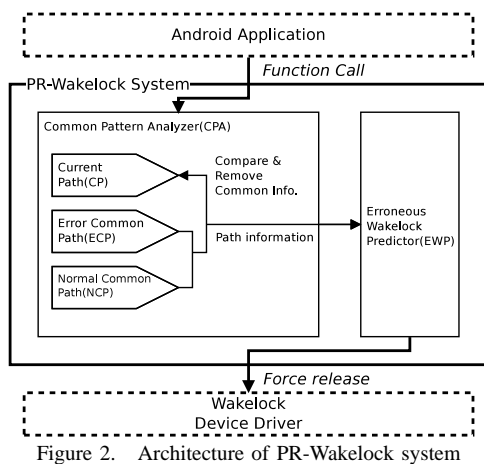


Figure 2. Architecture of PR-Wakelock system

of software bugs, the system remains in an active mode with power-saving mode disabled as shown in Fig. 1(b). In this case, unnecessary power consumption is expected during  $T_a$  period. Such inadequate situations of Wakelock have been reported occasionally in many Android applications because Wakelock can be accessed by any application through user’s API [11]. Unfortunately, a systematic safeguard to prevent such undesirable situation does not exist in the Android platform. To lessen this kind of Wakelock error, analyzing method of the program at compile time was proposed recently [11]. The compile-time analysis, however, has limitations because every possible software flow cannot be analyzed at compile time without actual running it.

To overcome this limitation, this paper proposes a run-time algorithm to prevent power waste caused by unreleased Wakelock. Fig. 1(c) shows the benefits of the proposed algorithm in this paper. Even if Wakelock is not released at  $t_2$ , runtime algorithm proposed in this paper can detect unreleased Wakelock and forcibly release it at  $t_3$ , which prevents waste of power consumption during this unnecessary wake-up period,  $T_b$ .

This paper describes the key idea and algorithm of the run-time Wakelock predictor. However, since this work is still at a work-in-process stage, the implementation result is not shown in this paper. Instead, offline analysis software will prove the feasibility of the proposed idea.

This paper consists of five sections including this introduction. In Section II, the basic concept of PR-Wakelock system is proposed. For evaluating the performance of PR-Wakelock system, offline simulation results are shown in Section III. Finally, conclusions and future works are described in Section IV and Section V, respectively.

## II. IMPROVED WAKELOCK SYSTEM

As described above, since unreleased Wakelock may cause a huge waste of power in the Android platform, this paper proposes an improved Wakelock system, which is called PR-Wakelock (Predict & self-Release Wakelock) as shown in Fig. 2. PR-Wakelock consists of two major modules; *Common Pattern Analyzer* (CPA) and *Erroneous Wakelock Predictor* (EWP).

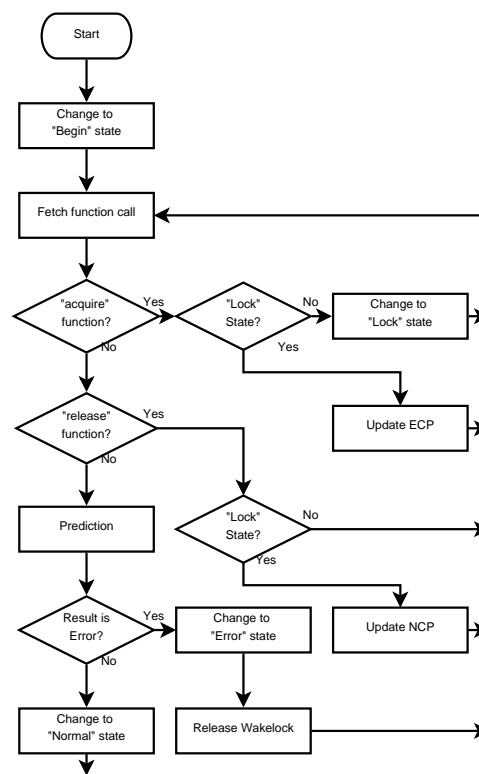


Figure 3. Flowchart of prediction algorithm

In this system, analyzing the sequence of *Wakelock-acquire* and *Wakelock-release* function-call is the key feature for predicting the erroneous Wakelock operation. CPA updates saved common paths when *Common Path*(CP) is revised. It checks sequence of the Wakelock function. If the *Wakelock-acquire* and *Wakelock-release* functions are invoked in order, CP stands for normal path because Wakelock is released normally after acquired. However, if *Wakelock-acquire* function of Wakelock is invoked twice in order, CP is at an erroneous path because Wakelock is locked until the next request. The algorithm of the PR-Wakelock system is presented in Fig. 3. For predicting erroneous Wakelock operation, EWP monitors function-calls at Lock or Normal state, and it performs prediction whenever function-call is detected. If EWP predicts that the current CP might be a part of an erroneous Wakelock sequence, the current Wakelock is forcibly released by EWP.

### A. Common Path Analyzer

To analyze the usage of Wakelock in an application software, it is necessary to monitor and record the executing pattern of application software in real-time manner. The execution patterns are recorded as a unit of *Path*, defined as a “sequence of function calls” in this paper. CPA module in Fig. 2 forms a Path from the executing sequence of function-calls in real-time. Since only the usage of Wakelock is important in this paper, CPA module in Fig. 2 starts recording the executing sequence of function-calls whenever *Wakelock-acquire* function is invoked, and stops recording if another *Wakelock-acquire* or *Wakelock-release* function is invoked. The recorded Path in real-time is saved as CP (Current Path). Besides this CP, in CPA module, Error Common Path (ECP) and Normal Common Path (NCP) are also managed. NCP means that a

		0	A	G	C	A	T
0	0	0	0	0	0	0	0
G	0	0	1	1	1	1	1
A	0	1	1	1	2	2	2
C	0	1	1	2	2	2	2

Figure 4. Example of Analyzing Subsequence

part of software sequence that has both *Wakelock-acquire* and *Wakelock-release* function. This NCP stands for normal situation of Wakelock system because *Wakelock-release* function is invoked after *Wakelock-acquire* function. On the contrary, ECP represents a part of software sequence that doesn't have *Wakelock-release* after *Wakelock-acquire*. Using these three Paths', CPA analyzes the characteristics of software modules and provides information to determine erroneous situation of Wakelock.

CPA analyzes the recorded CP when another *Wakelock-acquire* or *Wakelock-release* function is invoked again. If *Wakelock-release* function ends the current recording sequence, the path stored in CP is considered as a normal Wakelock pattern and updated into NCP. If another *Wakelock-acquire* function is invoked, on the other hand, this path is considered as an erroneous case of Wakelock and updated in ECP. To extract the feature of executing pattern, *Common Path*, defined as a subsequence of function calls, is extracted from CP after applying Longest Common Subsequence (LCS) algorithm to NCP and ECP that stores the latest common path of normal and erroneous case, respectively [12], [13]. CPA also compares the subsequences between ECP and NCP in order to remove the common path between them because both ECP and NCP may have the Wakelock-related path in common. The LCS-based common path analyzing algorithm used in CPA is depicted in Fig. 4, in which two subsequence, AGCAT and GAC, are compared. LCS algorithms can build the table in Fig. 4 using a dynamic programming technique [12], [13]. Each alphabet in the figure means function name. The circled numbers are common functions, which is related to common path. As a result, CPA extracts AC as a common path from these two sequences. On the other hand, numbers with rectangle are different path position. These can be selected for removing common path from original paths. After removing common path from NCP and ECP, the remaining subsequence, GAT and G, are updated to NCP and ECP, respectively. CPA manages NCP and ECP in this manner and supply to the EWP so that it can detect erroneous situation.

### B. Erroneous Wakelock Predictor

During recording the path into CP, EWP shown in Fig. 2 predicts a possible Wakelock errors whenever CP is updated. EWP determines whether current usage of Wakelock might be error or not based on the similarity between CP and either NCP or ECP. The similarity between the two paths are defined as the length of common path that can be determined by LCS algorithm. By comparing the similarity, EWP decides if the current executing sequence might be an error if the length of common path between CP and ECP is longer than the length

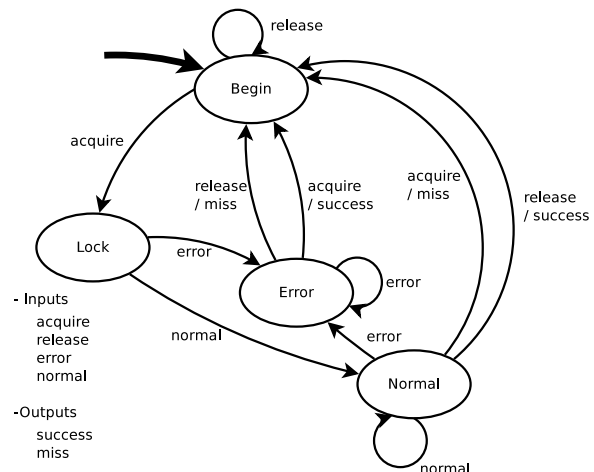


Figure 5. State of PR-Wakelock System

between CP and NCP as shown in (2). In (2)  $\mathcal{L}$  denotes length of sequence and  $\sqcap$  operator denotes LCS algorithm.

$$\mathcal{L}(\sqcap(CP, NCP)) + \Delta < \mathcal{L}(\sqcap(CP, ECP)) \quad (2)$$

However, the subsequence stored in CP is not a complete sequence but a portion of it, because CP stores only executing sequence path after *Wakelock-acquire* function is invoked. This means that CP may contain insufficient information to predict accurately. To increase the accuracy of prediction regardless of incomplete information of CP, predictor maintains a threshold value,  $\Delta$ , in (2) adaptively. To correct current information, EWP monitors function calls continuously until second Wakelock function is invoked, even though EWP decided current state. And EWP decides whether prediction is correct or not after it confirms second Wakelock function. At the beginning,  $\Delta$  starts from zero. If the prediction is confirmed as wrong, threshold value increases.

Fig. 5 shows the state diagram of predictor. EWP starts from *Begin* state, and returns to *Begin* state when the path from CP is completed recording. If Wakelock is acquired at sometime, EWP moves to *Lock* state and starts prediction to decide whether CP is error or not. EWP can predict *Normal* state despite of CP is error actually, because CP is not enough long to predict correctly. So EWP continues prediction until another Wakelock function is invoked even though current state is *Normal*. On the other hand, if EWP decides CP is error, state moves to *Error* state. And then PR-Wakelock system forcibly releases Wakelock.

### III. SIMULATION RESULTS

To verify the feasibility of the proposed Wakelock system, a simulation system was implemented on Samsung Galaxy Note (SHV-E160K), which is based on Android 4.0.3 (Ice Cream Sandwich). In order to test the algorithm, three Android applications were implemented for simulating erroneous situation. First application was based on a socket application, which connected to a server and periodically sent data through network. In this case, Wakelock was acquired when this application tried to connect the socket, and released when transmission of data was finished normally. Second application was based on I/O operation of file system that tries to write

TABLE I. SIMULATION RESULTS

	Socket program	I/O program	Music player
Average	77.45	86.44	74.58
Maximum	84	90.28	79.31
Minimum	56.5	81.21	68.10
Std. Dev.	10.69	2.72	3.28
$\Delta$ (Final value)	15	17	23.2

random data to SD Card. Wakelock was used between the open and close functions of the *Writer* class. Last one was the music player application. That played MP3 music when user requested, and Wakelock was used to prevent stopping of the music when the system entered the power-saving mode. In all three applications, invocation of the *Wakelock-release* function was omitted randomly with uniform distribution.

During the software execution, the function-calls are logged using the *Debug* class provided by Android API. The function name, thread ID, and calling time were logged in order to analyze path information. Since the purpose of the test-bed system was to demonstrate the feasibility of the proposed algorithm, the actual analysis was not performed in real-time yet. Rather, the logged information was analyzed by the separate analysis and prediction software that was implemented on a Linux platform.

The prediction accuracy of the implemented CPA and EWP was evaluated through 10 executions of each application programs and the results were shown in Fig. 6 and Table I. It showed about 86.44% of predictions are valid on flash I/O based application. Because similarity between NCP and ECP is different according to application,  $\Delta$  in (2) was also different.

#### IV. CONCLUSION

This paper proposes PR-Wakelock, which is an improved Wakelock algorithm for an Android platform. The proposed algorithm analyzes the behavior of Wakelock through the function-call analysis and predicts possible errors of Wakelock. Once improper operation of Wakelock is found, Wakelock is automatically released and CPU enters power-saving mode to reduce power consumption. Three Android softwares are implemented for simulating erroneous Wakelock behaviors, and offline analyzing software was also implemented on Linux platform for predicting error of Wakelock. According to the offline analysis, the prediction ratio of PR-Wakelock system was up to 86.44%. So, result shows that Wakelock error detection from function-call path analysis is feasible enough to apply on a real device.

#### V. FUTURE WORKS

PR-Wakelock system proposed in this paper has not been implemented yet. However, it can be implemented on Android system in order to monitor the operation of Wakelock in real-time, and, in turn, reduce the power consumption of Android-based mobile system. For this, CPA module can be implemented using *Debug* class device driver to monitor function calls without modifying Android framework itself. With implementing PR-Wakelock, the actual power consumption can be analyzed in the future work.

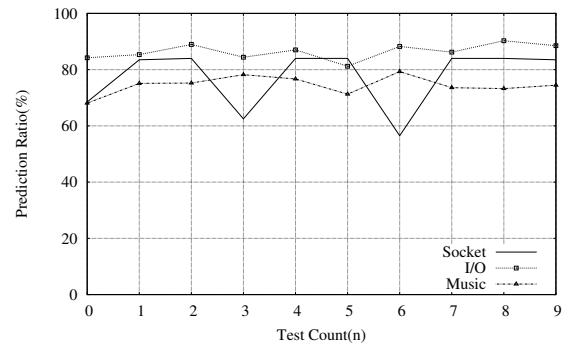


Figure 6. Simulation Results

#### REFERENCES

- [1] J.-B. Lee, M.-J. Kim, S. Yoon, and E.-Y. Chung, "Application-support particle filter for dynamic voltage scaling of multimedia applications," *IEEE Trans. Comput.*, vol. 61, no. 9, Sep. 2012, pp. 1256–1269.
- [2] X. Chen, C. Xu, and R. P. Dick, "Memory access aware on-line voltage control for performance and energy optimization," in *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2010, Nov. 2010, pp. 365–372.
- [3] W.-Y. Liang, S.-C. Chen, Y.-L. Chang, and J.-P. Fang, "Memory-aware dynamic voltage and frequency prediction for portable devices," in *14th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications*, 2008, pp. 229–236.
- [4] J. Kim, S. Yoo, and C.-M. Kyung, "Program phase and runtime distribution-aware online dvfs for combined vdd/vbb scaling," in *Design, Automation Test in Europe Conference Exhibition, 2009, DATE '09*, Apr. 2009, pp. 417–422.
- [5] —, "Program phase-aware dynamic voltage scaling under variable computational workload and memory stall environment," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 30, no. 1, Jan 2011, pp. 110–123.
- [6] L. Cai, N. Pettis, and Y.-H. Lu, "Joint power management of memory and disk under performance constraints," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 25, no. 12, Dec 2006, pp. 2697–2711.
- [7] W.-K. Lee, S.-W. Lee, and W.-O. Siew, "Hybrid model for dynamic power management," *IEEE Trans. Consum. Electron.*, vol. 55, no. 2, May 2009, pp. 656–664.
- [8] C. Gniady, A. R. Butt, Y. C. Hu, and Y.-H. Lu, "Program counter-based prediction techniques for dynamic power management," *IEEE Trans. Comput.*, vol. 55, no. 6, Jun. 2006, pp. 641–658.
- [9] Y.-S. Hwang, S.-K. Ku, and K.-S. Chung, "A predictive dynamic power management technique for embedded mobile devices," *IEEE Trans. Consum. Electron.*, vol. 56, no. 2, Jul. 2010, pp. 713–719.
- [10] "PowerManager.WakeLock—Android Developer." [Online]. Available: <http://developer.android.com/reference/android/os/PowerManager.WakeLock.html>. [retrieved: Feb, 2013]
- [11] A. Pathak, A. Jindal, Y. C. Hu, and S. P. Midkiff, "What is keeping my phone awake?: characterizing and detecting no-sleep energy bugs in smartphone apps," in *Proceedings of the 10th international conference on Mobile systems, applications, and services MobiSys '12*, 2012, pp. 267–280.
- [12] D. S. Hirschberg, "Algorithms for the longest common subsequence problem," *Journal of ACM*, vol. 24, no. 4, Oct. 1977, pp. 664–675.
- [13] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, "Introduction to Algorithms, 2nd Ed." The MIT Press, 2001.

## Using an Individual-Based Model of Uneven-Aged Forests for Studying Trade-off Between Timber Production and Deadwood Preservation

Bruno Bonté  
 Irstea, LISC and Irstea, EM  
 Clermont-Ferrand, France  
[bruno.bonte@irstea.fr](mailto:bruno.bonte@irstea.fr)

Valentine Lafond and Thomas Cordonnier  
 Irstea EM  
 Grenoble, France  
[\[first-name\].\[last-name\]@irstea.fr](mailto:[first-name].[last-name]@irstea.fr)

Jean-Denis Mathias  
 Irstea, LISC  
 Clermont-Ferrand, France  
[jean-denis.mathias@irstea.fr](mailto:jean-denis.mathias@irstea.fr)

**Abstract**—Integrating the multi-functional role of forests in forestry practices constitutes a challenging example of complex system management. Usually, win-win situations between functions are seldom and trade-offs have to be considered. This paper proposes a framework to study dynamical relationships between two important functions in forests: timber production and biodiversity conservation. We built an individual-based model of uneven-aged forests that explicitly takes into account timber harvesting options and dead wood dynamics. Dead wood compartment was selected because it represents a relevant surrogate for biodiversity in forests. We used dynamics metrics based on viability theory framework to evaluate simulations that contrasted different thinning intensities and thinning frequencies. Thanks to this model and the metrics used, we are able to discuss optimal strategies for preserving biodiversity while guaranteeing timber production.

**Keywords**—*Individual-based model; complexity; forest management; multi-functional performance.*

### I. INTRODUCTION

Multi-functional management is a challenging issue, which can be generalized for most social-ecological systems. The multi-functional paradigm of forest management is nowadays very common. Forests have multiple roles and provide multiple services for society, including not only timber production but also species conservation. Proposing tools that allow forest managers to fulfill different functions in a given place is therefore of growing concern [1][2]. However, models cannot represent all functions and thus, simulation studies can hardly determine the best choice to make [1]. Nevertheless, models can be used to better understand or evaluate different situations regarding to a specific point of view.

In this paper, we try to provide a common framework to researchers and decision makers, who want to evaluate forest management units regarding to two functions: timber production and biodiversity preservation. Our aim is to Specify entirely the model and the indicators used, in order to enables researchers to modify both, silvicultural scenarios, and forest dynamics model parameters. Hence, it will be possible to evaluate and compare different contexts, based on

the common rigorous and explicit theoretical framework provided by the model specification.

Consequently, the objective of this paper is: i) to propose an individual-based model that enables the evaluation of forest management within the framework of multi-functionality; ii) to discuss the trade-off between timber production and biodiversity preservation using a risk-based performance criteria to evaluate forestry practices. From the point of view of renewable resource management, we use the viability theory framework of resilience [3] to establish system performance criteria representing ecological biodiversity resilience [4]. This kind of approach is also related to the concept of permanence already used to evaluate forest capacity to ensure several functions [5].

The paper is split in four parts. In Section II, we present the general method adopted. In Section III, we present and specify the model used to simulate forest dynamics. In Section IV, we present and specify the forest management algorithm. Finally, in Section V, we present simulations results.

### II. MATERIAL AND METHOD

Forest dynamics modelling and simulation have been widely studied during these past years, providing spatial models of forests, validated for several tree species and forest management types. Those models can be used to test and understand trade-offs offered by some given forestry practices, regarding to some given functionalities [6].

However, these models cannot take into account all the functions of the forest, which can be defined at different spatial scales. For this reason, instead of trying to find the best trade-off between ecosystem services [1][6], we rather try to identify the solutions that better provide all services considered, i.e., solutions that are not pareto-dominated. The set of solutions found can then be used as a reference to make decision. Following this approach, we do not need to use commensurable criteria to evaluate the different services provided by the forest. Particularly, it enables us to use dynamics metrics adapted to each service considered. We can thus represent the fact that, for instance, the timber production function can be averaged over time, whereas the

insect habitat providing function must be permanently supplied.

Consequently, our methodological proposal is to design a simple individual-based model to simulate trajectories of these indicators for forest stands managed under different silvicultural scenarios. Applying dynamics metrics to these trajectories, we are able to compare the criteria values for the different scenarios.

We first identify the indicators to be monitored, in relation to the timber production and biodiversity preservation functions. The indicator chosen to monitor timber production is the harvested timber volume ( $V_t$ ) at each time  $t$ . Only the mean value of this indicator is important here: the dynamics of stand timber production is not important because wood supply permanency should be evaluated at a wider spatial scale than the one considered in silvicultural management. Consequently, the performance criteria considered for timber production is the mean harvested timber volume over a given period ( $\bar{V}_r$ ). We chose two indicators related to the biodiversity preservation function. In a review dedicated to indicators of biodiversity for ecologically sustainable management, Lindenmayer et al. [7] suggests two categories of indicators: taxon-based indicators based on the presence/absence of some species, and structure-based indicators based on the structure of habitats at the stand or landscape scale. Here, we decided to focus our work on structure-based indicators. We focus on two habitats known to be linked to the biodiversity of saproxylic beetles (beetles that depend on dead wood) [8][9] [10]: the average dead wood volume in the stand, denoted  $I_1$ , and the volume of adult dead trees in the stand (i.e., number of dead trees with diameter greater than 17.5cm), denoted  $I_2$ .

Both indicators ( $I_1$  and  $I_2$ ) can be monitored over time. We use resilience criteria for providing a way to compare dynamic trajectories of the indicators. For each structural indicator  $I_1$  and  $I_2$ , we consider a minimum value  $I_1^{\min}$  and  $I_2^{\min}$  and we compute a resilience criteria  $\gamma(I_k)_{k=\{1,2\}}$ , as the percentage of time, during which the minimum acceptable value is not reached, over a given time period  $T$  (1).

$$\gamma(I_k) = \frac{E \left[ \int_0^T \mu(I_k(t)) dt \right]}{T} \quad (1)$$

where ,

$$\begin{aligned} \mu(I_k(t)) &= 1 \text{ if } I_k(t) > I_k^{\min} \\ &= 0 \text{ otherwise} \end{aligned}$$

Mathematical expected value  $E \left[ \int_0^T \mu(I_k(t)) dt \right]$  is computed as an average over several replications of the model, which involves stochastic processes. This resilience criterion highlights the ability of the system to continuously provide saproxylic habitats despite of cyclic timber removals.

### III. MODELLING FOREST DYNAMICS

We consider a mono-specific uneven-aged forest. We use a spatial individual-based model, which takes into account interactions between trees. It is a demographic model, which simulates the processes of growth, mortality and recruitment. We choose a diameter-based dendrometric model because it has a low computational cost, compared to more complex individual-based models. The growth sub-model is based on the simplification of the *Samsara2* model developed for Norway spruce and Silver fir in French mountain forest [11]. Other processes are calibrated to obtain qualitative results similar to the *Samsara2* model.

A tree  $k$  (adult or juvenile) is characterised by its position in the 2D space (we write  $x_k \in R^2$ ) and its diameter at breast height:  $d_k \in R$ . We also consider coarse woody debris characterised by a volume  $v_{w,k}$ , a diameter  $d_{w,k}$  and an age  $a_{w,k}$ . Coarse woody debris can be either a dead tree (wind-throw, other natural death, ..) or parts of a tree that have been removed. We do not consider this distinction in this study. System state at time  $t$  is then characterised by the tree population of  $N_t$  trees  $(x_k, d_k)_{1 \leq k \leq N_t}$  and the coarse woody debris population of  $N_w$  coarse woody debris  $(v_{w,k}, a_{w,k}, d_{w,k})_{1 \leq k \leq N_w}$ .

Thereafter, we use basal area instead of diameter without any loss of information. The basal area is a value commonly used in forest management and measures the cross-section of tree trunks and stems at 1.30m high. Basal area of tree  $k$  is noted  $g_k$  and computed as specified in (2).

$$g_k = \frac{\pi d_k^2}{4} \quad (2)$$

A diameter-volume function noted  $V(d)$  enables us to convert diameter values to volume values. We used a simple classical volume tariff called Algan n°9 for diameters greater than 15cm and linearised this tariff for diameters smaller than 15cm. This function is parametrised with a circumference metric decrease of 6 cm per meter.

#### A. Competition

For all processes, a perfect one-sided competition  $C(x_k, d_k)$  expressed in ( $m^2/ha$ ) was considered, as specified in (3).

$$C(x_k, d_k) = \sum_{i|d_i \geq d_k} g_i \omega(x_i - x_k) \quad (3)$$

where  $\omega$  is a light interception kernel designed as a spatial kernel of parameter  $\sigma$ :

$$\begin{cases} \omega(x' - x) = \frac{1}{\pi\sigma^2}, & \text{if } \|x' - x\| \leq \sigma \\ \omega(x' - x) = 0, & \text{otherwise} \end{cases} \quad (4)$$

### B. Growth process

The growth process equation used for all trees (diameter greater than 7.5cm) is given by the basal area growth, computed as a potential growth  $Pot$ , multiplied by a reduction function  $Red$  (5 to 7). The potential function used is the potential growth used in the *Samsara2* model [11]. The reduction function  $Red$  has been chosen for its simplicity and calibrated to obtain similar simulation results as the *Samsara2* model using perfect one-sided competition.

$$\frac{dg_k}{dt}(x_k, d_k) = Pot(x_k, d_k) * Red(x_k, d_k) \quad (5)$$

with

$$Pot(x_k, d_k) = r_0 \frac{d_k^{r_1}}{1 + r_2 e^{r_3 d_k}} \quad (6)$$

and,

$$Red(x_k, d_k) = u_0 e^{-u_1 C(x_k, d_k)} \quad (7)$$

where,  $(r_i)_{(0 \leq i \leq 3)}$ , and  $(u_i)_{(0 \leq i \leq 1)}$  are species specific parameters related respectively to potential growth and competition.

### C. Mortality process

A logistic model is used for mortality. Probability of mortality per year  $P$  is given in (8).

$$P(x_k, d_k) = \max(P_1(x_k, d_k), P_2(x_k, d_k)) \quad (8)$$

This mortality function is composed of two functions  $P_1$  and  $P_2$  designed respectively for young trees and old trees.  $P_1$  is based on the French forestry inventory data from IFN (the regression method is not detailed here). Forestry inventories mostly contain small and medium diameter trees. According to literature, large tree mortality is higher than the mortality of small trees. Thus function  $P_2$ , is used to correct mortality rate for large diameter trees.  $P_1(x_k, d_k)$  and  $P_2(x_k, d_k)$  are given in (9) and (10).

$$P_1(x_k, d_k) = \frac{1}{1 + e^{p_0 + p_1 \cdot d_k - p_2 \cdot d_k^2 - z \cdot C(x_k, d_k)}} \quad (9)$$

$$P_2(x_k, d_k) = \frac{1}{1 + e^{-p_3 - p_4 d_k^2}} \quad (10)$$

where  $(p_i)_{(0 \leq i \leq 4)}$  and  $z$  are species specific parameters related respectively to potential death and death linked to competition.

### D. Recruitment process

Trees are recruited with a diameter of 7.5 cm. The probability that a tree is recruited in a position  $x'$  depends on the past environmental conditions at this position. We consider the number of seeds that have been spread at this position and the competition that this position has been subjected to over the past twenty-five years. At time  $t_0$  each tree  $k$  can recruit a young tree situated in a free position  $x'$  with the probability  $R(x', x_k, d_k, t_0)$  computed as in (11).

$$R(x', x_k, d_k, t_0) = \frac{b \cdot g_k \sum_{t=t_0-24}^{t_0} \omega_D(x' - x_k) e^{-sC(x', 7.5)}}{25} \quad (11)$$

where  $b$  is the potential recruitment rate;  $s$  is the sensitivity to light;  $C(x', 7.5)$  is the local competition for light on the position  $x'$  for a tree of diameter inferior than 7.5 cm as presented in (3);  $\omega_D$ -function is a dispersal function computed as a spatial kernel of parameter  $\sigma_D$  as the  $\omega$  function presented in (4).

### E. Coarse woody debris generation and decomposition

Each time a tree  $(x_k, d_k)$  dies, a new coarse woody debris  $i$  is generated with:  $v_{w,i} = V(d_k)$ ,  $d_{w,i} = d_k$  and  $a_i = 0$ .

Once a coarse woody debris has been generated, it is subjected to natural decomposition and its volume decreases according to (12).

$$\frac{dv_{w,k}}{dt} = -\alpha \cdot v_{w,k} \quad (12)$$

where  $\alpha$  is the decomposition rate. Furthermore, any coarse woody debris with  $v_{w,k} < 0.0001 \text{ m}^3$  is removed from the system. Age of the dead wood piece increases with time as described in (13).

$$\frac{da_{w,k}}{dt} = dt \quad (13)$$

## IV. MODELLING FOREST MANAGEMENT

In this part, we model management actions on the forest stand. The minimum diameter for removal is noted  $D_{min}$ . In the following, we note  $V_r$  the total volume removed. We consider a simplified view of forestry practices performed in the case of uneven-aged forest management (see [11] for a more complete simulation of uneven-aged management). A silvicultural scenario is defined by a removal intensity (removal stops when a total of  $\Delta ba$  per year has been removed) and the minimum interval during two removals ( $\Delta t$ ). To these forest management factors, we add 2 factors for dead wood management:

•  $p_a$  is the proportion of the dead wood basal area removed at each removal. This is related to the possibility of removing some recently dead trees. This parameter is interesting regarding conservation strategy (i.e., dead tree retention or not);

•  $p_e$  is the proportion of wood volume exported when a tree is removed. Meaning that  $(1-p_e)$  stays on the stand. This parameter can be related to biomass energy practices (i.e., whole tree extraction or not).

#### A. Removal decision

At time  $t$ , we note  $V_{cwd}$  the set of deadwood removable during the previous  $\Delta_t$  time period:

$$V_{cwd} = \{(v_{w,k}, a_{w,k}, d_{w,k}) \mid d_{w,k} > D_{min} \text{ and } a_{w,k} < \Delta_t\} \quad (14)$$

Then the basal area available for removing can be approximated by the value  $BA_a$ :

$$BA_a = \sum_{i \mid d_i > D_{min}} g_i + p_a \sum_{j \mid (v_{w,j}, a_{w,j}, d_{w,j}) \in V_{cwd}} \pi \frac{d_{w,j}^2}{4} \quad (15)$$

The removal decision is the following:

“A removal is performed at time  $t$  if and only if at this time  $BA_a \geq \Delta ba$  AND the time spent since the previous removal is greater than  $\Delta_t$ .”

If  $BA_a < \Delta ba$  when the time step since the previous removal is  $\Delta_t$ , the removal period is increased continuously until the date  $t_0$ , for which  $BA_a \geq \Delta ba$ . Then, the same procedure occurs at time  $t_0 + \Delta_t$ .

#### B. Removal process

First, dead trees to remove are successively withdrawn from the set  $V_{cwd}$  by decreasing order of diameter, until the proportion  $p_a$  of the total basal area of  $V_{cwd}$  is reached. Whenever a dead tree  $k$  is removed at time  $t$ , we observe the updates described by (16) to (18).

$$v_{w,k}(t+dt) = (1-p_e)v_{w,k}(t) \quad (16)$$

$$d_{w,k}(t+dt) = D_{min} \quad (17)$$

and

$$V_r(t+dt) = V_r + p_e v_{w,k}(t) \quad (18)$$

Then,  $\Delta ba$  living trees with diameter greater than  $D_{min}$  are randomly selected and harvested until  $\Delta ba$  is reached, unless  $\Delta ba$  has already been reached with dead trees. The probability  $P_k$  that an individual  $k$  is

removed at each successive step is given in (19) (see [11] for a justification):

$$P_k = \left( \frac{d_k}{\max_{i \in S} d_i} \right)^2 \quad (19)$$

where  $S = \{(x_i, d_i) \mid d_i \geq D_{min}\}$  is the set of trees that have not been removed. When a tree  $k$  is removed, a new dead tree  $l$  is created. Its volume is  $v_{w,l} = (1-p_e)V(d_k)$ , and its diameter is  $D_{min}$ . Then, the removed volume is increased as follows:

$$V_r(t+dt) = V_r(t) + p_e V(d_k) \quad (20)$$

where  $V(d_k)$  is given by the volume computation function.

## V. RESULTS

#### A. Experimental design

We performed simulations with parameter values described in Table 1. Initial states are obtained by simulating 2000 years with no management, followed by 1750 years with the chosen management strategy. It enables us to reach pseudo-equilibriums of managed forests. With this experimental design, we tested three different values for harvested quantities, presented as a yearly basal area harvested ( $\Delta ba$ ), as well as for the minimum removal period ( $\Delta_t$ ). Tested values are presented in Table 2. Nine different strategies are defined from the combination of these factors. In addition, a scenario with no removal is presented. Twenty replications were performed for each removal strategy.

TABLE I. VALUES OF THE PARAMETERS

$D_{min}$	$p_a$	$p_e$	$d_0$	$d_1$	$d_2$	$d_3$	$r_0$	$r_1$	$r_2$
27.5	0.9	0.9	-4.37	-0.014	$4.10^{-4}$	-9	9.64	0.24	1.11

$r_3$	$u_0$	$u_1$	$b$	$s$	$\sigma$	$\sigma_0$	$\alpha$
-0.10	0.85	0.035	11	0.05	10	15	0.07

TABLE II. VALUES TESTED FOR THE TWO SIVICULTURAL FACTORS

$\Delta ba^{min}$ ( $m^2 \cdot ha^{-1} \cdot year^{-1}$ )	0.2	0.4	0.6
$\Delta_t$ (year)	5	10	15

For each combination of factors values, we computed the timber production criterion  $\bar{V}_r$  and the two biodiversity preservation criteria  $\gamma(I_r)$  and  $\gamma(I_d)$ . As said previously, the first indicator ( $I_1$ ) is the total volume of deadwood on the stand. It can be computed at any time step as presented in (21).

$$I_1 = \sum_{0 \leq k \leq N_w} v_{w,k} \quad (21)$$

A minimum value of 15  $m^3/ha$  of dead wood in European forest stands seems to be reasonable regarding to actual knowledge [12]. Consequently,  $I_1^{min}$  is equal to 15  $m^3$ .



The second indicator ( $I_2$ ) does not correspond to a specific recommendation, but enables us to discuss about a qualitative aspect of the dead wood stock. As dead wood diversity plays an important role in saproxylic beetles diversity [10], we monitor the volume of adult dead trees on the stand:

$$I_2 = \sum_{\{k|d_{w,k} > 17.5\}} v_{w,k} \quad (22)$$

We chose a limit of 4 m<sup>3</sup> corresponding to the mean volume of two adult dead trees. Consequently,  $I_2^{min}$  is equal to 4m<sup>3</sup>.

*B. Simulation results*

The mean timber production and the resilience of the dead wood stock  $\gamma(I_1)$  (% of time in which dead wood volume is greater than 15 m<sup>3</sup>) for each removal strategy are presented in Figure 1. The scenario without any removal is represented in this Figure as a green diamond. We can see that in this case the production is null and the dead wood stock resilience is 100%, meaning that the volume of the dead wood stock ( $I_1$ ) is always greater than 15 m<sup>3</sup> ( $I_1^{min}$ ) for this scenario.

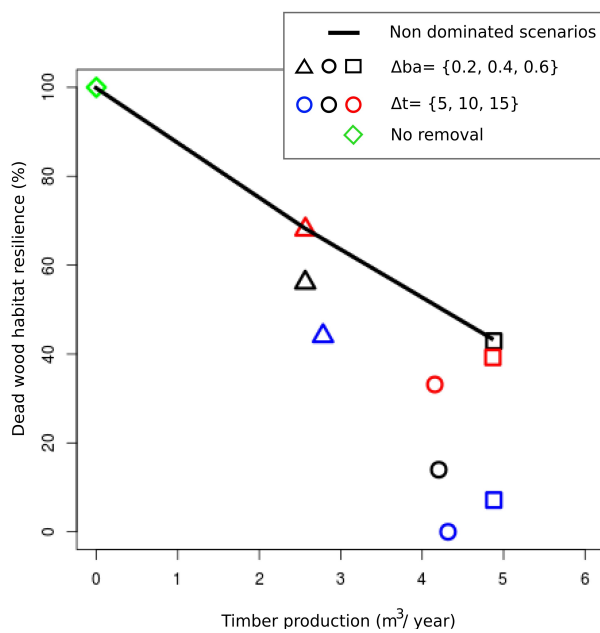


Figure 1. Timber mean production and dead wood habitat resilience for different removal strategies.

We first notice that increasing the time lap between removals has a low impact on production (less than 5%), but increases dead wood habitat resilience. It is mainly due to the fact, that increasing time lap between removals enables the system to regenerate deadwood, despite of the regular timber removals. However, for  $\Delta ba = 0.6$ , high time laps ( $\Delta_t = 10$  and  $\Delta_t = 15$ ) between timber removals, lead to equivalent value of the resilience criteria.

Then, increasing the harvest basal area leads to a non-linear increase of the timber production, whereas it decreases

the dead wood stock resilience indicator  $\gamma(I_1)$  in most cases. However, increasing the harvest basal area from 0.4 m<sup>2</sup>.ha<sup>-1</sup>.y<sup>-1</sup> to 0.6 m<sup>2</sup>.ha<sup>-1</sup>.y<sup>-1</sup> does not increase the production criteria, as much as increasing it from 0.2 m<sup>2</sup>.ha<sup>-1</sup>.y<sup>-1</sup> to 0.4 m<sup>2</sup>.ha<sup>-1</sup>.y<sup>-1</sup>. The first reason is that the higher the removal intensity is, the younger (and thus the smaller) the trees are. Consequently, for the same harvest basal area, the timber volume removed is lower. The second reason is that the whole basal area of 0.6 m<sup>2</sup>.ha<sup>-1</sup>.y<sup>-1</sup> cannot be harvested at every removal period. Consequently the removal period is increased, in respect to the removal algorithm, until the harvest basal area is available. This is the reason why, even if the mean timber production is higher, resilience indicator  $\gamma(I_1)$  for  $\Delta ba = 0.6$  is higher than that for  $\Delta ba = 0.4$ . Interval between removals is increased and thus resilience is higher.

Figure 2 shows the same results but considering the volume of adult dead trees, instead of the whole dead wood volume (% of time in which adult dead trees ( $d_{w,k} > 17.5$  cm) volume is greater than 4m<sup>3</sup>). Once again, the resilience indicator is 100% in the scenario with no removal. Like  $\gamma(I_1)$ ,  $\gamma(I_2)$  increases when the time lap between removal increases. However, unlike  $\gamma(I_1)$ , excepted for the five year removal period where the same phenomenon occurs, adult dead trees resilience is higher for an average harvest basal area (circles), than for a high removal (squares). This is due to the fact that, when high values of basal area are removed at the same time, most of adult trees are removed, and thus the forest is mostly composed of young trees.

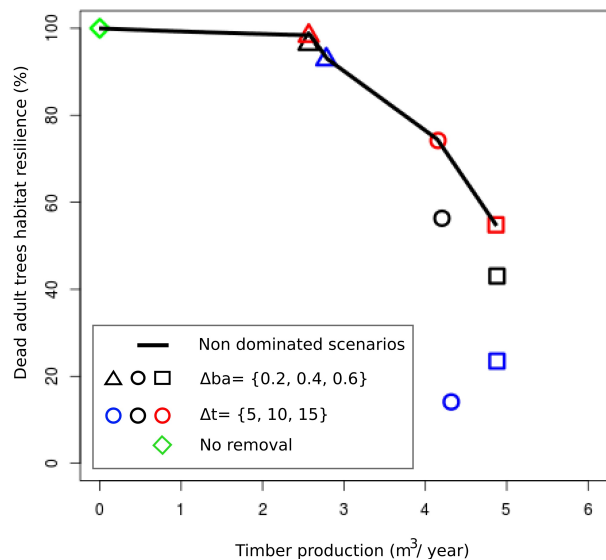


Figure 2. Timber mean production and adult dead trees habitat resilience for different removal strategies.

Finally, Figures 1 and 2 enable us to identify silvicultural scenarios that are non-dominated by neither a criterion nor the other. These scenarios are linked by black lines in both Figures. In regard to these two functions, preferred scenarios chosen should be on these lines. Furthermore, pareto-

dominated scenarios can be compared relatively to their distances to the non-dominated scenarios for each criteria.

## VI. CONCLUSION

Multi-functional forest management raises several issues, at the political point of view, to design forest management policies at regional scales, as well as at the technical point of view, to design forestry practices at the management unit scale.

We developed a simple individual-based model, in order to test the impact of forestry practices on indicators linked to timber production and biodiversity resilience. We showed that this model enables decision makers, to discuss the trades-off offered by different forestry practices, concerning timber production and some indicators of biodiversity preservation.

Beyond forest management, the multi-functional management issue can be extended to most social-ecological systems [13]. Considering the more general problem of multi-functional socio-ecosystem management, we think that this work provides an interesting point of view, where a generic model is designed to evaluate multi-criteria decisions when criteria are not commensurable and similar situations are numerous.

One of the main advantages of this model lies in the spatialization of processes. It will enable us to consider more complex forestry practices. Besides, in future work, we plan to propose and test adaptive forest managements, relying on simple indicators that the forest manager can easily estimate, such as the potential biodiversity index developed for assessing potential biodiversity in forest stands [2].

## REFERENCES

- [1] J. Kangas and A. Kangas, "Multiple criteria decision support in forest management - the approach, methods applied, and experiences gained", *Forest Ecology and Management*, vol. 207(1-2), Mar. 2005, pp. 133-143.
- [2] L. Larrieu and P. Gonin, "The potential bioersvity index (pbi) - a quick and simple method for assessing potential biodiversity in forest stands", *Revue Forestiere Francaise*, vol. 60(6), 2008, pp. 727-748.
- [3] S. Martin, "The cost of restoration as a way of defining resilience: A viability approach applied to a model of lake eutrophication", *Ecology and Society*, vol. 9(2), 2004, art. 8.
- [4] C. S. Holling, "Resilience and stability of ecological systems", *Annual Review of Ecology and Systematics*, vol. 4, 1973, pp. 1-23.
- [5] T. Cordonnier, B. Courbaud, F. Berger, and A. Franc, "Permanence of resilience and protection efficiency in mountain norway spruce forest stands : A simulation study", *Forest Ecology and Management*, vol. 256(3), 2008, pp. 347-354.
- [6] W.S. Schwenk, T.M. Donovan, W.S. Keeton, and J.S. Nunery, "Carbon storage, timber production, and biodiversity: Comparing ecosystem services with multi-criteria decision analysis", *Ecological Applications*, vol. 22(5), 2012, pp. 1612-1627.
- [7] D.B. Lindenmayer, C.R. Margules, and D.B. Botkin., "Indicators of biodiversity for ecologically sustainable forest management", *Conservation Biology*, vol. 14(4), 2000, pp. 941-950.
- [8] C. Bouget, A. Lassauce, and M. Jonsell, "Effects of fuelwood harvesting on biodiversity - a review focused on the situation in europe", *Canadian Journal of Forest Research*, vol. 42(8), 2012, pp. 1421-1432.
- [9] A. Lassauce, Y. Paillet, H. Jactel, and C. Bouget, "Deadwood as a surrogate for forest biodiversity: Meta-analysis of correlations between deadwood volume and species richness of saproxylic organisms", *Ecological Indicators*, vol. 11(5), 2011, pp. 1027-1039.
- [10] A. Brin, C. Bouget, H. Brustel, and H. Jactel, "Diameter of downed woody debris does matter for saproxylic beetle assemblages in temperate oak and pine forests", *Journal of Insect Conservation*, vol. 15 (5), 2011, pp. 653- 669.
- [11] V. Lafond, G. Lagarrigues, T. Cordonnier, and B. Courbaud, "Uneven-aged management options to promote forest resilience for climate change adaptation: effects of group selection and harvesting intensity", *Annals of Forest Science*, 2013, pp. 1-14, doi:10.1007/s13595-013-0291-y.
- [12] J. Muller and R. Butler, "A review of habitat thresholds for dead wood: a baseline for management recommendations in European forests", *European Journal of Forest Research*, vol. 129(6), 2010, pp. 981-992.
- [13] E. Ostrom, "A general framework for analysing sustainability of social- ecological systems", *Science*, vol. 325(5939), 2009, p.p. 419-422

# Stability Analysis of Global FCFS and Presorting Service Discipline

Willem Mélange, Joris Walraevens, Dieter Claeys, Bart Steyaert and Herwig Bruneel  
 Department of Telecommunications and Information Processing  
 Ghent University - UGent  
 E-mail: {wmelange,jw,dc,bs,hb}@telin.UGent.be

**Abstract**—In this paper, we consider a continuous-time queueing system with two different types (1 and 2) of customers with two dedicated servers (also named 1 and 2). This means server 1 (2) can only serve customers of type 1 (2). The goal of this paper is to determine the stability condition for our system with global first-come-first-serve (FCFS) and presorting service discipline, i.e., all arriving customers are accommodated in one single FCFS queue, regardless of their types, with an exception of the first  $N$  customers. For the first  $N$  customers the FCFS rule holds only within the types, i.e., customers of different types can overtake each other in order to be served. The motivation for our work comes from traffic and is to be able to give advise about the optimal length of filter lanes, i.e., lanes reserved for vehicles making a specific turn at a junction.

**Keywords**—queueing, stability, blocking, global FCFS, presorting.

## I. INTRODUCTION

The motivation for this work is an every day problem in traffic. Traffic jams might occur for multiple reasons. One reason are traffic junctions. Consider, for instance, the following situation: vehicles approach a junction with two possible destinations (1 and 2) as seen in Fig. 1. In traffic context, it is often not physically feasible to provide two separate lanes for each possible destination (as seen in Fig. 1(b)). If it would be, vehicles for both directions can be kept apart completely. The other extreme occurs much more frequently, namely, when there is one lane on the main road (Fig. 1(a)). In the case where there is only one lane on the main road, it is possible that vehicles on that road heading for destination 1 may be hindered or even blocked by vehicles heading for destination 2, even when the subroad leading to destination 1 is free, simply because cars that go to 2 are in front of them. In other words, there is a first-come-first-serve (FCFS) order on the main road regardless what destination they have. In the rest of this paper we will call this service discipline global FCFS (gFCFS). In queueing theory terms, a service discipline where there are 2 types of customers that are accommodated in a single queue and who are served in a FCFS manner regardless of their type. When we look at the case in Fig. 1(a), there is even a global first-in-first-out (gFIFO) order on the main road. At any given time, at most one server will be working. A possible way to minimize the impact of this blocking phenomenon is the use of filter lanes, i.e., lanes reserved for vehicles making a specific turn at a junction (as seen in Fig. 1(c)). It is clear that in this case, we cannot longer talk about gFCFS or gFIFO as service discipline. In the rest of the paper we will call this new service discipline, which can be seen as sort of relaxation of the

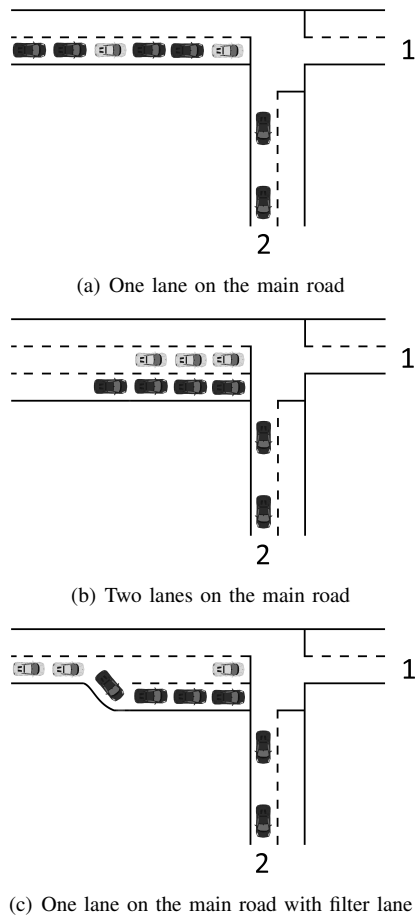


Fig. 1. Light grey vehicles with destination 1 and dark grey vehicles with destination 2 approaching a traffic junction

gFCFS service discipline, gFCFS with presorting (P-gFCFS). Again in queueing theory terms, a service discipline where there are 2 types of customers that are accommodated in a single queue and who are served in a FCFS manner regardless of their type with an exception for the first  $N$  customers. Thus, the customers can only be served if they are one of the first  $N$  customers in the system and there are no customers of the same type in front of them. Fig. 1(c) is an example of such a system with a P-gFCFS service discipline. In this specific case,  $N$  is equal to 4. If the vehicle is in the first 4 vehicles on the main road and there is no vehicle in front of this vehicle with the same direction, the vehicle will be able to drive without delay caused by other vehicles to the destination that vehicle desires.

However, if there are 4 vehicles with another destination in front of this vehicle, even when there are no vehicles in front of this vehicle with the same destination, the vehicle is not able to drive without delay caused by other vehicles to his destination (as seen in Fig. 1(c)). We refer to [1], [2] for a general overview and validation of modelling traffic flows with queueing models.

Analogously, at a security checkpoint (e.g., at an international airport or train station) people are usually body-searched by someone of the same gender. As a result, when a group of friends of the same gender arrive, the people of the opposite gender behind them may have to wait until the whole group has been checked, even when the other security person is available, at least when it is not allowed to overtake at the security checkpoint (which is often the case for security reasons). Here, we can also have an relaxation of the gFCFS service discipline. The security person can hand-pick a person from the waiting line to come to the front of the line. We also presume that the security person will only do this for one of the first  $N$  persons in the waiting line.

In [3] and [4], we already got some insight in the impact of the blocking phenomenon caused by a gFCFS service discipline on the performance of the involved systems (or a P-gFCFS with parameter  $N$  equal to 2). As stated earlier, in this paper, we want to relax the gFCFS rule and get some insight in the impact of this relaxation. In other words, we want to investigate the impact of the  $N$ -gFCFS service discipline on the performance of the involved systems.

The structure of the rest of the paper is as follows: we first tackle in Section II the (more simple) problem where the types in the arrival stream of customers are independent. In this Section, first the mathematical model is described in Subsection II-A. Next we analyse the stability condition in Subsection II-B. Then we tackle in Section III the (more difficult) problem where the types in the arrival stream of customers are dependent. The same structure is used as in the Section II where the types of customers in the arrival stream are independent. The paper continues with a discussion about the results and some numerical examples in Section IV. Finally, some conclusions are drawn and future research is suggested in Section V.

## II. UNCORRELATED TYPES IN THE ARRIVALS

We start with the case of uncorrelated types in the arrivals. This is for instance an adequate model for traffic junctions, as the destination of consecutive cars can largely be regarded as independent.

### A. Mathematical Model

We consider a continuous-time queueing model with infinite waiting room. There are two servers, where server 1 is working at rate  $\mu_1$  and server 2 at rate  $\mu_2$  (exponential service times). There are two types (classes) of customers. Each of the two servers is dedicated to a given class of customers. In this case, server 1 always serves customers of type 1 and server 2 always serves customers of type 2. The customers are served as follows: if both types are present in the first  $N$  customers in the system, the first customer of each type can be served by its server. The customers not in the first  $N$  customers are

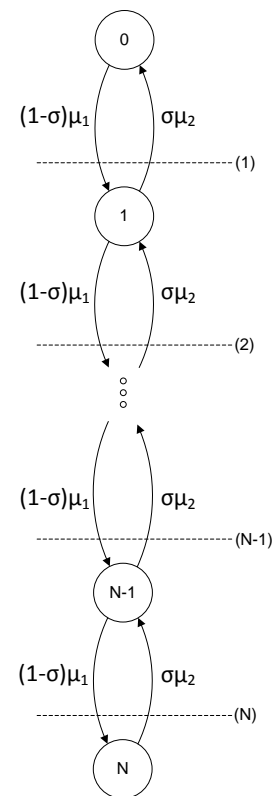


Fig. 2.  $(N + 1)$ -state Markov chain to determine the stability condition of the system with P-gFCFS with uncorrelated types in the arrival stream

served in a global FCFS way, i.e., they are blocked not only by customers of their own type in front of them, but also by customers of different type. We will call this service discipline global FCFS with presorting (P-gFCFS). For example, if the first  $i - 1$  customers are of type 1 and the  $i$ -th customer is of type 2, then this customer can be served by server 2 if  $i \leq N$ . However, if the first  $N$  customers are of type 1 and the  $N + 1$ -th customer is of type 2, then this customer cannot be served by server 2 even if the server is idle.

The customers enter the system according to a Poisson arrival process with mean arrival rate  $\lambda$ . With probability  $\sigma$ , the customer is of type 1 and with probability  $1 - \sigma$  the customer is of type 2.

### B. Analysis of the Stability Condition

When looking at the stability condition, we can presume that the system is constantly provided with new customers and the system will therefore be filled with at least  $N$  customers all the time. Note that we are only interested in the number of customers of type 1 and 2 in the first  $N$  customers of the system. Thus, the exact queueing order of the types of the first  $N$  customers is of no importance. These observations lead to the  $(N + 1)$ -state Markov chain in Fig. 2. The state  $m$ ,  $m$  customers of the first  $N$  customers are of type 2 (and thus  $N - m$  of type 1). The rate to go from state  $m$  to state  $m - 1$ , is  $\sigma\mu_2$ ; namely a rate  $\mu_2$  to end the service in state  $m$  of the customer with type 2 multiplied with the probability  $\sigma$  that the new  $N$ -th customer of our system is of type 1. Similarly, the rate to go from state  $m$  to state  $m + 1$ , is  $(1 - \sigma)\mu_1$ .

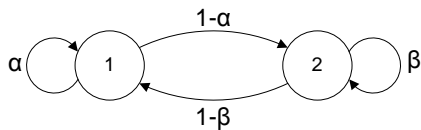


Fig. 3. 2-state Markov chain to determine the type of an arriving customer

Fig. 2 models the well-known birth-and-death process for a  $M|M|1|N$  queue [5] and the probability to be in state  $m$  is known to be given by

$$p(m) = \frac{\left(\frac{(1-\sigma)\mu_1}{\sigma\mu_2}\right)^m \left(1 - \frac{(1-\sigma)\mu_1}{\sigma\mu_2}\right)}{1 - \left(\frac{(1-\sigma)\mu_1}{\sigma\mu_2}\right)^{N+1}}. \quad (1)$$

Having obtained the  $p(m)$ 's, we can now move to the stability condition. Therefore we postulate that the average amount of work per unit time that enters the system ( $\gamma$ ) is smaller than the average amount of work the system can execute per unit time, i.e., the average amount of work the system would execute per unit time when it would be constantly provided with new customers. Here, the system is able to execute 2 units of work per unit of time when both servers are able to work (when the system is in one of the states 1 to  $N-1$ ). The system is able to execute 1 unit of work per unit of time when only one server is able to work (when the system is in state 0 or  $N$ ). The stability condition is thus

$$\gamma < p(0) + 2 \sum_{m=1}^{N-1} p(m) + p(N) \quad (2)$$

$$\gamma < \frac{\left(1 + \frac{(1-\sigma)\mu_1}{\sigma\mu_2}\right) \left(1 - \left(\frac{(1-\sigma)\mu_1}{\sigma\mu_2}\right)^N\right)}{1 - \left(\frac{(1-\sigma)\mu_1}{\sigma\mu_2}\right)^{N+1}}. \quad (3)$$

where  $\gamma$  (the average amount of work that enters the system per unit time) is given by

$$\gamma \triangleq \lambda \left( \frac{\sigma}{\mu_1} + \frac{1-\sigma}{\mu_2} \right). \quad (4)$$

Equation (3) can then be written as

$$\lambda < \frac{\left(\frac{\sigma}{\mu_1}\right)^N - \left(\frac{1-\sigma}{\mu_2}\right)^N}{\left(\frac{\sigma}{\mu_1}\right)^{N+1} - \left(\frac{1-\sigma}{\mu_2}\right)^{N+1}}. \quad (5)$$

which says that on average, there are not more arrivals than service completions.

### III. CORRELATED TYPES IN THE ARRIVALS

We now turn to the case that some correlation in the types of consecutively arriving customers is present. This can, for instance, be the case in the modelling of a security check point, where partners of different sex (negative correlation) may arrive more frequently, or groups of people, all of the same sex (positive correlation).

#### A. Mathematical Model

The model is the same as in Section II-A except for the arrival stream. Now the customers enter the system according to a Poisson arrival process with mean arrival rate  $\lambda$ . The type of the arriving customer is determined by a two-state Markov chain (see Fig. 3). If the previous customer is of type 1, then the customer is of type 1 with probability  $\alpha$  and of type 2 with probability  $(1-\alpha)$ . If the previous customer is of type 2, then the customer is of type 1 with probability  $(1-\beta)$  and of type 2 with probability  $\beta$ . Notice here already that we can transform  $\alpha$  and  $\beta$ , in two other parameters  $\sigma$  and  $K$  that have a more intuitive meaning. The transformations from  $(\alpha, \beta)$  to  $(\sigma, K)$  are

$$\sigma = \frac{1-\beta}{2-\alpha-\beta}, \quad (6)$$

$$K = \frac{1}{2-\alpha-\beta} \quad (7)$$

and from  $(\sigma, K)$  to  $(\alpha, \beta)$  are

$$\alpha = 1 - \frac{1-\sigma}{K}, \quad (8)$$

$$\beta = 1 - \frac{\sigma}{K}. \quad (9)$$

The intuitive meaning behind the parameter  $\sigma$  is the given relative frequency distribution of the type of the customers. The fraction of customers that are of type 1 (2) is  $\sigma$  ( $1-\sigma$  respectively). The parameter  $K$  on the other hand gives a clear indication about the correlation. The parameter is directly proportional to the mean number of customers of the same type that arrive back-to-back. More specifically, we have

$$\begin{aligned} E[\text{number of customers of type 1 arriving back-to-back}] \\ = \frac{1}{1-\beta} = \frac{K}{\sigma}, \end{aligned} \quad (10)$$

$$\begin{aligned} E[\text{number of customers of type 2 arriving back-to-back}] \\ = \frac{1}{1-\alpha} = \frac{K}{1-\sigma}, \end{aligned} \quad (11)$$

where  $E[\dots]$  stands for the the expected value of what's between brackets. Notice here that when  $K$  equals 1, the types of customers in the arrival stream are uncorrelated, and the model transforms to that of Section II.

#### B. Analysis of the Stability Condition

We take the same approach as in Section II-B. The Markov chain corresponding to Fig. 2 is more complicated in this case. Now we do not only have to keep track of the number of type 1 and 2 customers in the first  $N$  customers, but also of the type of the "last" customer in this set of customers. These observations lead to the  $2N$ -state Markov chain in Fig. 4, where in state  $(m, t)$ ,  $m$  customers of the  $N$  first customers are of type 2 (and thus  $N-m$  of type 1) and the "last" customer is of type  $t$ . Notice that we do not have states  $(0, 2)$  and  $(N, 1)$  since the "last" customer cannot be of type 2 (1) if all  $N$  customers are of type 1 (2).

The balance equations (see transitions through the dotted lines  $(1b)$  to  $((N-1)b)$  in Fig. 4) are

$$\mu_2 p(m, 1) = \mu_1 p(m, 2), \quad (12)$$

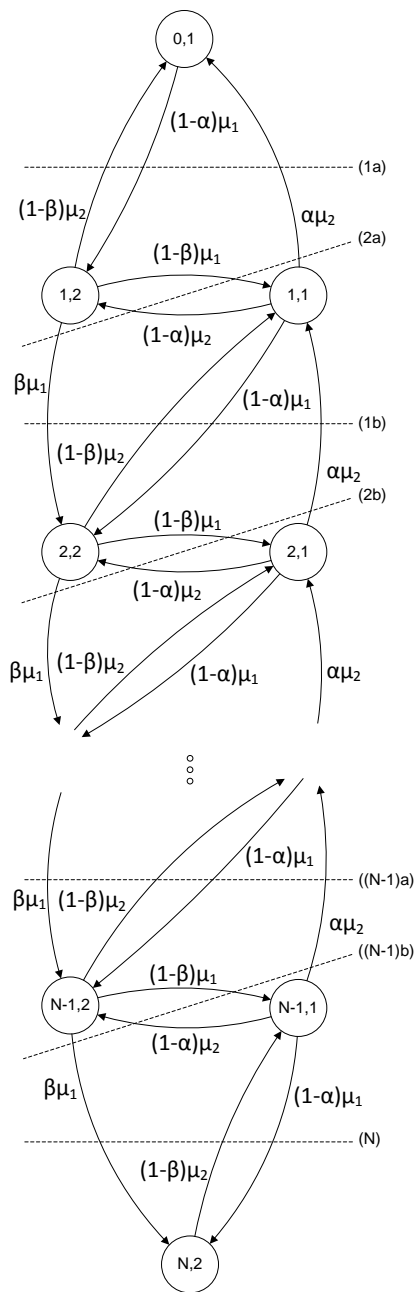


Fig. 4.  $2n$ -state Markov chain to determine the stability condition of the system with P-gFCFS with correlated types in the arrival stream

where  $m = 1, \dots, N - 1$  and (see transitions through the dotted lines (2a) to  $((N - 1)a)$  in Fig. 4)

$$\begin{aligned} & (1 - \beta)\mu_2 p(m, 2) + \alpha\mu_2 p(m, 1) \\ &= (1 - \alpha)\mu_1 p(m - 1, 1) + \beta\mu_1 p(m - 1, 2), \end{aligned} \quad (13)$$

where  $m = 2, \dots, N - 1$ . Equations (12) and (13) can be rewritten as

$$p(m, 1) = \frac{\mu_1}{\mu_2} p(m, 2), \quad (14)$$

$$\begin{aligned} & ((1 - \beta)\mu_2 + \alpha\mu_1) p(m, 2) \\ &= \left( (1 - \alpha)\frac{\mu_1^2}{\mu_2} + \beta\mu_1 \right) p(m - 1, 2) \end{aligned} \quad (15)$$

or even further as

$$p(m, 1) = \frac{\mu_1}{\mu_2} p(m, 2), \quad (16)$$

$$p(m, 2) = \frac{(1 - \alpha)\mu_1^2 + \beta\mu_1\mu_2}{(1 - \beta)\mu_2^2 + \alpha\mu_1\mu_2} p(m - 1, 2). \quad (17)$$

This yields for  $m = 1, \dots, N - 1$

$$p(m, 1) = \frac{\mu_1}{\mu_2} \theta_1^{m-1} p(1, 2) \quad (18)$$

$$p(m, 2) = \theta_1^{m-1} p(1, 2), \quad (19)$$

where

$$\theta_1 = \frac{(1 - \alpha)\mu_1^2 + \beta\mu_1\mu_2}{(1 - \beta)\mu_2^2 + \alpha\mu_1\mu_2}. \quad (20)$$

The balance equation corresponding with transition (1a) reads

$$(1 - \alpha)\mu_1 p(0, 1) = (1 - \beta)\mu_2 p(1, 2) + \alpha\mu_2 p(1, 1) \quad (21)$$

and using (12)

$$p(1, 2) = \frac{(1 - \alpha)\mu_1}{(1 - \beta)\mu_2 + \alpha\mu_1} p(0, 1). \quad (22)$$

Using (22) in (18) and (19), we find

$$p(m, 1) = \frac{\mu_1}{\mu_2} \theta_1^{m-1} \frac{(1 - \alpha)\mu_1}{(1 - \beta)\mu_2 + \alpha\mu_1} p(0, 1), \quad (23)$$

$$p(m, 2) = \theta_1^{m-1} \frac{(1 - \alpha)\mu_1}{(1 - \beta)\mu_2 + \alpha\mu_1} p(0, 1). \quad (24)$$

The last balance equation corresponding with transition (N) leads to

$$p(N, 2) = \frac{(1 - \alpha)\mu_1^2 + \beta\mu_1\mu_2}{(1 - \beta)\mu_2^2} p(N - 1, 2), \quad (25)$$

where we used (12) to eliminate  $p(N - 1, 1)$ . The normalization condition

$$\sum_{m=0}^N (p(m, 1) + p(m, 2)) = 1, \quad (26)$$

where  $p(0, 2) = p(N, 1) = 0$  by definition, finally yields  $p(0, 1)$ . Just as in Section II-B we can write the stability condition as the inequality that the average amount of work per unit time that enters the system ( $\gamma$ ) is smaller than the average amount of work the system can execute per unit time. In this case, the stability condition is given by

$$\gamma < p(0, 1) + 2 \sum_{m=1}^{N-1} (p(m, 1) + p(m, 2)) + p(N, 2) \quad (27)$$

$$\gamma < \frac{(1 + a\theta_1^{N-1})(\theta_1 - 1) + 2b(\theta_1^{N-1} - 1)}{(1 + a\theta_1^{N-1})(\theta_1 - 1) + b(\theta_1^{N-1} - 1)} \quad (28)$$

with

$$a = \frac{(1 - \alpha)\mu_1}{(1 - \beta)\mu_2}, \quad (29)$$

$$b = \left( \frac{(1 - \alpha)\mu_1}{(1 - \beta)\mu_2 + \alpha\mu_1} \right) \left( \frac{\mu_1}{\mu_2} + 1 \right). \quad (30)$$

To make the results more intuitive we use the transformations from equations (8) and (9). Here, we also already see that not

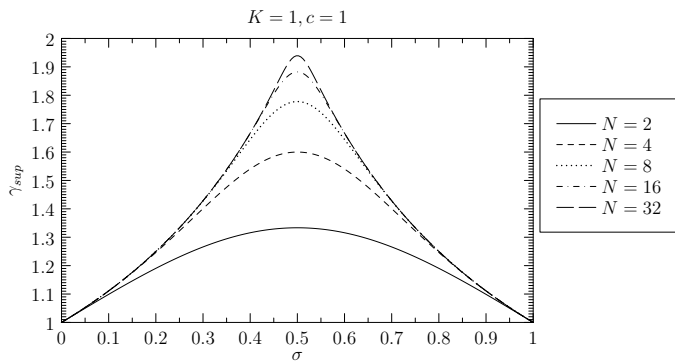


Fig. 5.  $\gamma_{sup}$ , least upper bound of the set of  $\gamma$  values where the system is stable, versus parameter  $\sigma$  with  $K = 1$  and  $c = 1$

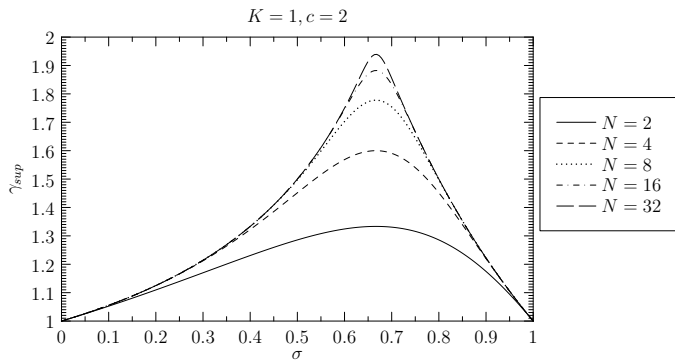


Fig. 6.  $\gamma_{sup}$ , least upper bound of the set of  $\gamma$  values where the system is stable, versus parameter  $\sigma$  with  $K = 1$  and  $c = 2$

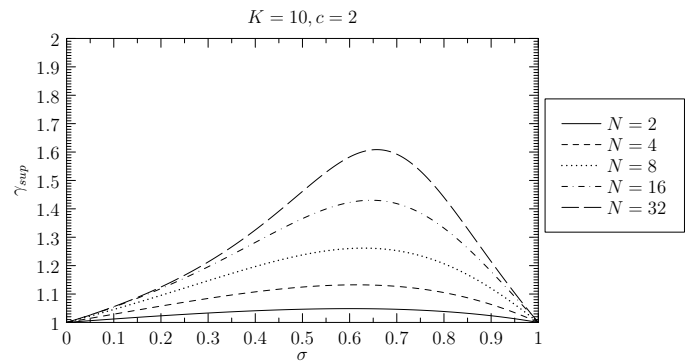


Fig. 7.  $\gamma_{sup}$ , least upper bound of the set of  $\gamma$  values where the system is stable, versus parameter  $\sigma$  with  $K = 10$  and  $c = 2$

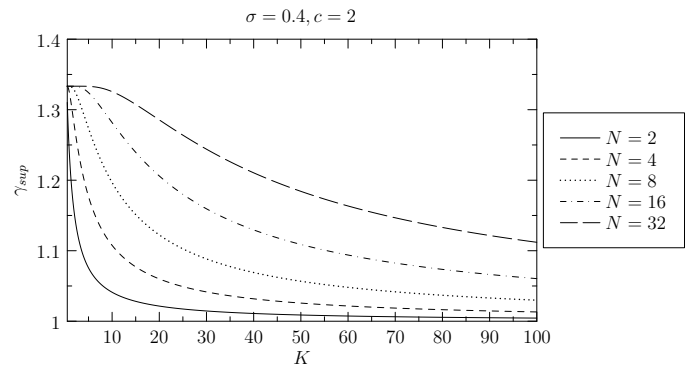


Fig. 8.  $\gamma_{sup}$ , least upper bound of the set of  $\gamma$  values where the system is stable, versus parameter  $K$  with  $\sigma = 0.4$  and  $c = 2$

the exact values of  $\mu_1$  and  $\mu_2$  are of importance but only the ratio. The stability condition becomes

$$\gamma < \frac{(1 + a\theta_1^{N-1})(\theta_1 - 1) + 2b(\theta_1^{N-1} - 1)}{(1 + a\theta_1^{N-1})(\theta_1 - 1) + b(\theta_1^{N-1} - 1)} \quad (31)$$

with

$$c = \frac{\mu_1}{\mu_2}, \quad (32)$$

$$\theta_1 = c \frac{K + (1 - \sigma)c - \sigma}{cK - (1 - \sigma)c + \sigma}, \quad (33)$$

$$a = c \frac{(1 - \sigma)}{\sigma}, \quad (34)$$

$$b = \left( \frac{(1 - \sigma)}{cK - (1 - \sigma)c + \sigma} \right) (c + 1). \quad (35)$$

Notice that when  $\theta_1 = 1$ , we need to use l'Hôpital's rule to determine the value of the right hand side of (31). Thus, when  $\theta_1 = 1$ , the stability condition is given by

$$\gamma < \frac{1 + aN\theta_1^{N-1} - (N-1)(a-2b)\theta_1^{N-1}}{1 + aN\theta_1^{N-1} - (N-1)(a-b)\theta_1^{N-1}}. \quad (36)$$

#### IV. DISCUSSION OF THE RESULTS AND NUMERICAL EXAMPLES

It is always interesting to look at the extreme situations. The first one is when  $N = 1$ . This means that we have a first-in-first-out queue regardless of the types of customers. The

stability condition (31) becomes

$$\gamma < 1 \quad (37)$$

This is what we would also intuitively expect. The amount of work the system can execute per unit time is 1 (right-hand side of equation (37)). In other words, at any given time only one server is working. The second extreme situation is when  $N = \infty$ . In this case, separate queues for each server are present. We have to split up this situation in three cases. The first case is when  $\theta_1 = 1$ , rewritten as  $\frac{\sigma}{\mu_1} = \frac{1-\sigma}{\mu_2}$  (independent of  $K$ ) or the load is balanced between both servers. The stability condition (31) becomes

$$\gamma < 2 \quad (38)$$

This is, again, what we would also expect intuitively. The amount of work the system can execute per unit time is 2 (right-hand side of equation (42)). In other words, at any given time both servers are working provided the system is constantly provided with new customers. Notice that this is only possible when the load is balanced. The second case is when  $\theta_1 > 1$ , rewritten as  $\frac{\sigma}{\mu_1} < \frac{1-\sigma}{\mu_2}$  or server 1 has a heavier load. The stability condition (31) becomes

$$\gamma < \frac{a\theta_1 - a + 2b}{a\theta_1 - a + b} \quad (39)$$

or further simplified as

$$(1 - \sigma)\lambda < \mu_2. \quad (40)$$

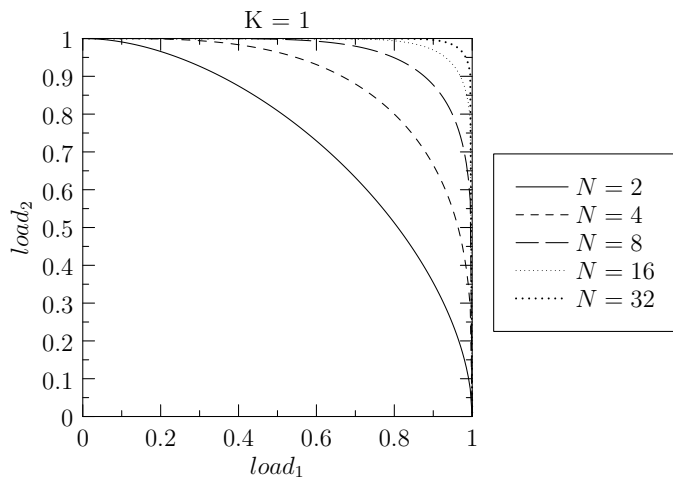


Fig. 9.  $load_2$  versus  $load_1$  with uncorrelated types of customers in the arrival stream ( $K = 1$ )

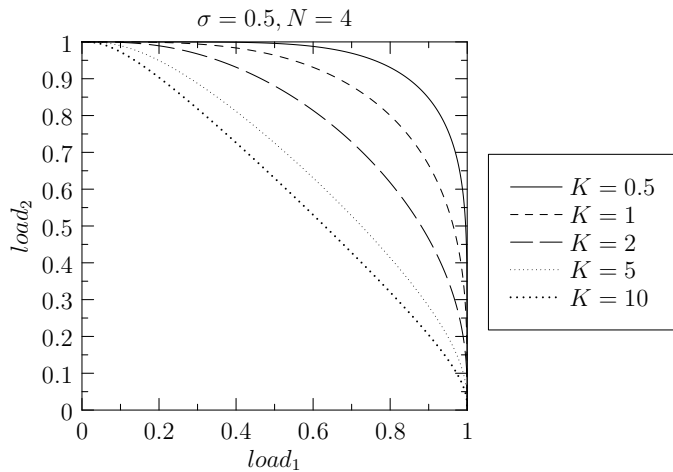


Fig. 10.  $load_2$  versus  $load_1$  with  $N = 4$  and  $\sigma = 0.4$

A similar result can be found in the third case. This is when  $\theta_1 < 1$ , rewritten as  $\frac{\sigma}{\mu_1} > \frac{1-\sigma}{\mu_2}$  or server 2 has a heavier load. The stability condition (31) becomes

$$\gamma < \frac{1 - \theta_1 + 2b}{1 - \theta_1 - b} \quad (41)$$

or further simplified as

$$\sigma\lambda < \mu_1. \quad (42)$$

In both cases, our stability condition is governed by the queue with the heavier load. This is what is expected, since when the queue with the heaviest load is stable, automatically the other queue with the lower load is also stable. Notice here already that when the load is balanced we get the maximum total load ( $\theta_1 = 1$ ).

We now look at intermediate values of  $N$ . Figures 5, 6 and 7 show the influence of the relative frequency distribution. In all three figures we have plotted  $\gamma_{sup}$  versus  $\sigma$ . Here,  $\gamma_{sup}$  is the least upper bound or supremum of the set of  $\gamma$  values where the system is stable and  $\sigma$  represents the relative frequency distribution of the type of the customers. Fig. 5 shows the case where we have a symmetric system ( $\mu_1 = \mu_2$ ) and the

types of the customers in the arrival stream are uncorrelated ( $K = 1$ ). In this case,  $\sigma = \phi$  and  $\sigma = 1 - \phi$  ( $0 < \phi < 1$ ) lead to the same results. The key observation to understand the latter is that for the operation of the system the exact types of the  $N$  first customers are irrelevant if both servers have the same service rate  $\mu$ . When the  $N$  first customers are all of the same type, the system only processes one customer anyway with the same service rate  $\mu$ , independent of the type of customer. In fact, a system with  $\sigma = 1 - \phi$  can be conceived as a system with  $\sigma = \phi$  whereby the names of the types 1 and 2 have been “swapped”. Thus, there exists symmetry in the customer types around the value  $\sigma = \frac{1}{2}$ . The impact of P-gFCFS is the largest when we reach the maximum for  $\gamma_{sup}$  at  $\sigma = \frac{1}{2}$ . In Fig. 6 and 7 this symmetry is broken since both customers no longer introduce the same average amount of work that enters the system. Fig. 6 shows the case where we have no longer a symmetric system ( $\mu_1 = 2\mu_2$ ) but the types of the customers in the arrival stream are still uncorrelated. Here, we see that the maximum is shifted. This maximum is now at

$$\sigma_{max} = \frac{\mu_1}{\mu_1 + \mu_2} \quad (43)$$

or rewritten

$$\frac{\sigma_{max}}{\mu_1} = \frac{1 - \sigma_{max}}{\mu_2} \quad (44)$$

which in words means that the maximum total load is reached when both customers introduce the same average amount of work that enters the system. In Fig. 7, the asymmetric case ( $\mu_1 = 2\mu_2$ ) where the types of the customers in the arrival stream are correlated, is plotted. Here, we notice that the correlation has indeed no influence on the  $\sigma_{max}$  corresponding with the maximum total load (in both figures 6 and 7, this maximum is at  $\sigma = \frac{2}{3}$ ). But the correlation has an influence on the value of the maximum. From figures 5-7 we see that impact of P-gFCFS is most noticeable when the relative load distribution is in balance. If the relative load distribution is totally out of balance the impact of P-gFCFS becomes negligible, which is also intuitively clear since we then approach a system with almost only one type of customers and thus a single server system.

Fig. 8 shows the influence of correlation. In this figure we have plotted  $\gamma_{sup}$  versus parameter  $K$ , which gives an indication about the correlation in the system as discussed in Section III-A. Here, we see clearly that the throughput is larger when the types of the customers alternate more in the arrival stream. When  $K = \frac{1}{2}$  (when the types arrive constantly alternating), we get the largest total load. Notice that this is not the same value for all different  $N$  but the difference is negligible. On the other hand when  $K = \infty$ , there only arrives one type of customer and we get a single server system and the limit is thus 1. Notice that correlation can have a devastating impact on the total load of the system with P-gFCFS

Fig. 9 and 10 show the influence of the load of one type of customer on the load of the other type of customer. In both figures we have plotted  $load_2 (= \frac{1-\sigma}{\mu_2})$  versus  $load_1 (= \frac{\sigma}{\mu_1})$ . The  $load_2$  in both figures is the least upper bound of the set of  $load_2$  values where the system is stable, for a given  $load_1$  value. In Fig. 9 we look at the case with uncorrelated types of the customers in the arrival stream ( $K = 1$ ). Here, we see that for  $N = 2$ ,  $load_1$  has a huge impact on  $load_2$ . This impact decreases when  $N$  becomes larger. In traffic context is



this exactly what we wanted to become with the filter lanes. We wanted to decrease the impact of the vehicles going to destination 1 on vehicles with destination 2 and visa versa. In Fig. 10, the types of the customers in the arrival stream are correlated and we have a parameter  $N = 4$  and  $\sigma = 0.4$ . In this figure we look at the impact of correlation. We can clearly see that in our system, correlation has a devastating impact. It even undoes the impact that or parameter  $N$  has on the system.

## V. CONCLUSIONS AND FUTURE RESEARCH

In this paper, we have analysed the stability condition of a continuous-time queueing model where two types of customers, both to be served by their own dedicated server, are accommodated in one common FCFS queue with an exception for the first  $N$  customers of the system. We have shown the positive impact on the total maximum load of the relaxation of this condition (the parameter  $N$  of our P-gFCFS service discipline). We have also deduced that we can achieve the largest maximum total load when the load is balanced between both servers. It is also around this value, our parameter  $N$  has the largest impact. When the loads are totally out of balance, our parameter  $N$  has almost no impact at all. We have also shown that if there is a lot of correlation between the types of the customers in the arrival stream, this has an devastating impact on our system and can even undo the positive impact of the parameter  $N$  (filter lanes). Finally we have also shown that in our system the load of one type of customer can have a big impact on the maximum allowable load of the other type. Here, the parameter  $N$  also helps to minimize the impact of one type of customer on the other.

Our future research goal is to go beyond the stability analysis of P-gFCFS. A first interesting extension could be the tail probabilities of the number of customers in the system (or at least some approximation). And especially in traffic context, it would be interesting to be able to set the parameter  $N$  so that only a certain percentage of the customers has to wait longer than a certain period of time.

## REFERENCES

- [1] T. Van Woensel and N. Vandaele, "Empirical validation of a queueing approach to uninterrupted traffic flows," *4OR, A Quarterly Journal of Operations Research*, vol. 4, pp. 59–72, 2006.
- [2] T. Van Woensel and N. Vandaele, "Modeling traffic flows with queueing models: A review," *Asia-Pacific Journal of Operational Research*, vol. 24, pp. 435–461, 2007.
- [3] W. Mélange, H. Bruneel, B. Steyaert, and J. Walraevens, "A two-class continuous-time queueing model with dedicated servers and global fcfs service discipline," in *Analytical and Stochastic Modeling Techniques and Applications*, vol. 6751 of *Lecture Notes in Computer Science*, pp. 14–27, Springer Berlin / Heidelberg, 2011.
- [4] H. Bruneel, W. Mélange, B. Steyaert, D. Claeys, and J. Walraevens, "A two-class discrete-time queueing model with two dedicated servers and global fcfs service discipline," *European Journal of Operational Research*, vol. 223(1), pp. 123–132, 2012.
- [5] L. Kleinrock, *Theory, Volume 1, Queueing Systems*. Wiley-Interscience, 1975.

## Development of Logistical Model Based on Integration of Ontology, Multi-Agent approach and Simulation

Konstantin A. Aksyonov, Eugene A. Bykov, Olga P. Aksyonova, Alena L. Nevolina

Department of Information Technology

Ural Federal University

Ekaterinburg, Russia

BPsim.DSS@gmail.com, speedmaster@inbox.ru, wiper99@mail.ru, mail@bpsim.ru

**Abstract**—The paper focuses on development of logistical model based on multi-agent model of a resource conversion process. The model is used for planning and dispatching of gas stations network. We compare existing approaches to multi-agent planning, including requirements and capabilities nets, multi-agent resource conversion process model, active and passive convertors model, agent-based simulation modeling, and their software implementations (Magenta, BPsim, Onto Modeler, AnyLogic). The most perspective approaches are implemented in Magenta and BPsim systems. However multi-agent approach based on distributed calculations has one disadvantage, which is frequent plan modification, resulting in certain instability of the model. To avoid this effect, we use situation filtration and diagnosis block. In order to take delays into consideration and analyze the bottlenecks in logistical chain, we use simulation modeling. We compare existing knowledge representation models, identify the benefits, and develop a methodology that considers all advantages.

*Keywords-logistics; decision support; simulation; expert system; frame; UML*

### I. INTRODUCTION

The following components are considered during system analysis of logistical, production, business systems, and business processes: mission, vision, strategy, processes [1]. Development of applied decision support systems based on hybrid (dynamic and intelligent) models requires the following functions [1]:

- Design of conceptual model of the problem domain,
- Problem domain knowledge definition and knowledge-based output,
- Definition of dynamic resource conversion processes,
- Development of hierarchical process model,
- Availability of situation and command definition languages (to define control model),
- Design of multi-agent models (availability of coalitions of intelligent agents). Models of intelligent agents correspond to the models of decision-making people,
- Integration of conceptual, simulation modeling, expert systems, and situational management,
- Integration with tools for Computer-Aided Software Engineering (CASE-tools).

Multi-agent systems engineering approaches can be distinguished into two types:

- Based on object-oriented methods and technologies and
- Application of traditional knowledge engineering methods.

An actual task is development of dynamic situations modeling system, based on object-oriented technologies.

Planning is one of applied directions of multi-agent systems. "An agent is an encapsulated computer system that is situated in some environment and that is capable of flexible, autonomous action in that environment in order to meet its design objectives" [2-3].

An example of multi-agent system applications to operation planning of a flexible industrial system is discussed by Jennings [2]. We can name the following advantages of multi-agent planning systems:

1. Formalization of decision making nodes (situation processing scenarios) in form of the agents, which relates to knowledge formalization,
2. Planner is embedded in between certain elements of a multi-agent system by means of interaction (negotiations) of these elements. Thus, it may modify the plan in case of delays or unexpected situations,
3. Network of interconnected agents coordinates its activity independently.

Additional advantage of multi-agent planning consists in availability of automatic notifications of process actors about events and changes at control object, which makes control transparent. Therefore, problem domain knowledge data is formalized during development and deployment of a multi-agent system, and the decision making process is automated. This all facilitates decision making activity.

Among the most considerable results, one may note the development and practical application of the Requirements and Capabilities (RC-) networks apparatus [4-5]. This approach adheres to the "classical" interpretation of a multi-agent system and is oriented to problem solving in computer networks.

Apparatus of RC-networks has been developed by Vittich and Skobelev [4] and implemented within Magenta technology [5]. It has been applied in a family of applied intelligent planning systems for the following objects [5]:

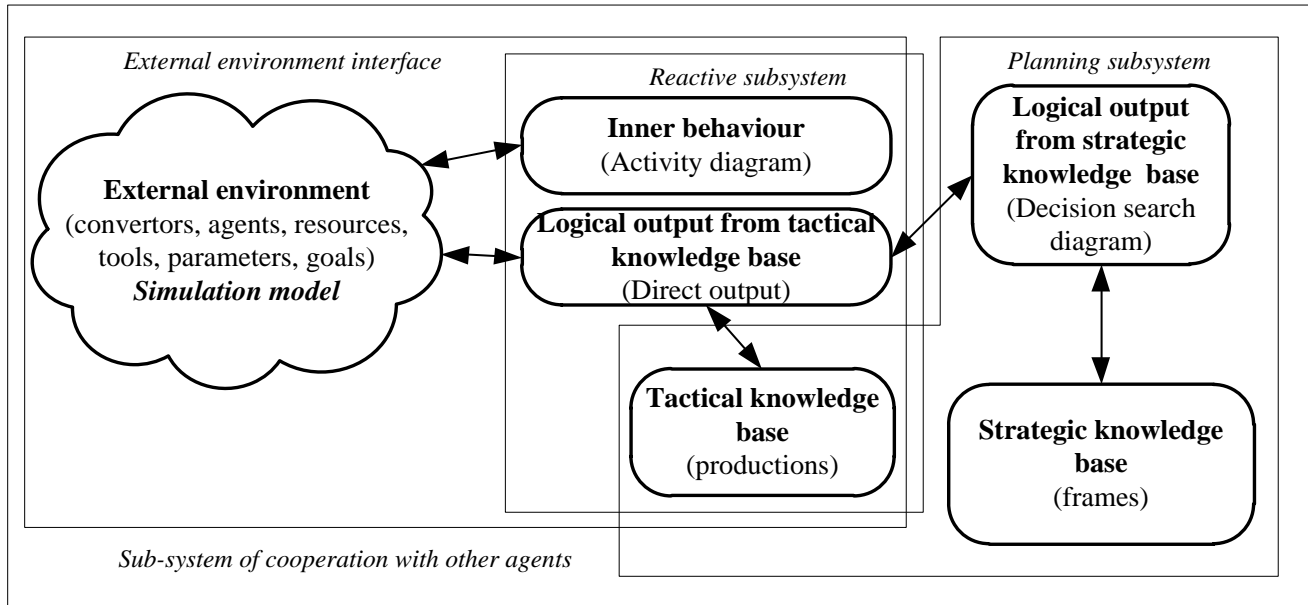


Figure 1. Multi-agent resource conversion process model structure

- An enterprise, controlling ocean tanker fleet,
- Cargo transportation company,
- Project managing company.

Multi-agent approach also found its use in simulation modeling. Thereby, the simulation modeling system AnyLogic supports modeling of reactive agents [6]. AnyLogic uses state chart extensions of Unified Modeling Language for Real-Time (UML-RT) for agent behavior formalization. An example of multi-agent system implementation within the dynamic expert system G2 [7] is presented in [8]. G2 makes use of the built-in simulation modeling sub-system ReThink [7].

The agent concept is re-considered within the framework of simulation modeling. Here, its communication features at the network protocols level, as well as capabilities of moving on the network, are becoming less important. At the same time, its intelligence and sociality are increasing. By intelligence we mean consideration of large volumes of data and knowledge and complexity of logical output machine implementation, at least for G2 and BPsim systems. Sociality includes modeling of social behavior, internal beliefs, desires, and agent goals.

## II. SIMULATION MODELING, KNOWLEDGE REPRESENTATION AND DESIGN OF DECISION SUPPORT SYSTEMS

Range of simulation modeling systems application substantially broadened during last decades. First, simulation modeling systems and software-based simulation models are built into enterprise control loops, integrated with sensors, controllers, corporate information systems via data exchange interfaces, thus, receiving data on current state of the managed object. Second, most simulation modeling systems

allow modification/refresh of initial data and adapt decision on the basis of current situation, during experiment execution. Third, most applied decision support systems, that are based on simulation modeling sub-system or simulation model, use heuristic-based sub-systems of decision optimization and improvement for formalization of decision making people knowledge and control algorithms.

Projects in the areas of hybrid decision support systems deployment for logistical, industrial management, and construction are described in [9-11]. Such systems integrate simulation model with heuristic or optimization blocks. Therefore, we are dealing with open integrated modules/systems for simulation modeling. Expert system elements, including a knowledge base and logical output machines, and ontologies are actively used in decision support systems as well.

Despite active application of conceptual modeling tools based on Unified Modeling Language (UML) in the area of information systems development, application of such tools in simulation modeling engineering is limited [12]. An advantage of conceptual and simulation modeling integration approach is the capability of rapid transition from conceptual models to the models of engineering and application (program implementation) [12]. To define a transition from conceptual models to simulation models one may use ontologies or knowledge representation models. Ontology development experience for logistical projects and supply chains is presented in [13]. RC-networks [4-5] use ontologies for formalization of decision making points in distributed networks of control object (logistical chain). The following agents are available in the cargo transportation project: truck, technical examination, order, gas station, driver.

Multi-agent resource conversion process model has been implemented in BPsim software suite [10-11]. Multi-agent resource conversion process model has been implemented as a result of integration of simulation, multi-agent and expert modeling [1, 20]. Multi-agent resource conversion process model structure is presented in Figure 1.

III. ANALYSIS OF EXISTING KNOWLEDGE REPRESENTATION MODELS

Comparison of typical knowledge representation models for the problem of knowledge extraction for the problem domain of logistics, business, Multi-agent Resource Conversion Processes (MRCP), and further application in software simulation models is presented in Table 1.

Comparison revealed that the most effective approach is a combination of production-based and frame-based approaches for knowledge formalization in the logistical problem domain. A disadvantage of frame-based approach is the complexity of logical output machine.

For system analysis, the UML class diagram may be used as a basis for the definition of frame-concepts structure. Further definition of conceptual graphs (semantics) and filling with data form the knowledge base. Decision search diagrams (extended UML sequence diagrams) are used for implementation of logical output machine visual builder. Sequence diagram graphically define sequence of method calls between classes while solving a specific problem (scenario). Such approach allows to visual definite the solution [14].

Depending on modeling aspects, various concepts of the problem domain use different elements of the resource conversion process model. Base classes (concepts) of the problem domain of simulation modeling of logistical and business processes include the following.

- *Business process* together with its temporal and cost features, as well as properties of inputs, outputs, and tools of multi-agent resource conversion process operations (ultimate units of a business process),
- *Locations* (inhabited and non-inhabited), containing geo-information – in multi-agent resource conversion processes may be represented with corresponding transacts,
- *Transportation vehicles* – in multi-agent resource conversion processes may be represented with corresponding transacts or tools,
- *Personnel* – in MRCP represented with tools or corresponding transacts,
- *Decision making people* – in MRCP represented with agents; heuristics are formalized in form of

agent rules,

- *Road network* and transportation routes – in MRCP represented by transaction processing routes between operations, defining transitions on logical chain,
- *Resources and cargo* (resource and cargo storage), used in business processes – in MRCP represented by corresponding transacts or resources,
- *Statistics* and experiment results accumulation classes are generated on the output of MRCP model.

Results of development of tools and methods of business processes modeling and software engineering for automation of the process starting with process formalization, development of dynamic models of the processes and decision making processes, running simulation experiments with the models, bottlenecks analysis, re-engineering, and optimization of the processes, until software engineering (database structure generation and software module prototypes implementation), are presented in [15]. Also, a semantic model, defining multi-agent resource conversion process, is presented in [15]. The model uses discrete event simulation modeling as a dynamic component and expert and multi-agent simulation of intelligent agents as an intelligent component. It includes the following components: *processes, operations, resources, control commands, mechanisms, resource sources and receivers, junctions, parameters, agents*. Information resources due to their features are defined separately. These are *Messages* and *Requests* for operation execution. Cause-and-effect relations between conversion elements and resources are defined with the *Relation* object.

A specialized problem-oriented language is used in dynamic situations modeling system BPsim.MAS [10] for definition of subject ontologies and cause-and-effect laws for definition of process dynamics. According to Guizzardi and Wagner [12], subject ontology of discrete-event simulation modeling uses a concept of “CasualLow”, which is equivalent to the “Action” concept in multi-agent resource conversion process model.

Semantic model became the basis and was extended with detailed definition of semantics from the point of view of simulation modeling and consideration of agents’ behavior. Actions may result in both parametric and structural modifications of MRCP model elements. Model is presented in Figure 2. This model is extended with elements of logistical projects ontology, presented in [13], and adapted to specific features of gas stations supply network project.

TABLE I. COMPARISON OF VARIOUS KNOWLEDGE REPRESENTATION MODELS

Requirements for the model	Productions	Semantic networks	Frames
Visibility	○	●	●
Hierarchical data definition	○	●	●
Simplicity of adding new knowledge	●	○	○
Consistency with object-oriented approach	○	○	●
Use of UML	○	○	●
Consideration for process dynamics	○	○	○

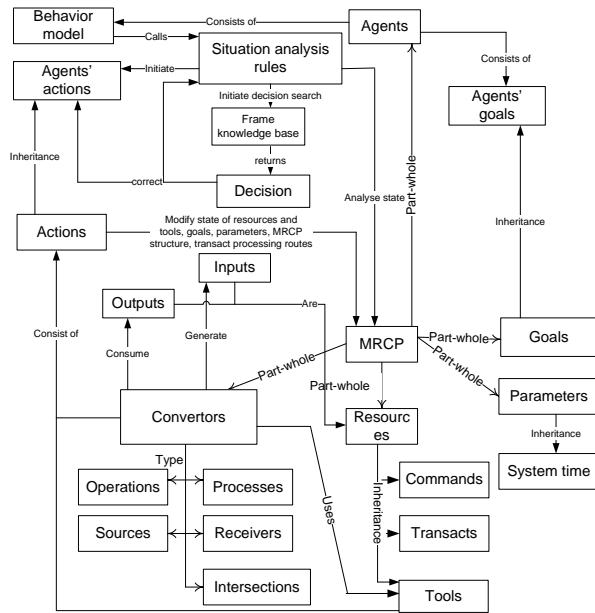


Figure 2. Semantic model of multi-agent resource conversion processes

#### IV. METHODOLOGY OF INFORMATION SYSTEMS ENGINEERING WITH AID OF BPSIM

The suggested methodology has the following specific features: 1. Simulation modeling for “as-to-be” model verification at re-engineering stage; 2. Capability of defining decision making people models. Software engineering process is separated into several stages. These stages are discussed further in relation to BPsim software.

First stage is research. Simulation model “as-is” is designed on this stage. Simulation experiments are carried out targeting identification of non-optimal process organization. Model results are used for design of the “as-to-be” model of the multi-agent resource conversion process, which in turn is used for software engineering. The following products from BPsim family are used on this stage: 1. Dynamic situations modeling system BPsim.MAS (“Multi-agent simulation”) is used for design of simulation model and running simulation experiments, 2. CASE tool BPsim.SD (“Software Developer”) is used for designing diagrams, defining software architecture. Note that the results of work in BPsim.MAS may be automatically (with aid of software assistant) transformed into BPsim.SD model, since operations, defined in multi-agent resource conversion process mode, are transformed into operations on Data Flow Diagrams (DFD) [12], agents – into external entities, and order of operations defines direction of data flow [15].

Second stage consists in the project study. Structural analysis results become the basis for object-oriented engineering when building information system model. An important feature of BPsim.SD is the capability to acquire use-case diagrams on the basis of the data from data flow diagrams: external entities of DFD diagrams are

transformed into roles, and processes – into use-cases. Sequence of data processing method calls in multi-agent resource conversion processes problem domain is reflected on sequence diagrams. Class diagrams are defined for the conceptual model specification of the problem domain. Class properties define the features of the problem domain objects, and methods describe operations of resources seizure, release and conversion, and the code from multi-agent resource conversion process model may be transferred into method definition [15].

The third stage consists in system development, while the fourth refers to system deployment. Effectively designed elements of the hybrid model of multi-agent resource conversion processes implemented in software, may be immediately included into decision support system or the corporate information system and used in decision making process, both as a whole or separately in form of calls to software assistants. Program model may be converted into the library of executable application and be used in control system. The most complex and costly intelligent components in decision support systems development are the logical output machine simulation modeling core, multi-agent system core and agent communications. Due to that, an actual problem is development of applied decision support systems based on ready-made intelligent components.

#### V. ANALYSIS OF EXISTING APPROACHES

Skobelev refers to Magenta [16] technology as to the first generation of multi-agent platforms and presents results of second generation platform development. One of the features of second generation multi-agent platform is the control over all kinds of mobile resources that have built-in GPS (Global Positioning System) and/or GLONASS (Global Navigation Satellite System) sensors. Part of this platform is deployed directly on the mobile devices. Advantages of second generation platform are illustrated in [16] on the truck control project for the European transportation network, since no similar system can “consolidate cargo, adapt delivery routes, plan deliveries and assign trucks on the basis of events flow, such as new order entry or changes in resource availability”. A conceptual approach to design of a similar system is presented in [17].

One of multi-agent planning platform requirements is support of the full control cycle that consists of the following stages [16]:

- 1) **Reaction to event** (discrete event control). Events (orders, delays, failures, etc.) are received in real time. They need to be planned, taking into account current plans, individual preferences and limitations of resources and orders,
- 2) **Dynamic planning** (re-planning / dispatching),
- 3) **Coordination and plan revision “on the fly”**. Changes are made to resource plans without stopping and restarting the main program, by modifying the schedule “on the fly”, making use of both free windows and adaptive repositioning of previously assigned orders. When planned and actual schedules differ, automatic re-planning and user coordination are required.

#### 4) Plan execution monitoring and control

Successful deployment of planning system highly depends on user "trust" factor. Users, typically logistical specialists or dispatchers, are responsible for planning and control of transportation process. User trust may be provided by offering aids to override the system operation and make modifications accordingly. Planning problem is usually complex, since certain events, situations and parameters can be omitted during development of intelligent planning system. Thus, the following additional requirements emerge:

##### 1) Negotiations on the plan with end users

##### 2) Manual plan editing, including manual planning

In order to design an efficient supply chain planning system one must control of supply and individual order terms; diagnose the bottlenecks of transportation networks and logistical centers. Thus, there is a requirement for analysis and modeling of logistical systems that are based on queuing systems.

The following approaches to multi-agent system and business process models are available: 1. Multi-agent resource conversion process model [1, 20], 2. Requirements and capabilities networks (RC-nets) by Vittich, Skobelev, and Rzevski [4-5, 16], used for development of intelligent multi-agent systems in logistics, 3. Models of Active and Passive Convertors (APC) by Moskalyov and Klebanov (Onto Modeler software) [19], 4. Approach, implemented in AnyLogic software [6, 9].

The results of these approaches analysis are presented in Table 2. The requirements for logistical system may be separated into 3 groups: knowledge formalization; analysis and simulation; full control lifecycle support.

As we can see from the table, the RC-net satisfies most requirements of logistical models and supports corresponding tasks. Also the multi-agent resource

conversion process model is perspective, if improved and perfected, due to detailed study of integration of conceptual, simulation, expert, multi-agent, and situation modeling.

Generally, there is the following difference between RC-networks based and multi-agent resource conversion processes apparatuses:

1. Approaches have different knowledge distribution and representation. Each agent of RC-network possesses only its own knowledge. In order to achieve the common goal of planning and control a feature of agent communication is required. Thus, the RC-network provides a decentralized control system. AnyLogic provides a set of blocks for simulation modeling of transportation systems, but lacks tools for knowledge (ontology) processing. Since all knowledge about the whole control object in multi-agent resource conversion processes and APC is stored in a common knowledge base, the multi-agent resource conversion process offers a more centralized control system. It receives information from distributed sources (fuels residues sensors, transportation vehicles monitoring systems, corporate information system).

2. Approaches differ by technical implementation. Last applications of RC-networks are oriented towards distributed calculations and networks. Software implementations of multi-agent resource conversion process model and APC model, as well as AnyLogic system, are all local simulation modeling systems..

3. Planning and dispatching tasks are solved with aid of RC-networks and multi-agent resource conversion process model. AnyLogic system is not oriented towards these problems, since the planning block needs to be implemented. APC approach cannot be used in planning tasks and logistical systems dispatching. Still, there are some disadvantages in the RC-networks approach [16]:

TABLE II. ANALYSIS OF APPROACHES AND DYNAMIC MODELS OF SITUATIONS

Features	MRCP	RC-net	APC	AnyLogic
1. Various resource types	●	●	●	●
2. Temporal parameters	●	●	●	●
3. Conflicts on common resources and tools	●	●	●	●
4. Discrete operations	●	●	●	●
5. Complex resources (transacts), transact queue	●	●	●	●
6. Situations analysis and decision search (heuristics block)	●	●	○	○
7. Agent communications	●*	●	○	●
8. Use of geo-data	●	●	○	●
9. Development of subject ontologies	●	●	●	○
10. Subject ontology of logistical problems	●	●	○	○
11. Software engineering (use of UML)	●	○	○	●
12. Support for distributed computing environments	○	●	○	○
13. Simulation modeling	●	○	●	●
14. Full control cycle				
14.1. Reaction to external event from the control object	●	●	○	○
14.2. Dynamic planning	●	●	○	○
14.3. Dispatching (re-planning)	●	●	○	○
14.4. Coordination and plan revision "on the fly"	●	●	○	○
14.5. Plan execution monitoring and control	●	●	○	○
14.6. Negotiations on the plan with end users	●	●	○	○
14.7. Manual plan editing	●	●	○	○

\* – feature under development

- a. Most scheduled trips are edited during execution (23% of all decisions are revised within the first hour after they are accepted, and only 10% of all decisions are alive for over 3 hours),
- b. Generally, there are 0 to 18% of proactivity phases that improve the schedule,
- c. System notifies logistical chain actors of all changes related to in, which causes certain tension.

Multi-agent resource conversion process model software implementation was further improved by agent heuristic blocks, implemented on the basis of production and frame-based expert sub-systems, and also situation diagnosis and filtration blocks, which altogether drastically solaces the disadvantages of logistical chains.

VI. APPLICATION OF BPSIM.MAS FOR MODELING OF LOGISTICAL PROCESSES

Currently, the intelligent planning system is under development. It is going to be deployed at logistical department of gas supplying company, located in Ekaterinburg, Russia, a major city in authors' area. The gas stations network of the supplier consists of 24 gas stations. Fuel to these stations is delivered by 12 fuel tankers. The total number of fuel tanks at all gas stations is 85 [10]. Interaction of intelligent system modules is presented in Figure 3.

Initially, a logistics specialist sets initial values. They include date of the plan, fuel tankers shift start time, distribution strategy for various fuel types. Initial data also includes residues at gas stations and technical condition of tanker fleet. After finishing the algorithm operation, users have an opportunity to modify the plan. Next, the plan is exported into simulation module and is corrected depending on simulation results. Simulation model also monitors residues on each gas station and fuel usage dynamics. Events, received from the control object, are processed in situation diagnosis block. Processing results determine, whether the plan needs to be modified either in part, in whole, or not at all. Sample situations include failure of the gas tanker, availability of new tanker, closure of oil farm, appearance of a new order.

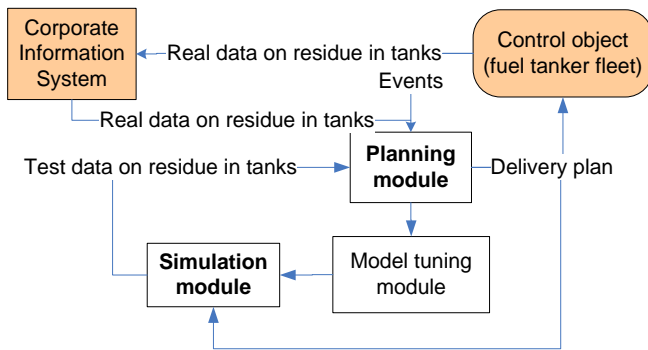


Figure 3. Interaction of planning system

VII. CONCLUSION AND FUTURE WORK

The paper discussed development of logistical model based on multi-agent model of a resource conversion process. Also the following problems have been solved:

A semantic model of multi-agent resource conversion processes has been designed,

A methodology for design of information systems, based on the model of multi-agent resource conversion processes has been presented,

A comparison of existing multi-agent approaches of information system development for the problems of planning and logistics has been presented.

We analyzed requirements and capabilities nets, multi-agent resource conversion process model, active and passive convertors model, agent-based simulation modeling, and their software implementations (Magenta, BPsim, Onto Modeler, AnyLogic). The most perspective approaches are implemented in Magenta and BPsim systems. However multi-agent approach based on distributed calculations has one disadvantage, which is frequent plan modification, resulting in certain instability of the model. To avoid this effect, we use situation filtration and diagnosis block. In order to take delays into consideration and analyze the bottlenecks in logistical chain) we use simulation modeling.

Despite the active application of UML-based conceptual modeling tools in the area of information systems development, their application in simulation modeling engineering is limited. An advantage of conceptual and simulation modeling integration is the availability of rapid transition from conceptual models to design and application models, and software implementation. Ontologies and knowledge representation models may be used for transitions from conceptual model to simulation model. The ontologies are used in Magenta and BPsim systems.

Further research includes:

1. Estimation of proactivity phases of various event types for the suggested approach,
2. Improvement of multi-agent resource conversion process model architecture,
3. Design of full knowledge base for situations,
4. Study of separate situations effect on plan stability,
5. Deployment of decision support system.

ACKNOWLEDGMENT

Work was carried out under the terms of contract № 02.G25.31.0055 (project 2012-218-03-167) and the Ural Federal University development program with the financial support of young scientists.

REFERENCES

[1] K. A. Aksyonov, "Model of a multi-agent resource conversion process and system analysis of organizational technical systems", Computer and information technology bulletin, vol. 6, pp. 38-45, Ekaterinburg, Russia, 2009.

[2] N. R. Jennings, "On agent-based software engineering," Artificial Intelligence, vol. 117, pp. 277-296, 2000. URL: <http://www.agentfactory.com/~rem/day4/Papers/AOSE-Jennings.pdf> [retrieved: June, 2013]

- [3] M. Wooldridge, Agent-based software engineering, IEE Proceedings on Software Engineering, 144 (1). 1997. pp. 26–37.
- [4] V. A. Vittich and P. O. Skobelev Multiagent Interaction Models for Constructing the Needs-and-Means Networks in Open Systems. Automation and Remote Control. Vol. 64, 2003, pp. 162-169. <http://link.springer.com/article/10.1023%2FA%3A1021836811441> [retrieved: June, 2013]
- [5] G. Rzevski, J. Himoff, and P. Skobelev, “MAGENTA Technology: A Family of Multi-Agent Intelligent Schedulers. International conference on multi-agent systems” Proceedings of Workshop on Software Agents in Information Systems and Industrial Applications 2 (SAISIA). Fraunhofer IITB, Germany, 2006. URL: <http://rzevski.net/06%20i-Scheduler%20Family.pdf> [retrieved: June, 2013]
- [6] Agent Based Modeling. Available at: <http://www.anylogic.com/agent-based-modeling> [retrieved: June, 2013]
- [7] Gensym G2 2011 Editions. Product page. Available at: <http://www.gensym.com/product/G2> [retrieved: June, 2013]
- [8] G. V. Rybina and S. S. Parondjanov “Modelling of intelligent agent interaction processes in multi-agent systems”. Artificial intelligence and decision making. no. 3, 2008, pp. 3-15. [http://www.aidt.ru/index.php?option=com\\_content&view=article&id=145](http://www.aidt.ru/index.php?option=com_content&view=article&id=145) [retrieved: June, 2013]
- [9] A. V. Borshev. “Application of simulation modeling in Russia”, Proceedings of 3<sup>rd</sup> Russian scientific conference IMMOD, St. Petersburg, Russia, 2007, pp. 11-16, <http://www.anylogic.ru/upload/iblock/981/98110c406433a4ed72833480ad775068.pdf> [retrieved: June, 2013]
- [10] K. A. Aksonov, E. A. Bykov, A. A. Skvortsov, O. P. Aksonova, and E. F. Smolij, “Intelligent system for scheduling transportation within gas stations network”, Proceedings of the 2012 Winter Simulation Conference (WSC 2012) December 2012, Berlin, Germany. URL: <http://informs-sim.org/wsc12papers/includes/files/pos194.pdf> [retrieved: June, 2013]
- [11] K. A. Aksonov, E. A. Bykov, K. Wang, and O. P. Aksonova, “Application of simulation-based decision support systems to optimization of construction corporation processes”, Proceedings of the 2012 Winter Simulation Conference (WSC 2012) December 2012, Berlin, Germany. URL: <http://informs-sim.org/wsc12papers/includes/files/pos172.pdf> [retrieved: June, 2013]
- [12] G. Guizzardi and G. Wagner, “Tutorial: Conceptual simulation modeling with Onto-UML”, Proceedings of the 2012 Winter Simulation Conference, (WSC 2012), December 9-12, 2012, Berlin, Germany. pp. 52-66, <http://informs-sim.org/wsc12papers/includes/files/inv284.pdf> [retrieved: June, 2013]
- [13] M. Kowalski, S. Zelewski, D. Bergenrodt, and H. Klupfel, “Application of new techniques of artificial intelligence in logistics: an ontology-driven case-based reasoning approach”, Proceedings of ESM’2012 (ESM - European Simulation and Modelling Conference) October 22-24, 2012, FOM University of Applied Sciences, Essen, Germany. pp. 323-328.
- [14] K. A. Aksonov, I. I. Sholina, and E. M. Safrygina, “Development and application of object-oriented modeling and decision support system for multi-agent resource conversion processes”, Scientific and technical bulletin of St. Petersburg State Technical University. Informatics, Telecommunication, Control., vol. 3 (80), pp. 87-97, Ekaterinburg, Russia, 2009.
- [15] K. A. Aksonov and I. A. Spitsina, “Method for engineering of enterprise information systems based on semantic models of multi-agent resource conversion processes and software”, Automatization and state-of-the-art technology, vol.9, pp. 22-30, Moscow, Russia, 2009.
- [16] P. O. Skobelev “Multi-agent technologies for resource management in real time”, Proceedings of the seminar “Mechanics, control and informatics”, Tarusa, Russia, 2-4 March 2011, available at [http://www.iki.rssi.ru/seminar/2011030204/presentation/20110303\\_03.pdf](http://www.iki.rssi.ru/seminar/2011030204/presentation/20110303_03.pdf) [retrieved: June, 2013]
- [17] M. Klump and G. Sandhaus “Dynamic scheduling in logistic with agent-based simulation”, Proceedings of ESM’2012 (ESM - European Simulation and Modelling Conference) October 22-24, 2012, FOM University of Applied Sciences, Essen, Germany. P.329-336.
- [18] V. A. Vittich and P. O. Skobelev, “Multi-agent interaction models for design of requirements and capabilities networks in open systems”, Automatics and Telemekhanics, vol. 1, pp. 177-185, Ekaterinburg, Russia, 2003
- [19] I. M. Moskalyov, “System for analysis and optimization of resource conversion processes”, PhD research work, Ural State Technical University, Ekaterinburg, Russia, 2006.
- [20] K. Aksonov, E. Bykov, L. Dorosinskiy, E. Smolij, and O. Aksonova, “Decision Support based on Multi-Agent Simulation Algorithms with Resource Conversion Processes Apparatus Application, Multi-Agent Systems”, In Modeling, Interactions, Simulations and Case Studies, Faisal Alkhateeb, Eslam Al Maghayreh and Iyad Abu Doush (Ed.), ISBN: 978-953-307-176-3, InTech, 2011, Available from: <http://www.intechopen.com/articles/show/title/decision-support-based-on-multi-agent-simulation-algorithms-with-resource-conversion-processes-appar>, pp. 301-326 [retrieved: June, 2013]



# An Eclipse Plug-in for Aspect-Oriented Bidirectional Engineering

Oscar Pulido-Prieto, Ulises Juárez-Martínez

Division of Postgraduate and Research Studies

Instituto Tecnológico de Orizaba

Orizaba, Veracruz, México

Email: opp026@gmail.com, ujuarez@ito-depi.edu.mx

**Abstract**—This paper describes a plug-in for the IDE Eclipse, which enables the generation of AspectJ and CaesarJ code from a UML class diagram, as well the capability to perform reverse engineering from these languages to obtain a system model in a UML class diagram representation. In addition, the capability of generating an XML representation for visualization and understanding of AspectJ and CaesarJ code is provided. This plug-in also provides a graphical interface for system design through UML.

**Keywords**-*AspectJ; CaesarJ; UML; Software Engineering; Reverse Engineering.*

## I. INTRODUCTION

Aspect-oriented programming (AOP) [1] involves the development of modular components, which simplifies their reuse and maintenance. AOP solves encapsulation problems in object-oriented programming (OOP) [1] and structured programming [1]; such problems consist of an inability to encapsulate elements whose functionality covers more than a single object when these elements are not related through inheritance, composition, or aggregation.

Nowadays, there are several commercial tools for code generation; these cover the object-oriented paradigm and the aspect-oriented paradigm, but only partially for the aspect-oriented paradigm. This necessitates the use of extensions that are difficult for novice developers to understand. The inability to make equivalences between different aspect-oriented languages is another problem; it decreases interoperability and delays paradigm consolidation.

In this paper, a plug-in for the Integrated Development Environment (IDE) Eclipse [2] is presented. This plug-in adds the capability of generating code in AspectJ and CaesarJ from a model generated with a Unified Modeling Language (UML) extension through stereotypes, and generates a model from both languages, which comprises direct engineering. In addition, the plug-in generates an eXtensible Markup Language (XML) representation of both languages to improve the application design, which comprises reverse engineering. The combination of both direct and reverse engineering constitutes what is known as bidirectional engineering.

This paper is organized as follows: In Section 2, previous studies related to this work are presented. In Section 3, the plug-in architecture is described. In Section 4, the plug-in capabilities are analyzed. In Section 5, a discussion of the

scope of this work is presented. In Section 6, a code generation example is provided. In Section 7, the conclusions are discussed. Finally, in Section 8, future work is described.

## II. RELATED WORK

In [3], a plugin for an Eclipse Integrated Development Environment was presented. This plug-in generates AspectJ and CaesarJ code from an XML Metadata Interchange (XMI) document generated with a modeling tool. This plug-in was developed using the Eclipse Modeling Framework (EMF), a Model-View-Controller (MVC)-based architecture and a meta-model for AspectJ and another for CaesarJ. Both meta-models were defined through Java annotations, which belong to the core of EMF (Ecore) and transform Java interfaces into EClasses. This transformation occurs through the XQuery query language to search for items that are required in the XMI document. This plug-in is only focused on code generation from UML class diagrams; the system modeling is carried out with an external tool and is exported to the XML format, and the plug-in just generates code.

In [4], a tool that allows the representation of CaesarJ source code in XML format, simplifying the design of code generators and code analyzers, was presented. This tool generates XML documents based on a Java file containing CaesarJ code. The generated XML document is validated using an XML Schema (XSD). As an XML management Application Programming Interface (API), Java API for XML Processing (JAXP) was also used because this API supports several processing standards, allows transformations of XML documents, and XML Schema support is a standard part of JAXP. An analysis of the different approaches for creating XML documents was made, and one representing the elements as an abstract syntax tree was selected. As a result, a minimal tool for the generation of an XML document from CaesarJ source code and for obtaining source code for an XML document was developed.

In [5], BON-CASE, a Computer-Aided Software Engineering (CASE) tool for generating Java code, was presented. This was accomplished from a modeling diagram Business Object Notation (BON) language. This tool was developed to solve the need for a system that allows the generation of source code efficiently from a system abstract model, and also to generate robust and accurate implementations. The authors highlighted the existing CASE tool features, which allow tests to be carried out and code to

be generated, but with limitations in terms of robustness and analysis of the correctness of a model. BON-CASE is an extensible CASE tool for the BON modeling language focused on contracts and frame specifications; it generates Java source code. BON-CASE generates low coupling components to integrate these components in other platforms. BON-CASE offers a formal specification backup, code generation extensible templates, and a partially validated meta-model. The authors mentioned that UML is a modeling language that generates system-independent views. UML allows the use of formal techniques through Object Constraint Language (OCL). BON, on the other hand, is based on formal techniques that allow an inherent design by contracts. BON-CASE generates code in the Java Modeling Language (JML).

In [6], meta-programming was defined as the act of writing programs that generate other programs; this approach is essential for automated software development. On this basis, the authors developed Meta-AspectJ, a meta-language that extends Java code to generate AspectJ code. Meta-AspectJ is based on generative programming and aspect-oriented programming to design a specific domain aspect-oriented code generator, which generates efficient code to solve AspectJ limitations in general purpose code generators. The authors mentioned the disadvantages of domain-specific APIs: compatibility regressive problems when libraries are updated and bad interaction when these are independently modified. Meta-AspectJ solves these problems through the use of annotation that does not interfere with code execution and syntax.

In [7], an analysis of aspect-oriented framework architecture was performed. The authors proposed that there is a problem that results from the inability to encapsulate nonfunctional requirements as these are scattered throughout the system; they mentioned that the main problems of aspect-oriented architecture are the language type used and source transformations.

In [8], Aspectra, a framework designed to carry out test entry in previously generated aspects to measure reliability, was developed. The authors mentioned that the development of aspect-oriented software improves the software quality, but does not provide the desired accuracy owing to programmer mistakes, difficulty in verifying the appearance of an error during unit tests, or these aspects not being implemented directly.

In [9], the specifications of a meta-model for aspect-oriented modeling based on extension mechanisms using UML 2.0 and XMI, software implementation to facilitate use in other software tools, were presented.

In [10], methods of automatic code generation of aspect-oriented models by Theme/UML were defined. Additional requirements to develop a code generator were cited as follows: a meta-build system model and generator specification, which in turn is made up of snippets and production rules.

In [11], problems resulting from the use of frameworks in the design of an application due to the difficulty of

demonstrating the design using UML were highlighted. To solve this problem; the authors proposed the creation of a framework called Aspect-Oriented Crosscutting Framework.

In [12], a method of aspect-oriented modeling, using the model-driven process that focuses on business applications, was described. The authors mentioned that there are no means of modeling crosscutting concerns in UML; this generates system designs with scattered artifacts; against this background, they proposed modeling crosscutting concerns in UML diagrams representing aspects, advice, and crosscuts as first-class models to produce an aspect-oriented model.

Table 1 shows a comparison of the contributions of each study described above. The target language and modeling language are shown, as well as the restrictions in terms of both design and programming. Among these studies, only the BON-CASE tool provides mechanisms for both direct and reverse engineering, but is limited to the Java language and the BON modeling language; the other works shown in Table 1 are design specifications. In the last row, our work is shown in order to highlight the current contribution with respect to previous studies.

### III. PLUG-IN ARCHITECTURE

Eclipse is an IDE based on plug-ins that encourages the use of the Model-View-Controller architectural pattern.

In the Model layer, a meta-model through EMF is defined, as well as an XSD template and an eXtensible Stylesheet Language Transformations (XSLT) template for transformations between aspect source code and its XML representation. With EMF, a set of interfaces was generated to obtain a general meta-model and two specific meta-models for AspectJ and CaesarJ, as is shown in Figure 1(1).

In the Controller layer, the behavioral mechanism for the plug-in is defined. The controller has two parsers, which transform source code into an XML representation: one for AspectJ and the other for CaesarJ. This is possible by using an Abstract Syntax Tree, which decomposes the source code into nodes and then generates the XML representation, as is shown in Figure 1(2). On the other hand, an XSLT is used to parse the XML representation into source code. The controller also has an XMI analyzer to generate a Java model from the XMI document, as is shown in Figure 1(3); every class is tailored with one meta-model interface. Later, a template of Java Emitter Templates (JET) is used in order to generate source code from meta-model instances, either AspectJ or CaesarJ, as is shown in Figure 1(4).

In the View layer, a graphical interface to allow system modeling is defined, and then this interface is exported into source code or an XML representation through the XMI standard. The use of XMI furthermore provides an export and import mechanism to facilitate work with other tools. For generating a model representation, the plug-in takes source code in AspectJ and CaesarJ, and parses it for generating the XML representation. Then, an XMI document is generated with an XSLT and, finally, the model is loaded in the main window.

TABLE I. COMPARISON OF RELATED WORK

Author	Output Language	Modeling Language	Implementation	Independent Tool	Reverse and Direct Engineering
Rosas-Sánchez	Java, AspectJ, CaesarJ	UML	Yes	No	Direct
Salinas-Mendoza	Java, CaesarJ, XML	Not apply	Yes	Yes	Not applicable
Paige	Java	BON	Yes	Yes	Both
Huang	AspectJ	Java Annotations	No	No	Direct
Constantinides	Not apply	Not apply	No	No	Not applicable
Xie	Not apply	Not apply	Yes	Yes	Not applicable
Evermann	AspectJ	UML	No	Not apply	Not applicable
Hetch	Not apply	Theme/UML	No	Not apply	Not applicable
Júnior	AspectJ	UML	No	Not apply	Not applicable
Mosconi	Not specified	UML	No	Not apply	Not applicable
Pulido-Prieto	Java, AspectJ, CaesarJ	UML	Yes	No	Both

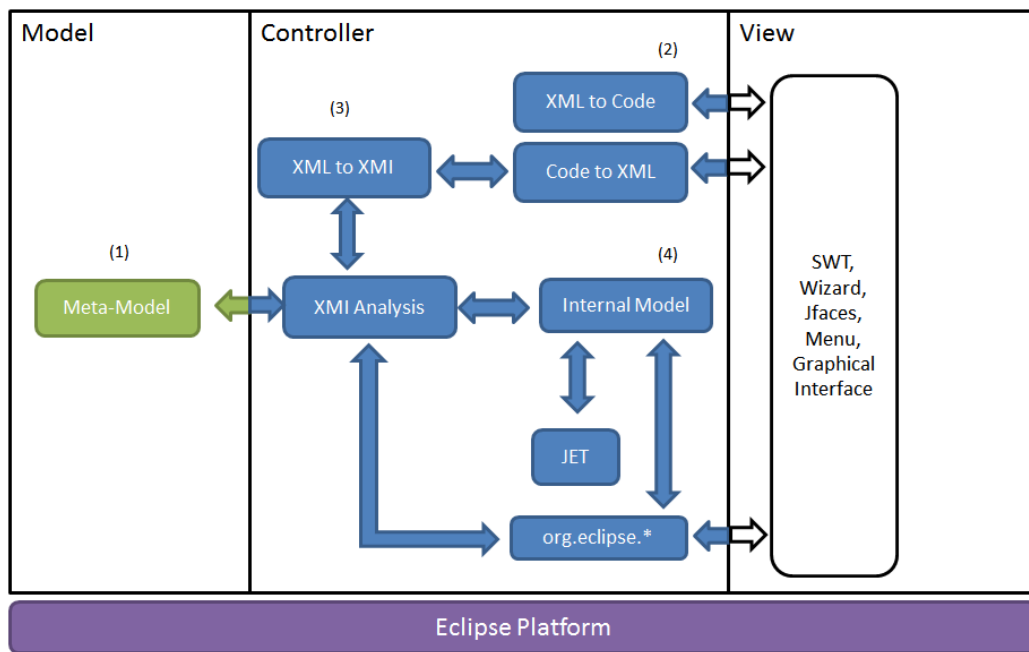


Figure 1. Plug-in logical architecture

In Figure 2, the plug-in’s physical architecture is shown. The plug-in consists of five components, two of which are *parsers*, one for AspectJ and the other for CaesarJ; another one is the *view*, which allows the design of a system class diagram and generates the XMI representation or takes an XMI document and generates a class diagram; another is the *javamodel*, which allows an XMI document to be read, validated, and for source code to be generated through JET templates; and last, there is the *controller*, which allows communication among the other components.

IV. PLUG-IN CAPABILITIES

This plug-in allows the generation of aspect-oriented source code from UML class diagrams, to achieve a reverse engineering process in order to obtain the model from source code and generate an XML representation of it. Figure 3 shows how the graphic interface is implemented.

This project extends the previous work support [4]; in these work, a prototype for CaesarJ source code

transformation into XML was developed. The support for bit operations has been added, as well as full support for signature patterns, support for more than two mixins, class element access was restructured, and anonymous class definition and class definition inside of blocks, such as classes, initializers, methods, iterations, or bifurcations, were added.

On the other hand, an AspectJ parser was developed, which works similarly to the project described in [4]. Besides, full support for signature patterns, bit operations, static crosscutting support, and error and warning declarations such as superclasses, interfaces, methods, and field inter-type declarations were added.

Additional work from a previous study was also incorporated [5]; it includes source code generation from a model. In addition, a graphical interface was implemented to avoid the external tool requirement for model generation; thus, the reverse engineering capability was also added, whereby the model is obtained from source code.

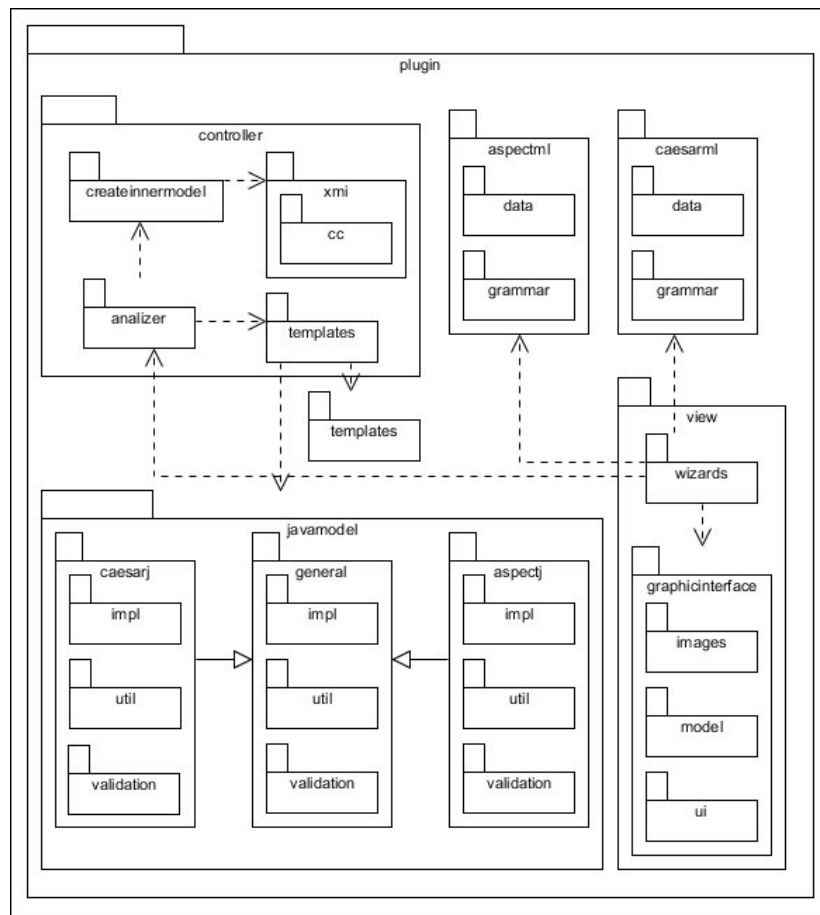


Figure 2. Plug-in physical architecture

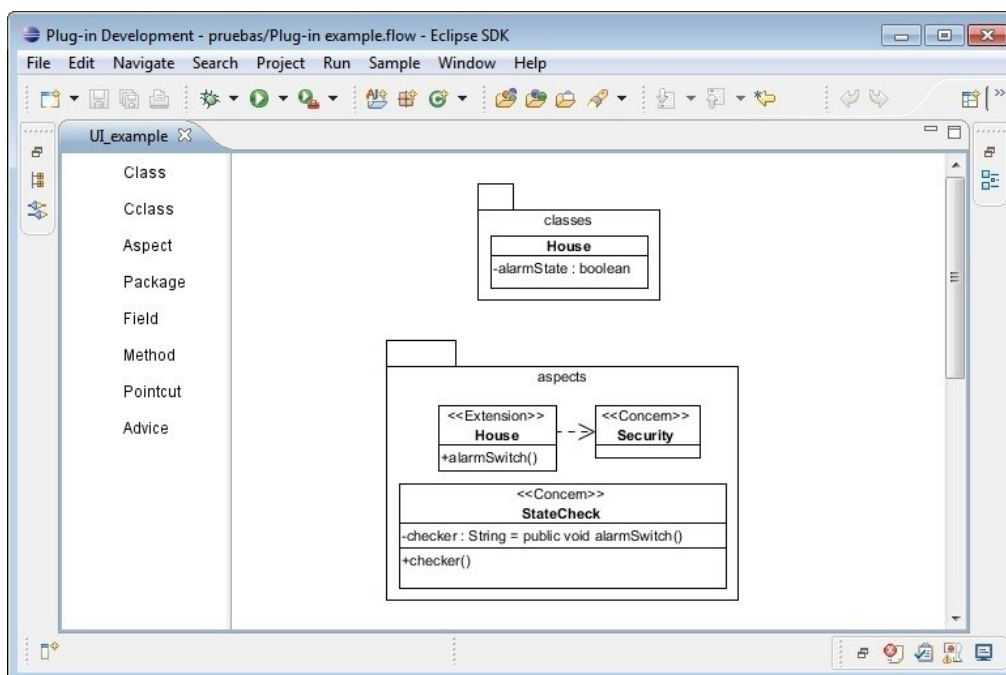


Figure 3. Plug-in graphical interface

Since a reverse engineering process is performed, this plug-in allows equivalences to be made between AspectJ and CaesarJ when source code is generated from a model. This is achieved with limitations in implementation in both languages, where CaesarJ lacks static crosscuttings and AspectJ lacks wrappers and mixins. If a model obtained from AspectJ implementation with static crosscuttings is used to generate CaesarJ source code, a wrapper is used to generate classes that have these static crosscuttings. When a model is obtained from CaesarJ implementation with wrappers and mixins, it uses an inner subclass of wrapper class for the wrapper. If a class has more than one superclass indicating a mixin composition, a warning message is shown, indicating modeling inconsistency.

## V. DISCUSSION

With this plug-in, an XML representation of Java, AspectJ and CaesarJ is generated. In the cases of Java and AspectJ, this plug-in lacks support for generics, but uses enhanced *for* and *multi-catch* exceptions to improve the equivalences between AspectJ and CaesarJ. Another limitation is a lack of support to store and represent methods and initializer implementation; thus, when the reverse engineering process is realized, a code loss occurs since the abstraction level is higher. Method, constructor, and initializer implementation through activity diagrams is proposed, where the flux for each method, constructor, and initializer is modeled.

## VI. EXAMPLE

In this section, a code generation example is presented, where the class *House*, with one field and without methods, is extended through a *Security* aspect, which provides one method for the alarm turn-on and turn-off, while another aspect called *StateCheck* provides a pointcut, which applies over the *alarmSwitch()* method. Modeling of this is shown in Figure 3.

For AspectJ, the *Extension* stereotype indicates that static crosscuttings will be added. The resulting source code is shown in Figure 4.

```

1 package aspects;
2 import classes.House;
3 public aspect Security {
4     public void House.alarmSwitch() {
5         //TODO Auto-generate code
6     }}

```

Figure 4. AspectJ source code for aspect Security.

An XML representation for the AspectJ code is shown in Figure 5, where line 3 is equivalent to line 1 in Figure 4, line 4 is equivalent to line 2 in Figure 4, line 5 is equivalent to line 3 in Figure 4, and lines 6 to 9 are equivalent to line 4 in Figure 4.

For CaesarJ, the *Extension* stereotype indicates a wrapper declaration. The resulting source code is shown in Figure 6.

An XML representation for the CaesarJ source code is shown in Figure 7, where line 3 is equivalent to line 1 in

Figure 6, line 4 is equivalent to line 2 in Figure 6, line 5 is equivalent to line 3 in Figure 6, lines 6 to 8 are equivalent to line 4 in Figure 6, and lines 9 to 13 are equivalent to line 4 in Figure 6.

```

01 <aspect-source-program>
02   <aspect-class-file name="Security.aj">
03     <package-decl name="aspects"/>
04     <import module="classes.House"/>
05     <aspect name="Security" visibility="public">
06       <intertype-declaration target="House"
07         kind="method">
08         <method name="alarmSwitch"
09           visibility="public">
10           <type name="void" primitive="true"/>
11           <formal-arguments/>
12         </block/>
13       </method/>
14     </intertype-declaration/>
15   </aspect/>
16 </aspect-class-file/>
17 </aspect-source-program/>

```

Figure 5. XML representation for Security AspectJ code.

```

1 package aspects;
2 import classes.House;
3 public cclass Security {
4     public cclass SecurityWrappsHouse wraps House {
5         public void alarmSwitch() {
6             //TODO Auto-generate code
7         }}}

```

Figure 6. CaesarJ source code for aspect Security.

```

01 <java-source-program>
02   <java-class-file name="Security.java">
03     <package-decl name="aspects"/>
04     <import module="classes.House"/>
05     <cclass name="Security" visibility="public">
06       <cclass name="SecurityWrappsHouse"
07         visibility="public">
08         <wraps name="House"/>
09         <method name="alarmSwitch"
10           visibility="public">
11           <type name="void"
12             primitive="true"/>
13           <formal-arguments/>
14         </block/>
15       </method/>
16     </cclass/>
17   </cclass/>
18 </java-class-file/>
19 </java-source-program/>

```

Figure 7. XML representation for Security CaesarJ code.

For AspectJ, the *Concern* stereotype indicates an aspect, if a class with that stereotype contains a String field with a

crosscutting primitive, a pointcut is generated in the source code; if a method name is equal to this field, three definition of advice are generated to this pointcut, as is shown in Figure 8. For CaesarJ, the Concern stereotype indicates a *cclass*, as is shown in Figure 9. As is shown in Figure 9 and Figure 10, the only differences between the two languages are in line 2: the *aspect* keyword for AspectJ and the *cclass* keyword for CaesarJ are placed there.

```

01 package aspects;
02 public aspect StateCheck {
03     public pointcut checker() :
04         call(public void alarmSwitch());
05     before() : checker() {
06         //TODO Auto-generate code
07     }
08     void around() : checker() {
09         //TODO Auto-generate code
10     }
11     after() : checker() {
12         //TODO Auto-generate code
13     }}

```

Figure 8. AspectJ source code for aspect StateCheck.

```

01 package aspects;
02 public cclass StateCheck {
03     public pointcut checker() :
04         call(public void alarmSwitch());
05     before() : checker() {
06         //TODO Auto-generate code
07     }
08     void around() : checker() {
09         //TODO Auto-generate code
10     }
11     after() : checker() {
12         //TODO Auto-generate code
13     }}

```

Figure 9. CaesarJ source code for aspect StateCheckAspect.

An XML representation of an aspect with pointcuts is shown in Figure 10 for AspectJ and Figure 11 for CaesarJ. The only differences between the two XML documents is in line 4: a node *aspect* is placed there for AspectJ and a *cclass* node is placed there for CaesarJ. For Figure 10 and Figure 11, lines 5 to 13 are equivalent to lines 3 and 4 for Figure 8 and Figure 9, lines 14 to 21 in Figure 10 and Figure 11 are equivalent to line 5 in Figure 8 and Figure 9, lines 25 to 33 in Figure 10 and Figure 11 are equivalent to line 8 in Figure 8 and Figure 9, and lines 35 to 43 in Figure 10 and Figure 11 are equivalent to line 11 in Figure 8 and Figure 9.

## VII. CONCLUSIONS

This paper presented a plug-in for AspectJ and CaesarJ code generation from a UML class diagram through their XML specification. This plug-in applies reverse engineering to obtain a UML class diagram from source code, obtain an

XML source code representation, or make equivalence between the two aspect-oriented languages. In addition, two previous studies were integrated and extended in order to obtain a functional tool for bidirectional software engineering; furthermore, the plug-in capabilities and limitations were presented.

```

01 <aspectj-source-program>
02   <aspectj-class-file name="C:\StateCheck.aj">
03     <package-decl name="aspects"/>
04     <aspect name="StateCheck" visibility="public">
05       <pointcut name="checker"
06         visibility="public">
07         <pointcut-arguments/>
08         <pointcut-expressions>
09           <pointcut-expression designator="call"
10             signature="public void
11               alarmSwitch()"/>
12         </pointcut-expressions>
13       </pointcut>
14       <advice spec="before">
15         <formal-arguments/>
16         <pointcut-expressions>
17           <pointcut-expression
18             reference="checker">
19             <arguments/>
20           </pointcut-expression>
21         </pointcut-expressions>
22       </advice>
23       <advice spec="around">
24         <type name="void" primitive="true"/>
25         <formal-arguments/>
26         <pointcut-expressions>
27           <pointcut-expression
28             reference="checker">
29             <arguments/>
30           </pointcut-expression>
31         </pointcut-expressions>
32       </advice>
33       <advice spec="after">
34         <formal-arguments/>
35         <pointcut-expressions>
36           <pointcut-expression
37             reference="checker">
38             <arguments/>
39           </pointcut-expression>
40         </pointcut-expressions>
41       </advice>
42     </aspect>
43   </aspectj-class-file>
44 </aspectj-source-program>

```

Figure 10. XML representation for StateCheck AspectJ source code.

## VIII. FUTURE WORK

As future work, method and initializer body definition will be implemented through activity diagrams, and AspectJ

annotations and inclusion of generics for this language will be analyzed.

```

01 <java-source-program>
02   <java-class-file name="C:\StateCheck.java">
03     <package-decl name="aspects"/>
04     <cclass name="StateCheck" visibility="public">
05       <pointcut name="checker"
06         visibility="public">
07         <pointcut-arguments/>
08         <pointcut-expressions>
09           <pointcut-expression designator="call"
10             signature="public void
11               alarmSwitch()"/>
12         </pointcut-expressions>
13       </pointcut>
14       <advice spec="before">
15         <formal-arguments/>
16         <pointcut-expressions>
17           <pointcut-expression
18             reference="checker">
19             <arguments/>
20           </pointcut-expression>
21         </pointcut-expressions>
22       </block/>
23     </advice>
24     <advice spec="around">
25       <type name="void" primitive="true"/>
26       <formal-arguments/>
27       <pointcut-expressions>
28         <pointcut-expression
29           reference="checker">
30           <arguments/>
31         </pointcut-expression>
32       </pointcut-expressions>
33     </block/>
34   </advice>
35   <advice spec="after">
36     <formal-arguments/>
37     <pointcut-expressions>
38       <pointcut-expression
39         reference="checker">
40         <arguments/>
41       </pointcut-expression>
42     </pointcut-expressions>
43   </block/>
44 </advice>
45 </cclass>
46 </java-class-file>
47 </java-source-program>

```

Figure 11. XML representation for StateCheck AspectJ source code.

#### ACKNOWLEDGMENT

The authors thank CONACyT (Consejo Nacional de Ciencia y Tecnología) and DGEST (Dirección Nacional de Educación Superior Tecnológica) for the support granted for the completion of graduate study.

#### REFERENCES

- [1] G. Kiczales, et al, "Aspect-oriented programming," in European Conference on Object-Oriented Programming (ECOOP'97), vol. 1241, June. 1997. pp. 220-242. doi: 10.1007/BFb0053381
- [2] L. Vogel, "Eclipse IDE," vogella.com, 3rd edition, pp. 492.
- [3] E. Y. Rosas-Sánchez, "Desarrollo de un plug-in para el ide eclipse para la generación de código de aspectj y caesarj basado en xmi y perfiles de uml," Master's thesis, Instituto Tecnológico de Orizaba, 2011.
- [4] A. Salinas-Mendoza, "Caesarml: una representación basada en xml para código fuente de caesarj," Master's thesis, Instituto Tecnológico de Orizaba, 2011.
- [5] R. Paige, L. Kaminskaya, J. Ostroff, and J. Lancaric, "Boncase: an extensible case tool for formal specification and reasoning," Journal of Object Technology, vol. 1, no. 3, Aug 2002, pp. 77-96.
- [6] S. Huang, D. Zook, and Y. Smaragdakis, "Domain-specific languages and program generation with meta-aspectj," ACM Transactions on Software Engineering and Methodology (TOSEM), vol. 18, no. 2, Nov. 2008, article no. 6, pp. 1-32, doi:10.1145/1416563.1416566.
- [7] C. Constantinides, A. Bader, T. Elrad, P. Netinant, and M. Fayad, "Designing an aspect-oriented framework in an object oriented environment," ACM Computing Surveys (CSUR), vol. 32, no. 1es, p. 41, Mar. 2000, pp. 1-12, doi:10.1145/351936.351978.
- [8] T. Xie and J. Zhao, "A framework and tool supports for generating test inputs of aspectj programs," in Proceedings of the 5th International Conference on Aspect-oriented Software Development, Mar. 2006, pp. 190-201, doi:10.1145/1119655.1119681.
- [9] J. Evermann, "A meta-level specification and profile for aspectj in uml," in Proceedings of the 10th International Workshop on Aspect-oriented Modeling, Mar. 2007, pp. 21-27, doi:10.1145/1229375.1229379.
- [10] M. Hecht, E. Piveta, M. Pimenta, and R. Price, "Aspect oriented code generation," 20. Simpsio Brasileiro de Engenharia de Software (SBES'06), Florianopolis, SC, Brazil, 2006, pp. 209-223.
- [11] J. Júnior, V. Camargo, and C. Chavez, "Uml-aof: a profile for modeling aspect oriented frameworks," in Proceedings of the 13th Workshop on Aspect-oriented Modeling, 2009, pp. 1-6, doi:10.1145/1509297.1509299.
- [12] M. Mosconi, A. Charfi, J. Svacina, and J. Wloka, "Applying and evaluating aom for platform independent behavioral uml models," in Proceedings of the 2008 AOSD Workshop on Aspect-oriented Modeling, 2008, pp. 19-24, doi:10.1145/1404920.1404924.

## Strategic Goal Oriented Supplier Selection

Chang Joo Yun, Chung-Hsing Yeh

Faculty of Information Technology

Monash University

Clayton, Australia

{chang.joo.yun; chunghsing.yeh}@monash.edu

**Abstract**—This paper develops a novel approach for selecting a strategic supplier that best meets both long-term strategic goals and short-term requirements of an auto parts manufacturing company in a supply chain. The relative importance value of the supplier selection criteria to the strategic goals is first assessed by identifying how the criteria contribute to the goals. Based on the relative importance value of the criteria to the goals, the criteria weights are obtained. This strategic goal oriented weighting method is incorporated into a Multi-criteria Decision Making (MCDM) model. The potential contribution of the suppliers to the strategic goals is evaluated by the MCDM model. An empirical study on the supplier selection problem of the auto parts company is conducted to demonstrate the applicability and effectiveness of the approach. The approach has general application in selecting a strategic supplier for a supply chain with various strategic goals and supplier selection criteria.

**Keywords**—strategic suppliers; strategic goal oriented weighting method; Multi-criteria Decision Making.

### I. INTRODUCTION

Supplier selection has become a crucial issue for the success of a company in a supply chain, since good suppliers can considerably help the company improve its price competitiveness [1] and enhance the quality of its products [2]. A conventional approach to supplier selection is that a company contacts many available suppliers, sends requests for quotations for a specific order, evaluates them in terms of their quote and performance, and selects the best-performing one for the given order [3].

Selecting the most suitable supplier among a large number of potential suppliers with different levels of performance is inherently a Multi-criteria Decision Making (MCDM) problem [4]. The supplier selection problem thus requires making trade-offs between conflicting quantitative and qualitative criteria in order to evaluate the performance of the suppliers. Various supplier selection approaches have been proposed to meet the short-term requirements of customer orders. To select top-performing suppliers that best meet the short-term customer orders, the selection criteria such as price, quality, delivery, service, and technical capability have long been used [5][6] [7][8].

To ensure long-term survival and growth, companies increasingly want to develop long-term relationships with competitive suppliers rather than dealing with many suppliers on an order-by-order basis. It is difficult and

inefficient for the companies to maintain partnerships with all available suppliers, as it causes the excessive transaction and management costs, and difficulties for key supplier development. To develop the long-term relationships with suppliers, it is of strategic importance for a company to select the strategic supplier that can best meet long-term expectations of the company in addition to satisfying the short-term requirements of individual orders. No existing approaches, however, are available to choose a strategic supplier for meeting both long-term strategic goals and short-term requirements of companies. To fill this important gap, this paper presents a new approach for selecting the long-term strategic supplier that best contributes to both long-term strategic goals and short-term requirements of a company. To illustrate this approach, an MCDM model with a new weighting method is developed to evaluate the potential contribution of suppliers to the strategic goals by examining the relative importance value of the supplier selection criteria to the goals.

In Section 2, we first present the strategic supplier selection problem of an auto parts manufacturing company in South Korea (referred to as the KAP company). We then develop a new strategic supplier selection approach in the context of KAP in Section 3. In Section 4, we conduct an empirical study to illustrate the approach. Finally, we discuss key findings and practical implications of the study in Section 5.

### II. THE SUPPLIER SELECTION PROBLEM IN AN AUTO PARTS MANUFACTURING COMPANY

The KAP company is a first-tier auto parts manufacturer in South Korea. KAP produces auto parts such as sun roof, rail roof, pedal, fuel tank, suspension, axle housing, wheel housing, frame parts, lamp parts, and FRT cowl parts. KAP receives materials and components from about 160 suppliers for manufacturing auto parts, and supplies the auto parts to four major Korean auto companies including Hyundai, Kia, Renault Samsung, and GM Daewoo as a first vendor. KAP has eight domestic factories located all over the country and the purchasing division of each KAP factory selects its suppliers close to the factory for transportation cost reduction. Thus, most of the 160 suppliers of KAP are small and spread all over the country. This creates many low quality suppliers and causes the excessive transaction expenses for KAP to deal with many minor suppliers. Very often a customer order is met by many suppliers with small quantities. This makes



the qualified suppliers have less opportunity to develop as a major supplier by constantly receiving many orders with large quantities. In this regard, KAP wants to develop a small number of long-term suppliers. In addition, developing long-term relationships with competitive suppliers gives KAP substantial benefits such as product cost reduction, quality insurance, product commonality and risk reduction of product liability. To address this issue, there is an urgent need for KAP to have a strategic plan for developing long-term strategic suppliers.

Table 1 shows the strategic goals  $G_k$  ( $k=1, 2, \dots, 10$ ) and their weights  $w_k$  given by KAP in relation to its strategic plan for supplier selection. KAP considers achieving profitability ( $G_1$ ), coercive power for supplier management ( $G_4$ ) and long-term relationship between KAP and its suppliers ( $G_5$ ) as the most important goals with regard to supplier selection. Development of large suppliers ( $G_2$ ) and minimisation of product costs ( $G_3$ ) are also important strategic goals, as they affect long-term profit and growth of KAP. KAP prefers to work with major suppliers for ease of communication ( $G_8$ ), since they have advanced transaction systems and well-trained employees to make transaction processes easier for KAP.

The strategic goals are not directly used as criteria for evaluating the performance of suppliers, as it is difficult to assess and measure the performance of suppliers with respect to these goals. Thus, the supplier selection criteria  $C_j$  ( $j=1, 2, \dots, 14$ ), which are measurable quantitatively or assessable qualitatively and are independent of each other, are used to achieve the strategic goals, as shown in Table 2. Each supplier selection criterion is linked to the different strategic goals in terms of its relative importance value to the goals, as shown in Fig. 1. The total turnover and profits criteria ( $C_1$  and  $C_2$ ) significantly affect the achievement of profitability ( $G_1$ ), development of large suppliers ( $G_2$ ), and minimisation of product cost ( $G_3$ ). The types of equipment and number of equipment ( $C_3$ , and  $C_4$ ) are used to achieve part commonality ( $G_7$ ).  $C_1$ ,  $C_2$ ,  $C_3$ , and  $C_4$  are quantitative criteria and their values are obtained from KAP.

TABLE I. STRATEGIC GOALS AND THEIR WEIGHTS

Strategic goal		Weight
$G_1$	Profitability	0.2
$G_2$	Development of large suppliers	0.1
$G_3$	Minimisation of product costs	0.1
$G_4$	Coercive power for supplier management	0.2
$G_5$	Long-term relationship between KAP and its suppliers	0.2
$G_6$	Quality improvement of KAP and its suppliers	0.025
$G_7$	Part commonality	0.025
$G_8$	Ease of communication	0.1
$G_9$	Mitigation of patent disputes	0.025
$G_{10}$	Maximisation of return on investment	0.025
Total		1

To meet the strategic goals involving minimisation of product cost ( $G_3$ ), long-term relationship ( $G_5$ ), quality improvement ( $G_6$ ), part commonality ( $G_7$ ) and maximisation of return on investment ( $G_{10}$ ), qualitative criteria  $C_5$ ,  $C_6$ ,  $C_7$ , and  $C_8$  are used. These four qualitative criteria are assessed by KAP, based on its own assessment sheet with a score range of 0-100. For  $C_5$ , the lower the item price, the higher the score. Six selection criteria  $C_9$ ,  $C_{10}$ ,  $C_{11}$ ,  $C_{12}$ ,  $C_{13}$ , and  $C_{14}$  are related to the strategic goals including coercive power ( $G_4$ ), long-term relationship ( $G_5$ ), ease of communication ( $G_8$ ), and mitigation of patent disputes ( $G_9$ ).  $C_9$  and  $C_{10}$  are measured in percentage and qualitative criteria  $C_{12}$ ,  $C_{13}$ , and  $C_{14}$  are measured based on a 1-10 rating scale by KAP.

TABLE II. SUPPLIER SELECTION CRITERIA

Criteria	Measure
$C_1$	The total turnover of the supplier Dollar amount
$C_2$	The total profits of the supplier Dollar amount
$C_3$	The number of types of equipment of the supplier to produce KAP item Number
$C_4$	The total number of equipment of the supplier to produce KAP item Number
$C_5$	Average item price Score (0-100)
$C_6$	Average item quality (Error and shortage rate and late delivery rate history in previous transaction with KAP) Score (0-100)
$C_7$	Quality management system (Quality, resource, facility, and process) Score (0-100)
$C_8$	Technical skills Score (0-100)
$C_9$	The proportion of turnover of the supplier made from KAP Percentage
$C_{10}$	The proportion of profit of the supplier made from KAP Percentage
$C_{11}$	The total partnership year between KAP and the supplier Year
$C_{12}$	Personal connection, regionalism and kinship (blood relation) between KAP and the supplier Rating scale (1-10)
$C_{13}$	Proactive manner for communication, and transaction Rating scale (1-10)
$C_{14}$	Compatible strategic objectives of the supplier with KAP Rating scale (1-10)

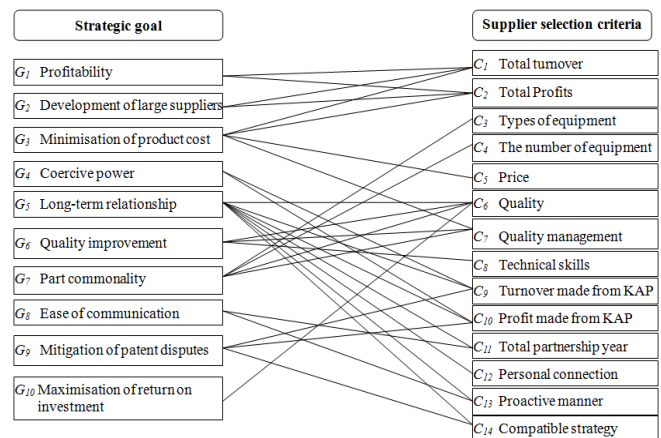


Figure 1. The relationship between the strategic goals and supplier selection criteria.

Based on the strategic goals and the supplier selection criteria, KAP wants to find which supplier can best perform for meeting its long-term strategic goals and short-term requirements. To address this important issue of KAP, we develop a strategic supplier selection approach.

III. THE STRATEGIC SUPPLIER SELECTION APPROACH

To evaluate the potential contribution of suppliers to the strategic goals, we develop an MCDM model with a strategic goals oriented weighting method. Evaluating suppliers can be formulated as an MCDM problem. The MCDM problem involves a finite set of  $I$  suppliers  $S_i$  ( $i=1, 2, \dots, I$ ) with respect to a set of  $J$  supplier selection criteria  $C_j$  ( $j=1, 2, \dots, J$ ). Assessments are conducted by decision makers to determine the weight vector  $W=(w_1, w_2, \dots, w_j)$  and the decision matrix  $X=\{x_{ij}, i=1, 2, \dots, I; j= 1, 2, \dots, J\}$ . The weight vector  $W$  represents the weights of criteria  $C_j$  and the decision matrix  $X$  represents the performance ratings ( $x_{ij}$ ) of suppliers  $S_i$  with respect to criteria  $C_j$ . The supplier selection problem has been solved by various MCDM methods [9][10][11][12][13]. Especially, the SAW method, also known as the weighted sum method, is the widely-used and well-known MCDM method [14]. The basic logic of SAW is to obtain a weighted sum of the performance ratings of each alternative with respect to criteria. This method is well justified theoretically and easily understood by a decision maker [14]. We use the SAW method to aggregate the performance ratings and criteria weights for evaluating suppliers.

In MCDM, different criteria weighting methods have been widely used to determine the importance of criteria [15][16][17]. We develop a new strategic goal oriented weighting method to determine the criteria weights by considering the strategic goals of KAP. To calculate the weights of the supplier selection criteria  $C_j$  ( $j=1, 2, \dots, J$ ) (given in Table 2), the relative importance value  $r_{jk}$  ( $k=1, 2, \dots, K$ ) of the supplier selection criteria  $C_j$  to the strategic goals  $G_k$  ( $k=1, 2, \dots, K$ ) (given in Table 1) is assessed by investigating how the criteria contribute to the goals of KAP, as shown in Fig. 1. A point estimate measurement, such as a five-point Likert type scale [17] has been widely used for making subjective assessments, as it is intuitively easy for a decision maker to use in expressing the subjectiveness and imprecision of the decision maker's evaluation [18]. The relative importance value of the supplier selection criteria with respect to the strategic goals is assessed by using a set of five linguistic terms {Not Important (NI), Somewhat Important (SI), Moderate (M), Important (I), Very Important (VI)}, which is associated with a corresponding set of numerical values {1, 2, 3, 4, 5}. The assessment is made by KAP, based on its current strategic needs and operational settings.

With the linguistic terms, the relative importance value of the supplier selection criteria with respect to the strategic goals is given from the knowledge of experts in KAP. Table 3 shows the relative importance value for determining criteria weights under various strategic goals. For example, the total turnover of the supplier ( $C_1$ ) is considered to be important (I) to achieve profitability ( $G_1$ ) and reduction of

product cost ( $G_3$ ). KAP considers that suppliers making high revenues can contribute more to profitability and cost reduction of KAP. Partnership year ( $C_{11}$ ) is considered to be very important (VI) to achieve long-term relationship between KAP and its suppliers ( $G_3$ ) and ease of transaction ( $G_8$ ). Partnership year ( $C_{11}$ ) that KAP has built with its suppliers can be a good indicator of buyer-supplier relationships. If KAP builds long-term relationships with its suppliers, transaction between KAP and its suppliers will become easier and more cost-effective to each other. The turnover and profits made from KAP criteria ( $C_9$  and  $C_{10}$ ) are the most important supplier selection criteria to achieve coercive power ( $G_4$ ) and long-term relationship ( $G_3$ ) of KAP. This coincides with KAP's opinions that they want to be suppliers' major customer taking most of their orders in order to hold a dominant position over a long-period of time. The types of equipment and number of equipment criteria ( $C_3$  and  $C_4$ ) are considered to be not important (NI) to achieve profitability ( $G_1$ ), product cost reduction ( $G_3$ ), quality improvement ( $G_6$ ), ease of transaction ( $G_8$ ), and mitigation of patent disputes ( $G_9$ ), while it significantly contributes to expansion of the size of the supplier ( $G_2$ ).

Based on the relative importance value of the supplier selection criteria with respect to the strategic goals, criteria weights  $w_j$  are calculated. The decision matrix  $X$  representing the performance ratings ( $x_{ij}$ ) of each supplier  $S_i$  ( $i=1, 2, \dots, I$ ) with respect to the supplier selection criteria  $C_j$  is assessed by KAP. These performance ratings are given based on both quantitative and qualitative assessments. The quantitative assessments are conducted based on objective data, while the qualitative assessments need the decision makers' subjective judgments to rate the performance of suppliers with respect to the supplier selection criteria.

TABLE III. THE RELATIVE IMPORTANCE VALUE OF THE SUPPLIER SELECTION CRITERIA TO THE STRATEGIC GOALS

Criteria	Strategic goal									
	$G_1$	$G_2$	$G_3$	$G_4$	$G_5$	$G_6$	$G_7$	$G_8$	$G_9$	$G_{10}$
$C_1$	I	VI	I	SI	SI	M	NI	NI	NI	M
$C_2$	I	VI	I	SI	SI	M	NI	NI	NI	M
$C_3$	NI	I	NI	SI	SI	NI	M	NI	NI	SI
$C_4$	NI	I	NI	SI	SI	NI	M	NI	NI	SI
$C_5$	VI	SI	VI	NI	M	SI	NI	NI	NI	SI
$C_6$	SI	M	M	NI	M	VI	M	NI	SI	SI
$C_7$	SI	M	I	NI	SI	VI	M	NI	SI	SI
$C_8$	SI	M	SI	NI	SI	I	NI	NI	M	M
$C_9$	M	NI	NI	VI	VI	NI	NI	NI	I	I
$C_{10}$	M	NI	NI	VI	VI	NI	NI	NI	I	I
$C_{11}$	NI	NI	NI	M	VI	NI	NI	VI	SI	I
$C_{12}$	NI	NI	NI	M	I	NI	NI	M	SI	NI
$C_{13}$	NI	NI	NI	M	I	NI	NI	I	NI	NI
$C_{14}$	NI	NI	NI	M	I	NI	NI	SI	I	I

Given the relative importance value  $r_{jk}$  of the supplier selection criteria  $C_j$  and the decision matrix  $X$ , an MCDM model with a strategic goal oriented weighting and SAW method is developed to evaluate the potential contribution of suppliers to the strategic goals as follows:

Step 1: Calculate the strategic goal oriented criteria weights  $w_j$  ( $j=1, 2, \dots, J$ ) of the supplier selection criteria  $C_j$  ( $j=1, 2, \dots, J$ ) by aggregating the weight  $w_k$  of each strategic goal  $G_k$  ( $k=1, 2, \dots, K$ ) and relative importance value  $r_{jk}$  of each selection criterion  $C_j$  to each goal  $G_k$  by

$$w_j = \frac{\sum_{k=1}^K w_k r_{jk}}{\sum_{j=1}^J \sum_{k=1}^K w_k r_{jk}} \quad (1)$$

Step 2: Normalise the performance ratings  $x_{ij}$  in the decision matrix  $X$  of supplier  $S_i$  for criteria  $C_j$  to make them compatible across the criteria by

$$y_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^I x_{ij}^2}} \quad (2)$$

Step 3: Calculate the contribution scores  $T_i$  ( $i=1, 2, \dots, I$ ) of supplier  $S_i$  by aggregating the strategic goal oriented criteria weights  $w_j$  and the normalised performance ratings  $y_{ij}$  by

$$T_i = \sum_{j=1}^J w_j y_{ij} \quad (3)$$

#### IV. THE EMPIRICAL STUDY

An empirical study is conducted to demonstrate how the strategic supplier selection approach works for KAP in order to meet its both long-term strategic goals and short-term requirements. To evaluate the potential contribution of suppliers to the strategic goals, an MCDM model with the strategic goal oriented weighting method is used. Ten suppliers are used as samples for the empirical study.

Based on the relative importance value criteria (given in Table 3) of the supplier selection criteria (given in Table 2) with respect to the strategic goals (given in Table 1), criteria weights  $w_j$  are calculated by Eq. (1). Table 4 shows the weights of the supplier selection criteria. The highest weight 0.094 is given for the proportion of turnover and profit of the supplier made from KAP ( $C_9$  and  $C_{10}$ ). KAP has assigned the highest weight 0.2 to coercive power for supplier management ( $G_4$ ) and long-term relationship between KAP and its suppliers ( $G_5$ ), and the proportion of turnover and profit of the supplier made from KAP ( $C_9$  and  $C_{10}$ ) are the most important selection criteria of suppliers to achieve coercive power for supplier management ( $G_4$ ) and long-term relationship between KAP and its suppliers ( $G_5$ ) of KAP. The number of types and total number of equipment of the supplier to produce KAP item ( $C_3$  and  $C_4$ ), on the other hand, gains the lowest weight 0.053.

TABLE IV. THE WEIGHTS OF THE SUPPLIER SELECTION CRITERIA

Criteria	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	$C_6$	$C_7$
Weight	0.077	0.083	0.053	0.053	0.082	0.065	0.062
Criteria	$C_8$	$C_9$	$C_{10}$	$C_{11}$	$C_{12}$	$C_{13}$	$C_{14}$
Weight	0.056	0.094	0.094	0.080	0.066	0.068	0.067

In the supplier selection problem, the performance ratings of ten suppliers with respect to the quantitative and qualitative selection criteria (given in Table 2) are assessed by KAP as shown in Table 5. The basic units of  $C_1$  and  $C_2$  are one million dollars. Average item price ( $C_3$ ) and quality ( $C_6$ ) criteria are adjusted by taking the reversal of the original data to make a consistent comparison across all criteria. The performance ratings of ten suppliers are then normalised by Eq. (2).

Given the criteria weights and normalised performance ratings of suppliers, the contribution scores of ten suppliers are calculated by Eq. (3). The highest and lowest contribution scores of the suppliers are 0.3164 and 0.2860 respectively. Table 6 shows the normalised performance ratings with respect to 14 selection criteria, and contribution scores (given in the last row) of the ten suppliers. Supplier  $S_2$  is a highest rated supplier that best contributes to both long-term strategic goals and short-term requirements of KAP.

To help better illustrate how the strategic goals affect the strategic supplier selection, sensitivity analysis has been conducted by adjusting the strategic goal weights. Sensitivity analysis for strategic goals  $G_1$ ,  $G_4$ , and  $G_5$  has been used as an example. Figs. 2, 3, and 4 show how priorities of ten suppliers can change, based on the different weight of the strategic goals  $G_1$ ,  $G_4$ , and  $G_5$  by sensitivity analysis. The x-axis shows the weight of each strategic goal and y-axis displays the contribution score of ten suppliers.

TABLE V. THE PERFORMANCE RATINGS OF TEN SUPPLIERS

	$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	$S_6$	$S_7$	$S_8$	$S_9$	$S_{10}$
$C_1$	59	86	80	105	108.7	100	123	80	90	76
$C_2$	6	5	4.7	15	12	20	18	21.5	17	25
$C_3$	20	69	123	48	101	84	21	95	44	61
$C_4$	2500	5160	1800	2000	980	1200	400	980	912	1900
$C_5$	90	99	95	99	100	90	70	90	85	55
$C_6$	80	50	95	75	70	80	85	80	30	80
$C_7$	95	99	80	55	10	90	80	80	70	75
$C_8$	90	80	75	30	25	95	90	80	80	70
$C_9$	0.98	0.86	0.8	0.69	0.95	0.5	0.5	0.94	0.9	0.7
$C_{10}$	0.75	0.79	0.5	0.9	0.9	0.7	0.83	0.6	1	0.7
$C_{11}$	23	20	22	9	18	9	16	10	15	7
$C_{12}$	10	7	10	8	7	9	8	9	8	10
$C_{13}$	9	7	10	6	7	9	10	4	6	9
$C_{14}$	8	5	5	10	10	9	3	5	7	9

TABLE VI. THE NORMALISED PERFORMANCE RATINGS, CONTRIBUTION SCORES AND RANKINGS OF SUPPLIERS

	$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	$S_6$	$S_7$	$S_8$	$S_9$	$S_{10}$
$C_1$	0.202	0.294	0.274	0.359	0.372	0.342	0.421	0.274	0.308	0.260
$C_2$	0.119	0.099	0.093	0.297	0.238	0.396	0.356	0.426	0.336	0.495
$C_3$	0.085	0.294	0.525	0.205	0.431	0.358	0.090	0.405	0.188	0.260
$C_4$	0.361	0.744	0.260	0.288	0.141	0.173	0.058	0.141	0.132	0.274
$C_5$	0.322	0.354	0.340	0.354	0.358	0.322	0.251	0.322	0.304	0.197
$C_6$	0.339	0.212	0.402	0.318	0.296	0.339	0.360	0.339	0.127	0.339
$C_7$	0.389	0.405	0.327	0.225	0.041	0.368	0.327	0.327	0.286	0.307
$C_8$	0.379	0.337	0.316	0.126	0.105	0.400	0.379	0.337	0.337	0.295
$C_9$	0.387	0.340	0.316	0.273	0.375	0.198	0.198	0.372	0.356	0.277
$C_{10}$	0.304	0.320	0.203	0.365	0.365	0.284	0.336	0.243	0.405	0.284
$C_{11}$	0.457	0.398	0.437	0.179	0.358	0.179	0.318	0.199	0.298	0.139
$C_{12}$	0.365	0.255	0.365	0.292	0.255	0.328	0.292	0.328	0.292	0.365
$C_{13}$	0.359	0.279	0.399	0.239	0.279	0.359	0.399	0.159	0.239	0.359
$C_{14}$	0.338	0.211	0.211	0.423	0.423	0.381	0.127	0.211	0.296	0.381
Score	0.3155	0.3164	0.3115	0.2872	0.2980	0.3122	0.2860	0.2933	0.2890	0.3009

Fig. 2 shows that  $S_6$  is superior to suppliers  $S_2$ ,  $S_1$  and  $S_3$  in terms of the contribution score when the weight of profitability ( $G_1$ ) is higher than about 0.45, while suppliers  $S_2$ ,  $S_1$ , and  $S_3$  are preferable to  $S_6$  when the weight of profitability ( $G_1$ ) is less than about 0.16. Assigning the higher weight on profitability ( $G_1$ ) increases the probabilities of  $S_6$  to be selected. The contribution scores of suppliers  $S_6$ ,  $S_2$ ,  $S_1$ , and  $S_3$  are always more dominant than suppliers  $S_{10}$ ,  $S_5$ ,  $S_8$ ,  $S_9$ ,  $S_4$ , and  $S_7$ , regardless of the weight of profitability ( $G_1$ ) due to the fact that the performance ratings of suppliers  $S_6$ ,  $S_2$ ,  $S_1$ , and  $S_3$  with respect to profitability ( $G_1$ ) are much superior to suppliers  $S_{10}$ ,  $S_5$ ,  $S_8$ ,  $S_9$ ,  $S_4$ . This indicates that suppliers  $S_6$ ,  $S_2$ ,  $S_1$ , and  $S_3$  contribute more to the profitability ( $G_1$ ) strategic goal than suppliers  $S_{10}$ ,  $S_5$ ,  $S_8$ ,  $S_9$  and  $S_4$ . Fig. 3 shows that two suppliers  $S_2$  and  $S_1$  are always superior to others. This indicates that suppliers  $S_2$  and  $S_1$  contribute more to the coercive power ( $G_4$ ) strategic goal than suppliers  $S_3$ ,  $S_6$ ,  $S_{10}$ ,  $S_5$ ,  $S_8$ ,  $S_9$ ,  $S_4$  and  $S_7$ . Fig. 4 shows that  $S_1$  is superior to supplier  $S_2$  when the weight of  $G_5$  is higher than about 0.25, while supplier  $S_2$  or  $S_6$  can be chosen when weight of  $G_5$  is less than about 0.25. Suppliers  $S_1$ ,  $S_2$ ,  $S_3$ , and  $S_6$  are always more dominant than others. This indicates that they contribute more to the long-term relationship ( $G_5$ ) strategic goal than others. It is noteworthy that depending on the different weights of the strategic goals, priorities of suppliers change. Thus, KAP will be able to use these different weighting combination results for its various strategic plans. The result of the sensitivity analysis demonstrates that determining the different weights of strategic goals changes criteria weights, thus affecting strategic supplier selection of KAP.

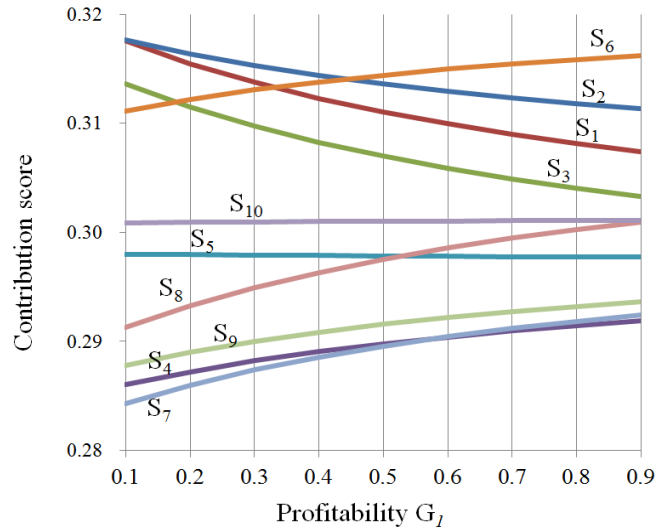


Figure 2. Sensitivity analysis based on profitability ( $G_1$ ).

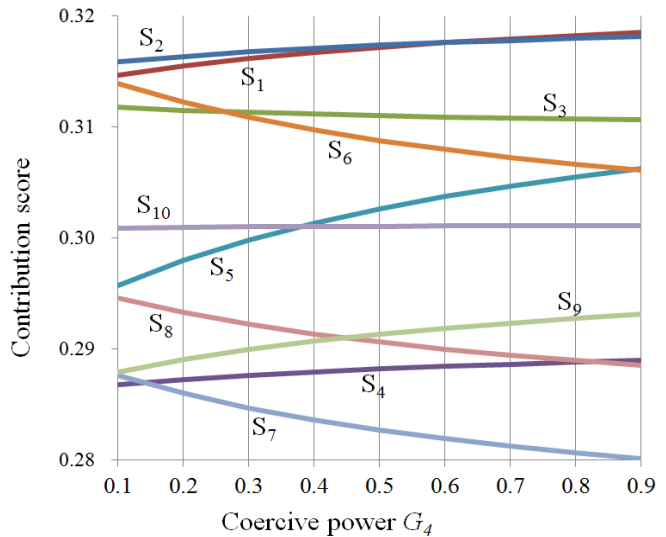


Figure 3. Sensitivity analysis based on coercive power ( $G_4$ ).

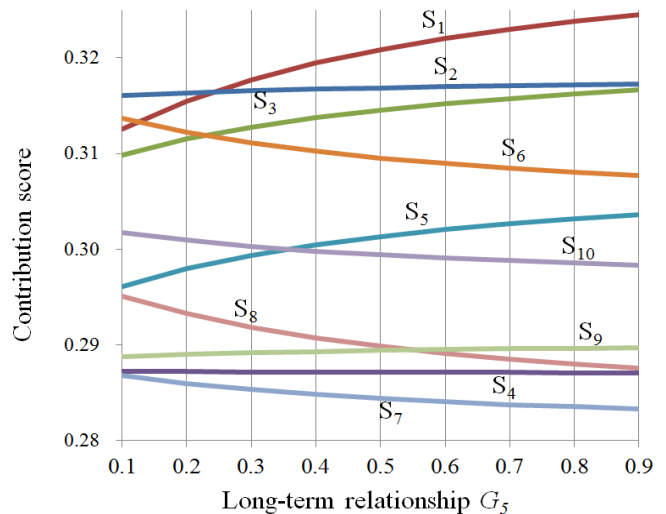


Figure 4. Sensitivity analysis based on long-term relationship ( $G_5$ ).

## V. CONCLUSION

Selecting the supplier that best meets both specific long-term expectations and short-term requirements of a company requires a comprehensive approach. To address this decision problem, we have developed a new strategic supplier selection approach for meeting both long-term strategic goals and short-term requirements of customer orders of a company. A new weighting method has been developed to evaluate the potential contribution of the suppliers to the strategic goals of a company by examining how the selection criteria contribute to the goals. The strategic supplier selection approach developed in this paper provides a structured approach for a company to choose the long-term strategic supplier most contributing to its long-term survival and growth, and short-term requirements. Although the approach is exemplified with the strategic supplier selection problem in the auto parts manufacturing company, it has general application in selecting a strategic supplier in different supply chains with various strategic goals and short-term requirements.

## REFERENCES

- [1] S. H. Ghodsypour and C. O' Brien, "The Total Cost of Logistics in Supplier Selection, under Conditions of Multiple Sourcing, Multiple Criteria and Capacity Constraints," *International Journal of Production Economics*, vol. 73, 2001, pp. 15-27.
- [2] R. Ramanathan, "Supplier Selection Problem: Integrating DEA with the Approaches of Total Cost of Ownership and AHP," *Supply Chain Management: An International Journal*, vol. 12, 2007, pp. 258-261.
- [3] W. Ho, P. K. Dey, and M. Lockström, "Strategic Sourcing: a Combined QFD and AHP Approach in Manufacturing," *Supply Chain Management: An International Journal*, vol. 16, 2011, pp. 446-461.
- [4] S. C. Ting and D. I. Cho, "An Integrated Approach for Supplier Selection and Purchasing Decisions," *Supply Chain Management: An International Journal*, vol. 13, 2008, pp. 116-127.
- [5] S. H. Cheraghi, M. Dadashzadeh, and M. Subramanian, "Critical Success Factors for Supplier Selection: an Update," *Journal of Applied Business Research*, vol. 20, 2004, pp. 91-108.
- [6] G. W. Dickson, "An analysis of vendor selection systems and decisions," *Journal of Purchasing*, vol. 2, 1996, pp. 5-20.
- [7] P. Parthiban, H. Abdul Zubar, and P. Katarar, "Vendor Selection Problem: A Multi-criteria Approach based on Strategic Decisions," *International Journal of Production Research*, vol. 51, 2012, pp. 1535-1548.
- [8] C. A. Weber, J. R. Current, and W. C. Benton, "Vendor Selection Criteria and Methods," *European Journal of Operational Research*, vol. 50, 1991, pp. 2-18.
- [9] P. Agarwal, M. Sahai, V. Mishra, M. Bag, and V. Singh, "A Review of Multi-criteria Decision Making Techniques for Supplier Evaluation and Selection," *International Journal of Industrial Engineering Computations*, vol. 2, 2011, pp. 801-810.
- [10] L. De Boer, E. Labro, and P. Morlacchi, "A Review of Methods Supporting Supplier Selection," *European Journal of Purchasing and Supply Management*, vol. 7, 2001, pp. 75-89.
- [11] W. Ho, X. Xu, and P. K. Dey, "Multi-criteria Decision Making Approaches for Supplier Evaluation and Selection: A Literature Review", *European Journal of Operational Research*, vol. 202, 2010, pp. 16-24.
- [12] H.-J. Shyr and H.-S. Shih, "A Hybrid MCDM Model for Strategic Vendor Selection," *Mathematical and Computer Modelling*, vol. 44, 2006, pp. 749-761.
- [13] M. Sonmez, "A Review and Critique of Supplier Selection Process and Practices," Working Paper [1], Jan 2006, Business School, Loughborough University, Loughborough.
- [14] C. L. Hwang and K. Yoon, *Multiple Attribute Decision Making - Methods and Applications: A State-of-the-Art Survey*, Springer-Verlag, New York, 1981.
- [15] M. Pöyhönen and R.P. Hämmäläinen, "On the Convergence of Multiattribute Weighting Methods," *European Journal of Operational Research*, vol. 129, 2001, pp. 569-585.
- [16] J.-J. Wang, Y.-Y. Jing, C.-F. Zhang, and J. H. Zhao, "Review on Multi-criteria Decision Analysis Aid in Sustainable Energy Decision-making," *Renewable and Sustainable Energy Reviews*, vol. 13, 2009, pp. 2263-2278.
- [17] C.-H. Yeh and Y.-H. Chang, "Modeling Subjective Evaluation for Fuzzy Group Multicriteria Decision Making," *European Journal of Operational Research*, vol. 194, 2009, pp. 464-473.
- [18] C.-H. Yeh, R.J. Willis, H. Deng, and H. Pan, "Task Oriented Weighting in Multi-criteria Analysis," *European Journal of Operational Research*, vol. 119, 1999, pp. 130-146.

## Enacting a Requirement Engineering Process with Meta-Tools: an Exploratory Project

Sana Damak Mallouli

Centre de Recherche en Informatique  
University of Paris 1 Pantheon Sorbonne  
Paris, France  
sana.mallouli@gmail.com

Saïd Assar

Institut Mines-Telecom,  
Telecom Ecole de Management  
Evry, France  
said.assar@it-sudparis.eu

**Abstract**— An engineering process can hardly be described rigorously because of its dynamic and decision-oriented nature. Applying goal-oriented modeling for representing such processes is a promising approach. However, goal-oriented process models have no clear operational semantics, and building an enactment engine for such models is a highly complex task. To avoid building such tools in an ad-hoc manner and to overcome maintainability and portability issues, we investigate in this paper the practical feasibility of meta-CASE and CAME based approaches for constructing an enactment tool for a goal-oriented model. We describe and analyze a project in which MetaEdit+, a leading meta-tool, is tested. Our aim is to evaluate the meta-tool approach and to explore its possibilities in terms of process model enactment. During this project, a requirement elicitation process is used as a preliminary test bed.

*Keywords*-meta-modeling; intentional process modeling; meta-CASE; MetaEdit+; execution semantics.

### I. INTRODUCTION

Goal modeling is a prominent design paradigm in various domains such as business process modeling [1], method engineering [2], and requirements engineering [3]. In the method engineering context, the highly dynamic and decision-oriented nature of engineering processes has led to describe such processes using goal-oriented models [4]. The main advantage of goal-oriented process modeling is its ability to go beyond the simple modeling of sequences of activities as proposed by other notations such as BPMN (Business Process Model and Notation) or SPEM (Systems Process Engineering Meta-model). However, how to enact goal-oriented models and how such enactment can be specified and implemented are still open research questions.

The general topic of this paper is how to construct a tool that provides enactment mechanisms for goal-oriented process models. In previous research projects, our research team has engineered multiple tools for earlier event-oriented methods [5], context-oriented [6], and goal-oriented methodological processes [7]. As these tools were built in an ad-hoc manner, recurrent problems have shown up. Maintainability is a key issue: as product and process meta-models are hard coded, it is very difficult to make the tools evolve when the meta-models are changed, even slightly. Portability is a bottleneck too: any evolution of the underlying technology makes the tool rapidly obsolete unless

large updates are made to the code. Therefore, we have decided to investigate the possibility of using method engineering tools to solve these issues.

Meta-CASE and Computer Aided Method Engineering (CAME) technologies were introduced in the 90's as an answer to the general problem of providing software support to method and tool customization and/or creation [8]. These meta-tools were expected to facilitate and to accelerate the production of method toolset, and to overcome maintainability and portability drawbacks. As any other software artifacts, tools and meta-tools need to be assessed and evaluated [9]. However, research works that review meta-CASE/CAME tools are rather limited. Prominent evaluations are early works by Martiin & al. [10][11]. Although more recent, works in [12][13] are not fundamentally different from earlier studies. In [12], the evaluation framework is inspired by Martiin & al. [10], and the results are limited to general appreciations. The study in [13] is more detailed, but some studied tools are technically not available anymore (e.g., Mentor [6]). The novelty lies in the evaluation framework; it is based on ISO 9126 quality model with more usability and portability concerns.

What emerge from these previous studies is that when it comes to process enactment, there is clearly not enough empirical knowledge about the use of meta-tool in real cases. Authors claim that support for process modeling is insufficient, and that mechanisms for expressing process enactment are limited or inexistent. So the goal of this paper is to experiment meta-CASE technology for enacting a goal-oriented process, and to explore problems and issues related with such an approach. More precisely, this paper deals with the following research question: To which extent do meta-tools provide a viable and satisfying solution to the problem of building a maintainable and portable enactment engine for a goal-oriented formalism?

To answer this question, we empirically evaluate through a lab project MetaEdit+, a well-known meta-tool. This meta-tool was selected to run the project because it is recognized as a leading method engineering environment with high level of technical maturity [13][14]. The project consisted in using MetaEdit+ to define a goal-oriented formalism and to construct an enactment tool to support it. Evaluation relied on inspecting the obtained system at the end of the project, and on analyzing data gathered along the engineering process. The evaluation is based on a quality framework that

is synthesized from related works analysis. The framework is organized around three perspectives (Table 1): the proposed *formalism* for expressing the process enactment semantics, the *meta-tools* available for specifying the process, and the obtained *target tool* for process enactment.

TABLE I. META-CASE EVALUATION FRAMEWORK

Perspective	Criteria
Formalism	<ul style="list-style-type: none"> <li>Level of <b>expressiveness</b> for process specification</li> <li><b>Cognitive effort</b> to specify the process part</li> </ul>
Meta-tool	<ul style="list-style-type: none"> <li><b>Suitability</b> for process specification</li> <li><b>Ease of use</b> for process specification</li> </ul>
Target tool	<ul style="list-style-type: none"> <li>Level of <b>maintainability</b> gains for the target meta tool</li> <li>Level of <b>portability</b> gains for the target meta tool</li> </ul>

The rest of the paper is structured as follows. In Section 2, we briefly present works that are related to our research. Section 3 describes the Map formalism on which is based the project. In Section 4, the project itself is described. Section 5 presents the outcomes of the project in terms of meta-models and tool specifications, describes the obtained target tool and illustrates with a requirements elicitation process taken from the literature. Section 6 reports our evaluation of the project's outcomes using the quality framework. The paper ends with concluding remarks in which we discuss the results in regard with the research question, and we formulate revised research questions.

II. RELATED WORK

Providing process specification and enactment support is very important for Software and Method engineering fields. In this context, the Moskitt4ME approach [23] is a relevant solution where process models are initially expressed using SPEM meta-model, then transformed into BPMN models. The BPMN process description can then be enacted using an off-the-shelf execution engine called "Activiti Engine".

Our research question is similar to the issue addressed in Moskitt4ME approach. However, our solution has two main particularities: (1) first, we use a goal-oriented process modeling language to capture fine grained software engineering processes and to go beyond what SPEM based approaches can do; (2) second, our target is to design a CAME tool itself and to adopt meta-models driven development in order to guarantee maintainability and portability.

Another recent work dealing with enactment process is presented in [24]. This work proposes xSPEM tools (i.e., executable SPEM) for editing, simulating and verifying SPEM process models. The approach is based on dynamic meta-modeling (i.e., the definition of domain specific languages with behavioral semantics) in order to simulate and validate models. The basic idea is to extend the meta-model and assimilate its execution semantics to that of state-based machines and workflows.

xSPEM allows the simple modeling of sequences of activities. Thus, it is not well suited to represent engineering processes. For this reason, we are investigating goal-oriented process enactment in this paper.

III. THE MAP FORMALISM

Our research work concerns the Map formalism, a goal-oriented model that is particularly well adapted for representing engineering processes. Based on the intention paradigm [15], the Map captures the intentions that a process is expected to fulfill, together with a set of available strategies to realize these intentions (Fig. 1a). Each intention can be realized by one or more strategy, and the process is represented as a labeled graph with intentions as nodes and strategies as edges [16]. An edge enters a node if its strategy can be used to achieve the intention of the node. A section of the Map is a triplet composed of a source intention, a target intention and a strategy (e.g., <J, S<sub>JK1</sub>, K> in Fig. 1a).

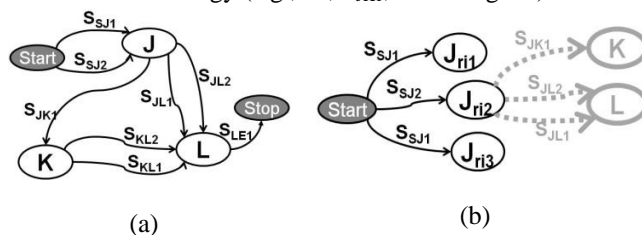


Figure 1. A map example (a), and an illustration of achieved intentions and candidate sections (b)

**Map operational semantics.** Beyond precedence relationships between intention achievements, the operational semantics of the Map are decision-oriented. The combination of a past intention achievement, a strategy and an intention (i.e., the triplet <J<sub>ri2</sub>, S<sub>JL1</sub>, L> in Fig. 1b) is called a *candidate section* and is a fundamental concept for expressing the operational semantics. At each execution of a section, a new set of candidate sections (i.e., sections that can be executed in the next step) is computed. Given a certain state of the working products (i.e., the set of product instances,) and the history of achieved intentions, the candidate sections computation is done by checking which sections match with scheduling possibilities [17].

IV. PROJECT OVERVIEW

The chosen meta-tool for the project is MetaEdit+ [18]. Figure 2 presents an overview of the project's main steps.

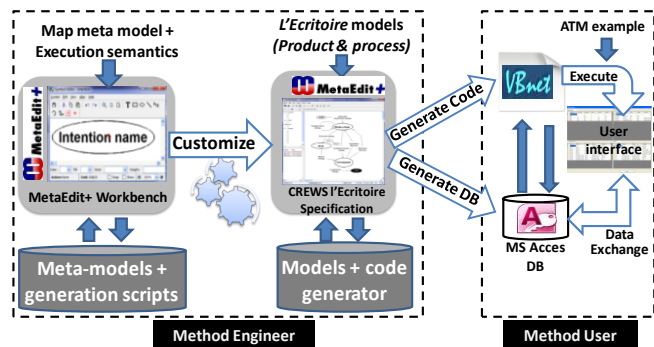


Figure 2. Overview of the engineering process underlying the exploratory project

The method engineer defines meta-models (product and process) using MetaEdit+ Workbench toolset and the

GOPRR (Graph, Object, Property, Relationship, Role) meta-modeling concepts, and specifies generation rules with Merl scripting language. A customized version of MetaEdit+ editor is automatically constructed (i.e., target tool) this tool includes code generation functionalities. The method user can then use this tool to define a product and a process model, and generates an enactment engine.

The project was run with the requirement elicitation process underlying CREWS *L'Ecritoire* [19][7]. Goal discovery and scenario authoring are complementary activities with CREWS *L'Ecritoire*: once a goal is discovered, a scenario is authored as a possible concretization of the goal, and can then be followed by further goal discovery from the authored scenario.

These goal-discovery/scenario-authoring sequences are repeated to incrementally populate the requirement chunks hierarchy (a *Requirement Chunk* (RC) is defined as a <goal, scenario> couple). The underlying RC elicitation process is described as a map.

This project was conducted by a group of three persons: a senior researcher, a PhD student and a master student. The senior researcher controlled the project, collected data and provided support all along the project in case of specific problems. The PhD student defined the Map formalism with MetaEdit+ and provided assistance in expressing and

validating Map operational semantics. The tasks of the MSC student were to specify the code generator in MetaEdit+ scripting language for the target platform (VB.NET and MS Access), and to run CREWS *L'Ecritoire* case example.

## V. PROJECT OUTPUT

### A. Meta-Modeling

For the product part, as the Map formalism does not specify product models, we defined a standard E/R style product meta-model. This meta-model is simple and is not presented here for the sake of space. Figure 3 presents the meta-model for the process part expressed in GOPRR. Boxes represent objects, diamonds represent relationships, and circles are roles. *Intention* and *Strategy* concepts are defined as objects. They are connected by two relationships (*Strategy-to-Intention* and *Intention-to-Strategy*) and, beyond name and identifier, have certain properties necessary for expressing operational semantics (i.e., "State" attributes). "Product fragment", "Product consumption" and "Product generation" are specific attributes which are added to the Map definition. They are necessary for expressing the behavior of a map when it is executed [20], and play an important role in computing candidate sections.

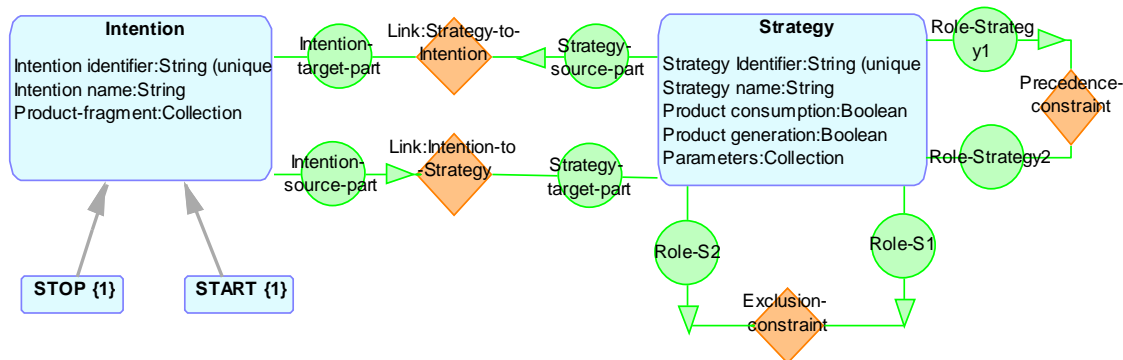


Figure 3. Map process meta-model defined in GOPRR

### B. Target Tool for CREWS *L'Ecritoire*

The product model is simple; it contains three objects (i.e., a requirement chunk, a goal and a scenario) and three links (And, Or, Refine). For the sake of space, details are not presented here, but interested readers can refer to [16].

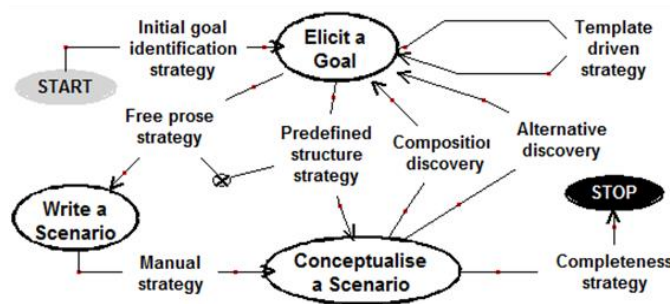


Figure 4. CREWS *L'Ecritoire* process model defined in the target tool

Figure 4 presents the requirement chunks elicitation process expressed as a map. Three intentions are defined together with a set of strategies to achieve them; this model is a slightly simplified version of the original process in [16].

### C. Enactment specification

Enactment of a specific map is handled by the generated Map engine. This engine is a VB.NET program which manipulates an MS Access DB. The executable code of the Map engine and the relational structure of the DB are product dependant, and will be constructed by the code generator according to the product model (i.e., requirement chunks) and to the process model (i.e., the map). Indeed, for each method, a specific Map engine is generated; however, the code generator is generic.

Figure 5 shows the structure of the engine's DB that is generated for the CREWS *L'Ecritoire* case. Tables in the "Process" group are generic and contain the map, those in the "Trace" group will contain the results of the execution



(product dependant), and those in the "Product" group will contain product instances (i.e., requirement chunks).

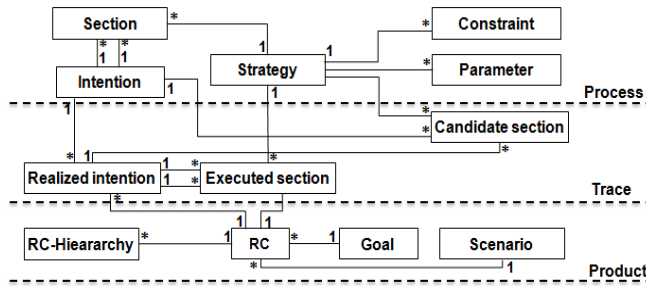


Figure 5. Database structure of the Map engine for CREWS L'Ecritoire

The dependency of the Map engine towards the product structure impacts also the generated algorithm for candidate sections computing, which is an essential element in the enactment process. This calculation is based on the content of the engine's DB. For each realized intention Ri in the trace, a set of candidate sections is calculated by linking Ri to the corresponding intention I in the process map, and identifying all connecting strategies and connected intentions which form a possible path from Ri.

Figure 6 presents a screen shot of Map engine execution for CREWS L'Ecritoire case. From left to right top down, the 1st window contains the list of executed sections, the 2nd contains the stack of achieved intentions, the content in the 3rd window is static as it contains the sections of the process map under execution, and the last window contains the dynamically computed list of candidate sections.

## VI. EVALUATION AND DISCUSSION

The quality framework (cf. Table 1) will now be used to evaluate and discuss the project results and outcomes. Results are synthesized in Table 2.

### A. Formalism and meta-tool

For the process part, suitability and usability are clearly insufficient. Operational semantics are specified through a collection of scripts written in Merl (MetaEdit+ scripting language). These scripts define navigation logic through the meta-models and output (i.e., generate) instruction in target code (i.e., VB.NET). This step is very complex, and must be handled in two sub steps: write, run and debug first the target code for a sample Map engine DB, and when this code is satisfying, adapt it and integrate it into Merl generation scripts. Errors in the generated code are very hard to be directly corrected in the generating Merl scripts. Although the meta-tool provides a debugging facility, it is largely insufficient when the generated code becomes complex. The cognitive effort for correlating the generated code with procedures for generating such code (i.e., Merl scripts writing) was so high that the whole project team needed to cooperate frequently on this task.

### B. Target tool

According to the project team perception, the obtained Map engine operates correctly and is easy to use, as the possibilities for interactions are limited to choosing a candidate section and inputting some data. However, the user interface is much less attractive although the Map enactment is made sufficiently clear through textual display windows (Fig. 6). This can be enhanced; it will however make the generated code and the generation scripts much more complex and less portable.

Concerning target tool maintainability, the scoring is mitigated. Any change to the Map formalism which does not impact the operational semantics is easily handled with MetaEdit+'s graphical interface and with small updates to the generated code. However, if the modification changes the operational semantics, the impact on the generated code can then be much larger.

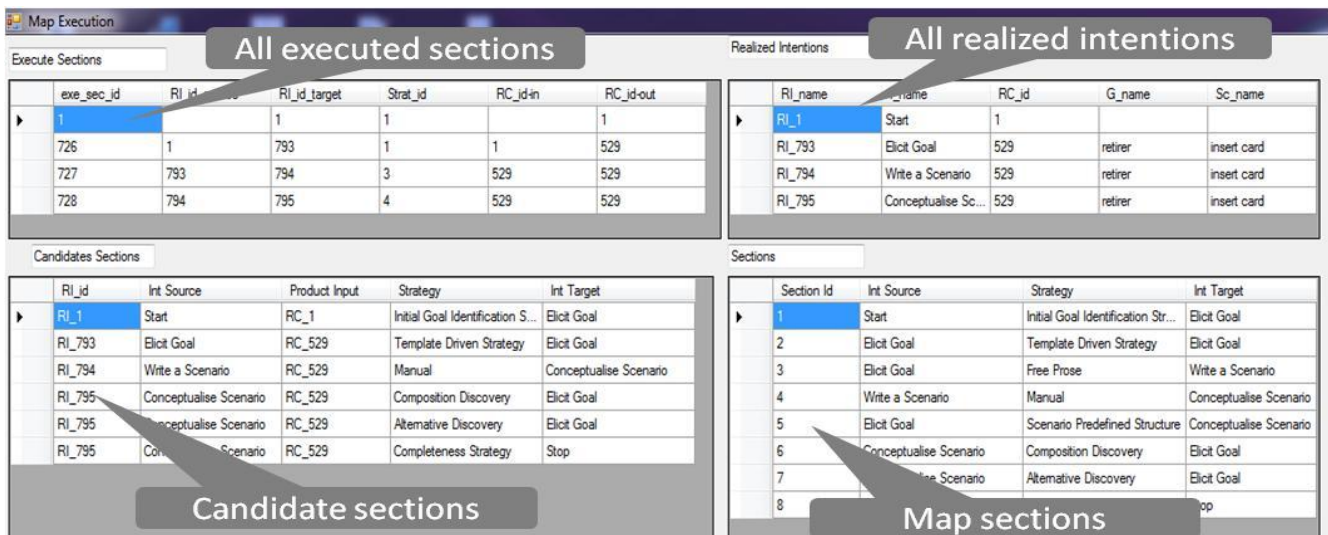


Figure 6. Interface for the generated target tool

The portability issue is strategically more important as it impacts tool mid and long term existence. The scoring here is unsatisfactory because, although it is fully possible to update the generating scripts and the generated code to accommodate a new version of the target platform (e.g., a new version of VB) or a different platform (i.e., MySQL DBMS), the cost of this task is high. However, compared with an ad-hoc approach, the project is insufficient to get a real insight about the comparative advantage of using meta-CASE in terms of portability.

TABLE II. PROJECT EVALUATION RESULTS

Perspective	Criteria	Evaluation result
Formalism	Expressiveness	▪ <u>Insufficient</u> : execution semantics is expressed using a large set of procedural scripts
	Cognitive effort	▪ <u>Very high</u> : the language designer must mentally correlate code generation actions (i.e., Merl scripts) with generated code (i.e., VB.net code)
Meta-tool	Suitability	▪ <u>Low</u> : a simple text editor for developing generation scripts, there is no model nor a GUI
	Ease of use	▪ <u>Low</u> : the validation of a script is very hard, with only a limited debugging facility
Target tool	Maintainability	▪ <u>Satisfactory</u> in case of modifications to the PLM that do not alter its execution semantics; <u>unsatisfactory</u> otherwise, because of complex changes to code generation scripts
	Portability	▪ <u>Unsatisfactory</u> : moving to a different target platform induces high level of changes to code generation scripts

## VII. CONCLUDING REMARKS

In this paper, we have reported an experimental development project dedicated to evaluating the contribution of meta-CASE to the enactment of Map, a goal-oriented process modeling formalism. As "building a system in and of itself does not constitute research", the contribution to basic research is "the synthesis of new concepts in a tangible product" [21]. What we learned from this experience is resumed in the two following. First, providing mechanisms for enacting a goal-oriented process is possible only if operational semantics can – at the conceptual level – be expressed using the execution paradigm underlying a certain target platform. This is line with recent work about the necessity to raise the level of abstraction of compilation techniques up to the modeling phase [22]. Second, in the Map case, building a generic Map engine is impossible as the algorithm for candidate section computation is product dependant. Meta-CASE technology provides thus an interesting possibility to overcome this problem by combining CASE customization and the design of adequate code generators. This opens the door to multiple applications in the field of Domain Specific Languages, CASE tool construction and process enactment.

### A. Threats to validity

Some threats to the validity of this study are: the limitation to only one project, the dependence on participants' background, and the mix between development team and observation team.

### B. Research question revisited

In meta-CASE actual technology, process operational semantics are expressed in a procedural manner in code generating scripts, and process enactment is obtained through the execution of generated code. These specifications are loosely correlated with the meta-modeling part, they are not specified in a declarative manner and they cannot be represented graphically. They are difficult to build and to validate. Indeed, like in compiler technology, the validity of the code generator cannot be demonstrated unless some formal notations are used, or the code generator itself is built using some generic tool. Thus, we revise our research question into: How to express in a rigorous manner process operational semantics at the meta-modeling level of abstraction in order to facilitate its validation and evolution? This is the subject of our actual research work. We are investigating graphical notations for expressing meta-models' operational semantics.

## REFERENCES

- [1] I. Bider and P. Johannesson, "Goal-oriented business process modeling: Guest Editorial," in *Business Process Management Journal*, vol. 10, no. 6, 2005, pp. 621–623.
- [2] J. Ralyté and C. Rolland, "An Approach for Method Reengineering," in H. S.Kunii, S. Jajodia, and A. Sølvberg, "Conceptual Modeling," ER 2001, Berlin/Heidelberg, LNCS, Springer, vol. 2224, 2001, pp. 471–484.
- [3] A. Lapouchnian "Goal-oriented requirements engineering: An overview of the current research," University Of Toronto, Canada, 2005.
- [4] S. Si-Said and C. Rolland, "Formalising Guidance for the CREWS Goal-Scenario Approach to Requirements Engineering," in H. Jaakkola, H. Kangassalo, and E. Kawaguchi, "Information Modelling and Knowledge Bases," IOS Press, vol.10, 1999, pp. 172–190.
- [5] C. Rolland and al., "The RUBIS system," in T.W. Olle, A.A. Verrijn-Stuart, and L. Bhabuta, "Computerized Assistance During the Information Systems Life Cycle," North-Holland, 1988, pp. 193–239.
- [6] S. Si-Said, C. Rolland, and G. Grosz, "MENTOR: A Computer Aided Requirements Engineering Environment," in CAiSE'96, P. Constantopoulos, J. Mylopoulos, and Y. Vassiliou, LNCS, Springer, Berlin/Heidelberg, vol. 1080, 1996, pp. 22–43.
- [7] C. Souveyet and M. Tawbi, "Process centred approach for developing tool support of situated methods," in DEXA'98, G. Quirchmayr, E. Schweighofer, and T.J.M. Bench-Capon, LNCS, Springer, Berlin/Heidelberg, vol.1460, 1998, pp. 206–215.
- [8] S. Kelly and K. Smolander, "Evolution and issues in metaCASE," *Information and Software Technology*, vol. 38, no. 4, 1996, pp. 261–266.
- [9] J.P. Gray, A. Liu, and L. Scott, "Issues in software engineering tool construction," *Information and software technology*. 2000, Vol. 42, no. 2, pp. 73–77.

- [10] P. Marttiin, M. Rossi, V.-P. Tahvanainen, and K. Lyytinen, "A comparative review of CASE shells: A preliminary framework and research outcomes," in *Information & Management*, vol. 25, no. 1, 1993, pp. 11–31.
- [11] P. Marttiin, F. Harmsen, and M. Rossi, "A functional framework for evaluating method engineering environments: the case of Maestro II/Decamerone and MetaEdit+," in S. Brinkkemper, K. Lyytinen, and R.J. Welke, "Method engineering: Principles of method construction and tool support," IFIP, Chapman & Hall, London, 1996, pp. 63–86.
- [12] I. Van de Veerd and M. Saeki, "An evaluation of computerized tools for method construction," Institute of Electronics, Inf. and Comm. Engineers (IEICE), Tokyo, Japan, 2007.
- [13] A. Niknafs and R. Ramsin, "Computer-Aided Method Engineering: An Analysis of Existing Environments," in Z. Bellahsene and M. Léonard, CAiSE'08, LNCS, Springer, Berlin/Heidelberg, vol. 5074, 2008, pp. 525–540.
- [14] S. Kelly and J.-P. Tolvanen, "Domain-specific modeling: enabling full code generation," Wiley-Interscience, IEEE Computer Society, N.J. Hoboken, 2008.
- [15] C. Rolland, "Capturing System Intentionality with Maps," in J. Krogstie, A.L. Opdahl, and S. Brinkkemper, *Conceptual Modelling in IS Eng.*, Springer, 2007, pp. 141–158.
- [16] C. Rolland, N. Prakash, and A. Benjamen, "A Multi-Model View of Process Modelling. Requirements Engineering," *Requirements Engineering*, vol. 4, no. 1, 1999, pp. 169–187.
- [17] M.H. Edme, "A proposal for intentional modeling and guiding of information system usage (in French)," PhD thesis, University of Paris 1 La Sorbonne, France, 2005.
- [18] MetaCASE: <http://www.metacase.com/> [retrieved: April 25th, 2013]
- [19] C. Rolland, C. Souveyet, and C.B. Achour, "Guiding goal modeling using scenarios," *IEEE Transactions on Software Engineering*, vol. 24, no. 12, 1998, pp. 1055–1071.
- [20] S. Assar, S.D. Mallouli, and C. Souveyet, "A behavioral perspective in meta-modeling," in 6th Int. Conf. on Software and Data Technologies (ICSOFT), Sevilla, Spain, 2011.
- [21] J.F. Nunamaker, M. Chen, and T.D.M. Purdin, "Systems development in information systems research," *Journal of Management Information Systems*, vol. 7, no. 3, 1990, pp. 89–106.
- [22] R. Bendraou, J.M. Jezéquel, and F. Fleurey, "Achieving process modeling and execution through the combination of aspect and model-driven engineering approaches," *Journal of Software: Evolution and Process*, vol. 24, no. 7, November 2012, pp. 765–781.
- [23] M.Cervera, M. Albert, V. Torres, and V. Pelechano, "The MOSKitt4ME Approach: Providing Process Support in a Method Engineering Context," in P. Atzeni, D. Cheung, and S. Ram, *Conceptual Modeling (ER)*, LNCS, Springer Berlin Heidelberg, 2012, pp. 228–241.
- [24] R. Bendraou, B. Combemale, X. Cregut, and M.-P. Gervais, "Definition of an Executable SPEM 2.0," *Proc. 14th Asia Pacific Software Engineering Conf. (APSEC)*, 2007, pp. 390–397.

# Finding an Optimal Model for Prediction of Shock Outcomes through Machine Learning

Sharad Shandilya

Xuguang Qi

Kayvan Najarian

Rosalyn H Hargraves

School of Engineering

Virginia Commonwealth University

Richmond VA

shandilya.sharad@gmail.com

Michael C Kurz

Department of Emergency Medicine,

Virginia Commonwealth University,

Richmond, VA

mkurz@vcu.edu

Kevin R Ward

Department of Emergency Medicine,

Michigan Critical Injury and Illness

Research Center,

University of Michigan,

Ann Arbor, MI

keward@med.umich.edu

**Abstract**—Predicting defibrillation success is of paramount importance to resuscitating a victim of cardiac arrest. Several studies have attempted to extract/discover predictive features from electrocardiogram signals. Till date, no method has been accepted or implemented in the field, primarily due to low accuracy and/or specificity. We process a relatively large database of signals and report performance of an integrative Machine Learning model through multiple measures. 358 signals, with 140 leading to return of spontaneous circulation through defibrillation attempts and the rest, 218 signals, leading to unsuccessful defibrillations were used to train and test the model on non-overlapping sample sets. Techniques from machine learning, non-linear dynamics and signal processing were applied to extract features and subsequently classify them. In this study, we identify opportunities for reducing variance in the predictive model and propose a method for searching the optimal model. The accuracy and Receiver Operating Characteristic area of the proposed model are 78.8% and 83.2%, respectively. These compare with 74% and 69.2% accuracy and Receiver Operating Characteristic area for the leading 'Amplitude Spectrum Area' measure. The performance of the model will further improve with addition of other physiologic signals, as previously shown in a study by our research group. The model shows great potential to be viable in the clinical setting.

**Keywords**-predictive model; overfitting; machine learning; defibrillation success; parameter search.

## I. INTRODUCTION

In the United States, 300,000 lives are annually lost due to cardiac arrest. Survival rates for out-of-hospital cardiac arrest patients are very low [2]. Ventricular Fibrillation (VF) is a common arrhythmia in cardiac arrest [3]. Coronary artery perfusion provided by Cardio-Pulmonary Resuscitation (CPR) prior to defibrillation has been shown to improve chances for Return of Spontaneous Circulation (ROSC) [4]. Defibrillation is a procedure that delivers an electrical current that depolarizes a critical mass of the myocardium simultaneously. Defibrillation increases the possibility of the sino-atrial node regaining control of the rhythm. Coronary artery perfusion provided by CPR prior to defibrillation has been shown to improve chances for ROSC [4]. Repetitive unsuccessful shocks can reduce chest

compression time and cause injury to cardiac tissue, impacting heart function upon survival. Even worse, unsuccessful shocks can cause VF to deteriorate into asystole or Pulseless Electrical Activity (PEA), which are more difficult to resuscitate [5]. A victim's chances of survival worsen by 10% for every minute of VF that remains untreated [4].

Hence, increasing efficacy of defibrillation attempts is of principal importance. To achieve this, we develop an integrative decision-support model that guides the interventionist by learning from real-time information extracted from the patient.

Fourier Transform (FT) based methods [1] assume a linear and deterministic basis for decomposing signals. As another limitation, FT decomposition yields a time averaged frequency estimate of the original signal. These limitations, except for the *deterministic* and *non-chaotic* assumption, are overcome by our use of the Wavelet Transform (WT) [6]. Furthermore, we use a dual-tree decomposition algorithm for the complex WT, nearly eliminating shift-variance, which is a limitation of DWT.

Additionally, the QPD-PD method [6] is able to characterize chaotic signals while allowing for stochasticity/non-determinism. In the same study, it was shown that features calculated through QPD-PD, along with those from WT, represent a powerful set whose knowledge can be integrated with a Machine Learning (ML) algorithm for improved performance.

Furthermore, we propose a novel method of selecting the optimal ML model to boost generalize-ability on blind data.

## II. OBJECTIVES

Thus far, no automated model or feature has been accepted in the field for decision-support during cardiac arrest due to low specificity at desired sensitivity and/or low overall accuracy. By relaxing the assumptions made by methods previously tried, as described in the previous section, we aimed to build a model with high sensitivity and specificity. Model (parameter) selection is another important endeavor in boosting performance, especially when employing heuristic-based ML algorithms.

A high specificity translates to a reduced number of unnecessary shocks, which cause thermal injury to the heart in addition to adding to the time lost.

Section 3.1 gives an overview of the methods and the novel techniques proposed. Section 3.2 details the data used to build and test the model. Section 3.3 pertains to filtering as a data pre-processing step. Section 3.4 describes the QPD-PD method used to characterize data prior to calculation of features. Section 3.5 describes the proposed method for model selection. Sections 3.6 reports how the model's performance was tested. Section 4 contains results of a comparative study with the leading Amplitude Spectrum Area (AMSA) measure [1]. Section 5 states the conclusions from this study.

### III. METHODOLOGY

#### 3.1 Overview

Data was characterized through techniques from non-linear dynamics, autoregressive modeling, time-frequency decomposition. Novel time-series features were devised in order to distinguish pre-defibrillation VF signals yielding a successful defibrillation from those that did not. The method Quasi-Period Density Prototype Distance (QPD-PD) derives stochastic quasi-periods through time-delay embedding.

Supervised feature selection was performed to identify the most discriminative features. Selection was performed in a nested fashion so as to maintain blindness to the test folds. In the same nested fashion, model selection was performed for different combinations of parameters. The overall approach is discussed in [6]. In this paper, we propose a novel method of model selection and term it "High-Platform Method". Additionally, the assumptions that underlie the QPD-PD model are formally tested and reported. Model performance is reported on a newly acquired large (for the problem context and when compared to other studies) dataset provided by Zoll Medical Corporation [Chelmsford, MA]. Simultaneous 10-fold cross-validation was used to evaluate the model. Matlab® software was utilized for all signal-processing needs. Figure 1 illustrates the high-level steps of the methodology.

Blocks A1, A2 and B1, B2 represent pre-processing of signals. C1, C2, and C3 represent extraction of features/characteristics from the data/signals. Blocks D1 and D2 represent the ML phase of the system where features are selected, model selection is performed by the proposed method, and the inducted ML model is tested.

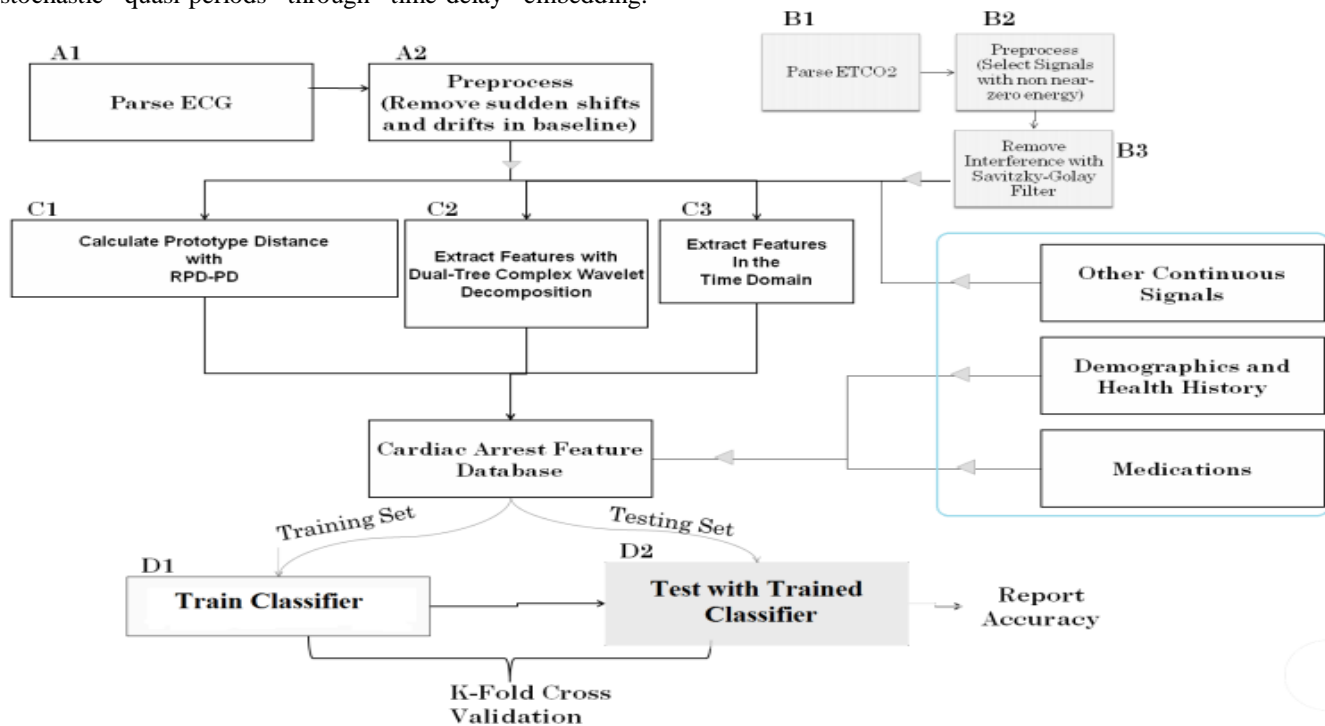


Figure 1. Overview of the System

#### 3.2 The Data

The data processed in this study contained a total of 358 defibrillation counter-shocks on 153 subjects. Every single one of these shocks was labeled / annotated by Dr. Michael Kurz. Among these shocks, 140 were successful and 218 were unsuccessful. The same dataset was used for training

and testing of both VCU algorithm and the AMSA method as described by [1]. Successful defibrillation was defined as a period of greater than 15 seconds with narrow QRS complexes under 150 beats per minute with confirmatory evidence from the medical record or electrocardiogram

(ECG) that a return of spontaneous circulation (ROSC) has occurred.

### 3.3 Pre-Processing

Signals were filtered by utilizing the method proposed in [7]. The method performs custom filtering and was designed by observing properties of the ECG signals in the database.

### 3.4 Feature Extraction

This QPD-PD method focuses on distributions of pseudo-periodicity, while accounting for stochastic character of the signal. Parameter selection and feature calculation are geared for classification. Figure 2 shows the quasi-period density plots for each class. The periods were then convolved with the exponential function in order to quantify the difference between the two densities. For the density represented by  $Q$ , we propose to calculate a new probability density function for  $Q$  by convolving as follows

$$Q * \text{Exp}(s) = \int_0^P Q(p) \text{Exp}(s - p) dp \quad (1)$$

where  $\text{Exp}$  is the exponential function,  $s$  is the shift,  $p$  is a specific quasi-period,  $P$  is the largest period recorded. Dual-tree complex wavelet transform and other time-series features were also calculated as described in [6].

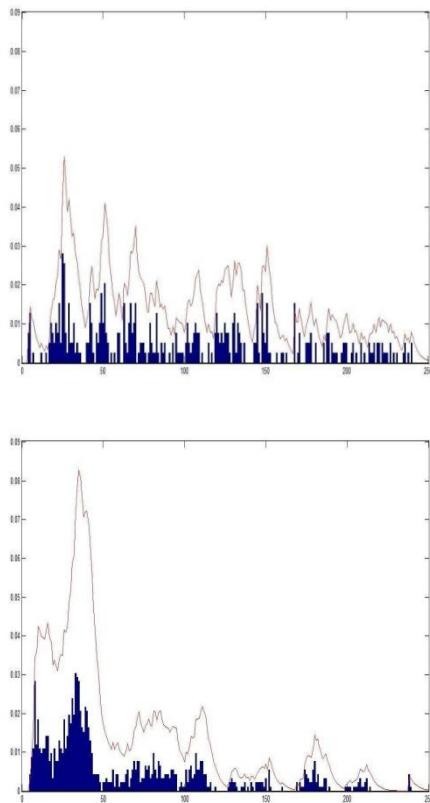


Figure 2. Quasi-Period Density. X-axis: Quasi-Periods; Y-axis: Amplitude Density. QPD for a successful shock (above) and QPD for an unsuccessful shock (below). Blue bars represent the normalized amplitude for each pseudo period. Red line represents QPD convolved with the exponential function. If most of the amplitude is clustered in neighboring Quasi-Periods, as is the case above, the convolution helps accentuate that fact (higher peak in the line-plot).

### 3.5 Feature and Model Selection

Feature selection, performed with cross-validation on the whole dataset, creates a positive bias in accuracies by indirectly using information from the test set. As such, feature selection must be performed within the training set that is generated for each run of  $k$ -fold cross-validation. However, using the entire training set leads to over-fitting within the training set, which creates a negative bias in accuracies when each test fold is passed through the model [8]. To prevent this, and to also select parameters for the learning algorithm in a nested fashion, we employ a twice-nested version of cross-validation.

For parameter tuning, we propose the ‘High-platform’ method. Figure 3 represents a plot of median accuracy for different combinations of parameters. We assume that close values of parameters create models that are conceptually and performance-wise similar. The conjecture is that picking a model that does both, maximizes the median accuracy and belongs to a ‘high-performing neighborhood’, would preclude an overfitted model that may have a high accuracy, but would be adjacent to other poorly performing models. After combinations of all parameters are tried and median-accuracies are recorded, we pick a model that

- 1) Exists within the neighborhood that has the highest mean median-accuracy, and
- 2) Has the highest median-accuracy within that neighborhood.

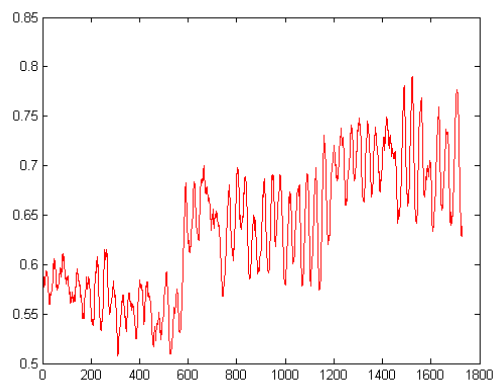


Figure 3. Finding a High-Platform. Four parameters (Learning Rate, Momentum, Hidden Neurons, and Epochs) for a Neural Network are varied. X-axis: Unique combinations of parameters. Y-axis: Median cross-validated accuracy for each combination. A region, such as the one between 1200 and 1450, with the highest mean median-accuracy is chosen.

Each neighborhood is defined by a fixed set of values for the subset of parameters that we want to optimize. Then averaging the accuracy over a neighborhood yields the ‘platform’, which amounts to nullifying the effect of varying values of the remaining parameters from the superset. For instance, optimizing a total of four parameters would involve the following. After all combinations of possible values of the 4 different parameters are tried,

- 1) Calculate the average accuracy for each unique combination of the first three parameters
- 2) Find and fix the combination that has the highest average accuracy,
- 3) Then, vary values of the fourth parameter and select the model with the highest accuracy.

To aid comprehension, this is similar to a *pseudo* "best-first" approach for parameter tuning while keeping the order of parameters the same as above. In this procedure, an exhaustive search would be performed for two parameters, then the third and fourth are chosen one at a time. It would progress as follows:

- 1) Try all possible combinations of the first two parameters, fix the best one, and call it Opt2 ,
- 2) Try all values for the third parameter (for a fixed value of the fourth parameter) and note the best one,
- 3) Try all values of the fourth parameter (for a fixed value of the third parameter) and note the best one,
- 4) From steps 2 and 3 above, pick the model with higher accuracy and call it Opt3,
- 5) Optimize on the remaining parameter, yielding Opt4.

This is similar to, but not exactly, best-first search because values for the third and fourth parameters are selected apriori during the procedure. Instead, the proposed high-platform method searches for the best model by taking one or more steps back from exhaustive search. For one-step-back, it reduces variance at both, penultimate and ultimate levels. Optimizing at the penultimate level is done by averaging variation in performance induced by varying values of the last parameter. For a greater reduction in variance, the search would take two-steps-back and average the variation in performance induced by all combinations of the remaining *two* parameters.

### 3.6 Classification and Validation

Cross-validation (CV) is a ubiquitous method that creates non-overlapping training and testing sets from the data set in order to train and test a given model. It is essential in avoiding a positive bias in accuracy when the entire data set is used for both training and testing. CV is desirable because it allows utilization of the entire dataset in order to validate a method, as opposed to having a single training set and a single test set. Multiple comparisons of the proposed model and AMSA method [1] were performed using 10-fold cross validation.

## IV. RESULTS AND DISCUSSION

Overall accuracy for the proposed model was 78.8%. We compared the proposed model with the current state-of-the-art AMSA method as follows.

- Comparison at 80% sensitivity: In this study, the two algorithms, proposed model and AMSA, were trained to provide sensitivity of 80%. In this case, our model provided an accuracy of 74% and specificity of

70.2%. For the same level of sensitivity, AMSA provided an accuracy of 53.6% and specificity of 36.7%.

- Comparison at 90% sensitivity: A similar analysis was conducted, except that both algorithms were trained to provide a sensitivity of 90%. our method provided an accuracy of 68.4% and specificity of 54.6%. For the same level of sensitivity, AMSA provided an accuracy of 43.3% and specificity of 13.3%.
- Using Receiver Operating Characteristic testing, the Area Under the Curve (AUC) for proposed method is 83.2% while this number for AMSA is 69.2%. These results are similar to the ones reported in [6].

For a given desired sensitivity, the proposed model can provide a significantly higher accuracy and specificity. Notably, within the range of 80-90% of sensitivity, the model provides about 40% higher specificity. This means that when trained to have the same level of sensitivity, the model will have far fewer false positives (unnecessary shocks). Furthermore, the significantly higher AUC of the selected model suggests significantly higher reliability than AMSA.

## V. CONCLUSION AND FUTURE WORK

We have developed a novel algorithm for predicting successful defibrillation of VF and then selecting an optimal model. The model is built upon knowledge extracted with signal-processing, non-linear dynamical and machine-learning methods. The selected ML model shows viability for decision-assistance in clinical settings. Our approach, which has focused on integration of multiple features through machine learning techniques, suits well to boosting predictive accuracy and model robustness.

Improvements will be sought in feature selection through novel methods. Performance of ML algorithms from disparate paradigms will be compared after model selection with a fixed (selected) feature subset.

## REFERENCES

- [1] G. Ristagno, A. Gullo, G. Berlot, U. Lucangelo, F. Geheb, and J. Bisera, "Prediction of successful defibrillation in human victims of out-of-hospital cardiac arrest: a retrospective electrocardiographic analysis," *Anaesth Intensive Care*, vol. 36, 2008, pp. 46-50.
- [2] G. Nichol, E. Thomas, and C. W. Callaway, "Regional variation in out-of-hospital cardiac arrest incidence and outcome," *J Am Med Assoc*, vol. 300, 2008, pp. 1423-1431.
- [3] V. M. Nadkarni et al, "First documented rhythm and clinical outcome from in-hospital cardiac arrest among children and adults," *JAMA*, vol. 295, 2006, pp. 50-57.
- [4] T. D. Valenzuela, D. J. Roe, S. Cretin, D. W. Spaite, and M. P. Larsen, "Estimating effectiveness of cardiac arrest interventions: a logistic regression survival model," *Circulation*, vol. 96, 1997, pp. 3308-3313.
- [5] H. Strohmenger, "Predicting Defibrillation Success," *Cardiopulmonary Resuscitation*, vol. 14, 2008, pp. 311-316.
- [6] S. Shandilya, K Ward, M Kurz, K Najarian. "Non-Linear Dynamical Time-Series Characterization for Prediction of

- Defibrillation Success through Machine Learning", BMC Informatics and Decision Making, 2012.
- [7] S. Shandilya, M. C. Kurz, K. R. Ward, and K. Najarian, "Predicting defibrillation success with a multiple-domain model using machine learning," IEEE Complex Medical Engineering, 2011, pp. 22-25.
- [8] R. Kohavi and G. John, "Wrappers for feature subset selection", Artificial Intelligence, vol. 97, 1997, pp. 273-324.



# A Design of Hybrid Automatic Repeat Request Scheme based on FlexRay used for Smart Hybrid Powerpack

Leilei Shi, Jekwang Choi and Hunmo Kim

Department of Mechanical Engineering  
Sungkyunkwan University  
Suwon, Korea

skkushileilei@126.com, bugs2020@naver.com, kimhm@me.skku.ac.kr

**Abstract**—FlexRay is an automotive communication protocol, which is developed by core members of the FlexRay consortium (BMW, Bosch, DaimlerChrysler, Freescale, GM, NXP Semiconductors and Volkswagen). Nowadays, it is set to become widely used in the automotive industry where it will replace traditional networking schemes such as Controller Area Network (CAN). However, in the FlexRay, it just uses the Cyclic Redundancy Check (CRC) code to detect the errors. In this paper, we design of a Hybrid Automatic Repeat Request (HARQ) scheme which combines the CRC code with Reed-Solomon (RS) code. Smart Hybrid Powerpack (SHPP) is an electro-hydraulic system, which is combined with advanced technology like fault tolerant, intelligent control, smart phone remote control and monitoring. This paper will apply the FlexRay into the SHPP system using the HARQ scheme. Simulation result shows good performance by using HARQ scheme comparing with CRC code only.

**Keywords**-FlexRay; HARQ; SHPP

## I. INTRODUCTION

FlexRay is a communication network for distributed automotive systems. It is an option for upgrading existing network systems using traditional CAN in the automotive industry as well as other industrial control applications. Because of the dual channels for fault tolerance which can communicate the same information and improve the fault tolerance ability, it could also be used for new applications in industrial automation, where safety and reliability in a work environment can be guaranteed [1].

The FlexRay protocol has been designed to carry information at a rate of 10Mbits/s by each channel, which means that an equivalent data rate of 20Mbits/s can be achieved. The FlexRay frame format consists of header segment (5 bytes), payload segment (0-254 bytes) and trailer segment (3 bytes). The frame trailer segment contains a 24-bits CRC code. FlexRay frame format is shown in Fig. 1, which is a standard [2].

Hydraulic systems have been widely used in industrial applications because of their durability, high power-to-weight ratio, reliability and large force and torque, etc. [3]. Two types of hydraulic transmission system are commonly used in industry, namely valve controlled and pump controlled hydraulic system [4]. Valve control hydraulic

system has the main fault of the low energy efficiency. The loss of energy is due to the leakage from the pump bypass valves or transfer of energy into heat via throttle losses at the control valves.

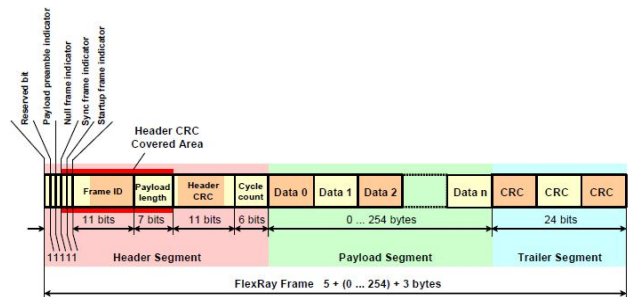


Figure 1. FlexRay frame format

At the range of pump controlled hydraulic systems, the concept of Electro Hydraulic Actuator (EHA) has been popular because that the EHA acts as a Power-Shift which shifts the power from high-speed electric motor to the high-force of hydraulic cylinder by bi-directional piston pump. Thus, the EHA creates a sleeker, cleaner way to produce hydraulic power with higher energy efficiency [3].

SHPP is an electro-hydraulic system which combines the EHA system with advanced technology like fault tolerance, intelligent control, smart phone remote control and monitoring. The whole structure of SHPP system is shown in Fig. 2.

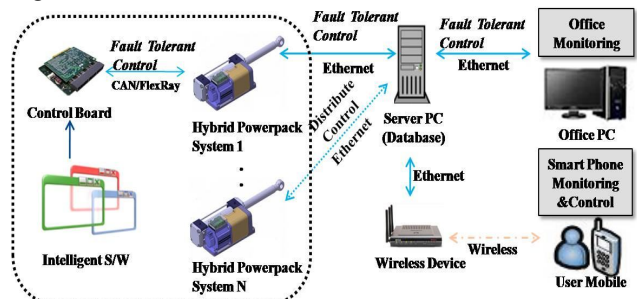


Figure 2. SHPP system structure

In the SHPP, for sensor part, some sensor fault tolerant technology will be done. The sensors include position sensor,

pressure sensor and temperature sensor. For network part, FlexRay error control coding will be made in this SHPP system. Some intelligent control methods like fuzzy control, adaptive fuzzy control will be tried in the SHPP. These control methods are smart technology recovering from faults like noises, disturbances and so on. The most advanced technology is that smart phone will act a good role in the system, including monitoring and control by the smart phone. All these advanced technology will make the SHPP system smart, intelligent, safe and accurate.

In this paper, we suggest using the FlexRay in the SHPP and the system can be more reliable and safe. Furthermore, based on the FlexRay protocol we design of a HARQ scheme as the fault tolerant logic to improve the error correction ability. The rest of this paper is organized as follows: Section 2 will introduce proposed HARQ scheme for SHPP system. Section 3 will show the MATLAB simulation result. Conclusions of this paper will be represented in the last section.

## II. DESIGN OF HARQ SCHEME FOR SHPP

### A. CRC Code and RS Code

In the HARQ scheme, CRC code and R-S code are used to design the fault tolerant logic. CRC code is used to detect errors after R-S decoding. In this paper, we use the CRC-24 code as the error detecting code. The generator polynomial  $g(x)$  is written as (1) shows to us [2].

$$g(x) = x^{24} + x^{22} + x^{19} + x^{18} + x^{16} + x^{14} + x^{13} + x^{11} + x^{10} + x^8 + x^7 + x^6 + x^3 + x + 1 \quad (1)$$

R-S code is one of the error correcting codes. It can efficiently correct not only random errors but also burst errors. R-S code is a non-binary cyclic code with symbols made up of  $m$ -bit sequences, where  $m$  is any positive integer having a value greater than 2. R-S  $(n, k)$  codes on  $m$ -bit symbols exist for all  $n$  and  $k$  for the following relation as (2) [5].

$$0 < k < n < 2^m + 2 \quad (2)$$

In (2),  $k$  is the number of data symbols being encoded, and  $n$  is the total number of code symbols in the encoded block. The relation between  $n$  and  $k$  can be given as (3) [5].

$$(n, k) = (2^m - 1, 2^m - 1 - 2t) \quad (3)$$

In (3),  $t$  is the symbol-error correcting capability of the code. It means that the code is capable of correcting any combination of  $t$  or fewer errors and  $t$  can be expressed as (4) [5].

$$t = (n - k) / 2 \quad (4)$$

In this paper, we use RS (15, 9) code with 4-bit in each symbol; so, the error probability of this R-S code is 3 symbols, which means it can correct 3 symbol errors or fewer symbol errors successfully.

### B. Proposed HARQ Scheme

The algorithm of the proposed HARQ scheme is shown as seen in Fig. 3. First, the signal from main controller will be changed to data. In the receiver, the R-S code will decode the coded data and then do CRC decoding. By using the R-S (15, 9) code, 3 symbols errors or less than 3 symbols errors will be corrected perfectly. Then, the CRC codes will check whether the R-S codes corrected all the errors successfully. If there is no error detected by CRC codes, the signals will be sent to the SHPP.

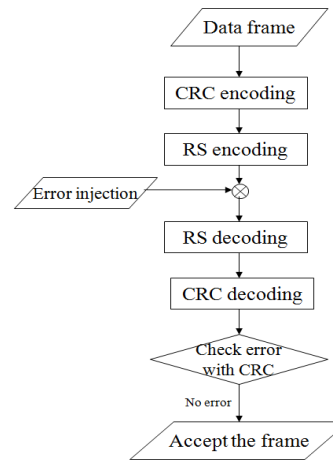


Figure 3. Proposed HARQ scheme

### C. System Modeling of Brushless DC Motor Drive System

In this paper, we only use the brushless DC Motor (BLDCM) drive modeling. After that, we will adapt to EHA system based on this paper result. First step is to make the BLDCM drive modeling for error control coding between Main controllers and Motor drive. The BLDCM system model is developed considering armature voltage and load torque as two inputs, angular velocity and Motor position as two outputs. In order to simplify the model, all the stator phase windings are assumed to have equal resistance per phase and constant self and mutual inductances, iron loss is negligible, motor flux is unsaturated and power semiconductor devices are ideal. The line to line voltage equation in the motor can be represented as, [6]

$$V(t) = I(t)R + L \frac{dI(t)}{dt} + e_b(t) \quad (5)$$

where  $V(t)$  is input voltage of phase,  $I(t)$  is armature current of phase,  $e_b(t)$  is back-emf of phase,  $R$  is armature resistance of phase,  $L$  is armature inductance of phase. The BLDCM generates three phase voltages by the line to line voltage

equation. The phase currents, phase voltages and phase back-emfs are assumed to be equal. The total resistance opposing the phase current will be twice the resistance per phase.

For linear analysis, we assume that the torque developed by the motor is proportional to the current the armature current and the air-gap flux. Thus, the motor torque equation can be expressed as, [6]

$$T_M(t) = K_T I(t) \quad (6)$$

where  $T_M(t)$  is the motor torque in N-m,  $K_T$  is the torque constant in N-m/A,  $I(t)$  is the armature current in A.

Since the sum of all the opposing torques due to mechanical elements of motor and load torque is equal to the torque developed by the motor, the cause-and-effect equations for the motor circuit can be written as, [6]

$$T_M(t) = T_L(t) + J_M \frac{dw(t)}{dt} + B_M w(t) \quad (7)$$

$$e_b(t) = K_b w(t) \quad (8)$$

where  $T_M(t)$  and  $T_L(t)$  are torque developed by motor and load torque,  $w(t)$  is the rotor angular velocity in rad/s,  $K_b$  is back-emf constant in N-m/A,  $e_b(t)$  is the back-emf in N-m/A.

### III. RESULTS

In this paper, motor control method is a PID control in the BLDCM drive system and we assume that speed control system is designed without load conditions because we focused on the network system between Main controller and BLDCM drive system. The PID controllers are most suited for control systems with fixed system dynamics and no parameter variations during the operating conditions. The PID controller can be easily implemented for the BLDCM drive system to improve its performance. The simulink model using PID controller is shown in Fig. 4.

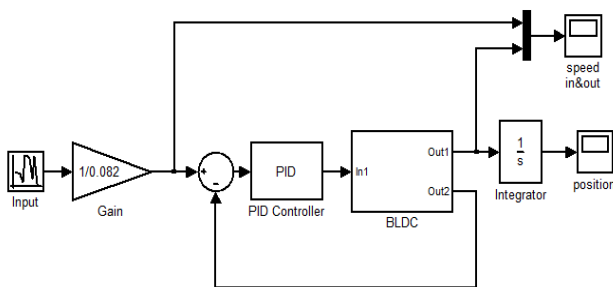


Figure 4. Simulink model using PID controller

To make the BLDCM drive system using PID control method, we decide parameters of the BLDCM. Parameters of the BLDCM are shown in Table I [6].

TABLE I. PARAMETERS OF THE BLDCM

Parameters	Values
Stator resister, $R(\Omega)$	0.57
Stator inductance, $L(\text{m-H})$	1.5
Inertia, $J_M(\text{kg-m}^2)$	0.000023
Torque constant, $K_T(\text{N-m/A})$	0.082
Back-emf constant, $K_b(\text{N-m/A})$	0.082

The gain parameters of PID controller are determined using the Ziegler Nichols Tuning method [7] and found to be  $K_p=0.2051$ ,  $K_i=29.52$  and  $K_d=0.00053326$ . The simulation of BLDCM drive control system is carried out using MATLAB software. We determine that the reference rotor speed is 414.63rad/s. The response of the BLDCM speed control system with PID controller is shown in Fig. 5.

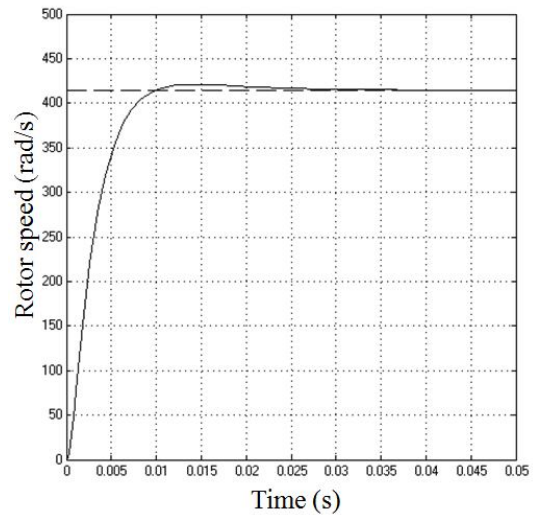


Figure 5. Response of the BLDCM control system with PID controller

For error control coding between the Main controller and the Motor drive system, we convert the rotor speed to the rotor speed voltage as (8) and we calculate the rotor voltage is 34V.

Fig. 6 shows the input voltage of the Speed controller, when the error is 3 symbols with CRC code only. The voltage is always 50V but the desired voltage is 34V. In the simulation, we assume the noise is forever existed and the error pattern is sure, which means even retransmission is made, the errors never change. When errors happen, CRC code can just detect the errors and retransmit the signal. It costs time and the errors could not be processed well forever.

Because of the shortage in the CRC code, when some errors are always existed in the channel, the retransmission is needed all the time. If the system is critical, it will have big problem. So we need some other methods to cover the CRC shortage. In the below, we add the RS code in the channel, which means HARQ scheme is used.

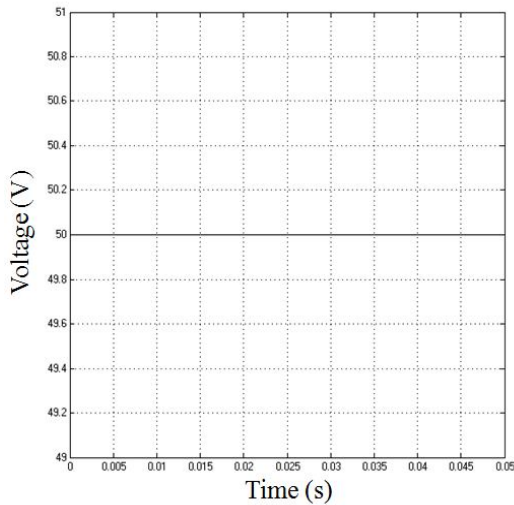


Figure 6. The input voltage of Speed controller (Error is 3 symbols with CRC code only)

Fig. 7 shows the simulink model of error control coding with HARQ scheme. The error control coding consists of mainly three parts that are Encoder, Channel and Decoder. Display 1 shows the corrected error symbols by RS code. Display 2 shows the CRC decoding result: “0” means no error and “1” means have errors.

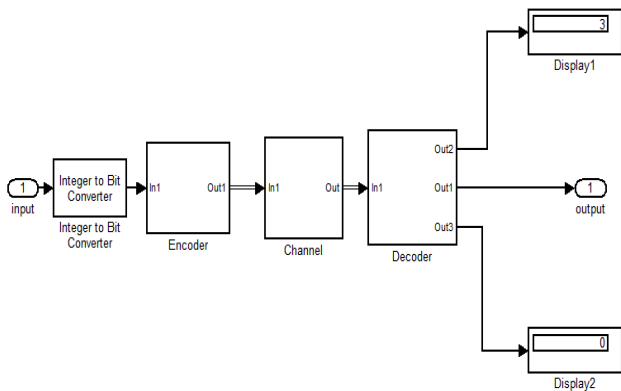


Figure 7. Simulink model of error control coding with HARQ scheme

Fig. 8 shows the input voltage of Speed controller, when the error is 3 symbols with HARQ scheme. The voltage is always 34V. In the simulation, we assume the error pattern is same with CRC code only case. Because the use of RS (15, 9) code, 3 symbol errors are corrected well and it doesn't need retransmission anymore.

Comparing Fig. 6 with Fig. 8, we know that HARQ scheme shows better performance than CRC code only. The error signal 50V is covered to 34V, which means the small error pattern is processed well by HARQ scheme. Furthermore, it does not need retransmission and can save time for the system.

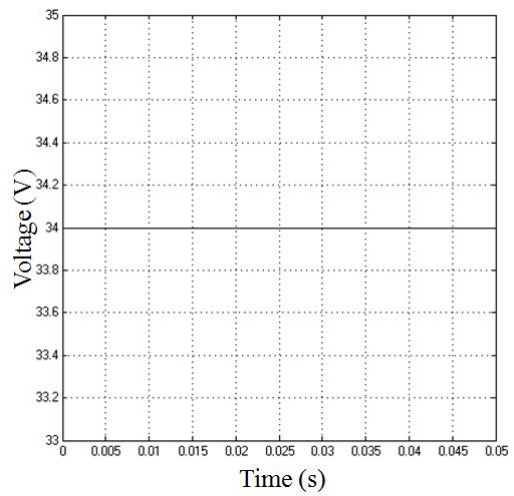


Figure 8. The input voltage of Speed controller (Error is 3 symbols with HARQ scheme)

#### IV. CONCLUSIONS

In this paper, we applied the HARQ scheme based on FlexRay used for SHPP system. In the system modeling part, BLDCM control system with PID controller is done. In the error control coding part, HARQ scheme performs well in the system fault tolerance. With the RS code, some burst error or small bits error can be corrected well. When errors are bigger than the ability which RS code can correct, we will use other methods like retransmission or second channel to cover the errors. Our final goal is to make the real SHPP system based on fault tolerant and use adaptive fuzzy control.

#### ACKNOWLEDGMENT

This research is supported by the Software Fusion Component Development of the Korea Evaluation Institute of Industrial Technology grant funded by the Korea government Ministry of Trade, Industry and Energy. The authors gratefully appreciate their support during this work.

#### REFERENCES

- [1] R. Shaw and B. Jackman, "An Introduction to FlexRay as an Industrial Network," IEEE Inter. Symp. On Industrial Electronics (ISIE 2008), IEEE Press, Jul. 2008, pp. 1849-1854.
- [2] FlexRay Consortium: FlexRay Communication System Protocol Specification Version 3.1, October 2010.
- [3] K. K. Ahn, D. N. C. Nam, D. Q. Truong, and J. I. Yoon, "Position control of Electro-hydraulic actuator (EHA) Using a novel back stepping controller," Journal of The Korean Society for Fluid Power & Construction Equipments, vol. 9, no. 3, Sep. 2012, pp. 16-22.
- [4] Y. A. Chinniah, "Fault Detection In the Electrohydraulic Actuator Using Extended Kalman Filter," PhD thesis, University of Saskatchewan, Saskatoon, Canada, March 2004.
- [5] L. L. Shi, Q. Li, M. B. Park, and H. M. Kim, "A Design of Fault Tolerant Logic on Experiments for Brake-by-Wire

- System,” Eurobake 2012 Conference Proceedings, FISITA, April 2012, ISBN: 978-0-9572076-0-8.
- [6] R. Shanmugasundram, K. M. Zakariah and N. Yadaiah, “Modeling, simulation and analysis of controllers for brushless direct current motor drives,” *Journal of Vibration and Control*, May 2012, pp. 1-15, doi: 10.1177/1077546312445200.
- [7] W. H. Kwun and O. K. Kwun, *Automatic Control Engineering*, First Edition, Fifth Print, CMG Publishing Group, 2009, ISBN: 978-89-7088-423-3.

# Sensor Web Deployment Using Informed Virtual Geographic Environments

Mehdi Mekni  
 University of Minnesota  
 Crookston, Minnesota, United States  
 mmekni@umn.edu

**Abstract**—In this paper, we propose a knowledge-based approach to support Sensor Web (SW) deployments in Informed Virtual Geographic Environments (IVGE) using multi-agent geo-simulation techniques. Sensor webs can be thought of as distributed network systems composed of hundreds of nodes deployed in geographic environments for in-situ sensing and data acquisition purposes. The deployment of sensor webs is by nature a spatial problem since nodes are highly constrained by the geographic characteristics of their environment. Therefore, there is a need for an efficient modelling paradigm to address the issue of SW deployment while taking into consideration the constraints of the geographic space and knowledge it provides to support their autonomous decision making capabilities. The proposed approach builds on top of our previous works on Informed Virtual Geographic Environments. It also relies on well-established theories on spatially reasoning agents and on qualitative reasoning about geo-simulation results. The applicability of our approach is illustrated using a scenario of a sensor web deployment for weather monitoring purposes.

**Keywords**—*Informed Virtual Geographic Environments; Sensor Web; Geographic Deployment; Multi-Agent Geo-Simulation.*

## I. INTRODUCTION

Sensor Webs (SW) are distributed network systems composed of hundreds of such sensor nodes [1] [2]. New capabilities such as micro-sensing and in-situ sensing as well as the wireless connection of these nodes open new possibilities for applications in various domains such as military, environment and disaster relief [3] [4]. The low per-node cost and the shrinking size of microprocessors in addition to the enhancement of their computation capacities, while decreasing their energy consumption, will allow dense distribution of these wireless networks of sensors and actuators [2]. SW can be thought of as a macro-instrument concept that allows for the spatio-temporal understanding of phenomena which take place in geographic environments through the coordinated efforts of a large number of sensing nodes of different types [2]. However, once SW are designed, the deployment of such complex systems is a real challenge because of the complexity and the large-scale of the geographic environment [4]. Sensor web deployment is by nature a spatial problem since nodes are highly constrained by the geographic characteristics of the environment. Even if it is practical to evaluate research

on the real hardware platform, it may not be practical to experiment in an appropriate environment. An example of this are sensor webs which operate on glaciers, remote wildlife habitats, volcanos, and other environments where in-situ sensing techniques are required and with which it is expensive or dangerous to experiment. Therefore, there is a need for an efficient modelling paradigm to address the issue of sensor webs deployment using actors representing sensor nodes evolving in and interacting with a representation of their geographic environment.

In order to address the above mentioned challenges, we propose a knowledge-based multi-agent geo-simulation approach to support the simulation of SW deployments in informed VGEs. A critical step towards the simulation of SW deployment is the creation of appropriate representations of the geographic space and of the sensors evolving in it, in order to efficiently support the sensors' spatio-temporal reasoning capabilities (Figure 1). Moreover, a VGE should provide sensor agents with knowledge about the virtual environment in which they evolve and with which they interact. A number of challenges arise when creating knowledge about the environment, among which we mention: 1) to represent knowledge using a standard formalism; 2) to provide agents with tools and mechanisms to allow them acquire knowledge about the environment; and 3) to infer and to predict based on premises and facts that characterise the geographic environment in order to support spatial agents' decision-making. This approach builds on our previous works on Informed Virtual Geographic Environments [5] [6] [7], on spatially reasoning agents [5] and on qualitative reasoning about geo-simulation results [8].

The rest of the paper is organised as follows. Section II presents related works on agent-based simulation tools for Sensor Web. Section III presents our approach and details its underlying components. Section IV illustrates our framework through a SW deployment scenario for weather monitoring purposes. Section V discusses the results and concludes with our future works.

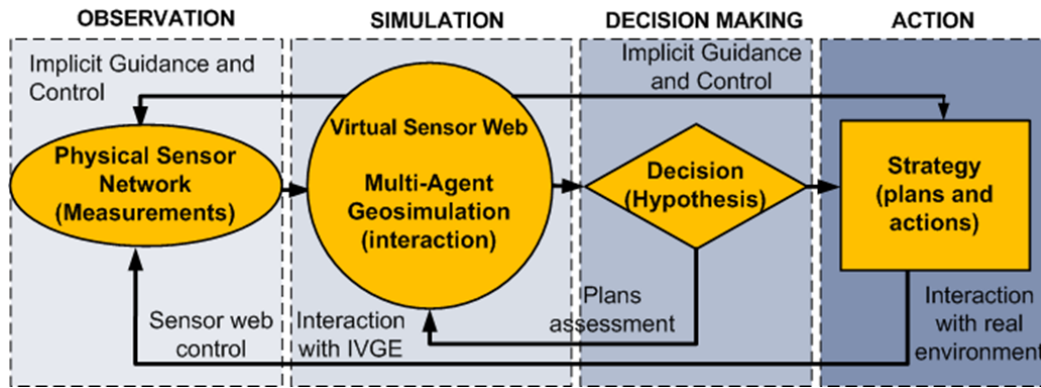


Figure 1. Sensor Web Multi-Agent Geo-Simulation Layered Approach: Observation; Simulation; Decision-Making; Action.

## II. RELATED WORK

According to our literature review, architectures for the management of sensor webs involving the Multi-Agent Geo-simulation paradigm do not exist. However, a few research projects have attempted to integrate the agent paradigm into sensor web architectures such as IrisNet [9], Abacus [10], Biswas and Phoha’s architecture [11], and SWAP [12].

Most of these architectures identify the need for distributed data collection and processing, and propose layered architectures to achieve this. In Abacus different agents in the processing layer detect and report alert conditions to a higher layer interacting with users [10]. IrisNet uses agents such as Sensor Agents (SA) and Sensor Organisers (SO) to collect and analyze data from sensors to answer specific classes of queries [9]. Biswas and Phoha’s approach uses agents in the service layer to analyze data from sensors and transfer it to the application layer [11]. All these approaches deal with data collection by providing a distributed infrastructure for publishing, discovering and accessing sensor resources. They also address the challenge of data fusion, to some extent, and aim to provide end-users with the information they need. These approaches share a common objective through the use of the agent-paradigm which is the distribution of tasks. However, these applications do not take complete advantage of the multi-agent systems approach. Indeed, they use reactive agents which are efficient for alerting purposes, but are neither able to perform situated behaviors nor autonomous decision-making. On the one hand, situated behaviors include performing spatial reasoning and taking advantage of the virtual environment’s description where sensor agents are located. On the other hand, autonomous decision-making includes managing sensor nodes in order to efficiently cover the area of interest while taking into account their limited capabilities as well as local spatial characteristics. We think that, in order to achieve intelligent and autonomous deployment of sensor webs, it is

essential to use a multi-agent geo-simulation approach in which agents are endowed with advanced capabilities such as perception, navigation, memory, and knowledge management. The knowledge management process includes the following tasks: (1) to represent knowledge about geographic environments using standard formalisms; (2) to allow spatial agents to acquire knowledge about the environment; (3) to allow agents to reason and to make decisions while taking into account knowledge about geographic environments.

As the above-mentioned architectures do not address the more challenging sensor web management issues, we propose the knowledge-based multi-agent geo-simulation model for the deployment of sensor webs in informed virtual geographic environments.

## III. A KNOWLEDGE-BASED MAGS APPROACH

The proposed approach (depicted in Figure 2) relies on the Multi-Agent Geo-Simulation paradigm in order to simulate the behaviour of a SW in a dynamic, complex, and large-scale virtual geographic environment. Sensors are

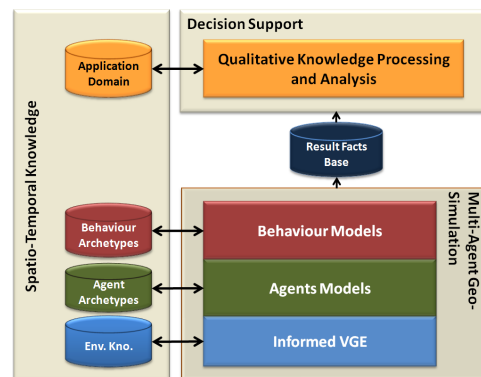


Figure 2. The proposed knowledge-based multi-agent geo-simulation model .

modeled as intelligent agents embedded in a virtual space where dynamic phenomena can occur. Sensor agents have reasoning capabilities allowing them to reason about the virtual space and to react to its dynamic phenomena. Spatio-Temporal knowledge is used for two main purposes. First, it is used during the geo-simulation to support agents reasoning capabilities. Second, it is used to analyze the results of the geo-simulation and to offer decision support to users. Finally, the results of the geo-simulation (which are inserted as facts in the Result Facts Base) are analyzed in order to offer decision support. In the following, we present these components.

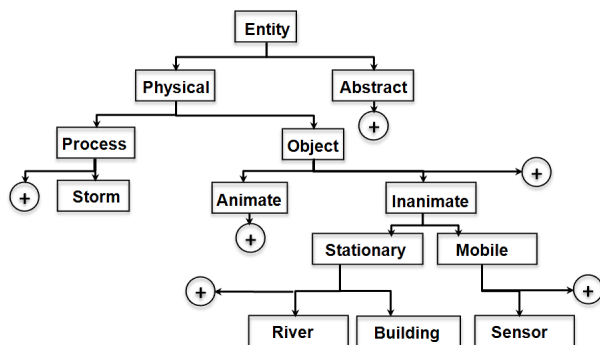


Figure 3. An example of the semantic type hierarchy of agent archetypes.

#### A. Multi-Agent Geo-Simulation

The idea behind the agent-based simulation approach is to move the most intensive processing out of the Physical Sensor Web (PSW) into a parallel Virtual Sensor Web (VSW) operating on a base station or a remote server. The objective is to reproduce, in a realistic manner, the real world in a virtual environment. Indeed, in this virtual environment, which imposes no limits on data processing, energy consumption and communication capabilities, it is possible to create a system for the deployment of the physical sensor web. In order to faithfully mimic the physical sensor web deployed in the area of interest, we need to simulate, in a realistic way, the physical sensor nodes as well as the geographic environment where they are located. Physical sensor web are represented in the virtual environment using software agents. An agent is a program with domain knowledge, goals and actions. An agent can observe and sense its environment as well as affect it. Agents' capabilities may include (quasi-) autonomy, perception, reasoning, assessing, understanding, learning, goal processing, and goal-directed knowledge processing. The reproduction of the geographic environment in which physical sensor nodes are deployed is based on reliable data obtained from Geographic Information Systems (GIS). The concept of Multi-Agent Geo-Simulation (MAGS) evolves from such type of simulations involving

software spatial agents immersed in a virtual geographic environment.

#### B. Spatio-Temporal Knowledge

As we mentioned so far, spatio-temporal knowledge is used in our approach 1) to support agents decision making during the geo-simulation and 2) to analyze the results of the geo-simulation in order to offer decision support. In the following we respectively present the representation formalism and the categories of spatio-temporal knowledge used in the proposed approach.

1) *Representation formalism:* We use Conceptual Graphs (CGS) to represent spatio-temporal knowledge and to support spatio-temporal reasoning. CGs were introduced by Sowa [13] as a system of logic based on Peirce's existential graphs and semantic networks of artificial intelligence. They provide extensible means to capture and represent the semantic of real-world knowledge and have been implemented in a variety of projects for information retrieval, database design, expert systems, qualitative simulations, and natural language processing. However, their application to dynamic geographic spaces modeling and analyzing is an innovative issue. More details about CGs and their theoretical foundations can be found in [13], among others.

Syntactically, a conceptual graph is a network of concept nodes linked by relation nodes. Concept nodes are represented by the notation [Concept Type: Concept Instance] and relation nodes by (Relationship-Name). The formalism can be represented in either graphical or character-based notations. In the graphical notation, concepts are represented by rectangles, relations by circles and the links between concepts and relation nodes by arrows. The character-based notation (or linear form) is more compact than the graphical one and uses square brackets instead of boxes and parentheses instead of circles. Some examples are presented in the following sub-section.

2) *Knowledge Categories:* We distinguish three levels of spatio-temporal knowledge: 1) Knowledge about the environment, 2) Knowledge about actors and their behaviours and 3) knowledge about the application domain (Figure 2)

- **Knowledge about the environment:** We define the notion of knowledge about the environment (Environment Knowledge (EK) for short) as a specification of a conceptualization of the environment characteristics. Hence, EK is a description of the spatial concepts (geographic features) and relationships (topologic, semantic) that may exist in a geographic environment. In multi-agent geo-simulation, EK is a specification used for enabling knowledge exploitation for spatial agents. Practically, EK is an agreement to use spatial concepts (i.e., ask queries and make assertions), spatial



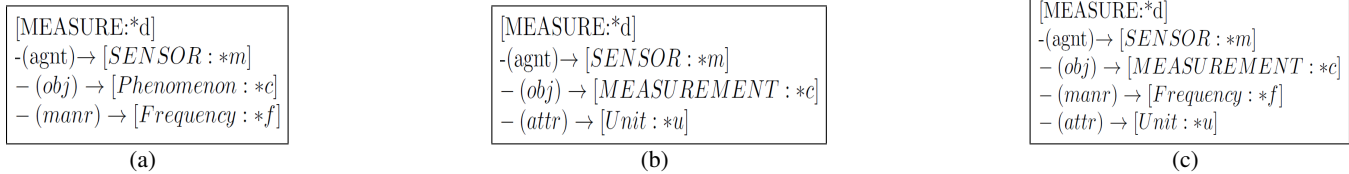


Figure 4. Specification of rules using the Prolog+CG syntax.

relationships (i.e., describe actions and behaviors), in a way that is consistent so we can share knowledge with and among spatial agents. Our aim is to improve the perception-decision-action loop on which relies most of the existing agents' models.

- **Knowledge about Actors and Behaviours Archetypes:**

Dealing with the specification of agents' attributes and associated spatial behaviors may be complex and time and effort consuming. Agents' characterization aims to specify: the agent archetype, its super-types and sub-types according to the semantic type hierarchy; and the behavior archetype that an agent archetype is allowed to perform within the informed VGE. Figure 3 shows an example of a semantic type hierarchy of agent archetypes. Entity is an abstract node and Storm, River, Building and Sensor are instance nodes (leaves) of this agent archetype lattice. A key characteristic of agent archetype is inheritance. Agents belonging to one or several agent archetypes inherit the characteristics associates with these agent archetypes.

For example, let us consider two agent archetypes *Temp-sensor* and *Press-sensor* respectively sensing temperature and pressure. The *Temp-sensor* is characterised by a measurement frequency *f*. On the other hand, *Press-sensor* is characterised by a one meter circular sensing field. Consider now *TP-sensor* a multi-functional sensor which inherits from *Temp-sensor* and *Press-sensor*. Thanks to the inheritance property provided by agent archetypes, this agent performs measurements at a frequency *f* within a circular sensing area of one meter. Since our research addresses the simulation of spatial behaviors, it has been influenced by some basic tenets of active theory [14]. In particular, our approach to manage environment knowledge rests on the commitments in active theory that: (1) activities are directed toward objects, zones, or actors; (2) activities are hierarchically structure; and (3) activities capture some context-dependence of the meaning of information. Theoretically, the common philosophy between our knowledge-based approach and activity theory is a view of the geographic environment from the perspective of an agent interacting with it [18]. Practically, the most important borrowings from activity theory are the

idea that [15]: (1) the semantic of behaviors and objects are inseparable; and (2) behaviors, objects, as well as agents are hierarchically structured. Let us define the following behavior archetypes that we associate with the Sensor agent archetype as follows: (a) "an agent \**m* which is a sensor measures an object \**c* which is a phenomenon with a frequency \**f*" (See Figure 4(a)); (2) "an agent \**m* which is a sensor measures an object \**c* which is a measurement of unit \**u*" (See Figure 4(b)). Since the above description is equal or more specific than the antecedent of the following behaviour, it can be inferred, by deduction, as shown in Figure 4(c).

- **Knowledge about the Application Domain:** The above mentioned levels of knowledge are used during the geo-simulation to support agents in their decision making. In contrast, knowledge about the application domain is mainly used to qualitatively analyze the results of the geo-simulation and is thus more linked to decision support. In the context of SW deployment, nodes are aimed to collect measurements about phenomena of interest which vary according to the application domain (military, environmental surveillance, etc.). Knowledge about the application domain defines phenomena of interest in a particular application domain. In our approach, we use the concept of spatio-temporal situations [19] to model and reason about phenomena of interest. A spatio-temporal situation represents a state, an event or a process situated in space and time and involving various objects of the world. Examples of spatio-temporal situations can be a sensor which brake down for a certain period of time at a certain spatial area (state), the start of rain in certain spatial area (punctual event) or a durative heavy rain in a given area (process). A spatio-temporal situation has a semantic type (rain, network breakdown, etc.), a start and end times and is located in space. Knowledge about the application domain defines spatio-temporal situations of interest according to their temporal (state, punctual event or durative process) and semantic characteristics. For example, the semantic punctual event "Start of rain" can be defined as the fact of water level exceeding a given threshold. Relationships between spatio-temporal situations (temporal and spatial) are also specified in the application domain knowledge, which enables defining

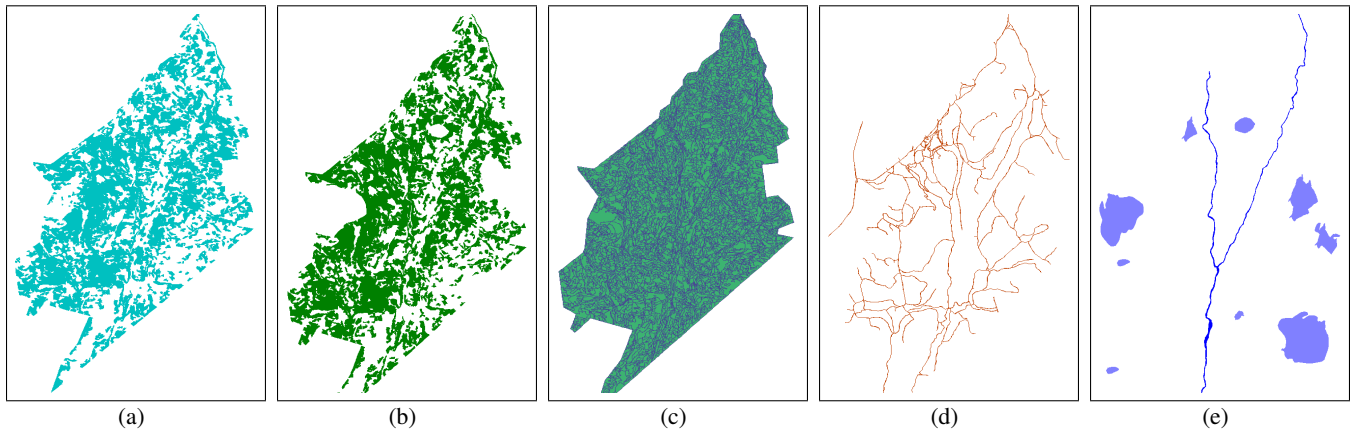


Figure 5. Various geometric and semantic layers related to the Montmorency experimental forest, St.Lawrence Region, Quebec, Canada: (a) and (b) two types of vegetation characterizing the land cover; (c) water resources including rivers and lakes; and (d) road network

complex phenomena. For example, a situation of storm can be defined as a situation of heavy rain followed by / accompanied with a situation of strong wind.

C. Decision support

The decision support component analyzes the result of the geo-simulation using application domain knowledge in order to identify situations of interest to the user. This data analysis process is implemented using the approach proposed in [19]. Details of this approach are beyond the scope of this paper. We only illustrate the principle using the simple example showed in Figure 6. In this example, the situation of interest

there is a flood situation if the water-level exceeds 0,15 meter. Otherwise, there is no flood situation. The decision support component uses this knowledge in order to analyze the facts collected by agents during the simulation (Result Facts Base). Particularly, the two facts illustrated in Fig. 3 are respectively interpreted as the start of flood in (Area A, t= 14:35) and the end of flood in (Area A, 22:12) (punctual events). Finally, start and end flood situations are used to identify the flood situation itself as a process located in Area A during the time interval [14:35, 22:12]. Obviously, detection of real complex situations requires taking into consideration other aspects (measurement errors, conflict of measurements between several sensors, etc.) that are beyond the scope of this paper.

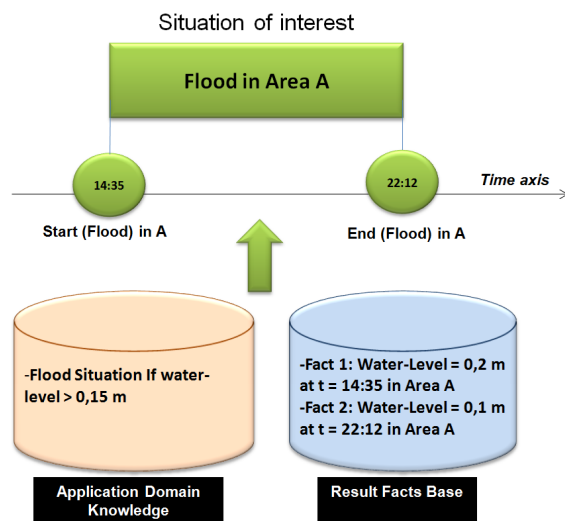


Figure 6. An example of a simulation result analysis and processing

is Flood. The application domain knowledge specifies that

IV. EXPERIMENTAL RESULTS

In order to illustrate our knowledge-based multi-agent geo-simulation framework, we propose to simulate a sensor web deployed in an IVGE representing the experimental forest of Montmorency (Quebec, Canada) for weather monitoring purposes (Fig. 4 and Fig. 5). This scenario shows how agents adapt their spatial behaviors with respect to knowledge they acquire from the IVGE using our environment knowledge management along with their perception capabilities. The objective of the simulated sensor web is to identify and monitor a simulated storm evolving in the IVGE. In order to reason about knowledge, we used a platform to deal with CGs manipulation called Amine [20]. Amine platform provides a pattern-matching and rule based programming paradigm embedded in Prolog+CG language which is basically an object-oriented and conceptual graphs-based extension of Prolog language.

A few agent archetypes representing different kinds of

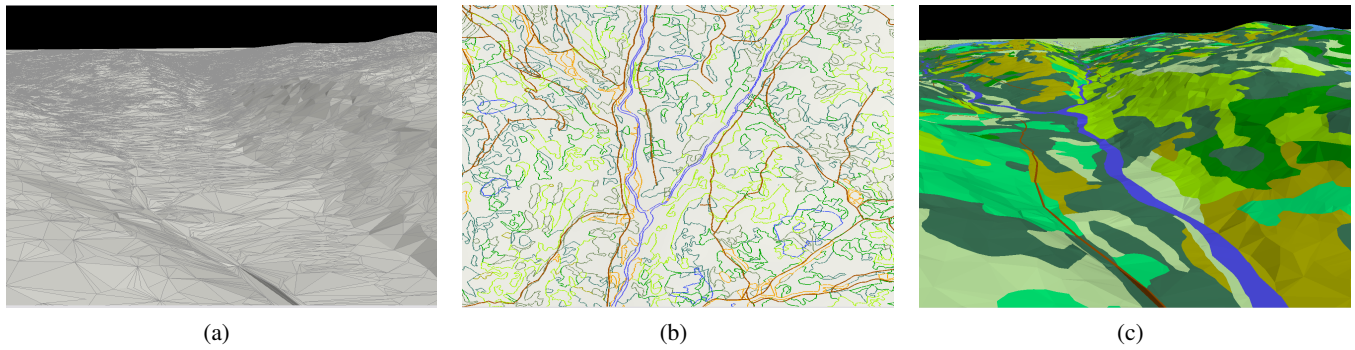


Figure 7. IVGE representing the Montmorency experimental forest; (a) the 2.5D triangulated elevation map; (b) the semantic maps; and (c) the 3D unified map

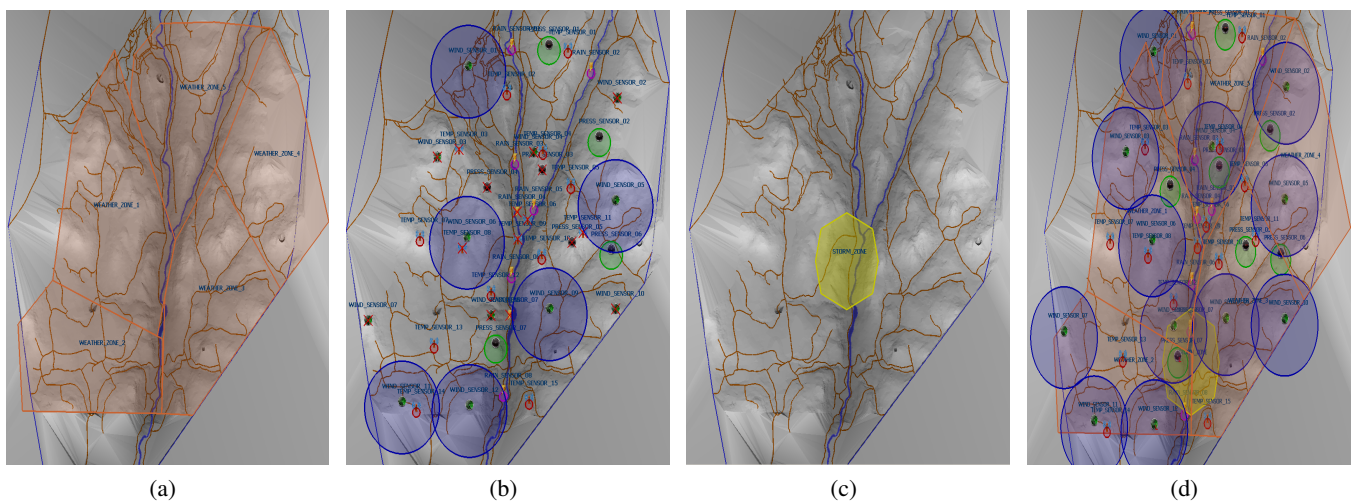


Figure 8. Simulation results: (a) WeatherZone agents covering the monitored geographic area; (b) the simulated sensors autonomously reaching their deployment position in the IVGE. The blue circle corresponds to the perception field for active sensors; and (c)

sensors are involved in this scenario. These sensors are first randomly deployed in the IVGE. Then, each sensor computes a path in order to reach its deployment position while taking into account the geographic environment characteristics. When sensors reach their final destinations, some of them stay active while other switch to idle in order to preserve the overall energy of the sensor web. Active sensors make measurements at a frequency  $f$  in order to monitor weather conditions. Active and idle sensors as well as the measurement frequency are specified in the simulation scenario created by the MAGS's user.

The simulated storm appears after a time frame  $t$  from the beginning of the simulation. If an active sensor perceives the storm agent, it directly accesses to its properties and extracts the information that it monitors depending on its kind of sensor, i.e. temperature, pressure, wind speed and direction, or humidity. If a difference above a certain threshold  $\Delta$  is

observed, the sensor proceeds as follows: (1) it accelerates its measurements frequency, (2) it adds a new fact that keeps track of the event with its timestamp in the Result Facts Base; and (3) it sends a message to weak up idle all sensors of the same kind which are situated in a certain estimated distance. As the simulation time goes by and the storm agent evolves in the IVGE, most of initially idle sensors become active to sense the observed phenomenon. When the storm agent is out of the perception field of the sensor, this latter senses a new difference between the past and the current measurement. It notifies the Result Facts Base by adding a new fact that keeps track of the new event with its timestamp; Idle and active sensors switch states in order to preserve their energy. In order to model the simulation described above, let us first consider the two agent archetypes ZONE and SENSOR. In contrast with the ZONE agent archetype, which represents a geographic area, the SENSOR agent archetype is associated with individual sensors

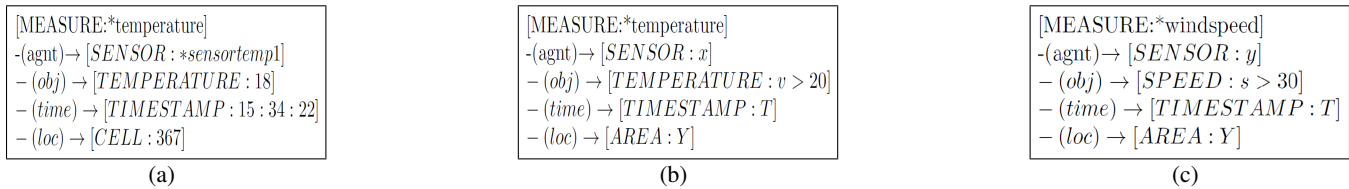


Figure 9. Behaviour archetypes expressed using the linear notation of Conceptual Graphs (CGs).

deployed in the informed virtual geographic environment. Let us consider the two agent sub-types *WEATHERZONE* and *STORMZONE*. Agents of type *WEATHERZONE* are stationary and represent meteorological conditions within the geographic area they cover. 5 instances of *WEATHERZONE* are created in order to approximately cover the monitored area (Fig. 6). In addition to their geometric characteristics, these agents encompass attributes which characterize the meteorological conditions such as temperature = 18 °C, pressure = 1010 hPa, humidity = 30. When a sensor detects a difference above the threshold, it adds a fact in the Results Facts Base using the CGs formalism. Consider the following example involving the *sensortemp1* adding a fact describing an observed difference of temperature measurement of value 18 at cell 367 at 15h : 34 : 22 (See Figure 9(a)).

Let us now define the following sub-types of *SENSOR* archetype: *TEMPSENSOR* for temperature measurement, *PRESSENSOR* for atmospheric pressure measurement, *WINDSENSOR* for wind speed and orientation measurement, and *HUMISENSOR* for humidity measurement (Figure 8). In this scenario, the situation of interest is the storm. Let us suppose that we need to describe the evolution of the storm during the geo-simulation for decision support purposes. Knowledge about the application domain allows specifying how the presence of a storm phenomenon can be detected in a certain area.

For simplification, let us consider the following (Prolog+CG) rule specifying that a storm is detected at time T in area A if there are (in the results facts base) facts describing that temperature exceeds 20 °C and wind speed exceeds 30km/h at the same time T and the same area A (See Figure 9(b) and Figure 9(c)). A storm detection event is triggered as soon as the conjunction of these two facts exists within the result facts base. The analysis of the simulation results conceptualize this event using the knowledge associated with the situation description shown in Figure 9.

## V. CONCLUSION AND PERSPECTIVES

Our environment knowledge management approach is original at various aspects. First, a multi-agent geo-simulation model which integrates an informed virtual geographic environment populated with spatial agents capable

of acquiring and reasoning about environment knowledge does not exist. Second, a formal representation of knowledge about the environment using CGs which leverages a semantically-enriched description of the virtual geographic environment has not yet been proposed. Third, providing agent with the capability to reason about a contextualized description of their virtual environment and phenomenon occurring within it during the simulation is also an innovation that characterizes our approach. We are currently working on the automated assessment of different simulation scenarios for sensor web deployment using our qualitative knowledge processing and analysis module. Indeed, users usually need to analyze and compare various scenarios in order to make informed decisions. This task may be complex and effort and time consuming. However, since our framework already supports the analysis of the multi-agent simulation results, it is easy to extend it to automatically assess scenarios.

## ACKNOWLEDGEMENT

This research was supported in part by the Grant in Aid provided by the University of Minnesota. The author would like to thank the reviewers for their valuable comments.

## REFERENCES

- [1] C.-H. Park, J. Cho, and D.-H. Kim, "Sensor web for supporting mobility in sensor networks," in *Computer Applications for Web, Human Computer Interaction, Signal and Image Processing, and Pattern Recognition*, ser. Communications in Computer and Information Science, T.-h. Kim, S. Mohammed, C. Ramos, J. Abawajy, B.-H. Kang, and D. Slezak, Eds. Springer Berlin Heidelberg, 2012, vol. 342, pp. 203–208.
- [2] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *IEEE Communication Magazine*, vol. 40, no. 8, pp. 102–114, 2002.
- [3] M. Bogdanovic, N. Veljkovic, and L. Stoimenov, "Spatial sensor web for the prediction of electric power supply system behaviour," in *Bridging the Geographic Information Sciences*, ser. Lecture Notes in Geoinformation and Cartography, J. Gensel, D. Josselin, and D. Vandenbroucke, Eds. Springer Berlin Heidelberg, 2012, pp. 81–98.
- [4] M. Mekni and P. A. Graniero, "A multiagent geosimulation approach for intelligent sensor web management," *IJDSN*, vol. 2010, pp. 10–26, 2010.

- [5] M. Mekni, N. Jabeur, and B. Moulin, "A generic model for situated holonic multi-agent systems: A case study on sensor web management," in *MDAI 2008: Proceedings of 5th International Conference on Modelling Decisions for Artificial Intelligence*, Sabadell, Spain, October 2008, pp. 35–46.
- [6] M. Mekni, "Abstraction of informed virtual geographic environments," *Geo-spatial Information Science*, vol. 15, no. 1, pp. 27–36, 2012.
- [7] M. Mekni and B. Moulin, "A multi-agent geosimulation approach for sensor web management," in *Proceedings of the 2008 Second International Conference on Sensor Technologies and Applications*, ser. SENSORCOMM '08. Washington, DC, USA: IEEE Computer Society, 2008, pp. 129–134.
- [8] H. Haddad and B. Moulin, "An agent-based geosimulation multidisciplinary approach to support scenarios evaluation in dynamic virtual geographic environments," in *Proceedings of the 2008 Spring simulation multiconference*, ser. SpringSim '08. San Diego, CA, USA: Society for Computer Simulation International, 2008, pp. 53–60.
- [9] P. Gibbons, B. Karp, Y. Ke, S. Nath, and S. Srinivasan, "Irisnet: an architecture for a worldwide sensor web," *Pervasive Computing, IEEE*, vol. 2, no. 4, pp. 22–33, 2003, 1536-1268.
- [10] I. Athanasiadis, M. Milis, P. Mitkas, and S. Michaelides, "Abacus: A multi-agent system for meteorological radar data management and decision support," in *ISESS'05: Proceedings of the Sixth International Symposium on Environmental Software Systems*, E. Jakeman & D.A. Swayne, Ed., Sesimbra, Portugal, 2005.
- [11] P. Biswas and S. Phoha, "A middleware-driven architecture for information dissemination in distributed sensor networks," in *Sensor Networks and Information Processing Conference*, 2004, pp. 605–610.
- [12] D. Moodley and I. Simonis, "New architecture for the sensor web: the swap framework," in *ISWC 2006: 5th International Semantic Web Conference*, Athens, GA, USA, 2006, p. 17.
- [13] J. Sowa, *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Course Technology, August 1999.
- [14] V. Kaptelinin, B. A. Nardi, and C. Macaulay, "Methods & tools: The activity checklist: a tool for representing the space of context," *interactions*, vol. 6, no. 4, pp. 27–39, Jul. 1999.
- [15] B. Nardi, *Context and Consciousness: Activity Theory and Human-Computer Interaction*. Mit Press.

## Block Algorithm and Its Implementation for Cholesky Factorization

Jianping Chen, Zhe Jin, Quan Shi, Jianlin Qiu, Weifu Liu

School of Computer Science and Technology

Nantong University

Nantong, P. R. China

e-mail: chen.jp@ntu.edu.cn, jin.z@ntu.edu.cn, shi.q@ntu.edu.cn, qiu.jl@ntu.edu.cn, liu.wf@ntu.edu.cn

**Abstract**—Block algorithm divides a large matrix into small blocks of submatrices to make full use of computer's caches. Blocked smaller submatrices can be directly loaded into the caches to compute. The efficiency of the computation is hence improved. The block algorithm for Cholesky factorization is studied in this paper. Detailed derivation and implementation of the block algorithm are presented. The computations of modules involved in the block algorithm are discussed. A C++ language program is developed. It is tested and compared with the standard (unblocked) version of the Cholesky factorization. The testing results show that the presented block algorithm outperforms the unblocked one in the cases of large matrices, with an increase of execution speed of 20%.

**Keywords**—Numerical computation; Cholesky factorization; matrix blocking; cache use.

### I. INTRODUCTION

Numerical computations of linear algebra problems including the Cholesky factorization play very important roles in many scientific researches and engineering applications. Today's computers are provided with two or more levels of caches between a high-speed CPU and a relatively low-speed memory. The caches operate at a speed close to that of the CPU. If the data computed currently or used frequently is placed in the caches, the access time of the data can be reduced significantly. The overall efficiency of the computation is hence increased greatly. However, the capacity of the cache memory is much smaller. If the amount of data being computed is very big, such as the case of the Cholesky factorization for a large matrix, it is impossible to load all the data into the cache. The cache can not be well utilized and the computation is less efficient. To solve the problem, the method of matrix blocking can be used [1, 2]. A large matrix is partitioned into blocks of submatrices of small sizes. The computations of these smaller submatrices are more likely to be carried out within caches. Li [3] has discussed the block techniques used in LAPACK (Linear Algebra PACKage, a software library for performing numerical linear algebra computations) and given the testing results on different machines. Andersen et al. [4] presented a recursive block algorithm for Cholesky factorization. It uses the technique of recursive blocking, where a big matrix is blocked recursively (each time split by half) until the final submatrices are small enough. With this

recursive blocking method, the blocking process can be made automatically while the block sizes in each level are different. At outer levels, the block sizes are still quite big, and at inner levels, the block sizes can be very small. Ng and Peyton [5] investigated the block algorithm of Cholesky factorization for the sparse matrix, where many of the matrix elements are zeros and special measures are taken to deal with them such as supernodes. The block algorithms mentioned above including the ones in the LAPACK were implemented by using the BLAS (Basic Linear Algebra Subroutine) routines [6], which is an open software library to perform basic linear algebra operations such as vector and matrix multiplications. On the other hand, these algorithms were developed and tested on the machines of more than 10 years ago, at which time the computers had small cache memories of 64K or 128K. Today's computers, including microcomputers, have much greater cache memories of more than 1M with very big main memories that reach the order of gigabytes. It is valuable to examine the effects of the block algorithms on these machines. The block algorithm of Cholesky factorization for a general matrix (not the sparse case) is investigated in this paper. The fixed blocking method is used, that is, a big matrix is divided into small blocks in a fixed size linearly part by part. Instead of using the BLAS subroutines, we do the implementation ourselves by C++ programming and discuss the implementation techniques involved in the block algorithm.

The remainder of the paper is organized as follows. Section 2 derives the block algorithm of Cholesky factorization. Section 3 discusses the implementation of the block algorithm and the computations of the modules involved. Presented in Section 4 are testing results and analysis. Finally, conclusions and considerations for future work appear in Section 5.

### II. BLOCK ALGORITHM FOR CHOLESKY FACTORIZATION

Cholesky factorization is used to solve linear systems of equations in the case that the coefficient matrix  $A$  is symmetric and positive definite:

$$AX = B. \quad (1)$$

$A$  can be factorized into the product of an upper-triangular matrix  $U$  with a lower-triangular matrix  $U^T$  ( $U^T$  is

the transpose of  $U$ ):

$$A = U^T U \quad (2)$$

with

$$U = \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ & u_{21} & \cdots & u_{2n} \\ & & \ddots & \vdots \\ & & & u_{nn} \end{pmatrix} \quad (3)$$

According to the standard algorithm of the Cholesky factorization [7], the matrix  $U$  is computed as follows.

$$u_{ii} = (a_{ii} - \sum_{k=1}^{i-1} u_{ki}^2)^{1/2} \quad i=1, 2, \dots, n \quad (4)$$

$$u_{ij} = (a_{ij} - \sum_{k=1}^{i-1} u_{ki} u_{kj}) / u_{ii} \quad j=i+1, i+2, \dots, n. \quad (5)$$

Now, we derive the block algorithm for Cholesky factorization. Since the matrix  $A$  is symmetric, only half of the data is needed. Here, the upper triangular part is used. We write (2) in a form of blocked submatrices:

$$\begin{pmatrix} A_{11} & A_{12} \\ & A_{22} \end{pmatrix} = \begin{pmatrix} U_{11}^T & \\ & U_{22}^T \end{pmatrix} \begin{pmatrix} U_{11} & U_{12} \\ & U_{22} \end{pmatrix}. \quad (6)$$

Assume that the order of  $A$  is  $n \times n$ . If we set the block size at  $d$ , then the order of  $A_{11}$  and  $U_{11}$  are  $d \times d$ ,  $A_{12}$  and  $U_{12}$  are  $d \times (n-d)$ , and  $A_{22}$  and  $U_{22}$  are  $(n-d) \times (n-d)$ . Among these blocked submatrices,  $A_{11}$ ,  $U_{11}$ ,  $A_{22}$  and  $U_{22}$  are upper triangular, and  $A_{12}$  and  $U_{12}$  are rectangular. Computing (6) in the blocked form, we obtain

$$A_{11} = U_{11}^T U_{11} \quad (7)$$

$$A_{12} = U_{11}^T U_{12} \quad (8)$$

$$A_{22} = U_{12}^T U_{12} + U_{22}^T U_{22}. \quad (9)$$

Eq. (7) is the Cholesky factorization of the submatrix  $A_{11}$ . It can be computed using the standard Cholesky factorization formulae in (4) and (5). If we properly choose the value of the block size  $d$ , i.e., the order of  $A_{11}$ , the computation is likely to be carried out within the cache. After  $U_{11}$  is solved,  $U_{12}$  is computed by (8):

$$U_{12} = (U_{11}^T)^{-1} A_{12}. \quad (10)$$

Let

$$A'_{22} = A_{22} - U_{12}^T U_{12}. \quad (11)$$

Eq. (9) becomes

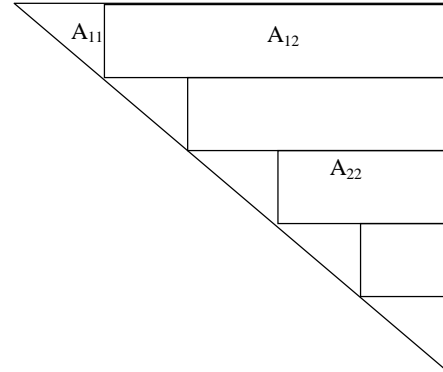


Figure 1. Blocking process of the block algorithm for Cholesky factorization

$$A'_{22} = U_{22}^T U_{22}. \quad (12)$$

Eq. (11) denotes a modification or update of  $A_{22}$  through  $U_{12}$ . The updated  $A_{22}$  is expressed as  $A'_{22}$ . Eq. (12) is the Cholesky factorization of  $A'_{22}$ . Usually,  $A'_{22}$  is still a big matrix. The same blocking process is then applied to  $A'_{22}$ . This process is repeated until the order of the final  $A'_{22}$  is less than or equal to  $d$ . At this time, the Cholesky factorization of the final  $A'_{22}$  is computed using the standard formulae of Cholesky factorization in (4) and (5). Fig. 1 illustrates the blocking process.

In summary, the block algorithm for Cholesky factorization can be described as follows.

Step 1: Block matrix  $A$  into submatrices  $A_{11}$ ,  $A_{12}$ ,  $A_{22}$  with a block size of  $d$ .

Step 2: Cholesky-factorize  $A_{11}$ :  $A_{11} = U_{11}^T U_{11}$ , yielding  $U_{11}$ .

Step 3: Compute the inverse of  $U_{11}^T$ , yielding  $(U_{11}^T)^{-1}$ .

Step 4: Compute  $U_{12}$ :  $U_{12} = (U_{11}^T)^{-1} A_{12}$ .

Step 5: Update  $A_{22}$ :  $A'_{22} = A_{22} - U_{12}^T U_{12}$ .

Step 6: If the order of  $A'_{22}$  is less than or equal to  $d$ ,

Cholesky-factorize  $A'_{22}$ :  $A'_{22} = U_{22}^T U_{22}$ , yielding  $U_{22}$ .

Otherwise, take  $A'_{22}$  as  $A$  and return to Step 1.

### III. IMPLEMENTATION OF BLOCK ALGORITHM

According to the description in above section, the computation of the block algorithm for Cholesky factorization consists of five modules, which include the Cholesky factorization of a small submatrix ( $A_{11}$ ), the computation of the inverse of a lower-triangular submatrix ( $(U_{11}^T)^{-1}$ ), the multiplication of a lower-triangular submatrix by a rectangular submatrix ( $(U_{11}^T)^{-1} A_{12}$ ), the multiplication of two rectangular submatrices ( $U_{12}^T U_{12}$ ), and the control of the repeated blocking process.

### A. Cholesky Factorization of Submatrix

The standard formulae in (4) and (5) are used to compute the Cholesky factorization of the small submatrix  $A_{11}$ . Since the matrix  $A_{11}$  is upper-triangular, the packed storage method can be applied to save the memory space. If a two dimension array is used to store the matrix, the number of the elements in the  $i$ th row of the array is  $n-i$  ( $i=0, 1, 2, \dots, n-1$ ). The outcome matrix ( $U_{11}$ ) is stored in the same way.

As the elements of the matrix are stored in the packed form, the subscript of each element in the two dimension array differs from its corresponding numbers of the row and the column in the original matrix. The new data structure is shown in Fig. 2.

Since the data structure is changed, the formulae in (4) and (5) need to be modified accordingly. The modified formulae are as follows.

$$u_{i1} = (a_{i1} - \sum_{k=1}^{i-1} u_{k(i-k)}^2)^{1/2} \quad i=1, 2, \dots, n \quad (13)$$

$$u_{ij} = (a_{ij} - \sum_{k=1}^{i-1} u_{k(i-k)}u_{k(j+i-k)}) / u_{i1} \quad j=1, 2, \dots, n-j \quad (14)$$

### B. Inverse Computation of Lower-Triangular Matrix

It is easy to prove that the inverse of a lower-triangular matrix is still a lower-triangular matrix and the values of its diagonal elements are the reciprocals of the values of the diagonal elements in the original matrix. So, the diagonal elements can be computed separately.

For convenience of the description, we use  $L$  to represent a lower-triangular matrix and  $B$  to represent its inverse matrix. Assume  $L=(l_{ij})_{d \times d}$  and  $K=(k_{ij})_{d \times d}$ . The diagonal elements of  $K$  can be easily computed as

$$k_{ii} = \frac{1}{l_{ii}} \quad i=1, 2, \dots, d. \quad (15)$$

The other elements can be computed by making use of the characteristics of the triangular matrix instead of using the normal method of Gaussian-Jordan elimination that is less efficient in this case. According to the theorem of linear algebra, the product of the multiplication of a matrix and its inverse matrix is a unit matrix, i.e., a matrix with its elements are zeros except the ones at the diagonal. So, we have

$$\sum_{k=j}^i (l_{ik} \times k_{kj}) = 0 \quad j < i. \quad (16)$$

Thus, the other elements of  $K$  are computed as follows.

$$k_{ij} = -\sum_{k=j}^{i-1} (l_{ik} \times k_{kj}) \quad i=1, 2, \dots, d-1; \quad j=1, 2, \dots, i-1. \quad (17)$$

a <sub>11</sub>	a <sub>12</sub>	í	a <sub>1 n-1</sub>	a <sub>1 n</sub>
	a <sub>21</sub>	a <sub>22</sub>	í	a <sub>2 n-1</sub>
		í	í	í
			a <sub>n-1 1</sub>	a <sub>n-1 2</sub>
				a <sub>n 1</sub>

Figure 2. The data structure corresponding to the packed storage

### C. Multiplication of Lower-Triangular Matrix by Rectangular Matrix

The order of the lower-triangular matrix ( $(U_{11}^T)^{-1}$ ) is  $d \times d$  and the order of the rectangular matrix ( $A_{12}$ ) is  $d \times (n-d)$ . The outcome matrix is also a rectangular matrix of  $d \times (n-d)$ . The data of the rectangular matrix  $A_{12}$  comes from the initial input matrix  $A$ . The computing result ( $U_{12}$ ) can be directly put into the output matrix of the block algorithm. For both of the two matrices, we need to determine their starting positions in the arrays.

If it is a normal  $n \times n$  two dimension array, we just need to add an offset  $d$  to the original starting position of each row. Assuming that the original matrix is  $R$  and the matrix actually used is  $P$ , the relationship is

$$P[i][j] = R[i][j + d]. \quad (18)$$

However, the matrix is now stored in the packed form. The offset to be added to the starting position of each row need be modified to  $d-i$ , where  $i$  is the number of the row. Therefore, the actual relationship between the two matrices is

$$P[i][j] = R[i][j + d - i]. \quad (19)$$

The computation method is simply the sequential multiplication of the elements of the corresponding row and column of the two matrices and the summation of the products.

### D. Multiplication of Two Rectangular Matrices

The two rectangular matrices multiplied are  $U_{12}^T$  and  $U_{12}$ , where the order of  $U_{12}^T$  is  $(n-d) \times d$  and  $U_{12}$  is  $d \times (n-d)$ . If  $n$  is very large,  $n-d$  is still very large. The two matrices can not be entirely loaded into the cache. Let  $S$  represent  $U_{12}^T$  and  $Y$  represent  $U_{12}$ . When doing the multiplication, we need to repeatedly multiply the elements of one row of  $S$  with the elements of a column in  $Y$ . In C++ language, the elements of an array are stored by rows. The fetch of elements in  $Y$  would be across the rows. Each part of data in  $Y$  would be



loaded in and out constantly between the cache and memory. This will slow the computation. To avoid the problem, we apply transposing to the matrix  $Y$  before computation. Then the computation becomes the multiplication of one row of  $S$  with a row in  $Y$  successively.

As we know, the matrices  $S$  and  $Y$  are transposes each other, i.e.,  $S = Y^T$ . After being transposed, the matrix  $Y$  becomes exactly the matrix  $S$ . Therefore, just one matrix  $S$  is needed in the computation. In addition, the result of the multiplication of two mutually transposed matrices is a symmetric matrix. Only half of the elements need to be computed. The operations of multiplication can be reduced by half.

#### E. Control of Repeated Blocking Process

The function of this part is to control the computing procedure of the block algorithm. It calls each of the modules in the algorithm to perform the corresponding computation. The preliminary work for the computation of each module, such as the matrix transposing and the matrix positioning, is also completed in this part. Recursion is used to implement the repeated blocking processing of the input matrix.

### IV. TESTING RESULTS AND ANALYSIS

Based on the discussion of the previous sections, a C++ language program is developed for the block algorithm of Cholesky factorization. The program is run and tested on a microcomputer. Its CPU is Core2 of 2.0GHz with a 2MB L2 cache and the main memory is 2 GB. The compiler used is Microsoft Visual C++ 6.0. Different block sizes and matrix orders are tried to test the execution time of the algorithm. A block size about 50 to 100 turns out to be the best choice for most cases. To make a comparison, the standard algorithm of Cholesky factorization is also programmed and run on the same computer. Table 1 gives the execution times of the block algorithm and the standard algorithm of Cholesky factorization for different matrix orders from 500 to 5000. The data in the fourth row of the table is the speed increase of the block algorithm over the standard algorithm in percentage.

TABLE 1. EXECUTION TIMES OF THE BLOCK ALGORITHM COMPARED THE STANDARD ALGORITHM (IN SECONDS)

Order of matrix	Standard Cholesky	Block Cholesky	Speedup in %
500	0.26	0.35	-34.6
1000	2.39	2.77	-15.9
2000	24.37	24.05	1.3
3000	88.42	82.16	7.1
4000	226.61	194.38	14.2
5000	474.4	382.13	19.4

As the data in the table shows, the block algorithm is superior to the standard algorithm in the case of a large matrix. With the increase of the matrix size, the advantage becomes more and more remarkable. The increase of the execution speed is about 20% when the matrix order reaches 5000.

The arithmetic complexity (total number of float multiplications and divisions) of the direct computation of Cholesky factorization using the equation (4) and (5) is about  $n^3/6$  for large values of  $n$  where  $n$  is the order of the matrix [7]. The block algorithm blocks the matrix for  $n/d$  times where  $d$  is the block size. At each time, the arithmetic operations contain the direct computation of Cholesky factorization of a matrix ( $d \times d$ ), the inverse computation of a lower-triangular matrix ( $d \times d$ ), the multiplication of a lower-triangular matrix ( $d \times d$ ) with a rectangular matrix ( $d \times (n-d)$ ), and the multiplication of two rectangular matrices ( $(n-d) \times d$  and  $d \times (n-d)$ ), where  $d$  is fixed and  $n$  is reduced by  $d$  each time. It can be counted that the total number of the float multiplications and divisions is at the order of  $n^3/6 + (n/d+1)n^2/2$ . For a large value of  $n$ , the arithmetic complexities of the two algorithms are parallel.

The above testing results and complexity analysis indicate that the block algorithm makes a better use of the cache to improve the computation efficiency for large matrices. For matrices of small sizes, the block algorithm does not perform better than the standard algorithm. This is because the block algorithm is more complicated than the standard one. It has more control operations, which increase the overhead. In the case of a small matrix, the overhead takes a considerable portion of the execution time, making the blocking algorithm less effective.

On the other hand, comparing with the results of the earlier researches [3,4,5] that show that the block algorithms manifest the advantages starting from a matrix order of several hundreds, our results show that the block algorithm manifests the advantage starting from a matrix order of more than a thousand. This implies that with the increase of the cache memories of today's computers, larger matrices can be more easily computed without blocking.

### V. CONCLUSION AND FUTURE WORK

The presented block algorithm of Cholesky factorization makes full use of the caches of today's computers to improve the computing efficiency. A considerable increase of execution speed is achieved in the case of large sizes of matrices. It indicates that the matrix blocking is an effective technique for the computation of dense linear algebra. Further work is considered. As mentioned previously, the computations of the modules in the block algorithm contain many loops. The technique of loop unrolling can be used to reduce the execution time spent on the loop control. The method of the recursive blocking will be tried and investigated to make a close comparison of the two blocking techniques. We shall also make more testing to the algorithms on different machines.

## ACKNOWLEDGMENT

Financial support from the National Natural Science Foundation (No. 61171132), the Jiangsu Provincial Natural Science Foundation (No. BK2010280) and the Nantong Municipal Application Research Program (No. BK2011026) of P. R. China are acknowledged. The valuable comments and suggestions of the reviewers are also appreciated.

## REFERENCES

- [1] S. Toledo, "Locality of reference in LU decomposition with partial pivoting," *SIAM Journal of Matrix Analysis & Application*, vol. 18, April 1997, pp. 1065-1081.
- [2] J. Dongarra, *Numerical Linear Algebra for High Performance Computers*. Philadelphia: SIAM, 1998.
- [3] Y. Li, "Block algorithms and their effect in LAPACK," *Journal of Numerical Methods and Computer Applications*, vol. 22, March 2001, pp. 172-180.
- [4] B. S. Andersen, J. Wasniewski, and F. G. Gustavson, "A recursive formulation of Cholesky factorization of a matrix in packed storage," *ACM Trans. Math. Softw.*, vol. 2, June 2001, pp. 214-244.
- [5] E. G. Ng and B. W. Peyton, "Block sparse Cholesky algorithms on advanced uniprocessor computers," *SIAM Journal on Scientific Computing*, vol. 14, May 1993, pp. 1034-1056.
- [6] Y. Li and P. Zhu, "Methods and techniques for speeding up BLAS," *Journal of Numerical Methods and Computer Applications*, vol. 3, Sep. 1998, pp. 227-240.
- [7] R. L. Burden and J. D. Faires. *Numerical Analysis (Seventh Edition)*. Beijing: Higher Education Press, 2001.
- [8] E. Elmroth, F. Gustavson, and I. Jonsson, "Recursive blocked algorithms and hybrid data structures," *SIAM Review*, vol. 46, Jan. 2004, pp. 33-45.
- [9] X. Wang and B. Feng, "Improved algorithm of matrix multiplication based on memory," *Journal of Northwest Normal University*, vol. 41, Jan. 2005, pp. 22-24.
- [10] S. Xu, *Program Collection for Commonly Used Algorithms (in C++ Language)*. Beijing: Tsinghua University Press, 2009.
- [11] K. Ji, J. Chen, Z. Shi, and W. Liu, "Study and implementation of block algorithm for matrix triangular factorization," *Computer Application and Software*, vol. 27, Sep. 2010, pp. 72-74.

# Multiagent Genetic Optimisation to Solve the Project Scheduling Problem

Konstantin Aksyonov, Anna Antonova

Department of Information Technologies

Ural Federal University, UrFU

Yekaterinburg, Russia

wiper99@mail.ru, antonovaannas@gmail.com

**Abstract**—This paper considers a project scheduling problem belonging to a class of multiobjective problems of complex systems control, whose decision search time grows exponentially depending on the problem dimension. In this paper, a survey of a modified genetic algorithm application to the project scheduling problem is presented. We propose a multiagent genetic optimisation method based on evolutionary and multiagent modelling by implementing different decision searching strategies, including a simulation module. The multiagent simulation module is intended to evaluate chromosome fitness functions and describe the dynamic nature of own and subcontracted resources allocation. The proposed multiagent genetic optimisation method, the MS Project resource reallocation method, and a heuristic simulation method have been compared whilst addressing a real-world scheduling problem. The comparison has shown: firstly, the unsuitability of the MS Project planning method to solve the formulated problem; and secondly, both the advantage of the multiagent genetic optimisation method in terms of economic effect and disadvantage in terms of performance. Some techniques to reduce the impact of the method's disadvantage are proposed in the conclusion, as well as the aims of future work.

*Keywords*-project scheduling; genetic algorithms; simulation; subcontract work optimisation.

## I. INTRODUCTION

The scheduling problem is one of the key problems in the management of organisational and technical systems. Inefficient scheduling can lead to financial losses, quality of service losses, and loss of competitiveness for the company. Companies from various scopes are faced with the scheduling problem, e.g. industrial and project companies, shopping centres, hospitals, and call centres.

There are several scheduling problem statements depending on the application sphere: operations calendar planning [1]–[5], limited resources assignment to a set of tasks [6]–[8], and the traveling salesman problem [9].

Classical scheduling problem solving methods have a number of disadvantages. Thus the use of combinatorial methods and mathematical programming leads to high computational resources utilisation when addressing large-scale problems. In addition, these methods are applied poorly to the problem with dynamically changing constraints. Simulation takes into account the dynamic nature of the problem, but leads to a random search process, which does

not guaranteed optimal decision finding. The use of genetic optimisation allows the shortcomings of the previous methods to be overcome [9]. The application of genetic optimisation to the scheduling problem with defined constraints is widely considered in the literature [1]–[9].

This paper focuses on the project scheduling problem with the use of evolutionary computation [12] and simulation. The remainder of the paper is organised as follows: Section 2 provides an overview of the application of the genetic algorithm to the scheduling problem. Section 3 formulates the project scheduling problem with time constraint. Section 4 introduces the genetic algorithm based on an annealing simulation and novelty search. Section 5 presents the algorithm of the multiagent genetic optimisation program based on the integration of evolutionary computation and multiagent simulation. Section 6 evaluates the practical implementation of the multiagent genetic optimisation program to solve the real-world scheduling problem. Section 7 concludes this paper and explores future work.

## II. RELATED WORK

In general, the scheduling problem is connected to the problem of seeking an operations sequence that satisfies the constraints and optimises the objective functions. The renewable resources (such as staff or equipment) are usually considered when studying the scheduling problem. For certain tasks (for example production planning) non-renewable resources should additionally be determined [1].

In the various scheduling problem studies different constraint sets are considered depending on the specific task. Four constraint types were identified in [3]: resource, precedence, physical layout, and information constraints. The time constraint type should be added to the constraint types list when analysing workflow inside a project development company. Time limitation is associated with a time frame for the early and late starts of the operations.

All constraints, except precedence ones, were studied by Brezuliani et al. [6]. Precedence and resource constraints were considered by Okada et al. [1], Klimek [2], Abdel-Khalek et al. [4], and Dhingra and Chandna [7]. Resource and information constraints were studied by Yang and Wu [8]. Resource, precedence, and time constraints were considered by Karova et al. [5]. Scheduling with a resource constraint to determine a public transport route was presented by Osaba et al. [9].

The optimisation objects are also different in the studies reviewed. The classical objective function of working time (makespan) minimisation was considered by Sriprasert and Dawood [3] and Osaba et al. [9]. The objective function of constraints violation penalty minimisation was considered by Karova et al. [5], and Yang and Wu [8]. Both mentioned objective functions were considered by Okada et al. [1], Brezuliani et al. [6], and Dhingra and Chandna [7].

There are different ways of conducting an objective function evaluation: analytical methods, simulation, artificial neural networks, fuzzy systems, and component modelling. Analytical methods are the most widely used; the drawback of this approach being the lack of the analysis of the complex system dynamic behaviour. This drawback is overcome by using a simulation model to evaluate the objective function by Osaba et al. [9]. The integration of evolutionary modelling and simulation can limit the random search space and enhance heuristic optimisation by taking into account the dynamically changing constraints of the scheduling problem.

The reviewed studies do not consider subcontracted workforce optimisation, while this problem is real to developers and even to mass production enterprises. The subcontracted workforce's optimisation problem is connected to subcontracted scheduling in order to maximise the utilisation of the company's own resources. In the literature, a problem regarding the appropriate selection of subcontractors using artificial intelligence methods was studied by Chen et al. [10]. A subcontract optimisation technique based on a simulation and heuristics was suggested by Aksyonov and Antonova [11]. The current article considers new subcontract optimisation techniques with the use of a genetic algorithm.

### III. PROBLEM STATEMENT

Let us consider the problem of unique projects scheduling aimed at the calendar planning of operations. All project operations have to be carried out in combination with a set of time constraints. The set of time constraints is defined through negotiations with customers. In the case of the organisation's own lack of resources, subcontracted resources have to be involved to meet the time constraints.

The objective functions of the considered problem are: 1) subcontract cost minimisation; and 2) minimisation of own resources total downtime. The second objective function is associated with the fixed labour cost in the project companies. If the salary is fixed then downtime is also paid, which is not profitable for the company.

For the project scheduling problem considered in this study, the following assumptions have been made:

1. A single project consists of a number of operations with a known processing time, early and late start dates, labour input, and labour cost.
2. The operation requires the availability of renewable resources (own or subcontracted workforces).
3. Non-renewable resources are not considered in the scheduling.
4. Operations cannot be interrupted.
5. Subcontractors can be involved to perform part of the operation.

6. Subcontractors can be interrupted and the operation can continue with the use of the company's own resources in the event of the appearance of own available resources.
7. Subcontractors are available every day on request in unlimited quantities (for example when working with different subcontractors).

Let us describe the problem of the project portfolio scheduling with the use of the following designations.

**Indices:**

- $i$ : project index,  $i=1, 2, \dots, P$ .
- $j$ : operation index,  $j=1, 2, \dots, Op_i$ .
- $w$ : department index,  $w=1, 2, \dots, V$ .
- $t$ : time index,  $t=0, 1, 2, \dots, T$ .

**Decision variables:**

$TB(i,j)$ : set of start dates of operations.

**Initial parameters:**

$ES(i,j)$ : early start date of the operation  $(i,j)$ .

$LS(i,j)$ : last start date of the operation  $(i,j)$ .

$SL_w$ : amount of persons in the department  $w$ .

$SLO(i,j,w)$ : amount of workforces (persons) needed in department  $w$  to fulfil the operation  $(i,j)$ .

$SS(i,j)$ : operation  $(i,j)$  subcontract cost per day.

**Parameters obtained in the decision-making process:**

$Active(i,j,t)$ : a sign of the operation  $(i,j)$  execution at the time  $t$ .

$$Active(i, j, t) = \begin{cases} 1, & \text{if operation } (i, j) \text{ is executed at the moment } t \\ 0, & \text{otherwise} \end{cases}$$

$RD(t,w)$ : resources from department  $w$  demand to fulfil the active operations at the time  $t$ .

$$RD(t, w) = \sum_{i=1}^P \sum_{j=1}^{Op_i} [Active(i, j, t) \cdot SLO(i, j, w)]$$

$VF(t,w)$ : amount of free workforces of department  $w$  at the time  $t$ .

$$VF(t, w) = \begin{cases} SL_w - RD(t, w), & \text{if } RD(t, w) \leq SL_w \\ 0, & \text{otherwise} \end{cases}$$

$V_{SC}(i,j)$ : volume of subcontracting workforces on operation  $(i,j)$ .

**Problem description:**

$$OF_1 = \sum_{i=1}^P \sum_{j=1}^{Op_i} (SS(i, j) \cdot V_{SC}(i, j)) \rightarrow \min, \quad (1)$$

$$OF_2 = \frac{\sum_{t=0}^T \sum_{w=1}^V VF(t, w)}{T \cdot V} \rightarrow \min, \quad (2)$$

$$TB(i, j) \in [ES(i, j); LS(i, j)] \quad \forall i, \forall j \quad (3)$$

Objective function (1) minimises the total subcontracting cost. Objective function (2) minimises the own resources total downtime. Constraint (3) saves the time frame of the operations start.

#### IV. GENETIC ALGORITHM BASED ON ANNEALING SIMULATION AND NOVELTY SEARCH

The genetic algorithms (GA) is one of the evolutionary approaches that can be used to solve complex system management problems in a short time [12]. The technique of the GA application includes the following steps: 1) selecting the method of encoding the problem decision (phenotype) into a chromosome (genotype); 2) definition of the evaluation method of the chromosome fitness function (FF); 3) the genetic operator's description; and 4) the initial population generation and GA work. The modification of the GA on the basis of an annealing simulation and novelty search is considered in the article in order to enhance the quality of the scheduling problem's decisions.

##### A. Chromosome Encoding

There are various techniques for decision encoding presented in the literature: operations sequence encoding [2][5][9], operations precedence encoding [1][3], operations start dates encoding [4][6], the resource assignments on the operation encoding [6]–[8]. We use the encoding of the operation start dates shifting because this technique supports time constraints, is not redundant, and is simple to implement.

The GA chromosome encodes the operations' start dates shifting from the initial work plan to the right or left on the time axis via binary code (0/1). The shifting range is two weeks on either side of the initial operation start date. The chromosome size is  $5 \cdot r$  genes, where  $r$  – number of analysed operations, 5 – number of the genes needed to encode a single operation shifting (4 genes to encode  $2^4=16$  shifting days and 1 gene to encode the shifting direction).

##### B. Genetic Algorithm Modification

A novelty concept is a major GA concept. This concept is connected with the emergence of new elements and interactions in the environment during evolution. Two novelty types are distinguished in [13]: 1) *combinatorial novelty* when the new species emerge by combining the existing species; and 2) *creative novelty* when the new species are not reproducible by a combination of the species. The validity of the fundamental feasibility of the second novelty type is still open.

Let us consider the case of a combinatorial novelty search as an adaptation mean in an open system. To implement this approach we modify a simple GA by introducing the concept of "decision originality" as a measure of the decision fitness to the environmental conditions [13]. The decision-chromosome's originality in the population is determined via the numerical transformation of the Hamming distance matrix.

Let us define the Hamming distance matrix as follows:

$$H = (h_{ij})_{i=1, j=1}^N, \quad (4)$$

where  $h_{ij}$  – Hamming distance between the  $i$ -th and  $j$ -th chromosomes ( $Ch_i$  and  $Ch_j$ ), equal to the number of positions

at which the corresponding gene values are different in chromosomes  $Ch_i$  and  $Ch_j$ ;  $N$  – number of chromosomes.

We associate the matrix  $H$  with the matrix of originality weights  $W$  defined as follows:

$$W = (w_{ij})_{i=1, j=1}^N, \quad (5)$$

where  $w_{ij}$  – weight of the corresponding value of the Hamming distance determined as a quadratic function, increasing in the range from  $l$  to  $R$  as element  $h_{ij}$  is changed in the range  $0$  to  $L$ :

$$w_{ij} = \frac{R-1}{\sqrt{L}} \cdot \sqrt{h_{ij}} + 1, \quad (6)$$

where  $L$  – the chromosome size;  $R$  – the maximum weight of the chromosome in the pair,  $R > 0$ .

The two strategies of chromosome crossing have been described using the concept of originality. The first strategy – *the originality search strategy (OSS)* [14] – focuses on the combinatorial search for the new decisions in the population by crossing chromosomes that have different encoding. The second strategy – *the maximum search strategy (MSS)* [12] – focuses on the targeted search for the best chromosomes by crossing chromosomes that are the most adapted to the environment. The fitness of the  $i$ -th chromosome to the environment is evaluated by the fitness function  $FF_i$ ,  $i=1..N$ .

Let us define the chromosome crossing probability matrices on the basis of the proposed strategies as follows:

$$P_{OSS} = (p_{ij}^{OSS})_{i=1, j=1}^N, \quad p_{ij}^{OSS} = \frac{w_{ij}}{\sum_{j=1}^N w_{ij}}, \quad (7)$$

$$P_{MSS} = (p_i^{MSS})_{i=1}^N, \quad p_i^{MSS} = \frac{FF_i}{\sum_{i=1}^N FF_i}. \quad (8)$$

In formulas (7) and (8) the matrices cells are filled by probabilities values in accordance with the roulette law [12]. In the case of OSS strategy, the weight of the chromosome originality serves as a measure of chromosome importance. In the case of MSS strategy, the chromosome FF serves as a measure of chromosome importance.

An annealing simulation algorithm (ASA) [15] is intended to implement the proposed chromosome crossing strategies during the GA work. This algorithm is based on the analogy of the metal annealing process, which results in the appearance of new metal properties. The technique for ASA and GA integration is proposed below.

Step 1. Set the annealing simulation algorithm parameters: the initial value of the parameter  $t_z$ ; the value of the parameter  $\alpha$ , that controls the rate of annealing temperature decrease,  $0 \leq \alpha \leq 1$ .

Step 2. Set the GA parameters: the number of generations  $K$ ; the chromosome size  $L$ ; the likelihood of the genetic operators being applied. Set the number of the current population  $Z$ :  $Z=1$ . Generate the initial population.

Step 3. Apply the genetic operators to the current population  $Z$  with a probability depending on the value of parameter  $t_z$ . Increase the number of the current population  $Z=Z+1$ . Change the value of parameter  $t_z$  [15]:

$$t_{z+1} = t_z + \alpha \cdot t_z. \quad (9)$$

Step 4. Check the condition of the GA ending:  $Z>K$ . If the condition is satisfied then go to Step 5, otherwise return to Step 3.

Step 5. Stop.

The probability of the genetic operator's application is defined on the basis of the annealing simulation, in order to reflect the operator's dynamic nature.

### C. Crossover Operator

The probability of selecting the first and second parents from the current population  $Z$  for the crossover operator (CO) is described below. The probability of selecting the first parent has to take into account both random selection and targeted selection based on the MSS strategy (8). The probability of random selection should be reduced in the population's evolution, and the probability of the MSS strategy should be increased. This fact is reflected in the probability formula of selecting the first parent  $i$  in the population  $Z$ :

$$P_i^Z(CO) = \frac{1}{N} \cdot \left( 1 - \exp\left(-\frac{1}{t_z}\right) \right) + p_i^{MSS} \cdot \exp\left(-\frac{1}{t_z}\right). \quad (10)$$

The probability of selecting the second parent has to take into account the OSS and MSS strategies. The probability of the OSS applying (7) should be reduced in the population's evolution, and the probability of the MSS applying (8) should be increased. This circumstance is reflected in the probability formula of selecting the second parent  $j$  for the first parent  $i$  in population  $Z$ :

$$P_j^Z(CO) = p_{ij}^{OSS} \cdot \left( 1 - \exp\left(-\frac{1}{t_z}\right) \right) + p_j^{MSS} \cdot \exp\left(-\frac{1}{t_z}\right). \quad (11)$$

### D. Mutation and Inversion Operators

The applied probability of the mutation operator (MO) in population  $Z$  is described below. This formula has to take into account the probability reducing during evolution in order to save genetic material [14]:

$$P_z(MO) = P_0(MO) \cdot \left( 1 - \exp\left(-\frac{1}{t_z}\right) \right), \quad (12)$$

where  $P_0(MO)$  – the initial value of the mutation operator applied probability.

The applied probability of the inversion operator in population  $Z$  is described by an analogy with the mutation operator applied probability.

### E. Fitness Function

The following fitness function considers both objective functions (1) and (2) described in Section 3:

$$FF = \omega_1 \cdot \left( \frac{OF_1^{init}}{OF_1} \right) + \omega_2 \cdot \left( \frac{OF_2^{init}}{OF_2} \right) \rightarrow \max, \quad (13)$$

where  $\omega_1, \omega_2$  – weight coefficients,  $\omega_1 + \omega_2 = 1$ ;  $OF_1^{init}, OF_2^{init}$  – objective functions initial values obtained by expert evaluation of the operation start date.

Used FF is described with the use of the linear convolution of normalised heterogeneous criteria (1) and (2).

The multi-agent resource conversion processes (MRCP) model described in [16] is used to evaluate the chromosome FF. The MRCP model is proposed to perform the decision-making process. The decision variables and input parameters described in Section 3 are fed at the model input. The parameters obtained in the decision-making process are the model output. In the MRCP model we use agents in order to implement the resource allocation algorithm and use simulation in order to perform the operations execution. The resource allocation algorithm is described in [13] and allows executors of operations to be appointed in accordance with the assumptions made in Section 3.

## V. MULTIAGENT GENETIC OPTIMISATION PROGRAM

The multiagent genetic optimisation (MGO) program has been developed on the basis of a BPsim.MAS dynamic situations modelling system and BPsim.MSN technical and economic development system [16]. The MGO program is intended to solve the problem of simulation and evolution modelling integration. BPsim.MAS supports the MRCP model description via graphical notation of the resource conversion processes. BPsim.MSN [16] ensures the development of the decision search information technology (IT) based on the UML sequence diagrams [17] and Transact-SQL database management language [18].

A genetic optimisation IT has been designed on the basis of BPsim.MSN. The genetic optimisation IT is intended to aid GA parameters setting and GA execution. The MRCP model is intended to conduct chromosomes FF evaluation by carrying out an experiment with the model. The decoded chromosome phenotype (operations calendar planning) is fed to the model input. The FF evaluation in accordance with (13) is obtained in the model output. Agents in MRCP model are used to allocate the renewable resources (both own and subcontracted).

The decision maker carries out the problem statement and solution choice among the solutions obtained by the use of the MGO program. The algorithm of interaction between the decision maker and MGO program during the decision-making process is shown in Figure 1.

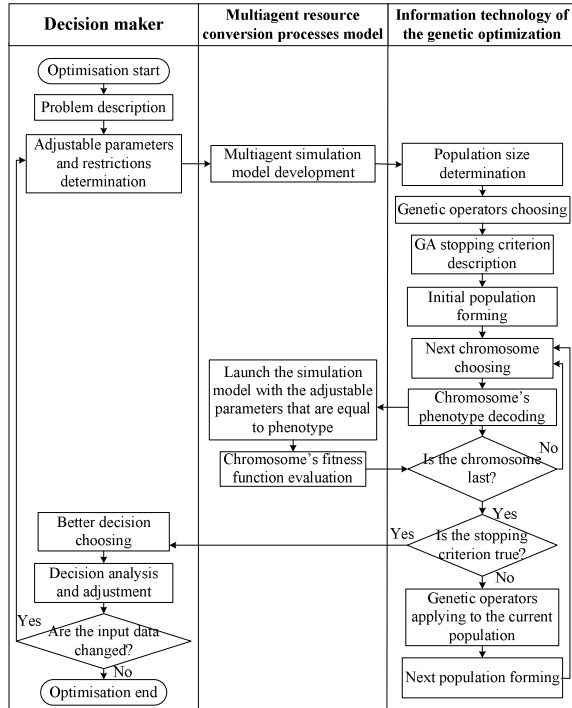


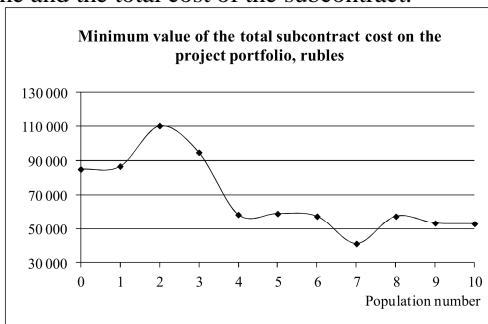
Figure 1. Interaction between decision maker and MGO program

The MGO program designed has a number of advantages compared to existing evolutionary scheduling optimisation software [2][4][8][9].

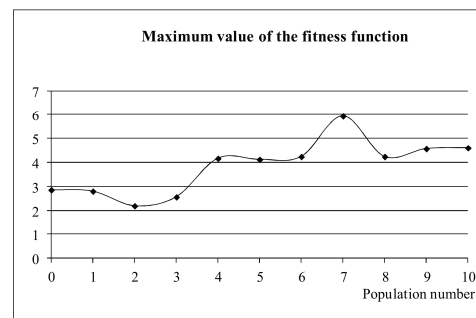
1. The integration of simulation, expert, multiagent, conceptual, and evolutionary approaches.
2. Description of the system models using MRCP and UML graphical notation.
3. The evolutionary and simulation models integration via wizard technology.

## VI. EXPERIMENTAL RESULTS

The application of the MGO program to solve the project scheduling problem is presented in this section. Let us consider a company «Telesystems» that consists of the project, manufacturing, and supply departments. The goal is the minimisation of the company department's total downtime and the total cost of the subcontract.



a.



b.

Figure 2. Dependencies of the fitness function and objective function values from the population number

A detailed statement of the problem is given in [11]. The MRCP model has been developed in order to evaluate the chromosome FF (13). The MRCP model implements the resources allocation model, which satisfies the assumptions determined in Section 3. The model adequacy has been proven in [11] through the evaluation of 5 projects. The following input information have been used in the model: 1) number of projects – 10 with 35 operations; 2) time interval  $T=430$  days (1 year and 3 months); and 3) time limit early and late start of the operations is determined by the shift in the provisional operations start dates for 2 weeks to the right or left along the time axis.

The following GA parameters have been determined in the course of the genetic optimisation IT work: 1) the population size – 10 chromosomes; 2) the chromosome size – 175 genes (5 genes to encode the 35 project operations); 3) the following genetic operators – reproduction based on roulette, five-point crossover with probabilities determined by (10) and (11), five-point mutation with an initial probability equal to 10% and dynamic probability determined by (12), inversion with initial probability equal to 5%; 4) algorithm stopping criterion – a change of 10 populations; 5) random initial population; and 6) following the ASA parameters values –  $t_{z0}=1$ ,  $\alpha=0.9$ ,  $K=10$ .

The dependencies of the chromosome FF and scheduling problem objective functions values from the population number have been obtained as a result of genetic optimisation using the developed MGO program. The change in the minimum value of the objective function (1) during genetic optimisation is shown in Figure 2.a. The change in the maximum value of the fitness function (13) during genetic optimisation is shown in Figure 2.b. The best decision has been achieved in the seventh generation.

The project scheduling problem for the «Telesystems» company also has been solved by use of the MS Project 2007 resources reallocation method and heuristic-simulation (HS) method described in [11].

MS Project 2007 provides the opportunity for resource reallocation (with smoothing) in order to avoid exceeding the own renewable resources availability. The percentage utilisation of the manufacturing department for the initial work plan for the «Telesystems» company is shown in Figure 3 via means of MS Project visualisation.

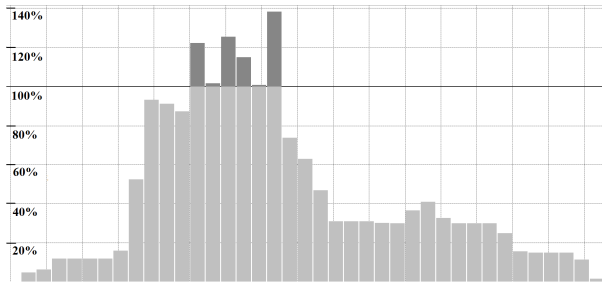
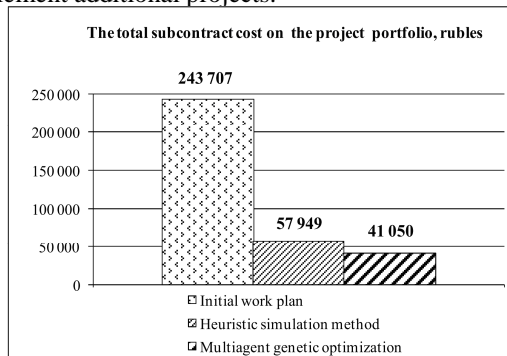


Figure 3. Percentage utilisation of the manufacturing department for initial work plan in MS Project

The initial work plan has been formed by decision making person. In the figure, the x-axis shows the time intervals (each of which lasts 12 days); the y-axis shows percentage utilisation. The resources availability overallocated (time intervals where the use of subcontract is necessary) is shown in figure as dark stripes above the horizontal line at the 100% utilisation level. The application of the MS Project resources reallocation method has allowed to reduce the total subcontract cost down to zero (that is the objective functions (1) and (2) have reached their optimal values). But the time constraints (3) have not been satisfied with the use of this method. In this way, the MS Project resources reallocation method is not suitable to considered scheduling problem.

The HS method is based on the analysis of the MRCP model output parameters. In the HS method, the following steps are performed [11]: 1) modelling results analysis – the subcontract cost and company resources utilisation; 2) search for bottlenecks associated with the operations that require high costs of subcontract; 3) shifting the start dates of operations on the period determined by HS information technology; 4) transfer the adjusted model at the experiments stage and experiments results evaluation.

Histograms of the objective functions (1) and (2) obtained by the MGO and HS methods are shown in Figures 4.a and 4.b compared with the initial work plan. The total subcontract cost and total downtime of the manufacturing department has been consistently reduced with the use of HS and MGO methods. All time constraints have been satisfied. The total downtime indicates on a reserve of own resources to implement additional projects.



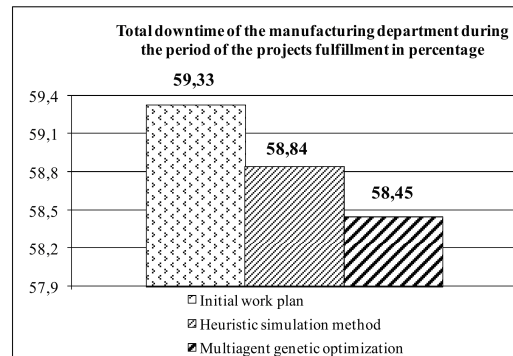
a.

Based on the analysis of the results it was concluded that the MGO method is more effective than the HS method in addressing the project scheduling problem in terms of economic effect. The total subcontract cost of the project portfolio has been reduced by 30% and the total downtime of the manufacturing department has been reduced by 1.5% for a six month period using the MGO method compared to the HS method. The total subcontract cost has been reduced by 6 times using the MGO method compared with the initial work plan. Applying the genetic optimisation based on the simulation and evolutionary modelling integration enhances the efficiency of the decision established by taking into account the dynamic resources allocation model in the simulation model and the fulfilment of the direct search in the decisions space by GA. The economic effect of applying the MGO program to solve the scheduling problem for the «Telesystems» company will be 405000 rubles per year, which is 9% higher than the economic effect from the use of the HS method to solve the same problem.

Let us compare the HS and MGO methods in terms of performance by measuring CPU time. The CPU time for the HS method  $T_{HSM}$  consists of the sum of the HS IT runtime  $T_{HSIT}$  and the model MRCP runtime  $T_{MRCP}$ . The sum is multiplied by the number of experiments  $X_{Iterations}$  conducted during the HS technology work. Thanks to the fact that  $T_{HSIT} \ll T_{MRCP}$  we can neglect the term  $T_{HSIT}$  and define  $T_{HSM}$  time as follows:  $T_{HSM} = X_{Iterations} \cdot T_{MRCP}$ .

The CPU time for the MGO method  $T_{MGO}$  consists of the sum of the genetic optimisation IT runtime  $T_{GOIT}$  and model MRCP runtime  $T_{MRCP}$ , which is multiplied by the chromosome number  $N$ . The sum is multiplied by the generations number  $K$ . Thanks to the fact that  $T_{GOIT} \ll T_{MRCP}$  we can neglect the term  $T_{GOIT}$  and define  $T_{MGO}$  time as follows:  $T_{MGO} = K \cdot N \cdot T_{MRCP}$ .

For the real-world scheduling problem the following parameter values have been used:  $X_{Iterations} = 3, K = 10, N = 10$ . In this case the HS method is more desirable in terms of performance and consumes 33 times less CPU time than the MGO method. This fact is connected to the use of the simulation model in GA for fitness function evaluation, which performed  $K \cdot N$  times. The CPU time of the MGO method is equal to 30 minutes for the real-world scheduling problem.



b.

Figure 4. Dependencies of the objective functions values from the decision seeking method



## VII. CONCLUSION AND FUTURE WORK

In this paper a multiagent genetic optimisation method used to solve the project scheduling problem has been described on the basis of the annealing simulation algorithm, novelty search algorithm, genetic algorithm, and multiagent simulation. The method combines three different decision seeking strategies: a random search strategy, originality search strategy, and maximum search strategy, in order to reflect the dynamic nature of the genetic operators applied. The proposed integration of evolutionary modelling and simulation limits the search space and adequately evaluates the dynamic fitness functions of the chromosomes. The method described has been implemented in a MGO program built on the basis of the BPsim.MAS multiagent modelling system and BPsim.MSN development system. The program integrates simulation, expert, multiagent, conceptual, and evolutionary modelling. The MGO method application to a real-world project scheduling problem has been compared with MS Project and HS methods. The MS project resource reallocation method has been found unsuitable for the scheduling problem considered because of the lack of constraints consideration. As a result of the comparison between MGO and HS methods, an improvement in decision quality under the constraints considered has been achieved using the MGO method.

The disadvantage of the MGO method is the high CPU time, which is 33 times higher than that of the HS method. This fact imposes constraints on the GA generation size (no more than 10 chromosomes) and GA iteration number (no more than 10 generations). Different ways to enhance the GA convergence applied should be considered in future work in order to meet the described constraints.

The aim of future research is to improve the rate of the proposed genetic algorithm convergence by applying elitism and taboo algorithms. The dependency between the decision search time and problem dimensions is assumed to be established. A comparison of the MGO method with the branch and bound method adapted to the problem considered is planned. Also consideration of non-renewable resources allocation in the simulation model is planned.

### ACKNOWLEDGEMENT

This work was performed under contract № 02.G25.31.0055 (project 2012-218-03-167).

### REFERENCES

- [1] I. Okada, X. F. Zhang, H. Y. Yang, and S. Fujimura, "A random key-based genetic algorithm approach for resource-constrained project scheduling problem with multiple modes," *Proc. Int. MultiConf. Engineers and Computer Scientists*, vol. 1, Hong Kong, 17–19 March, 2010.
- [2] M. Klimek, "A genetic algorithm for the project scheduling with the resource constraints," *Annals UMCS Informatica*, vol. 10, issue 1, pp. 117–130, 2010.
- [3] E. Sriprasert and N. Dawood, "Genetic algorithms for multi-constrained scheduling: an application for the construction industry," *Proc. CIB W78's 20th International Conf. Construction IT, Construction IT Bridging the Distance, CIB Report 284, Waiheke Island, New Zealand, 23–25 April 2003*, pp. 341–353.
- [4] H. Abdel-Khalek, M. H. Sherif, A. M. el-Lacany, and Y. Abdel-Magd, "Financing – scheduling optimization for construction projects by using genetic algorithms," *Proc. World Academy of Science, Engineering and Technology*, 2011, pp. 289–297.
- [5] M. Karova, J. Petkova, and V. Smarkov, "A genetic algorithm for project planning problem," *Proc. Int. Scientific Conf. Computer Science*, 2008, pp. 647–651.
- [6] A. Brezuliani, L. Fira, and M. Fira, "A genetic algorithm approach for scheduling of resources in well-services companies," *International Journal of Advanced Research in Artificial Intelligence*, vol. 1, no. 5, 2012.
- [7] A. Dhingra and P. Chandna, "A bi-criteria M-machine SDST flow shop scheduling using modified heuristic genetic algorithm," *International Journal of Engineering, Science and Technology*, vol. 2, no. 5, pp. 216–225, 2010.
- [8] F.-C. Yang and W.-T. Wu, "A genetic algorithm based method for creating impartial work schedules for nurses," *International Journal of Electronic Business Management*, vol. 10, no. 3, pp. 182–193, 2012.
- [9] E. Osaba, R. Carballedo, and F. Diaz, "Simulation tool based on a memetic algorithm to solve a real instance of a dynamic TSP," *Proc. IASTED Int. Conf. Applied Simulation and Modelling, Napoli, Italy, 25–27 June, 2012*, pp. 27–33.
- [10] M.-Y. Chen, H.-C. Tsai, and E. Sudjono, "Evaluating subcontractor performance using evolutionary fuzzy hybrid neural network," *International Journal of Project Management*, pp. 349–356, 2001.
- [11] K. A. Aksyonov and A. S. Antonova, "Application of simulation and intelligent agents to solve project management problem," *International Journal of Computer Science Engineering and Information Technology Research*, vol. 3, no. 1, pp. 321–330, 2013. Available from: [http://www.tjprc.org/view\\_archives.php?year=2013&id=14&jtype=2&page=3](http://www.tjprc.org/view_archives.php?year=2013&id=14&jtype=2&page=3). [Retrieved: June 2013]
- [12] D. Goldberg, *Genetic Algorithms*, Addison Wesley, 1989.
- [13] J. Lehman and K. Stanley, "Exploiting open-endedness to solve problems through the search for novelty," *Proc. Eleventh Intern. Conf. Artificial Life (ALIFE XI)*, Cambridge, MA, 2008.
- [14] L. A. Zinchenko, V. M. Kureichik, and V. G. Redko (Ed.), *Bionic Information Systems and their Practical Applications*, Moscow, FIZMATLIT, 2011.
- [15] W. L. Goffe, G. D. Ferrier, and J. Rogers, "Global optimization of statistical functions with simulated annealing," *Journal of Econometrics* 60, 1994, pp. 65–99.
- [16] K. Aksyonov et al., "Decision Support Systems Application to Business Processes at Enterprises in Russia," in *Efficient Decision Support Systems - Practice and Challenges in Multidisciplinary Domains*, C. Jao, Ed., InTech, 2011, pp. 83–108. ISBN: 978-953-307-441-2. Available from: <http://www.intechopen.com/articles/show/title/decision-support-systems-application-to-business-processes-at-enterprises-in-russia>. [Retrieved: June 2013]
- [17] The official UML web site, Available from: <http://www.uml.org>. [Retrieved: June 2013]
- [18] SQL Server language reference, Available from: [http://msdn.microsoft.com/en-us/library/ms166026\(v=sql.90\).aspx](http://msdn.microsoft.com/en-us/library/ms166026(v=sql.90).aspx). [Retrieved: June, 2013]

# Robust Optimization for Stochastic Wireless CDMA/TDMA Networks

Belarmino Núñez, Pablo Adasme, Ismael Soto

Departamento de Ingeniería Eléctrica,  
Universidad de Santiago de Chile,  
Avenida Ecuador 3519 Santiago, Chile

Emails: {belarmino.nunez,pablo.adasme,ismael.soto}@usach.cl

Abdel Lisser

Laboratoire de Recherche en Informatique,  
Université Paris-Sud XI,  
Bâtiment 650, Orsay Cedex, France

Email: abdel.lisser@lri.fr

**Abstract**—In this paper, we propose a distributionally robust formulation for packet transmission allocation in CDMA/TDMA networks. In particular, we adopt a utility-based framework where channel bit rates and packet experienced delays conditions are considered. Consequently, the total utility of the network subject to capacity and packet assignment constraints is maximized. For this purpose, we first formulate the problem as a (0-1) stochastic integer linear programming problem. Then, we transform the stochastic model into an equivalent deterministic formulation. Subsequently, we use the deterministic model to derive the distributionally robust counterpart. This is achieved while taking into account the set of all possible probability distributions for the input random parameters. Finally, we compare the optimal solutions of the stochastic and robust models. Our preliminary numerical results indicate that slight conservative solutions can be obtained when the instances dimensions increase.

**Keywords**—Stochastic programming; distributionally robust optimization; code division and time division multiple access, wireless networks.

## I. INTRODUCTION

Code division and time division multiple access (resp. CDMA, TDMA) are two wireless multi-carrier transmission schemes currently embedded into modern technologies such as Wifi and Wimax [14]. By multiple access, we mean several users can send signals simultaneously over a single wireless communication channel. In particular, CDMA uses a spread spectrum technology combined with coding schemes to allow multiple users on the same physical channel. On the other side, TDMA has the property of scheduling users in time by assigning all bandwidth channel capacity to only one user at a given time slot in order to transmit signals. Although, these transmission schemes work differently, the underlying purpose in both of them is nearly the same, i.e., to make an efficient use of resources such as power and bandwidth channel capacity of the network. Hybrid transmission schemes do also exist such as CDMA/TDMA [14]. In this paper, we use a utility-based framework and formulate a stochastic resource allocation model for a CDMA/TDMA network. Thus, we associate for each packet transmitted by a base station, a utility function depending on the corresponding channel bit rates and experienced delays according to a CDMA/TDMA network. Consequently, we maximize the total utility of the network subject to capacity and packet assignment constraints. The capacity constraint appears as a technological constraint in these type of networks since a particular base station can

not transmit packets in more than “ $M$ ” frames within a given time slot [11]. We derive a distributionally robust formulation for this stochastic problem. For this purpose, we first formulate the problem as a (0-1) stochastic integer linear programming problem. Next, we transform the stochastic model into an equivalent deterministic formulation. In particular, we transform the probabilistic constraints using the approach proposed by [1]. Afterward, we use the deterministic formulation to derive the robust counterpart [16]. This is achieved while taking into account the set of all possible probability distributions for the input random parameters. Finally, we compare the optimal solutions of the stochastic and robust models. Preliminary numerical comparisons are given.

The paper is organized as follows. In section II, we provide a brief state of the art concerning resource allocation in wireless CDMA/TDMA networks, stochastic programming and robust optimization. Section III presents the stochastic CDMA/TDMA model under study and its equivalent deterministic formulation. Then, in section IV, we derive the distributionally robust formulation. In section V, we provide preliminary numerical comparisons. Finally, in section VI, we conclude the paper.

## II. RELATED WORK

Resource allocation for wireless CDMA and TDMA networks has been studied separately (see e.g. [5], [12]) and jointly as well [11]. The latter approach, which we refer to as hybrid CDMA/TDMA transmission scheme, exploits delay tolerance of non-real time traffic in order to improve the use of bandwidth channel capacity of the network [17], [18]. This is possible since the base stations, which transmit signals to users, may have the possibility to choose different subchannels to transmit their signals. The resulting bit rate improvement is referred to as a multi-user diversity gain. Due to the inherent random nature of wireless channels in CDMA/TDMA networks, there have been proposed some few attempts while considering general stochastic approaches, e.g., [19]. However, to the best of our knowledge, none of them has yet considered stochastic programming as in [7] and/or distributionally robust optimization [16] approaches to deal with the inherent uncertainty of the problem.

Stochastic programming (SP) and robust optimization (RO) are two well known optimization techniques to cope with mathematical optimization problems involving uncertainty in

the input parameters. In SP, one often assumes that the probability distributions are discrete and known [2]. Two well known scenario SP approaches are the recourse model and the probabilistic constrained approaches [7], [15]. Different from SP, the RO framework assumes that the input random parameters lie within a convex uncertainty set and that the robust solutions must remain feasible for all possible realizations of the input parameters. Thus, the optimization is performed over the worst case realization of the input parameters. In compensation, we obtain robust solutions, which are protected from undesired fluctuations of the input parameters. In practice, this means that the objective function provides more conservative solutions. We refer the reader to [3], [4] for a more general understanding on RO. In particular, the distributionally robust optimization approach bridges the gap between the conservatism level of robust optimization and the stochastic programming approach. The conservatism can be measured by the loss in optimality in exchange for a robust solution which is more protected against uncertainty [3]. This means, the less conservative the robust solutions are, the better the RO approach. Thus, the distributionally robust approach optimizes the worst-case objective over a family of possible distributions. This approach was pioneered in [6], [9]. In particular, Scarf [6] proposes an application for a news-vendor problem while in [8], Yue et al. minimizes the worst case absolute regret for all distributions with certain mean and variance. In a similar vein, ElGhaoui et al. [13] proposed worst-case value at risk bounds for a robust portfolio optimization problem using only bounds on the means of the assets and their covariance matrix. Similarly, Natarajan et al. [10] derived a distributionally robust model for portfolio optimization, where the investor maximizes his worst case expected utility over a set of ambiguous distributions described by the knowledge of the mean, covariance and support information.

In this paper, we propose a (0-1) stochastic integer linear programming problem for wireless CDMA/TDMA networks. Next, we transform the stochastic model into an equivalent deterministic formulation we use to derive a novel distributionally robust counterpart [16]. This is achieved while taking into account the set of all possible probability distributions for the input random parameters.

### III. STOCHASTIC FORMULATION

We consider a wireless CDMA/TDMA network composed of a set of base stations  $\mathcal{B} = \{1, \dots, B\}$  and a set of packets  $\mathcal{N} = \{1, \dots, N\}$  waiting to be scheduled and transmitted by the different base stations in  $\mathcal{B}$ . We assume that the CDMA/TDMA network supports only non-realtime traffic. In practice, this means that the system can tolerate packet transmission delays. We also assume that the system is time slotted with each slot containing  $M$  frames. We consider the following (0-1) stochastic integer linear programming model for this problem and denote it hereby  $P_0$  as

$$\begin{aligned} \max_{\{x,y\}} \quad & E_{\xi} \left\{ \sum_{i=1}^N u_i(\xi) y_i + \sum_{i=1}^N \sum_{j=1}^B v_{ij}(\xi) x_{ij} \right\} \quad (1) \\ \text{s.t.} \quad & P_{\xi} \left\{ \sum_{i=1}^N m_{ij}(\xi) x_{ij} \leq M \right\} \geq 1 - \alpha; \forall j \in \mathcal{B} \quad (2) \end{aligned}$$

$$y_i + \sum_{j=1}^B x_{ij} = 1; \forall i \in \mathcal{N} \quad (3)$$

$$y_i, x_{ij} \in \{0, 1\}, \forall i, j \quad (4)$$

where the term  $E\{\cdot\}$  denotes mathematical expectation and  $P\{\cdot\}$  represents a probability constraint which should be satisfied at least for  $(1 - \alpha)\%$  of the cases where  $\alpha \in (0, 0.5]$  represents the risk. The matrix  $v = v(i, j)$  represents the utility gained by the system while transmitting packet  $i \in \mathcal{N}$  through the base station  $j \in \mathcal{B}$ . Similarly, the vector  $u = u(i)$  denotes the gained achieved by the system when packet  $i \in \mathcal{N}$  is not transmitted during a particular time slot. This is possible as we assume non-realtime traffic in the network. Finally, let matrix  $m = m(i, j)$  represent the number of required frames for transmission of packet  $i \in \mathcal{N}$  by base station  $j \in \mathcal{B}$  in a particular time slot. In  $P_0$ ,  $y \in \{0, 1\}^N$  and  $x \in \{0, 1\}^{NB}$  are binary decision variables we define as follows. Variable  $y_i$  equals one if packet  $i$  is not scheduled to be transmitted in the current time slot and equals zero otherwise. Similarly, variable  $x_{ij}$  equals one if packet  $i$  is scheduled to be transmitted within the current time slot by the base station  $j$  and equals zero otherwise. The objective function (1) maximizes the total utility of the system while constraint (2) imposes a maximum number of  $M$  frames for each  $j \in \mathcal{B}$ . Finally, constraint (3) indicates that each packet  $i \in \mathcal{N}$  must be transmitted or not by a unique base station in the system. Without loss of generality we assume that the matrix  $v = v(\xi)$  and the vectors  $u = u(\xi)$ , and  $m = m(\xi)$  are random variables distributed according to a discrete probability distribution  $\Omega$ . As such, one may suppose that  $v = v(\xi)$ ,  $u = u(\xi)$  and  $m = m(\xi)$  are concentrated on a finite set of scenarios as  $v_{ij}(\xi) = \{v_{ij}^1, \dots, v_{ij}^K\}$ ,  $u_i(\xi) = \{u_i^1, \dots, u_i^K\}$  and  $m_{ij}(\xi) = \{m_{ij}^1, \dots, m_{ij}^K\}$  respectively, with probability vector  $p^T = (p_1, \dots, p_K)$  such that  $\sum_{k=1}^K p_k = 1$  and  $p_k \geq 0$ . This allows us to formulate a deterministic equivalent formulation for  $P_0$  as follows [1]

$$\max_{\{x,y,z\}} \quad \sum_{k=1}^K p_k \left( \sum_{i=1}^N u_i^k y_i + \sum_{i=1}^N \sum_{j=1}^B v_{ij}^k x_{ij} \right) \quad (5)$$

$$\text{s.t.} \quad \sum_{i=1}^N m_{ij}^k x_{ij} \leq M + z_{jk} L; \forall j \in \mathcal{B}; k = 1 : K \quad (6)$$

$$\sum_{k=1}^K p_k z_{jk} \leq \alpha; \forall j \in \mathcal{B} \quad (7)$$

$$y_i + \sum_{j=1}^B x_{ij} = 1; \forall i \in \mathcal{N} \quad (8)$$

$$y_i, x_{ij}, z_{jk} \in \{0, 1\}, \forall i, j, k \quad (9)$$

where constraints (6)-(7) are the deterministic constraints replacing the probabilistic constraint (2) in  $P_0$ . Hereafter, we denote by  $P_1$  problem (5)-(9).

### IV. ROBUST FORMULATION

In this section, we derive a distributionally robust model for  $P_1$ . To this purpose, we assume that the probability distribution of the random vector  $p^T = (p_1, \dots, p_K)$  is not known and that it can be estimated by some statistical mean from some available

historical data. Thus, we consider the maximum likelihood estimator of the probability vector  $p^T$  to be the observed frequency vector.

In order to formulate a robust model for  $P_1$ , we write its objective function as follows

$$\min_{\{x,y\}} \max_{\{\pi \in H_\beta\}} \sum_{k=1}^K \pi_k \left( -\sum_{i=1}^N u_i^k y_i - \sum_{i=1}^N \sum_{j=1}^B v_{ij}^k x_{ij} \right) \quad (10)$$

and the left hand side of constraint (7) as the maximization problem

$$\max_{\{\pi \in H_\beta\}} \sum_{k=1}^K \pi_k z_{jk}, \quad \forall j \quad (11)$$

where the set  $H_\beta$  is defined as

$$H_\beta = \left\{ \pi_k \geq 0, \forall k : \sum_{k=1}^K \pi_k = 1, \sum_{k=1}^K \frac{|\pi_k - p_k|}{\sqrt{p_k}} \leq \beta \right\} \quad (12)$$

and  $\beta \in [0, \infty)$ . Now, let  $\delta_k = \pi_k - p_k$ , then the inner max problem in (10) can be written as

$$\max_{\{\delta\}} \sum_{k=1}^K (\delta_k + p_k) \left( -\sum_{i=1}^N u_i^k y_i - \sum_{i=1}^N \sum_{j=1}^B v_{ij}^k x_{ij} \right) \quad (13)$$

$$\text{s.t.} \quad \sum_{k=1}^K \frac{|\delta_k|}{\sqrt{p_k}} \leq \beta \quad (14)$$

$$\sum_{k=1}^K \delta_k = 0 \quad (15)$$

$$\delta_k \geq -p_k, \quad k = 1 : K \quad (16)$$

The associated dual problem is

$$\min_{\{w^1, \varphi^1, v^1\}} \sum_{k=1}^K p_k \left( -\sum_{i=1}^N u_i^k y_i - \sum_{i=1}^N \sum_{j=1}^B v_{ij}^k x_{ij} \right) + \sum_{k=1}^K p_k w_k^1 + \beta \varphi^1 \quad (17)$$

$$\text{s.t.} \quad \varphi^1 \geq \sqrt{p_k} \left( v^1 + w_k^1 - \sum_{i=1}^N u_i^k y_i - \sum_{i=1}^N \sum_{j=1}^B v_{ij}^k x_{ij} \right), \quad \forall k \quad (18)$$

$$\varphi^1 \geq -\sqrt{p_k} \left( v^1 + w_k^1 - \sum_{i=1}^N u_i^k y_i - \sum_{i=1}^N \sum_{j=1}^B v_{ij}^k x_{ij} \right), \quad \forall k \quad (19)$$

$$w_k^1 \geq 0, \quad \forall k \quad (20)$$

and  $\varphi^1, v^1, w^1$  are Lagrangian multipliers for constraints (14)-(16), respectively. Similarly, we obtain a dual formulation for

each  $j$  in (11) as follows

$$\min_{\{w^2, \varphi^2, v^2\}} \sum_{k=1}^K p_k z_{jk} + \sum_{k=1}^K p_k w_k^2 + \beta \varphi^2 \quad (21)$$

$$\text{s.t.} \quad \varphi^2 \geq \sqrt{p_k} (v^2 + w_k^2 + z_{jk}), \quad \forall k \quad (22)$$

$$\varphi^2 \geq -\sqrt{p_k} (v^2 + w_k^2 + z_{jk}), \quad \forall k \quad (23)$$

$$w_k^2 \geq 0, \quad \forall k \quad (24)$$

where  $\varphi^2, v^2, w^2$  are Lagrangian multipliers associated with its respective primal constraints. Now, replacing these dual problems in  $P_1$  gives rise to the following distributionally robust formulation we denote by  $P_R$

$$\max_{\{w^1, \varphi^1, v^1, w^2, \varphi^2, v^2, x, y, z\}} \sum_{k=1}^K p_k \left( \sum_{i=1}^N u_i^k y_i + \sum_{i=1}^N \sum_{j=1}^B v_{ij}^k x_{ij} \right) - \sum_{k=1}^K p_k w_k^1 - \beta \varphi^1 \quad (25)$$

$$\text{s.t.} \quad \varphi^1 \geq \sqrt{p_k} \left( v^1 + w_k^1 - \sum_{i=1}^N u_i^k y_i - \sum_{i=1}^N \sum_{j=1}^B v_{ij}^k x_{ij} \right), \quad \forall k \quad (26)$$

$$\varphi^1 \geq -\sqrt{p_k} \left( v^1 + w_k^1 - \sum_{i=1}^N u_i^k y_i - \sum_{i=1}^N \sum_{j=1}^B v_{ij}^k x_{ij} \right), \quad \forall k \quad (27)$$

$$w_k^1 \geq 0, \quad \forall k \quad (28)$$

$$\sum_{i=1}^N m_{ij}^k x_{ij} \leq M + z_{jk} L; \quad \forall j \in \mathcal{B}; k = 1 : K \quad (29)$$

$$z_{jk} \in \{0, 1\} \quad \forall j, k \quad (30)$$

$$\sum_{k=1}^K p_k z_{jk} + \sum_{k=1}^K p_k w_k^2 + \beta \varphi^2 \leq \alpha, \quad \forall j \in \mathcal{B} \quad (31)$$

$$\varphi^2 \geq \sqrt{p_k} (z_{jk} + v^2 + w_k^2), \quad \forall j, k \quad (32)$$

$$\varphi^2 \geq -\sqrt{p_k} (z_{jk} + v^2 + w_k^2), \quad \forall j, k \quad (33)$$

$$w_k^2 \geq 0, \quad \forall k \quad (34)$$

$$y_i + \sum_{j=1}^B x_{ij} = 1; \quad \forall i \in \mathcal{N} \quad (35)$$

$$y_i, x_{ij}, z_{jk} \in \{0, 1\}, \quad \forall i, j, k \quad (36)$$

In the next section, we provide numerical comparisons between  $P_1$  and  $P_R$ . This allows measuring the conservatism level of  $P_R$  w.r.t.  $P_1$ .

## V. NUMERICAL RESULTS

In this section, we present preliminary numerical results. A Matlab program is developed using Cplex 12.3 for solving  $P_1$  and  $P_R$ . The numerical experiments have been carried out on a Pentium IV, 1.9 GHz with 2 GB of RAM under windows XP. The input data is generated as follows. The probability vectors  $p^T$  is uniformly distributed in  $[0,1]$  such that the sum is equal to one. The parameter  $\alpha$  is set to 0.1. So far, each entry  $u_i^k, v_{ij}^k$  and  $m_{ij}^k, \forall i, j, k$  is generated randomly and uniformly distributed in  $[0,10], [0,20]$  and  $[1,5]$ , respectively.

The value of  $M$  is computed as  $M = \left( \sum_{i=1}^N m_{i1}^1 \right) * 0.5$ . So far, these random intervals are basically motivated on the range these parameters might be in realistic CDMA/TDMA systems. Nevertheless, more realistic input data will be used in a larger version of this paper.

In Table 1, the columns 1-3 give the size of the instances. Columns 4-5 provide the average optimal solutions over 25 different sample instances for  $P_1$  and  $P_R$ , respectively. Finally, column 6 gives the average gaps we compute for each instance as  $\frac{P_1 - P_R}{P_1} * 100$  %. These results are calculated for different values of  $\beta$ . From Table 1, we mainly observe that the solutions tend to be less conservative when the instances dimensions increase. In particular, we observe slight conservative solutions when the number of packets to be sent by the different base stations  $b \in \mathcal{B}$  increases. This is an interesting observation since the number of packets in real CDMA/TDMA networks is usually larger than the number of required base stations. Finally, we observe that by increasing the number of scenarios  $K$  in the robust model slightly affects the conservatism level. In

TABLE I. AVERAGE COMPARISONS OVER 25 INSTANCES.

Instance size			Avg. Opt. Sol.		Avg. Gap <sub>R</sub>
N	B	K	P <sub>1</sub>	P <sub>R</sub>	
$\beta = 5$					
10	5	10	121.4334	104.9270	13.54 %
20	10	20	245.7756	221.4753	9.87 %
50	10	30	593.6139	554.9500	6.51 %
50	10	40	580.7174	544.7319	6.19 %
100	20	50	1175.4	1128.4	3.99 %
$\beta = 10$					
10	5	10	120.3395	102.9955	14.37 %
20	10	20	245.1590	215.7783	11.97 %
50	10	30	595.3698	550.1536	7.59 %
50	10	40	582.5017	534.8705	8.17 %
100	20	50	1175.9	1111.1	5.50 %
$\beta = 50$					
10	5	10	123.6391	106.1063	14.14 %
20	10	20	245.5830	218.0361	11.20 %
50	10	30	594.9709	549.7108	7.60 %
50	10	40	579.8950	530.8381	8.45 %
100	20	50	1174.7	1110.7	5.44 %
$\beta = 500$					
10	5	10	122.7709	105.1106	14.34 %
20	10	20	245.7271	217.3286	11.52 %
50	10	30	594.5845	545.0882	8.32 %
50	10	40	579.4147	529.8945	8.54 %
100	20	50	1173.8	1105.4	5.82 %

order to see how the parameter  $\beta$  affects the optimal solutions given by  $P_R$ , in Figures 1, 2 and 3 we solve a small, a medium and a large size instance, respectively. These figures present the same information in their respective horizontal and vertical axes. In the horizontal axes, we give the value of beta while in vertical axes, we present the optimal solutions of  $P_1$ ,  $P_R$ , and the gaps obtained in each of these figure. These gaps are computed as  $\frac{P_1 - P_R}{P_1} * 100$  %. From these figures, we first confirm the fact that less conservative gaps are obtained for larger instances. In figure 1, we obtain a gap between 10 and 15% while in figures 2 and 3, they are between 6 and 8 %, and 4 and 6 %, respectively. Secondly, that the increase of parameter  $\beta$  affects the optimal solutions of  $P_R$  only when  $\beta \in [0, 30]$ . In view of this observation, in figure 4 we solve a new small size instance for these values of  $\beta$ . This table presents the same information as for figures 1, 2 and 3. Here, we mainly observe that the major fluctuations of the optimal solutions of  $P_R$  are due to small values of  $\beta \in [0, 5]$ . While small fluctuations are observed for values of  $\beta \in [5, 30]$ . For

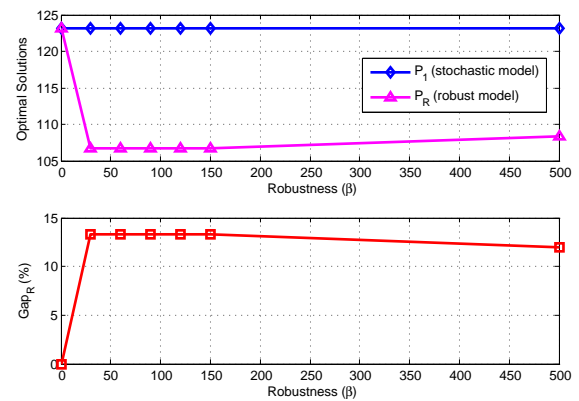


Fig. 1. INSTANCE # 1:  $N = 10, B = 5, K = 10$ .

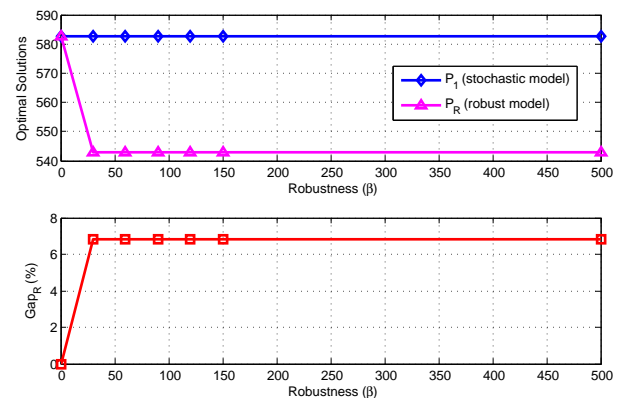


Fig. 2. INSTANCE # 2:  $N = 50, B = 10, K = 30$ .

example, when  $\beta$  goes from 0 to 5, we have a conservatism level gap increment of approximately 9% while for values of  $\beta$  that goes from 5 to 30, we obtain a gap increment of approximately 2,5%.

Finally, in figures 1, 2, 3 and 4 we observe similar gaps as in Table I. We recall that the gaps in these figures are obtained when using only one sample of the input data in  $P_R$ . This reflects, somehow, the accuracy of the gaps obtained with the proposed distributionally robust model.

## VI. CONCLUSION AND FUTURE WORK

In this paper, we proposed a distributionally robust formulation for packet transmission allocation in CDMA/TDMA networks where a utility function depending on channel bit rates and experienced delays is considered. Consequently, the total utility of the network subject to capacity and packet assignment constraints is maximized. To this purpose, we formulate the problem as a (0-1) stochastic integer linear programming problem which we transform into an equivalent deterministic model. Then, we use the deterministic model to derive the distributionally robust counterpart. This is achieved while taking into account the set of all possible probability distributions for the input random parameters. Then, we compared the optimal solutions of the stochastic and robust models. Preliminary numerical results indicate that the optimal solutions of the proposed robust model are much less conservative when

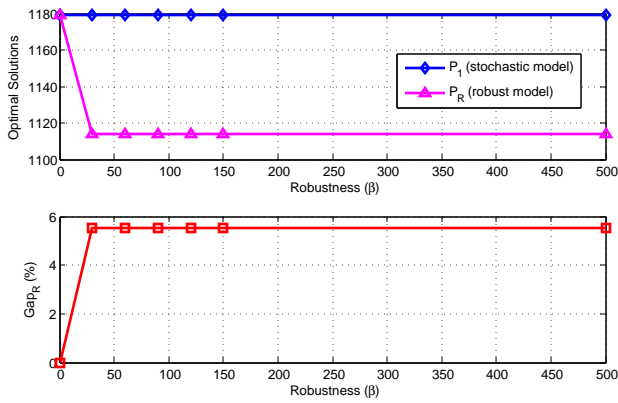


Fig. 3. INSTANCE # 3: N = 100, B = 20, K = 50.

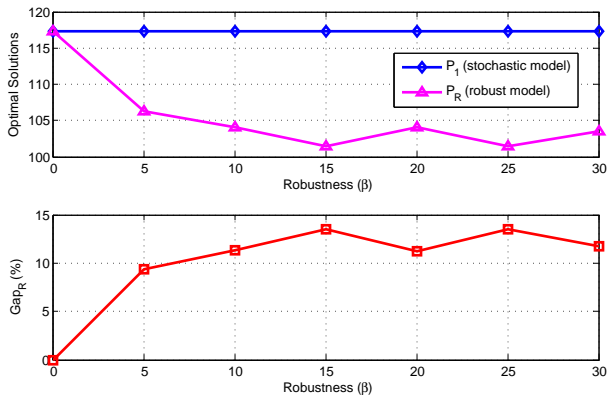


Fig. 4. INSTANCE # 4: N = 10, B = 5, K = 10.

the size of the instances increase. In particular, we emphasize the fact that slight conservative solutions are obtained when the number of packets to be sent by the different base stations increases w.r.t the number of bases stations. This is very important result in realistic CDMA/TDMA networks since the number of required base stations is often lower compared to the required bit rates throughput of the networks. Finally, we also highlight that by increasing the number of scenarios  $K$  in the robust model slightly affects the conservatism level. This is also another interesting result as one may use a much larger number of scenarios, which would allow approaching more realistic wireless networks.

As future research, we plan to consider other stochastic programming approaches while using continuous probability distributions and more realistic input data for CDMA/TDMA networks and also for different kinds of wireless networks such as orthogonal frequency division and space division multiple access transmissions schemes.

ACKNOWLEDGMENT

The authors Belarmino Núñez, Pablo Adasme and Ismael Soto would like to thank to the projects: Conicyt/Insertion project Number 79100020, “Center for multidisciplinary research on signal processing” (project Conicyt/ACT1120) and project USACH/Dicyt Number 061117S) for their partial financial support.

REFERENCES

- [1] A. Gaivoronski, A. Lisser, and R. Lopez, “Knapsack problem with probability constraints”, *Journal of Global Optimization*, Vol. 49, Number 3, 2011, pp. 397-413.
- [2] A. Shapiro, D. Dentcheva, and A. Ruszczyński, “Lectures on Stochastic Programming: Modeling and Theory”, SIAM Philadelphia, Series on Optimization, Vol. 9 of MPS/SIAM, Philadelphia, 2009.
- [3] D. Bertsimas and M. Sim, “The Price of Robustness”, *Operations Research*, Vol. 52, Number 1, 2004, pp. 35-53.
- [4] D. Bertsimas, D. Brown, and C. Caramanis, “Theory and applications of robust optimization”, *SIAM Review*, Vol. 53, Number 3, 2010, pp. 464-501.
- [5] D. Zhang, S. Oh, and N. Bhushan, “Optimal Resource Allocation for Data Service in CDMA Reverse Link”, *IEEE Transactions on Wireless Communications*, Vol. 6, Number 10, 2007, pp. 3648-3656.
- [6] H. Scarf, “A Min-Max Solution of an Inventory Problem”, *Studies in The Mathematical Theory of Inventory and Production*, 1958, pp. 201-209.
- [7] J. Birge and F. Louveaux, “Introduction to stochastic programming”, Springer-Verlag, New York, 1997.
- [8] J. Yue, B. Chen, and M. Wang, “Expected Value of Distribution Information for the Newsvendor Problem”, *Operations Research*, Vol. 54, Number 6, 2006, pp. 1128-1136.
- [9] J. Žáčeková, “On Minimax Solution of Stochastic Linear Programming Problems”, *Časopis pro Peštožaní*, Vol. 91, 1966, pp. 423-430.
- [10] K. Natarajan, M. Sim, and J. Uichanco, “Tractable Robust Expected Utility and Risk Models for Portfolio Optimization”, *Mathematical Finance*, Vol. 20, 2010, pp. 695-731.
- [11] K. Navaie and H. Yanikomeroglu, “Optimal Downlink Resource Allocation for Non-realtime Traffic in Cellular CDMA/TDMA Networks”, *IEEE Communications Letters*, Vol. 10, Number 4, 2006.
- [12] K. Shih, C. Chang, Y. Chen, and T. Chuang, “Dynamic bandwidth allocation for QoS routing on TDMA-based mobile ad hoc networks”, *Computer Communications*, Vol. 29, Number 9, 2006, pp. 1316-1329.
- [13] L. ElGhaoui, M. Oks, and F. Oustry, “Worst-case Value-at-Risk and Robust Portfolio Optimization: A Conic Programming Approach”, *Operations Research*, Vol. 51, 2003, pp. 543-556.
- [14] M. Sternad, T. Svensson, T. Ottosson, A. Ahlen, A. Svensson, and Anna Brunstrom, “Towards Systems Beyond 3G Based on Adaptive OFDMA Transmission”, *Proceedings of the IEEE*, Vol. 95, 2007.
- [15] R. Schultz, S. Leen, and M. Van Der Vlerk, “Two-stage stochastic integer programming: a survey”, *Statistica Neerlandica*, Vol. 50, 1996, pp. 404-416.
- [16] S. Liao, C. Delft, and J. Vial, “Distributionally robust workforce scheduling in call centres with uncertain arrival rates” *Optimization Methods and Software*, Vol. 28, Number 3, 2013, pp. 501-522.
- [17] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, and A. Viterbi, “CDMA/HDR: a bandwidth-efficient high-speed wireless data service for nomadic users”, *IEEE Communications Magazine*, vol. 38, 2000, pp. 70-77.
- [18] P. Viswanath, D. Tse, and R. Laroia, “Opportunistic beam-forming using dumb antennas”, *IEEE Transactions on Information Theory*, vol. 48, 2002, pp. 1277-1294.
- [19] X. Wang, D. Wang, H. Zhuang, and S. Morgera, “Fair energy-efficient resource allocation in wireless sensor networks over fading TDMA channels”, *IEEE Journal on Selected Areas in Communications*, Vol. 28, Number 7, 2010, pp. 1063-1072.

# Towards an Efficient Handling of the Maximum Triangle Packing Problem

Youcef Abdelsadek<sup>1,2</sup>, Francine Herrmann<sup>2</sup>

<sup>1</sup>Department of Informatics, Systems and Collaboration  
Public Research Centre-Gabriel Lippmann  
Belvaux, Luxembourg  
e-mail: abdelasad@lippmann.lu, otjacque@lippmann.lu

Imed Kacem<sup>2</sup>, Benoît Otjacques<sup>1</sup>

<sup>2</sup>Laboratoire de Conception, Optimisation et Modélisation  
des Systèmes (LCOMS EA 7306)  
University of Lorraine - Metz, France  
e-mail: francine.herrmann@univ-lorraine.fr,  
imed.kacem@univ-lorraine.fr

**Abstract**—This work addresses the problem of finding the maximum number of unweighted vertex-disjoint triangles in an undirected graph  $G$ . It is a challenging NP-hard problem in combinatorics and it is well-known to be APX-hard. We propose two heuristics for this problem. The first is based on local substitutions, while the second uses a surrogate relaxation of the related integer linear program which is analog to a specific knapsack problem. A computational comparison of the two heuristics using randomly generated benchmarks has shown that the first heuristic outperforms the second heuristic regarding the obtained packing solutions and the respective computation times.

**Keywords** - packing problems; maximum triangle packing; heuristics; computational study.

## I. INTRODUCTION

In the unweighted Maximum  $k$ -Set Packing ( $k$ -MSP) problem, a collection  $C$  of exactly  $k$  sized sets of a population  $P$  is given. The aim in the related optimization problem is to find the maximum number of pairwise disjoint sets. For example, let  $\Pi$  be an instance of a 2-MSP, where  $C = \{\{1, 2\}, \{1, 3\}, \{3, 4\}\}$  and  $P = \{1, 2, 3, 4\}$ . The optimal solution maximizing the number of pairwise disjoint sets for  $\Pi$  is  $\{\{1, 2\}, \{3, 4\}\}$ . The unweighted  $k$ -MSP is a well-known NP-complete problem in the complexity theory. Even when  $k = 3$ , the problem remains NP-hard and can be transformed to an instance of the 3-dimensional matching problem [1]. However, it is polynomial time solvable for the well-known 2-dimensional matching problem when  $k = 2$  [1].

In this paper, we will consider the Maximum Triangle Packing (MTP) problem. It can be formulated as follows. Let  $G$  be a graph such that  $G = (V, E)$  where  $V$  and  $E$  respectively denote the vertex set and the edge set for the graph  $G$ . For instance  $\Pi$  of the 3-MSP problem, the population  $P$  represents the set  $V$  while the collection  $C$  represents all the triangles (i.e., cliques of size 3) of the graph  $G$ . Indeed, a set  $S_i \in C$  which have  $x$ ,  $y$  and  $z$  as elements, if and only if,  $(x, y)$ ,  $(x, z)$  and  $(y, z)$  belongs to the set  $E$ . The goal in this case is to find the maximum number of vertex-disjoint triangles of the graph  $G$ , the so-called MTP problem. Notice that in this work we are not addressing the case of edges-disjoint triangles because we attempt to consider the disjoint 3-Set Cover (3-SC) problem [13]. The MTP problem can be found in numerous real-world problems, including scheduling, biology [2] and extraction of test forms [3].

The remaining paper's structure is as follows. In Section 2, we quote the relevant related work of the MTP problem cases. In Section 3, we give some basics and definitions illustrating them with an example. Section 4 describes the two proposed heuristics. In Section 5, we provide an experimental comparison of the aforementioned heuristics by randomly generated MTP instances and also detail how they are built. Section 6 concludes the paper and discusses further developments of this work.

## II. RELATED WORK

The interest of the scientific community regarding the MTP problem and  $k$ -MSP problem in general continues to grow almost daily. The MTP problem is NP-hard; Caprara and Rizzi [4] proved that the MTP problem is also APX-hard for graphs with a degree  $\geq 4$ , which means that we cannot approximate the MTP problem within any given approximation ratio (i.e., nonexistence of a polynomial time approximation scheme [1]) unless  $P = NP$ . However, it is solvable in polynomial time if the maximum graph degree is at most 3. In [5], it is showed that the MTP problem admits a  $(3 - (\sqrt{13} / 2) + \epsilon)$ -approximation algorithm when the maximum degree is at most 4, for any constant  $\epsilon > 0$ , whereas the  $((76 / 75) - \epsilon)$ -non-approximability is shown in [2]. In [6], the  $k$ -MSP problem (resp., MTP problem) admits a  $((k / 2) + \epsilon)$ -approximation algorithm (resp.,  $((3 / 2) + \epsilon)$ -approximation algorithm). Wang, Feng and Chen [9] gave an  $O(3.53^{3k})$ -time parameterized algorithm for the 3-MSP problem.

For the weighted case, a weight is assigned to each edge of a complete graph of size  $t * 3$ , the goal is to maximize the sum of edge weight selecting  $t$  vertex-disjoint triangles. Hassin and Rubinstein proposed a randomized  $((89 / 169) - \epsilon)$ -approximation algorithm [7] for the weighted case. An improvement of their work is given in [8], the approximation algorithm achieves an expected ratio of  $0.5257(1 - \epsilon)$ .

With respect to the constraint programming approach, an empirical comparison of three constraint models for the MTP problem is provided [10]. Additionally in [3], a branch and bound for the  $k$ -MSP problem is proposed. Recently, Bertsimas, Iancu and Katz [11] developed a pseudo-polynomial time approximation algorithm for the  $k$ -MSP problem within an approximation ratio of  $k^2 / (2k - 1)$ .

Due to the APX-hardness of the MTP problem, work tackling the non-approximability within some approximation ratio can give us a better grasp of its approximability.

### III. BASICS AND DEFINITIONS

In this section, we give some definitions which are necessary for the better understanding of the two proposed heuristics. The set of vertices  $\{v_1, v_2, \dots, v_n\} \in V$  has a size  $n$ , while the triangles set  $\{k_1, k_2, \dots, k_m\} \in K$  has a size  $m$ . The weight of a vertex  $v_i$  represents the number of triangles to which  $v_i$  belongs. These weights are called here  $\alpha_i$  such that  $i \in \{1, \dots, n\}$ . Figure 1 gives an example of a graph  $G$  and vertex weights. The weight of a triangle  $k_j = \{v_i, v_g, v_h\} \in K$ , is the sum of its vertices weights  $v_i, v_g, v_h \in V$ . We refer to the triangles weights as  $\beta_j$  such that  $j \in \{1, \dots, m\}$ . We say that two triangles  $k_j$  and  $k_{j'} \in K$  overlap, if they share a common vertex  $k_j \cap k_{j'} = v_i \in V$ . In Figure 1, the triangles  $k_1$  and  $k_2$  overlap in  $v_2$ . Moreover, the neighborhood degree  $\lambda_j$  of a triangle  $k_j \in K$  represents the number of triangles that overlap  $k_j$ , called in [10] collision count. For example, the triangle  $k_1$  has as a neighborhood degree value of 2 because it shares a vertex with  $k_2$  and  $k_3$ . Table 1 gives the triangle weight and neighborhood degree for the same graph  $G$ .

### IV. OUR HEURISTICS

Before describing the two proposed heuristics, we point out that each vertex of the input graph must belong to at least one triangle, without loss of generality. Moreover, in this study the outcomes of MTP instances are assessed regarding the number of obtained triangles, which means that the smallest gap between the optimal solution and the returned solution is better. We also pay attention to the computation time that it costs.

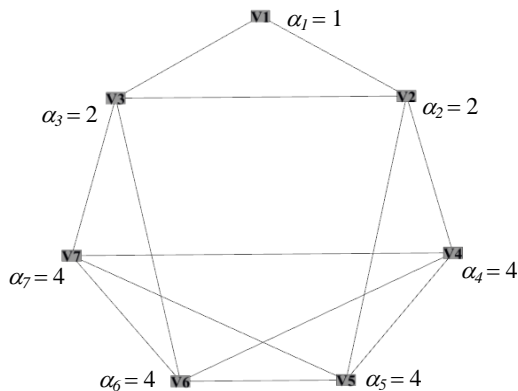


Figure 1. An example of a graph  $G$  and the vertex weights

TABLE I. TRIANGLE WEIGHT AND NEIGHBORHOOD DEGREE OF THE GRAPH  $G$

Triangle	The triangles set $K$						
	$k_1 = v_1 v_2 v_3$	$k_2 = v_2 v_4 v_5$	$k_3 = v_3 v_6 v_7$	$k_4 = v_4 v_5 v_7$	$k_5 = v_4 v_5 v_6$	$k_6 = v_6 v_7 v_5$	$k_7 = v_4 v_6 v_7$
$\beta_j$	5	10	10	12	12	12	12
$\lambda_j$	2	5	5	4	4	4	4

#### A. Local substitution-based heuristic

In this section, we explain the behavior of the first heuristic as illustrated in Algorithm 1.

The algorithm starts by computing the neighborhood degree  $\lambda_j$  for each triangle. The next step is to find the triangle with the smallest  $\lambda_j$  value that is done with *TriMin* (). The motivation behind this is to select the triangles which outline the graph structure with the aim that the selection in the rest of the graph (i.e., the inner parts of the graph) will not generate couples and singletons. Algorithm 1 avoids selecting a triangle (i.e., *Tmin*) which overlaps with disjoint triangles neglecting a local optimality. For that *MaxDisjoint* () returns *Neigh* the maximum number of disjoint triangles which overlap with a certain triangle  $k_j$ . Notice that the number returned by *MaxDisjoint* () is upper bounded by 3, that is because we cannot have 4 disjoint triangles which overlap a triangle. The final step is to remove the selected triangle(s) and its (their) neighborhood triangles with *RemoveNeigh* (). The process is repeated until there remains no triangles in  $\xi$ .

#### Algorithm 1 (Local substitution)

**Input:** the graph  $G$

**Output:** a packing of triangles  $PT$

- a)  $\xi \leftarrow K, PT \leftarrow \emptyset, Neigh \leftarrow \emptyset, Tmin \leftarrow \emptyset;$
- b) **While** ( $\xi \neq \emptyset$ )
- c)     Calculate the neighborhood degree  $\lambda_j$  for each
- d)      $j \in \{1, \dots, m\};$
- e)      $Tmin \leftarrow TriMin(\xi);$
- f)      $Neigh \leftarrow MaxDisjoint(Tmin);$
- g)     **if** ( $Neigh > 1$ )  $PT \leftarrow PT + Neigh, \xi \leftarrow \xi \setminus Neigh,$
- h)      $RemoveNeigh(Neigh);$
- i)     **else**  $PT \leftarrow PT + Tmin, \xi \leftarrow \xi \setminus Tmin,$
- j)      $RemoveNeigh(Tmin);$
- k)     **end while;**
- l)     return  $PT;$
- m)     **end.**

To better understand the aforementioned heuristic we take the graph in Figure 1 as an example and we briefly outline its steps. First of all, the set of solution  $PT$  is initialized to  $\emptyset$  and  $\xi$  is initialized to  $K$ . After that  $k_1$  is assigned to  $Tmin$ , because it has the smallest  $\lambda_j$  value and  $\{k_2, k_3\}$  to  $Neigh$ , this is due to the fact that  $k_2$  and  $k_3$  are disjoint and overlap  $k_1$ . Finally,  $\{k_1, k_2, k_3, k_4, k_5, k_6, k_7\}$  are removed from  $\xi$  and the algorithm returns  $PT = \{k_2, k_3\}$ , which is one of the optimal solutions.

#### B. The second heuristic

The MTP problem for a graph  $G = (V, E)$  can be stated as an Integer Linear Program (ILP), like in (1) where  $B \in \{0, 1\}^{n \times m}$  is the vertex-triangle Belonging matrix. The columns of the matrix  $B$  contains exactly 3 ones, which means that the vertices representing the 3 non-zero values belong to the triangle (column). Let  $x_k = 1$  if the  $k$ th triangle is selected in the returned solution, 0 otherwise. A surrogate relaxation of (1) is presented in (2).



$$\begin{aligned}
& \text{Maximize } \sum_{k=1}^m x_k \\
& \text{Subject to: } \sum_{k=1}^m B_{ik} x_k \leq 1 \quad \forall i = \{1, \dots, n\} \\
& \quad \quad \quad x_k \in \{0, 1\}^m
\end{aligned} \tag{1}$$

$$\begin{aligned}
& \text{Maximize } \sum_{k=1}^m x_k \\
& \text{Subject to: } \sum_{k=1}^m \beta_k x_k \leq \alpha_0 \\
& \quad \quad \quad \beta_k = \sum_{i=1}^n B_{ik} \alpha_i \quad \forall i = \{1, \dots, m\} \\
& \quad \quad \quad \alpha_0 = \sum_{i=1}^n \alpha_i \\
& \quad \quad \quad x_k \in \{0, 1\}^m
\end{aligned} \tag{2}$$

Detailed work on the multiple varieties of relaxations for the  $k$ -MSP problem is presented in [12]. In (1) a constraint is devoted to each vertex, whereas in (2) we attribute the vertex weight  $\alpha_i$  to each constraint. Hence, a reformulation of the constraint-weighted ILP leads to (2) where the ILP is variable-weighted. The relaxed problem is reduced to the one-dimensional knapsack problem (1-KP) [1] where the items represent the triangles and the knapsack capacity is  $\alpha_0$ . In this case, the optimal solution for the 1-KP is to select the items in the increasing weight order, because all the items have 1 as profit value. The solution of the relaxed problem yields an upper-bound on the optimal solution of MTP problem. The following Algorithm 2 allows us to derive a feasible solution based on the relaxed solution.

Before outlining the different steps of Algorithm 2, we define what a packing solution is. A packing solution exists if no triangle of a solution overlaps. Algorithm 2 takes as an input the graph  $G$  and the vector of weights  $\alpha$  and  $\beta$ . It starts by calling `SolveILP()`, which solves (2) in a linear time. After that, Algorithm 2 verifies whether the returned solution is a packing solution or not. In the negative case, it chooses the first encountered triangles, which overlaps and selects the one having the smallest triangle weight  $\beta_j$  value, maximizing the  $\alpha_0 - \beta_j$  for future iterations, this is done with `Intersect()`. To avoid that the selected triangle in Step e intersects again, we attribute an exclusive penalty-weight to the neighborhood of the selected triangle via `Penalty()`. The penalty is a large integer value which is assigned to a triangle weight  $\beta_j$  without affecting its vertices weights  $\alpha_i$  (i.e., not uniformly distributed to its vertices weights). This process is repeated until the returned solution is a packing solution and terminates.

To illustrate Algorithm 2, let us consider the example in Figure 1. Algorithm 2 starts with  $X = \{0, 0, 0, 0, 0, 0, 0\}$  and  $\alpha_0 = 1 + 2 + 2 + 4 + 4 + 4 + 4 = 21$ . Step c returns  $X = \{1, 1, 0, 0, 0, 0, 0\}$ . Step d checks if the returned solution is a packing solution. A negative response is returned because  $k_1$  overlaps  $k_2$  in  $v_2$ . `Intersect()` is called and returns  $k_1$  as the chosen triangle, due to  $\beta_1 < \beta_2$ . A penalty is attributed to each of  $k_2$  and  $k_3$  excluding them from the returned solution, because  $k_2$  and  $k_3$  are neighborhood triangles of  $k_1$ . Algorithm 2 starts a second iteration and has as a result of Step c  $X = \{1, 0, 0, 1, 0, 0, 0\}$ . A positive response is retu-

### Algorithm 2 (relaxed ILP)

**Input:** the graph  $G$ , the weight vectors  $\alpha$  and  $\beta$

**Output:** a packing of triangles  $X \in \{0, 1\}^m$

- a)  $\xi \leftarrow K, X \leftarrow \{0\}^m$ ;
- b) **Do**
- c)  $X = \text{SolveILP}(X, (2), \alpha, \beta)$ ;
- d) **if** ( $X$  is a packing) **goto** step h;
- e) **else** `Intersect` ( $X, \xi$ );
- f) `Penalty` ( $X, \beta$ );
- g) **While** ( $X$  is not a packing solution);
- h) **return**  $X$ ;
- i) **end.**

ned because  $k_1$  does not overlap  $k_4$ . Algorithm 2 terminates and returns  $\{k_1, k_4\}$  which is another optimal solution for this example.

## V. COMPUTANTIONAL COMPARISON

We now give some details of the experimental tests. We have implemented and tested the two aforementioned algorithms using different MTP instances and have compared their results to the solution returned by (1) via CPLEX 12.5 [14]. Due to the lack of benchmarks, the MTP instances are randomly generated in the following way: we start with a stable graph of size  $n$ , we fix a vertex  $v_i \in V$  and chose randomly two other vertices  $v_g$  and  $v_h \in V$ , creating the triangle  $k_j = \{v_i, v_g, v_h\} \in K$ . We repeat this process until all vertices participate in at least one triangle. The size of the MTP instances are up to  $m = 4182$  and  $n = 3000$ .

All the tests are done on Intel® Core™ i7-2600 CPU (3.40 GHz, 3.70 GHz), with 16GB of RAM. Algorithm 1 and Algorithm 2 are coded in the C Language. Table 2 summarizes the number of returned solutions along with the computation times. In the columns tuples (2, 3), (4, 5) and (6, 7) the number of returned triangles and the computation times of Algorithm 1, Algorithm 2 and CPLEX respectively are represented.

From Table 2, we observe that Algorithm 1 outperforms Algorithm 2 regarding the returned packing solutions for all generated MTP instances. However, both are dominated by CPLEX outcomes. We highlight that we interrupted CPLEX resolution after 48 hours for the last two MTP instances. Regarding the computational times, Algorithm 1 needs little CPU time compared to Algorithm 2 but both are still polynomial. These two heuristics provide results relatively quickly and are more suitable for user interaction system applications. In this context, the systems must return a solution with a certain quality of the approximation in a small amount of time.

We remark that the gap between the returned number of triangles by the two proposed heuristics and CPLEX solutions strongly depends on the order of the triangle set  $\xi$ . Indeed, in the case when two triangles have the same weight value, the first in the triangle set  $\xi$  is chosen. This choice allows to have deterministic solutions, if it happens to run the two heuristics several times in a row. An improvement of the two heuristics will be to sort the triangle set  $\xi$  in such a way that the most promising triangle will be placed first in  $\xi$ .

TABLE II. THE OBTAINED RESULTS OF ALGORITHM 1, ALGORITHM 2 AND CPLEX USING RANDOMLY GENERATED MTP INSTANCES

MTP Instances (vertices, triangles)	The results and their Computation Times (CT)					
	Algorithm 1		Algorithm 2		CPLEX	
	Sol	CT (sec)	Sol	CT (sec)	Sol	CT
(600, 671)	135	0.073	131	0.16	167	3.26 (sec)
(800, 834)	188	0.096	172	0.33	230	19.2 (sec)
(850, 869)	194	0.12	186	0.48	240	10.2 (sec)
(900, 978)	207	0.142	196	0.49	260	11 min
(1000, 1050)	228	0.188	217	0.64	291	9 min
(1000, 1132)	229	0.184	214	0.73	295	11 min
(1100, 1152)	251	0.206	246	0.9	319	10 min
(1150, 1225)	261	0.237	249	1	330	132 min
(1200, 1280)	270	0.270	259	1.1	346	101 min
(1400, 1495)	322	0.482	303	1.8	Interrupted	
(3000, 4182)	719	5.04	666	26.7	Interrupted	

Sol, sec and min represent solution, secondes and minutes respectively.

## VI. CONCLUSION AND FUTURE WORK

In this work, we deal with the problem of finding the maximum number of unweighted vertex-disjoint triangles in an undirected graph  $G$ . The problem is well-known to be NP-hard and APX-hard. Two heuristics have been proposed for this problem. To investigate the heuristic's performance, we provide a computational comparison of both of them to the CPLEX solution with randomly generated MTP instances. We have shown that the first heuristic outperforms the second regarding the obtained packing solutions and the respective computation times.

As a perspective for future work, we plan to use other metrics to compute the triangles and vertices weight improving also the penalty process. We also project to compare the two proposed heuristics with other approximation algorithms and to consider the weighted case of the MTP problem. The main motivation behind this work is to use the obtained solution for the MTP problem to consider the disjoint 3-SC problem [13]. In the disjoint 3-SC problem the optimality is the minimum of sets within a collection  $C$  of size at most 3 which cover the population  $P$ . We expect to address the disjoint 3-MSC problem applied on a graph  $G =$

$(V, E)$  where the vertices set  $V$  is covered by the fewest number of 3-clique, 2-clique and 1-clique (a  $k$ -clique is a clique of size  $k$ ) maximizing the 3-clique packing hopefully generating a cover with the fewest sets.

## ACKNOWLEDGMENT

We would like to thank the anonymous referees for their pertinent remarks which improved the presentation of this paper.

## REFERENCES

- [1] M. R. Garey and D. S. Johnson, "Computers and Intractability: A Guide to the Theory of NP-Completeness," W. H. Freeman and Co., New York, 1979.
- [2] M. Ashley, T. B.-Wolf, P. Berman, W. Chaovalitwongse, B. DasGupta, and M.-Y. Kao, "On approximation for covering and packing problems," in Journal of Computer and System Sciences, vol. 75, 2009, pp. 287–302.
- [3] D. I. Belov and R. D. Armstrong, "A constraint programming approach to extract the maximum number of non-overlapping test forms," in Computational Optimization and Applications, vol. 33, 2006, pp.319–332.
- [4] A. Caprara and R. Rizzi, "Packing triangles in bounded degree graphs," In Information Processing Letters, vol. 84, 2002, pp. 175-180.
- [5] G. Manić and Y. Wakabayashi, "Packing triangles in low degree graphs and indifference graphs," in Discrete Math., vol. 308, 2008, pp. 1455-1471.
- [6] C. A. J. Hurkens and A. Schrijver, "On the size of systems of sets every  $t$  of which have an SDR, with an application to the worst-case ratio of heuristics for packing problem," in SIAM J. DISC. Math., vol. 2, No. 1, 1989, pp. 68-72.
- [7] R. Hassin and S. Rubinfeld, "An approximation algorithm for maximum triangle packing," Discrete Applied Math., vol. 154, 2006, pp. 971-979.
- [8] Z.-Z. Chen, R. Tanahashi, and L. Wang, "An improved randomized approximation algorithm for maximum triangle packing," in Discrete Applied Math., vol. 157, 2009, pp. 1640-1646.
- [9] J. Wang, Q. Feng, and J. Chen, "An  $O^*(3.53^{3k})$ -time parameterized algorithm for the 3-set packing problem," in Theoretical Computer Science, vol. 412, 2011, pp. 1745-1753.
- [10] P. Prosser, "Triangle packing with constraint programming," In the 9<sup>th</sup> International Workshop on Constraint Modelling and Reformulation, 2010, pp. 1-15.
- [11] D. Bertsimas, D. A. Iancu, and D. Katz, "A new local search algorithm for binary optimization," INFORMS Journal on Computing, Articles in Advance, 2012, pp. 1-14.
- [12] R. Borndörfer and R. Weismantel, "Set packing relaxations of some interger programs," ZIB-Report SC 97-30, 1999, pp. 1-19.
- [13] R. Duh and M. Fürer, "Approximation of  $k$ -set cover by semi-local optimization," In Proceedings of the 29<sup>th</sup> annual ACM Sym. On Theory of Computing 1997, pp. 256-264.
- [14] www-01.ibm.com/software/commerce/optimization/cplex-optimizer/ [retrieved:10.05.2013].

# Trial Testing Efficiency of Algorithms for Task Execution in Multi-Processor Systems

## Experimentation System for Priority Scheduling and Open Pool of Tasks

Magdalena Respondek, Leszek Koszalka, Iwona Pozniak-Koszalka, and Andrzej Kasprzak

Department of Systems and Computer Networks

Wroclaw University of Technology,

Wroclaw, Poland

170966@student.pwr.wroc.pl, {leszek.koszalka, iwona.pozniak-koszalka, andrzej.kasprzak}@pwr.wroc.pl

**Abstract**—The need of an increased throughput has led to a new approach in the computer system design. In order to face the growing demands of a potential user, focus on multitasking and maximally enhance the capabilities- multiprocessor systems have been introduced. In such systems, two or more Central Processing Units (CPU) are working in parallel, sharing computer bus and communicating through shared memory. In this paper, the homogenous system, with three identical processors, having a common ready tasks' queue, is considered. It is modeled with the use of the adequate simulator. The challenge is to apply the best possible scheduling algorithms, as to provide an optimal system, which meets all the quality of service requirements. The conducted research on both- open and closed pool of tasks is fully described. The results are presented and thoroughly analyzed in order to choose the best possible algorithms for the discussed cases.

**Keywords**-multiprocessor system; tasks scheduling; algorithm; time efficiency; simulation

### I. INTRODUCTION AND MOTIVATION

The objective of this paper is to plan and, by simulating the multiprocessing system, to conduct such a research as to maximally optimize it in terms of time efficiency.

According to [1], there are usually considered 5 different criteria of system optimization from efficiency point of view:

- Maximum CPU utilization- each processor should be as busy as possible;
- Maximum throughput- number of processes finished in one CPU cycle- there should be more work done in less time;
- Minimum turnaround time- the time needed for the process to be executed (since its arrival to a system till completion);
- Minimum waiting time- time spent by the process in ready queue;
- Minimum response time- time between CPU request and the first answer.

In order to fulfill all of the above requirements, each of parallel processors has to have scheduling algorithm implemented. With the use of such algorithm, each CPU is able to schedule the execution of system processes and to properly manage the workload.

The research was a continuation of the studies already conducted on this matter – described in technical reports [2] and [3]. Additionally, besides examining, how basic algorithms behave in the controlled, closed pool of tasks, new research was made to analyze the topic under conditions close to real systems. Closer look was taken at the priority scheduling algorithm and the open pool of tasks.

The rest of the paper is organized as follows. Section II focusses on the experimentation system used in the research. In Section III, each of the used scheduling algorithms is fully described, along with its manner of working. Section IV is the main part containing the full studies, with the methods, research descriptions and exemplary results analyzed for two cases: Priority Scheduling and Open Pool of Tasks. The final remarks and plans for further research appear in Section V.

### II. EXPERIMENTATION SYSTEM

In order to investigate the properties of the scheduling algorithms for the multiprocessor's system, a proper tool had to be utilized. For this research the created and implemented three-processor experimentation system, called 'Simulator Pro 3', has been used. The basic scheme of the input – output system, is presented in Fig.1.

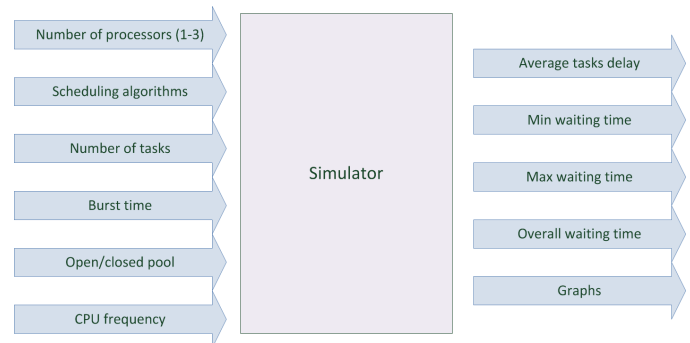


Figure. 1. Scheme of the simulator as input - output system.

The core of the system was the program described in [4]. 'Simulator Pro 3' was designed using C++ and must be launched on the MS Windows operating system. The system provides multiple options to a user, with a reasonable usage of which, the

complex research can be conducted. It can be noticed, that there are many options that can be modified in order to model the environment as specifically as it is possible. Hence, the constraints of the problem are those of a real-life system, including the number of processors along with their frequencies and pool of tasks, which system has to deal with (number of tasks, burst time, open/closed pool).

#### A. Input parameters

- *Number of processors:* 1-3;
- *Scheduling algorithms:* 6 different to choose from;
- *Number of tasks:* randomly or arbitrary chosen;
- *Burst time:* randomly chosen from 1-X range, where X is determined by the user;
- *Open/closed pool of tasks:* for the open pool additional number of processes can be entered into a queue;
- *Processes' priorities:* Importance of tasks represented by integer values;
- *CPU clock rate:* 1 - 4000MHz.

#### B. Output parameters

- *Average tasks delay:* Total waiting time divided by the number of tasks;
- *Min waiting time:* numbers of cycles representing the waiting time of the earliest executed task;
- *Max waiting time:* the longest time that the task had to wait for to be executed;
- *Overall waiting time:* waiting time obtained by all tasks;
- *Graphs:* bar charts representing waiting times of consecutive processes.

### III. IMPLEMENTED ALGORITHMS

In this research, the following requirements are taken into consideration. The access to CPUs is solved with a load sharing methodology, i.e., tasks are waiting in a ready queue, which is common for all three processors. Each CPU chooses the process from the queue to execute it according to a pattern, called scheduling algorithm.

In the modeled system, maximum of three tasks can be executed at the time. Each processor can have different scheduling algorithm implemented.

All the algorithms are described, e.g., in [5] and [6] and are depicted below:

#### A. First Come First Served

It is later referred to as FCFS. In this algorithm the tasks are executed in an order they request a CPU. That means that importance of a process is measured only by the time of its arrival. It can be managed by a FIFO (First in First Out) queue.

This type of algorithm is associated with the risk of convoy effect, where small tasks have to wait for the bigger one to be executed and get off of the processor.

#### B. Shortest Job First

Shortest Job First, later referred to as SJF, is the algorithm, in which the importance of the process is meant by the length of its next CPU burst. Hence, the processor chooses the task with the shortest duration time to be executed. When there are two processes with the same CPU burst length in a queue, FCFS scheduling is used.

Shortest Job First scheduling is hard to be implemented, because it is not possible to know the length of the next CPU burst, it can be only predicted.

There are two types of SJF algorithm: preemptive and non-preemptive. They are distinguishable only on the open pool of processes where the new task arrives to a queue while the other is being executed. If the duration of a new task is shorter than what is left of the currently executed one, the preemptive algorithm preempts the "old" process. That is why it can be also called Shortest Remaining Time First. In the same situation non-preemptive algorithm finishes the "old" task's execution and then can start the new one.

#### C. Round Robin

In Round-Robin (RR) type of scheduling there is a constant time quantum, after which the process is being preempted. Round-Robin works like FCFS scheduling, except that, after a time slice, CPU interrupts the execution of the process and takes on the new task from the queue. The "old" process is added to a queue's tail. Therefore in this algorithm, processes are not waiting long to be started, but because they are switched by the CPU, their delay time is long.

#### D. Priority Scheduling

In the Priority Scheduling, each process has a priority associated with it. Task with the biggest importance is the first one to be executed. Processes with equal priorities are scheduled according to a FCFS algorithm. SJF can be treated as a special case of Priority Scheduling, where priorities are corresponding with the tasks' lengths.

Priority algorithm can be either preemptive or non-preemptive. The main problem associated with this kind of scheduling is the starvation. On the open pool of task, when the new high-priority tasks are entering the system, the process with the lower significance may never be executed. The solution for this issue is aging where, after defined numbers of cycles, the priorities of long waiting tasks are increased.

### IV. INVESTIGATION

The research was mainly focused on two areas: (i) finding the best Priority Scheduling algorithm's parameters, and (ii) finding the best combination of algorithms in the open pool of tasks. After doing preliminary research, two complex experiments were conducted.

#### A Experiment #1 - Priority Scheduling

The first area of research was the Priority Scheduling algorithm. In this experiment not only the length or a number of

processes can be changed, but also the range of priorities and the aging step for algorithm.

1) Experiment Design

In this research there were taken three basic sets of data, each with 20 tasks. These processes' length was ranging from 1 to 10. First two sets had randomly chosen burst time. The last one's halves were sorted in ascending order.

Research was conducted on one processor only, for the clearer picture of Priority Scheduling manner of working. Once the best parameters were found, they could be applied for the further examination.

The experiments were conducted on every combination of sets and priorities range. For each composition of input parameters aging step had been changed gradually to observe the improvement.

The features of experiment design were taken as follows:

- 1 active processor;
- Non-preemptive Priority Scheduling applied;
- Three basic sets of tasks: SET 1 = 107, SET 2 = 134 tasks, and SET 3 = 110 tasks (see Table 1);
- Three ranges of priorities: 1-5, 1-10 and 1-20;
- Changing aging step (max value depending on the set).

TABLE 1. BASIC SET OF TASKS

No.	SET 1	SET 2	SET 3
1	9	6	1
2	2	10	2
3	6	3	3
4	8	8	4
5	2	5	5
6	4	8	6
7	7	10	7
8	9	7	8
9	9	5	9
10	4	8	10
11	5	5	1
12	6	5	2
13	4	9	3
14	9	3	4
15	8	8	5
16	3	10	6
17	7	5	7
18	2	2	8
19	1	10	9
20	2	7	10

2) Results

As the research was extensive, the results for two sets are presented in details: SET 1 of 107 tasks and SET 2 of 134 tasks - both for priority range of 1-5. Two cases are analysed in details (the average delay, the delay for consecutive tasks).

• Average delay

In Fig. 2 and Fig. 3, the averaged delay of all tasks in relation to aging step parameter is shown. Graphs are generalized and do not present the most essential data- for how long does the important tasks (with the high priority) have to wait for the execution.

What is more, when analyzing the graphs one can observe that the bigger number of tasks determines the longer average delay in the system. However, there is no clear consistency between the results for different length of processes. While the results are quite static for SET 1, for the second set of data delay time clearly rises gradually along with the aging step.

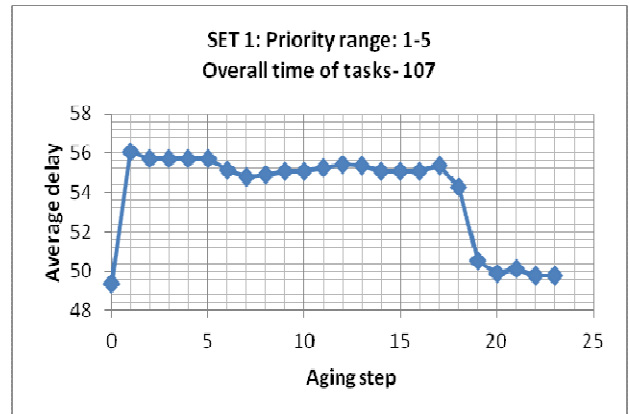


Figure 2. Average delay, SET 1, priorities 1-5.

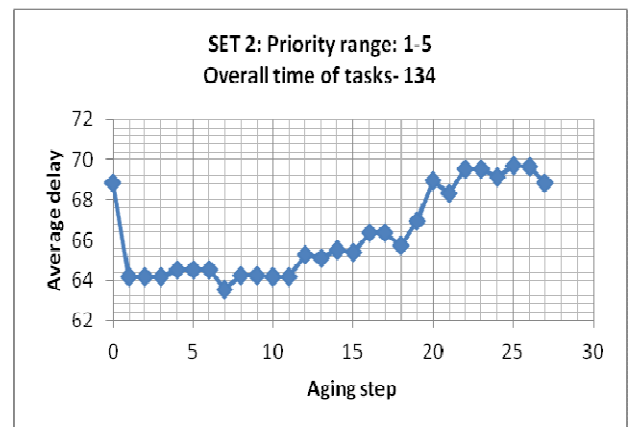


Figure 3. Average delay, SET 2, priorities 1-5.

Observing results presented in Fig.2 it can be noticed that the aging step of 20 has the same average delay that at the step of 0. It can be explained by reminding that no aging applied actually means that the value of aging step equals to a number of tasks in the system.

At the same time the best results are achieved by step 17 and equal to 14.1 cycles.

The shortest delay is observed for Fig.3 is the aging step of 7 and equals to approximately 64 cycles.

As no further conclusions can be drawn, more data is needed to analyse the priority problem. The outcome of this test can be fully reliable only when juxtaposed with the graph of delays for single tasks.

- **Delay for consecutive tasks**

The delay of consecutive tasks is a more complex and accurate way to present the results. The execution of a particular task can be observed and juxtaposed with the delay of other processes.

When analysing results (Fig.4) it appears that the bigger the aging step, the more rapid and uneven the graph. The maximum delay time is obviously the biggest for no aging applied, when the aging step is equal to a number of tasks (approx. 130 cycles for the task no. 11). The lower the aging step, the more the same task's delay decreases (by about 60 cycles in the minimum point).

When the aging step is big, the tasks are executed in an order of their priorities, no matter the size. It can cause the starving of the remaining processes (with lower importance).

On the other hand, small value of aging parameters cause that tasks are executed in the order of parameters but also causes a uniform distribution of processor's time. This solution may not be accepted, due to significant lowering the importance of the tasks.

### 3) Conclusion

To sum up the problem of choosing the optimal values of aging parameters, it can be said that the system must be well known, to model it properly. As it was observed on the above examination, both - set parameters and the length of tasks in the queue - have a high influence on the average time of tasks execution.

It appeared that the average delay of tasks for different aging parameters may not be accurate enough. On such representation, the data about the most important tasks is lost, so that the user may not have the good overall view of the system. It can help, however, when the large system is modelled and it is not possible to display characteristics of every task.

The more accurate and the more complicated at the same time is the graph representation of delay for every single task. It showed the remarkable influence of aging parameter on the system's time of processes execution. Due to this representation a connection between aging parameter and the accuracy of result (meant by the emphasis placed on priorities) was visible.

All in all, the means of choosing the best algorithm parameter is to find the best ratio between the aging parameter and pool of tasks' size, best fitted for the given requirements.

## B Experiment #2 - Open pool of tasks

The second experiment was focused on open pool of tasks. Usually, in the real-life system, the new tasks are constantly coming to the processor, requiring its time. Therefore, the closed pool system is not accurate enough for modelling the tasks execution. It can only show the overall view of the problem. The cases considered in this research are the rough approximation of a real-life dynamic system, which is hard to be predicted and requires an on-line tasks management [9].

One point has to be remembered when it comes to simulation of an open loop system - the data randomness. Due to this "issue" each experiment has to be repeated and the data must be averaged for more reliability of results.

### 1) Experiment Design:

The research was very exhaustive. As there were 6 different algorithms and 216 unique combinations for three processors should be considered. For each experimentation point (basic tasks plus new tasks introduced to the system) all those combinations had to be computed three times. At the end the mean and the standard deviation had been calculated for the average delay of the system.

The overall results for every combination had been compared and 20 - best and worst (in terms of tasks' delay) - schemes of algorithms had been found.

The features of experiment design were taken as follows:

- 6 algorithms- along with preemptive ones
- 2 sets: 10 and 20 basic tasks;
- Open pool of 10 or 20 tasks;
- Basic parameters of a system;
- 3 repetitions of each experimentation point;
- Clock rates of CPUs- 2000Hz.

### 2) Results

The outcome is displayed on the bar graphs representing the average value of processes' delay in the modeled system along with the standard deviation of results. The considered algorithms are symbolized by:

- FC - concerns FCFS;
- PN - non-preemptive priority scheduling;
- PW - preemptive priority scheduling;
- SN - non-preemptive SJF;
- SW - preemptive SJF;
- RR - Round-Robin.

The exemplary results will be shown for the two cases analyzed in details:

Case 1 : 10 basic tasks plus 10 new tasks,

Case 2 : 10 basic tasks plus 20 new tasks.

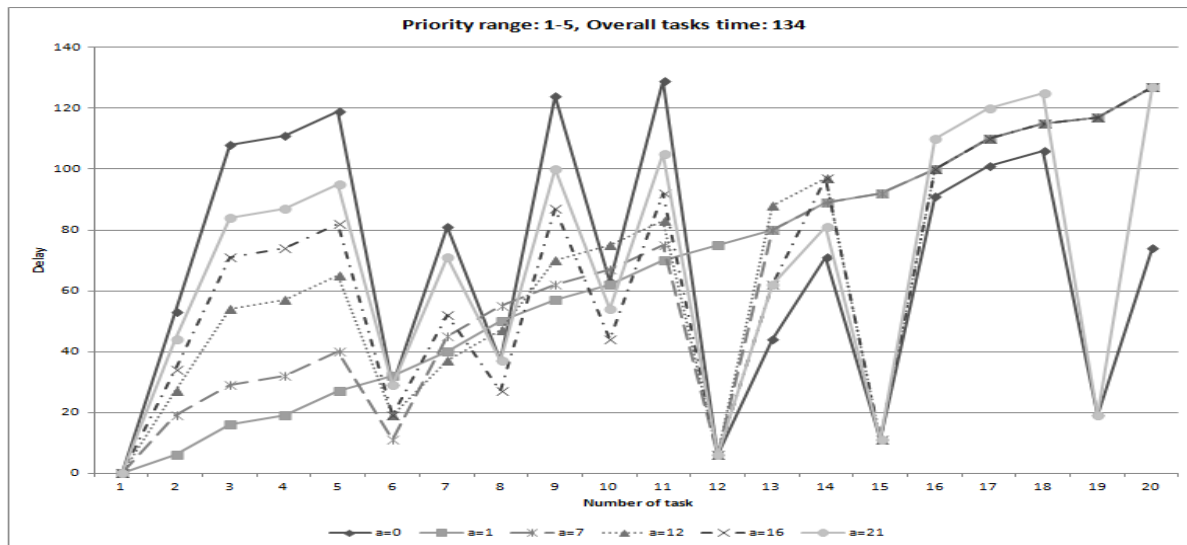


Figure 4. Delay for consecutive tasks; SET 2, priorities: 1-5.

• **Case 1: 10 basic tasks + 10 new tasks**

The outcome is shown in Fig.5. It can be observed that the best results for pool of 10 new tasks are achieved by mixed scheme of non-preemptive and preemptive SJF algorithm and equals to 2.65 cycles. The part of each best combination is the best SJF algorithm (either preemptive or non-preemptive). What can be surprising it is usually combined with the considered as not-so-good Priority and FCFS algorithms ([2] and [8]). Preemptive algorithm seems to be slightly better than non-preemptive one.

What more can be noticed, is that there is a very high variation of results, even for such a small pool of tasks. The deviation for best results shows that they are not so stable. So, it would be wiser to choose one of the 10 best combinations of algorithms, which have the lowest standard deviation.

The worst for this experimentation point is the combination of 2 Round- Robin algorithms with a preemptive priority scheduling and equals to 9.75 cycles. The result is more than three times worse that the best result. The least efficient of the bunch seems to be the Round-Robin algorithm, being a part of each combination with the biggest delays.

The variation of the bad result is not as high as for the good ones. It can be explained by the fact that the worst solutions usually give constant average delays while at the same time “good” combinations are similarly efficient.

• **Case 2: 10 basic tasks +20 new tasks**

The bigger the number of new tasks introduced to the system, the more consistent the outcome. It can be established by observing the standard deviation of results. There is also a bigger difference between the outcomes. The more efficient combination’s average delay equals to 1.77 cycles. The worst outcome reaches 7.53 cycles, which is more than 4 times longer. When looking at the results for the best algorithm combination

(Fig.6.), same tendencies as before can be observed. The lowest delays are obtained by the same algorithms. There are slight differences for the worst outcome. The main part of worst combination is the FCFS algorithm along with the Priority Scheduling. It is not possible to differentiate which is worse-preemptive or non-preemptive version of algorithm.

3) *Conclusion*

Because of the randomness of data introduced to the system, it is hard to find the one optimal solution for the given problem. The best or the worst outcome can only be approximated to a smaller group of possible algorithms. Because of the exhaustive research, it was hard to truly examine the problem. This part of a research could be an introduction for the further study. It gives the overall view of the best and worst algorithms.

V. CONCLUSION

The research presented in this paper allowed for only ‘local’ conclusions. At this stage of investigation, we can conclude about the remarkable influence of aging parameter on the system’s time of processes execution and initially recommend combinations with SW for open pool of tasks, and not recommend combinations with RR from the minimizing average delay point of view.

The created ‘Simulator Pro 3’ is being used as teaching and research tool in Electronics Faculty, Wroclaw University of Technology, Poland. It still gives an opportunity for further investigations. In the nearer future, the various scenarios with Priority Scheduling is planned to be applied. What is more, as it was described in Section IV, the problem of open pool of tasks is very complex and needs far more investigation to fully explore it.

The authors of this paper are planning to perform studies with the several other scheduling algorithms, e.g., based on evolutionary ideas, which were proven their efficiency in [9] and [10], and in own works [11] and [12].

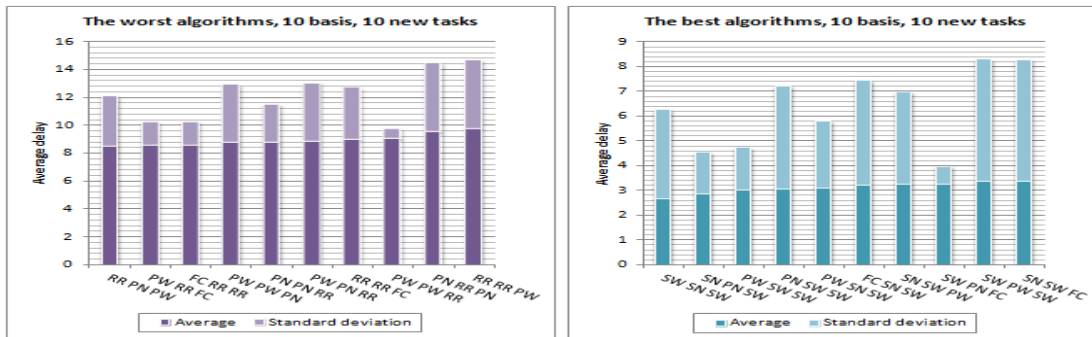


Figure 5. The worst and the best outcome for 10 basic and 10 new tasks set.



Figure 6. The worst and the best outcome for 10 basic and 20 new tasks set.

ACKNOWLEDGEMENT

This work was supported by the statutory funds of the Department of Systems and Computer Networks, Faculty of Electronics, Wrocław University of Technology, No S20010W4.

REFERENCES

- [1] A. Silberschatz, J. L. Peterson, and P. B. Galvin, Operating System Concepts, WNT, Warsaw, 2006.
- [2] M. Respondek, "Trial testing efficiency of tasks execution in three-processor system", Technical Report, Faculty of Electronics W4K2, Wrocław University of Technology, 2011.
- [3] M. Respondek, "Evaluation of algorithms for tasks execution in three-processor system with priority scheduling.", Research Report, Faculty of Electronics W4K2, Wrocław University of Technology, 2012.
- [4] B. Czajka, S. Zagorski, and L. Koszalka, "Three-processor computer simulation. The shortest tasks execution optimization", Research Report, Faculty of Electronics W4K2, Wrocław University of Technology, 2006.
- [5] B. Czajka and I. Pozniak-Koszalka, "Evaluation of tasks scheduling algorithms in multi-core and multi-queuing environments using system MESMS2", Proceedings to IARIA ICONS, IEEE CPS,2009, pp. 17-22.
- [6] Process management, Multiprocessor systems. Available on: <http://siber.cankaya.edu.tr> [retrieved: November, 2011].
- [7] F. Ramming, "Real time operating systems", University Paderborn, Heinz Nixdorf Institute, SBCCI'01. Available on <http://www.inf.ufrgs.br/~flavio/ensino/cmp502/TutorialRammig> [retrieved: August, 2012].
- [8] E .O. Oyetunji and A. E. Oluleye, "Performance assessment of some CPU scheduling algorithms", Research Journal of Information Technology, Vol. 1, August 2009, pp. 22-26.
- [9] V. Gaba and A. Prasha,, "Comparison of processor scheduling algorithms using genetic approach", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 2, Issue 8, August 2012, pp. 37-45.
- [10] D. M. Zydek and H. Selvaraj, "Fast and efficient processors allocation algorithm for torus-based chip multiprocessors," Journal of Computers & Electrical Engineering, Vol. 37, Issue 1, January 2011, pp. 91-105.
- [11] D. Krol, D. Zydek, and L. Koszalka, "Problem independent approach to multiprocessor dependent task scheduling", Journal of Electronics and Telecommunications, Vol. 58, Issue 4, 2012, pp. 369-381.
- [12] I. Poźniak-Koszalka, W. Proma, L. Koszalka, M. Pol, and A. Kasprzak, "Task allocation in mesh structure: 2Side leapFrog Algorithm and Q-learning based Algorithm", Lecture Notes in Computer Science, Springer, Vol. 7336, 2012, pp. 576-587.



# Preprocessing of Binary Executable Files Towards Retargetable Decompilation

Jakub Křoustek, Dušan Kolář  
Faculty of Information Technology, IT4Innovations Centre of Excellence  
Brno University of Technology  
Brno, Czech Republic  
{ikroustek, kolar}@fit.vutbr.cz

**Abstract**—The goal of retargetable machine-code decompilation is to analyze and reversely translate platform-dependent executable files into a high level language (HLL) representation. This process can be used for many different purposes, such as legacy code reengineering, malware analysis, etc. Retargetable decompilation is a complex task that must deal with a lot of different platform-specific features and missing information. Moreover, input files are often compressed or protected from any kind of analysis (up to 80% of malware samples). Therefore, accurate preprocessing of input files is one of the necessary prerequisites in order to achieve the best results. This paper presents a concept of a generic preprocessing system that consists of a precise signature-based compiler and packer detector, plugin-based unpacker, and converter into an internal platform-independent file format. This approach has been adopted and tested in an existing retargetable decompiler. According to our experimental results, the proposed retargetable solution is fully competitive with existing platform-dependent tools.

**Keywords**—reverse engineering, decompilation, packer detection, unpacking, executable file, Lissom

## I. INTRODUCTION

Reverse engineering is used often as an initial phase of a reengineering process. As an example we can mention reengineering of legacy software to operate on new computing platforms. One of the typical reverse-engineering tools is a machine-code decompiler, which reversely translates binary executable files back into an HLL representation, see [1], [2] for more details. This tool can be used for binary code migration, malware analysis, source code reconstruction, etc.

More attention is paid to retargetable decompilation in recent years. The goal is to create a tool capable to decompile applications independent of their origin into a uniform code representation. Therefore, it must handle different target architectures, operating systems, programming languages, and their compilers. Moreover, applications can be also packed or protected by so-called *packers* or *protectors*. This is a typical case of malware. Therefore, such input must be *unpacked* before it is further analyzed; otherwise, its decompilation will be inaccurate or impossible at all. Note: in the following text, we use the term *packing* for all the techniques of executable file creation, such as compilation, compression, protection, etc.

In order to achieve retargetable decompilation, its preprocessing phase is crucial because it eliminates most of the platform-specific differences. For example, this phase is responsible for a precise analysis of an input application (e.g., detection of a target platform). Whenever a presence of a packed code is detected, such application has to be unpacked.

Furthermore, the platform-dependent object file format (OFF) is converted into an internal uniform code representation. The final task of preprocessing is an information gathering, such as detection of originally used programming language, compiler, or its version. This information is valuable during the following phases of decompilation because different languages and compilers use different features and generate unique code constructions; therefore, such knowledge implies more accurate decompilation.

In this paper, we present several platform-independent preprocessing methods, such as language and compiler detection, executable file unpacking, and conversion. These methods were successfully interconnected, implemented, and tested in a preprocessing phase of an existing retargetable decompiler developed within the Lissom project [3].

The paper is organized as follows. Section II discusses the related work of executable file preprocessing. Then, we briefly describe the retargetable decompiler developed within the Lissom project in Section III. In Section IV, we give a motivation for a compiler and packer detection within decompilation. Afterwards, our own methods used in the preprocessing phase are presented in Section V. Experimental results are given in Section VI. Section VII closes the paper by discussing future research.

## II. RELATED WORK

There are several studies and tools focused on binary executable file analysis and transformation. Most of them are not focused directly on decompilation but some of these ideas can be applied in this field. Their major limitation for such usage is their bounding to one particular target platform.

In this section, we briefly mention several existing tools used for packer detection, unpacking, and OFF conversion.

### A. Compiler and Packer Detection

The knowledge of the originally used tool (e.g., compiler, linker, packer) for executable creation is useful in several security-oriented areas, such as anti-virus or forensics software [4]. Overwhelming majority of existing tools are limited to the Windows Portable Executable (WinPE) format on the Intel x86 architecture and they use signature-based detection. Almost all of these tools are freeware but not open source.

Formats of signatures used by these tools for pattern matching usually contain a hexadecimal representation of the first few machine-code instructions on the application's entry point (EP). EP is an address of the first executed instruction

within the application. A sequence of these first few instructions creates a so-called *start-up* or *runtime* routine, which is quite unique for each compiler or packer and it can be used as its footprint. Accuracy of detection depends on the signature format, their quality, and used scanning algorithm. Identification of sophisticated packers may need more than one signature.

Databases with signatures are either internal (i.e., pre-compiled in code of a detector), or stored in external files as a plain text. The second ones are more readable and users can easily add new signatures. However, detection based on external signatures is slower because they must be parsed at first. Some detection tools are distributed together with large, third-party external databases.

### B. Unpacking

Binary executable file packing is done for one of these reasons—code compression, code protection, or their combination. The idea of code compression is to minimize the size of distributed files. Roughly speaking, it is done by compressing the file's content (i.e., code, data, symbol tables) and its decompression into memory or into a temporal file during execution.

Code protection can be done by a wide range of techniques (e.g., anti-debugging, anti-dumping, insertion of self-modifying code, interpretation of code in internal virtual machine). It is primarily used on MS Windows but support of other platforms is on arise in the last years (e.g., gzexe and ElfCrypt for Linux, VMProtect for Mac OS X, multi-platform UPX and HASP).

Packers are proclaimed to be used for securing commercial code from cracking; however, they are massively abused by malware authors to avoid anti-virus detection. Decompilation of compressed or protected code is practically impossible, mainly because it is “just” a static code analysis and unpacking is done during the runtime. Therefore, it is crucial to solve this issue in order to support decompilation of this kind of code.

UPX is a rare case of packers because it also supports decompression. Unpacking is a very popular discipline of reverse engineering and we can find tools for unpacking many versions of all popular packers (e.g., ASPackDie, tEunlock, UnArmadillo). We can also find unpacking scripts for popular debuggers, like OllyDbg, which do the same job.

Currently, about 80% to 90% of malware is packed [5] and about 10 to 15 new packers are created from existing ones every month [6], more and more often using polymorphic code generators [7]. In past, there were several attempts to create generic unpackers (e.g., ProcDump, GUV32), but their results were less accurate than packer-specific tools. However, creation of single-purpose unpackers from scratch is a time consuming task. Once again, these unpacking techniques are developed primarily for MS Windows and other platforms are not covered.

### C. Object-File-Format Conversion

This part is responsible for converting platform-dependent file formats into an internal representation. We can find several existing projects focused on this task. They are used mostly for

OFF migration between two particular platforms and they were hand-coded by their authors just for this purpose. Therefore, they cannot be used for retargetable computing.

A typical example is the MAE project [8], which supports execution of Apple Macintosh applications on UNIX. Sun Microsystems Wabi [9] allows conversion of executables from Windows 3.x to Solaris. AT&T's FreePort Express is another binary translator of SunOS executables into the Digital UNIX format. More examples can be found in [10].

### III. LISSOM PROJECT'S RETARGETABLE DECOMPILER

The Lissom project's [3] retargetable decompiler aims to be independent on any particular target architecture, operating system, or OFF. It consists of two main parts—the preprocessing part and the decompilation core, see Figure 1. Its detailed description can be found in [11], [12].

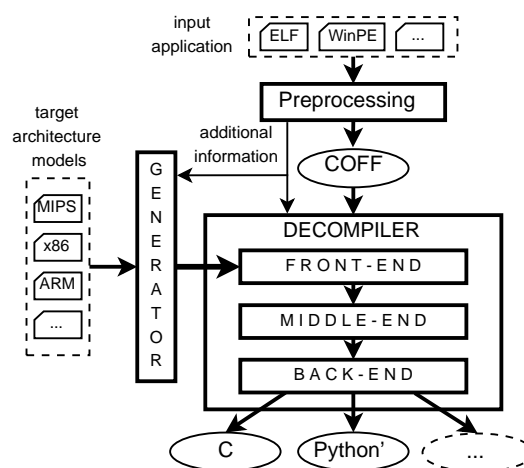


Fig. 1: The concept of the Lissom project's retargetable decompiler.

The preprocessing part is described in the following section. Basically, it unpacks and unifies examined platform-dependent applications into an internal Common-Object-File-Format (COFF)-based representation.

Afterwards, such COFF-files are processed in the decompilation core, which is partially automatically generated based on the description of target architecture. This decompilation phase is responsible for decoding of machine-code instructions, their static analysis, recovery of HLL constructions (e.g., loops, functions), and generation of the target HLL code. Currently, the C language and a Python-like language are used for this purpose and the decompiler supports decompilation of MIPS, ARM, and x86 executables.

### IV. MOTIVATION

The information about the originally used compiler is valuable during the decompilation process because each compiler generates a quite unique code in some cases; therefore, such knowledge may increase a quality of the decompilation results. One of such cases are so-called instruction idioms. Instruction idiom represents an easy-to-read statement of the HLL code that is transformed by a compiler into one or more machine-code instructions, which behavior is not obvious at the first sight. See [13] for an exhausting list of the existing idioms.

We illustrate this situation on an example depicted as a C language code in Figure 2. This program uses an arithmetical expression “ $-(a \geq 0)$ ”, which is evaluated as 0 whenever the variable  $a$  is smaller than zero; otherwise, the result is evaluated as  $-1$ . Note: the following examples are independent on the used optimization level within the presented compilers. All compilers generate 32-bit Linux ELF executable files for Intel x86 architecture [14] and the assembly code listings were retrieved via `objdump` utility.

```
#include <stdio.h>

int main(int argc, char **argv)
{
    int a;

    scanf("%d", &a);
    // Prints - "0" if the input is smaller than 0
    //       - "-1" otherwise
    printf("%d\n", -(a >= 0));

    return 0;
}
```

Fig. 2: Source code in C.

Several compilers substitute code described in Figure 2 by instruction idioms. Moreover, different compilers generate different idioms. Therefore, it is necessary to distinguish between them. For example, code generated by the GNU compiler GCC version 4.0.4 [15] is depicted in Figure 3. As we can see, the used idiom is non-trivial and its readability is far from the original expression.

```
Address  Hex dump  Intel x86 instruction
-----
; scanf
; Variable 'a' is stored in %eax
80483e2:  f7 d0      not  %eax
80483e4:  c1 e8 1f   shr  $31,%eax
80483e7:  f7 d8      neg  %eax
; Print result stored in %eax
; printf
```

Fig. 3: Assembly code generated by gcc 4.0.4.

The Clang compiler is developed within the LLVM project [16], [17]. Output of this compiler is illustrated in Figure 4. As we can see, Clang uses idiom, which is twice as long as the previous one and it is assembled by the different set of instructions. Therefore, it is not possible to implement one generic decompilation analysis. Such solution will be inaccurate and slow (i.e., detection of all existing idioms no matter on the originally used compiler).

Decompilation of instruction idioms (or other similar constructions) produces a correct code; however, without any compiler-specific analysis, this code is hard to read by a human because it is more similar to a machine-code representation than to the original HLL code. Compiler-specific analyses are focused on these issues (e.g., they detect and transform idioms back to a well-readable representation), but the knowledge of the originally used compiler and its version is mandatory.

Figure 5 depicts decompilation results for the gcc compiled code listed in Figure 3 (i.e., code generated by gcc 4.0.4). The Lissom retargetable decompiler was used for this task. As we

```
Address  Hex dump  Intel x86 instruction
-----
; scanf
; Variable 'a' is stored on stack at -16(%ebp)
8013bf:  83 7d f0 00  cmpl $0,-16(%ebp)
8013c3:  0f 9d c2     setge %dl
8013c6:  80 e2 01     and  $1,%dl
8013c9:  0f b6 f2     movzbl %dl,%esi
8013cc:  bf 00 00 00 00  mov  $0,%edi
8013d1:  29 f7       sub  %esi,%edi
8013d3:  ;...
8013d6:  89 7c 24 04  mov  %edi,4(%esp)
; Print result stored on stack at 4(%esp)
; printf
```

Fig. 4: Assembly code generated by clang 3.1.

can see, the expression contains bitwise shift and xor operators instead of the originally used comparison operator. This makes the decompiled code hard to read.

```
#include <stdint.h>
#include <stdio.h>

int main(int argc, char **argv)
{
    int apple;
    apple = 0;
    scanf("%d", &apple);
    printf("%d\n", -(apple >> 31 ^ 1));
    return 0;
}
```

Fig. 5: Decompiled source code (without compiler-specific analyses).

Furthermore, it is important to detect compiler version too. In Figure 6, we illustrate that the different versions of the same compiler generate different code for the same expression. We use gcc version 3.4.6 and the C code from the Figure 2.

```
Address  Hex dump  Intel x86 instruction
-----
; scanf
; Variable 'a' is stored on stack at -4(%ebp)
80483f3:  83 7d fc 00  cmpl $0,-4(%ebp)
80483f7:  78 09       js   8048402
80483f9:  c7 45 f8 ff ff ff  movl $-1,-8(%ebp)
8048400:  eb 07       jmp 8048409
8048402:  c7 45 f8 00 00 00 00  movl $0,-8(%ebp)
8048409:
; Print result stored on stack at -8(%ebp)
; printf
```

Fig. 6: Assembly code generated by gcc 3.4.6.

In this assembly code snippet, we can see that no instruction idiom was used. The code simply compares the value of a variable with zero and sets the result in a human-readable form. It is clear that the difference between the code generated by the older (Figure 6) and the newer version (Figure 3) of this compiler is significant. Therefore, we can close this section stating that information about the used compiler and its version is important for decompilation.

## V. PREPROCESSING PHASE OF THE RETARGETABLE DECOMPILER

In this section, we present a design of the preprocessing phase within the Lissom project retargetable decompiler. The

complete overview is depicted in Figure 7. The concept consists of the following parts.

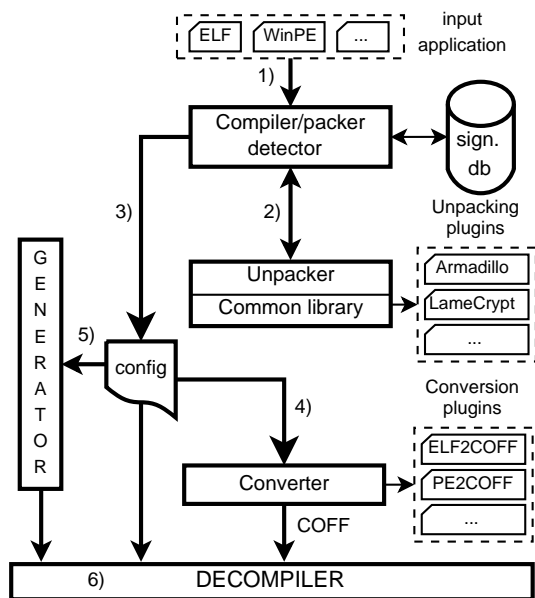


Fig. 7: The concept of the preprocessing phase.

At first, the input executable file is analyzed and the used OFF is detected. All common formats are supported (e.g., WinPE, UNIX ELF, Mach-O). Information about the target processor architecture is extracted from the OFF header (e.g., `e_machine` entry in ELF OFF) and it is used together with other essential information in further steps.

The next part of this step is a detection of a tool used for executable creation. This is done using a signature-based detection of start-up code as described in Section II. Example of such a start-up code can be seen in Figure 8. Signature for this code snippet is “5589E583EC18C704240100000FF15-----E8”, where each character represents a nibble of instruction’s encoding. All variable parts must be skipped during matching by a wild-card character “-”, e.g., a target address in the `call` instruction. This signature format is quite similar to formats used by other detectors listed in Section II.

Address	Hex dump	Intel x86 instruction
0040126c:	55	<b>push %ebp</b>
0040126d:	89e5	<b>mov %esp, %ebp</b>
0040126f:	83ec18	<b>sub \$0x18, %esp</b>
00401272:	c7042401000000	<b>movl \$0x1, (%esp)</b>
00401279:	ff1500000000	<b>call *0x0</b>
0040127f:	e8	...

Fig. 8: Start-up code for MinGW gcc v4.6 on x86 (`crt2.o`) generated by `objdump -d`.

Our signature format also supports two new features—description of nibble sequences with zero or more occurrences and description of unconditional short jumps. Example of the former one is “(90)”, denoting an optional sequence of `nop` instructions for x86 architecture. Example for the second one is “#EB”, denoting an unconditional short jump for the

same architecture, which size is specified in the next byte; everything between the jump and its destination is skipped. In Figure 9, we can find a code snippet covered by signature “#EB(90)40”.

Address	Hex dump	Intel x86 instruction
00401000:	eb 02	<b>jmp short &lt;00401004&gt;</b>
00401002:	xx xx	<i>; don't care</i>
00401004:	90	<b>nop</b>
00401005:	90	<b>nop</b>
00401006:	40	<b>inc %eax</b>

Fig. 9: Example of advanced signature format.

These features come handy especially for polymorphic packers [7] producing a large number of different start-up codes (e.g., Obsidium packer). Describing one version of such packer usually needs dozens of classical signatures. However, this number can be significantly reduced using the above-mentioned features.

Signatures within our internal database were created with focus on the detection of the packer’s version. This information is valuable for decompilation because two different versions of the same packer may produce diverse code constructions. The database also contains signatures for non-WinPE platforms; therefore, it is not limited like other tools. Finally, new signatures can be automatically created whenever the user can provide at least two files generated by the same version of packer. Presence of multiple files is mandatory in order to find all variable nibbles in the start-up code.

Whenever a usage of packer was detected in the first phase, the unpacking part is invoked. Unpacking is done by our own generic unpacker, which consists of a common unpacking library and several plugins implementing unpacking of particular packers. The common library contains the necessary functions for rapid unpacker creation, such as detection of the original entry point (OEP), dump of memory, fixing import tables, etc. Therefore, a plugin itself is very tiny and contains only code specific to a particular packer.

A plugin can be created in two different ways: either it can reverse all the techniques used by the packer and produce the original file, or the plugin can execute the packed file, wait for its decompression, and dump its unprotected version from memory to file. The first one is hard to create because it takes a lot of time to analyze all the used protection techniques. Its advantage is that unpacking can be done on any platform because the file is not being executed. That is the main disadvantage of the second approach. Such a plugin can be created quickly; however, it must be executed on the same target platform. In present, we support unpacking of several popular packers like Armadillo, UPX (Linux and Windows), NoodleCrypt and others in the second way. See Section VII for its future research.

After unpacking, the re-generated executable file is once more analyzed. In rare cases, second packer was used and we need to unpack this file once more. Otherwise, the analysis will try to detect the used compiler and its version, and generate a configuration file, which is used by other decompilation tools. This configuration file also contains information about

the target architecture, endianness, bitwidth, address of OEP, etc.

Afterwards, the platform-specific unpacked executable file is converted into an internal COFF-based representation. The converter is also realized in a plugin-based way and each plugin converts one particular OFF. Currently, we support ELF, WinPE, Mach-O, and several others OFF. See [10] for more details about this tool.

Using the information about the target architecture in the configuration file, the instruction decoder is automatically created by the generator tool [12]. Instruction decoder is the first part of the decompiler's front-end, which translates machine code instructions into a semantics description of their behavior.

Finally, the COFF executable file is processed in the generated decompiler according to the configuration file. Using the provided information about used compiler, it can selectively enable compiler-specific analyses (e.g., detection of instruction idioms, recovery of functions).

## VI. EXPERIMENTAL RESULTS

This section contains an evaluation of the previously described method of packer detection. The accuracy of our tool (labeled as "Lissom") is compared with the latest versions of existing detectors. Their short overview is depicted in Table I.

TABLE I: Overview of existing compiler/packer detection tools.

tool		signatures		
name	version	internal	external	total
Lissom	1.00	2181	0	2181
RDG Packer Detector [18]	0.6.9	?	10	?
ProtectionID (PID) [19]	0.6.4.0	499	0	499
Exeinfo PE [20]	0.0.3.2	667	7075	7742
Detection is Easy (DiE) [21]	0.6.4	?	1870	?
NtCore PE Detective [22]	1.2.1.1	0	2806	2806
FastScanner [23]	3.0	1605	1832	3437
PEiD [24]	0.95	672	1774	2446

All of these detection tools use the same approach as our solution—detection using signature matching. As we can see in Table I, most of them use a combination of pre-compiled internal signatures and a large external database created by the user community. The competitive solutions are limited to WinPE OFF and a number of their signatures varies between hundreds and thousands. The number of internal signatures is not always absolutely precise because some authors do not specify this number, like RDG or FastScanner. Therefore, we had to analyze such applications and try to find their databases manually (e.g., using reverse engineering). We were unable to find it in the RDG and DiE detectors. Our solution consists of 2181 internal signatures for all supported OFFs and we also support the concept of external signatures.

Using reverse engineering, we also figured out that several tools (e.g., PEiD) use additional heuristic technique for packer detection. These techniques are not focused on the start-up code or machine code at all. They are analysing several properties of the executable file (e.g., attributes of sections, information stored within file header) and they perform a detection of packer-specific behavior. Using this heuristic analysis, it is possible to detect even the polymorphic packers like Morphine encryptor.

Twenty WinPE packers (e.g., ASPack, FSG, Obsidium, UPX) and several their versions (105 different tools in total) were used for comparison of previously mentioned detectors. We used these packers for packing several compiler-generated executables—with different size (50kB to 5MB), used compiler, compilation options, and packer options. The purpose is that some packers create different start-up code based on the file size and characteristics (data-section size, PE Header flags, etc.). The test set consists of 5267 executable files in total. We prepared three test cases for the evaluation of the proposed solution.

At first, we evaluated the detection of packer's name. This type of detection is the most common and also the easiest to implement because generic signatures can be applied (i.e., signatures with only few fixed nibbles describing complete packer family). On the other hand, this information is critical for the complete decompilation process because if we are unable to detect usage of executable-file protector, the decompilation results will be highly inaccurate. The results of detection are compared in Figure 10.

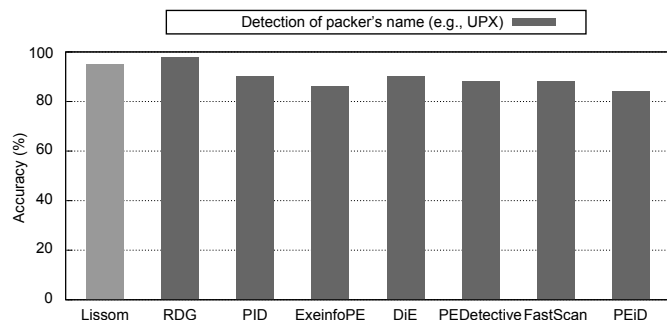


Fig. 10: Summary of packer detection (packer names).

According to the results, the RDG [18] detector has the best ratio of packer's name detection (98%), while our solution was second with ratio over 95%. All other solutions achieved comparable results—between 80% and 90%. We can also notice that larger signature databases do not imply better results in this category (e.g., Exeinfo PE). Such large databases are hard to maintain and they can produce several false-positive results because of too much generic signatures.

Afterwards, we tested the accuracy of tool's major version detection. In other words, this test case was focused on tool's ability to distinguish between two generations of the same tool (e.g., UPX v2 and UPX v3). This feature comes handy in the front-end phase during compiler specific analyses. For example, the compiler may use in its newer versions more aggressive optimizations that have a very specific meaning and they need a special attention by the decompiler (e.g., instruction idioms, loops transformation, jump tables), see Section IV for details. The results are depicted in Figure 11.

Within this test case, RDG and our solution once again achieved the best results (both scored 93%). Only ExeinfoPE and ProtectionID exceeded 80% success ration from the others.

Finally, we tested the ratio of precise packer's version detection. This task is the most challenging because it is necessary to create one signature for each particular version of each particular packer. This information is crucial for the unpacker because the unpacking algorithms are usually created

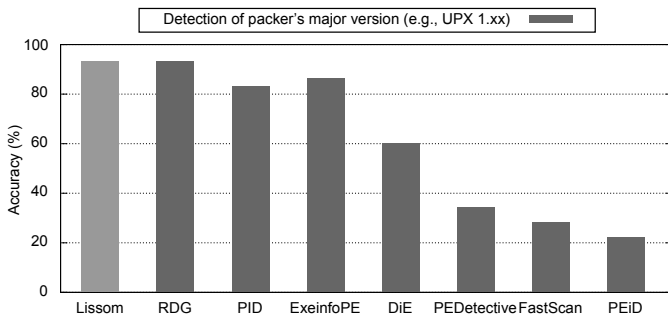


Fig. 11: Summary of packer detection (packer versions).

for one particular packer version and their incorrect usage may lead to a decompilation failure.

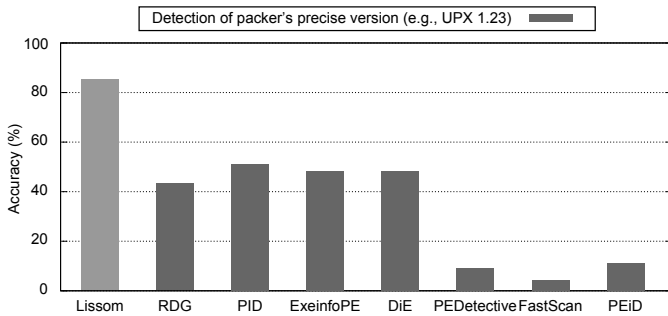


Fig. 12: Summary of packer detection (detailed detection).

Based on the results depicted in Figure 12, our detector achieved the best results in this category with 86% accuracy. The results of other solutions were much lower (51% at most). This is mainly because we focus primarily on detecting the precise version and we also support search in the entire PE file and its overlay and not just on its entry point.

## VII. CONCLUSION

This paper was aimed on architecture-independent preprocessing methods used within the existing retargetable decompiler. We introduced methods of packer detection, unpacking, and OFF conversion. Up to now, this concept has been successfully tested on the MIPS, ARM, and x86 architectures within the Lissom project's [3] retargetable decompiler.

We made several tests focused on accuracy of our solution and according to the experimental results, it can be seen that our concept is fully competitive with other existing tools.

We close the paper by proposing two areas for future research. (1) The unpacking phase can be enhanced by using retargetable simulators [25]. Such tools can emulate the target host system and, therefore, it will not be necessary to unpack executables on the same system as its origin. (2) We can further increase decompilation results by creation of new signatures and compiler-specific analyses (e.g., better loop statement recovery, detecting different types of function calls).

## ACKNOWLEDGMENTS

This work was supported by the project TA ČR TA01010667 System for Support of Platform Independent Malware Analysis in Executable Files, BUT FIT grant FIT-S-11-2 and FEKT-FIT-J-13-2000 Validation of Executable Code

for Industrial Automation Devices using Decompilation, by the project CEZ MSM0021630528 Security-Oriented Research in Information Technology, and by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070).

## REFERENCES

- [1] C. Cifuentes, "Reverse compilation techniques," Ph.D. dissertation, School of Computing Science, Queensland University of Technology, Brisbane, AU-QLD, 1994.
- [2] M. J. V. Emmerik, "Static single assignment for decompilation," Ph.D. dissertation, University of Queensland, Brisbane, AU-QLD, 2007.
- [3] Lissom, 2013, available on URL: <http://www.fit.vutbr.cz/research/groups/lissom/> [retrieved: 2013-04-22].
- [4] G. Taha, "Counterattacking the packers," in AVAR, 2007.
- [5] M. M. T. Brosch, "Runtime packers: The hidden problem?" in Black Hat, 2006.
- [6] K. Babar and F. Khalid, "Generic unpacking techniques," in 2nd International Conference on Computer, Control and Communication, 2009, pp. 1–6.
- [7] Y. Song, M. E. Locasto, A. Stavrou, A. D. Keromytis, and S. J. Stolfo, "On the infeasibility of modeling polymorphic shellcode," in 14th ACM Conference on Computer and Communications Security (CCS'07). ACM, 2007, pp. 541–551.
- [8] Apple Inc., "Macintosh application environment," 1994, available on URL: <http://www.mae.apple.com/> [retrieved: 2013-04-22].
- [9] D. R. P. Hohensee, M. Myszewski, "Wabi cpu emulation," Hot Chips VIII, 1996.
- [10] J. Křoustek, P. Matula, and L. Ďurfina, "Generic plugin-based convertor of executable file formats and its usage in retargetable decompilation," in 6th International Scientific and Technical Conference (CSIT'2011). Lviv Polytechnic National University, 2011, pp. 127–130.
- [11] L. Ďurfina et al., "Design of a retargetable decompiler for a static platform-independent malware analysis," International Journal of Security and Its Applications, vol. 5, no. 4, 2011, pp. 91–106.
- [12] L. Ďurfina, J. Křoustek, P. Zemek, and B. Kábele, "Detection and recovery of functions and their arguments in a retargetable decompiler," in 19th Working Conference on Reverse Engineering (WCRE'12). Kingston, Ontario, CA: IEEE Computer Society, 2012, pp. 51–60.
- [13] H. Warren, Hacker's Delight. Boston: Addison-Wesley, 2003.
- [14] Intel Corporation, "Intel 64 and ia-32 architectures software developer's manual volume 1: Basic architecture," 2011.
- [15] GNU Compiler Collection, 2013, available on URL: <http://gcc.gnu.org/> [retrieved: 2013-04-22].
- [16] Clang, 2013, available on URL: <http://clang.llvm.org/> [retrieved: 2013-04-22].
- [17] The LLVM Compiler System, 2013, available on URL: <http://llvm.org/> [retrieved: 2013-04-22].
- [18] RDG Packer Detector, 2013, available on URL: <http://rdgsoft.8k.com/> [retrieved: 2013-04-22].
- [19] ProtectionID, <http://pid.gamecopyworld.com/>, 2013, available on URL: [retrieved: 2013-04-22].
- [20] ExeinfoPE, 2013, available on URL: <http://www.exeinfo.xwp.pl/> [retrieved: 2013-04-22].
- [21] Die, 2013, available on URL: <http://hellspawn.nm.ru/> [retrieved: 2013-04-22].
- [22] NtCore PE Detective, 2013, available on URL: <http://www.ntcore.com/> [retrieved: 2013-04-22].
- [23] FastScanner, 2013, available on URL: <http://www.at4re.com/> [retrieved: 2013-04-22].
- [24] PEiD, 2013, available on URL: <http://www.peid.info/> [retrieved: 2013-04-22].
- [25] Z. Příkryl, "Advanced methods of microprocessor simulation," Ph.D. dissertation, Brno University of Technology, Faculty of Information Technology, 2011.

# EKOCA: Energy Aware Overlapping Multihop Clustering for Wireless Sensor Networks

Eman Ramadan, Moustafa A. Youssef, Magdy Abd-ElAzim Ahmed  
 Dept. of Computer and Systems Engineering  
 Alexandria University  
 Alexandria, Egypt  
 email: eman.ramadan@alexu.edu.eg, moustafa.youssef@ejust.edu.eg,  
 magdy\_aa@hotmail.com

Mohamed Nazih El-Derini  
 Dept. of Computer and Systems Engineering  
 Pharos University  
 Alexandria, Egypt  
 email: nazih.elderini@pua.edu.eg

**Abstract**—Among the algorithms developed for Wireless Sensor Networks is Overlapping Multihop Clustering. Overlapping clusters are useful in many network applications, such as intercluster routing, node localization and time synchronization protocols. In some environments, especially hostile, sensor nodes are always left unattended and there is no way to recharge them or exchange their batteries. Therefore, use of energy is a key issue in designing protocols for sensor networks in order to extend the network lifetime. In this work, we present a distributed energy aware overlapping multihop clustering algorithm based on the remaining energy of the sensors. Each node first elects itself as a cluster head according to a certain probability. During the next rounds, cluster head nodes select the new cluster head nodes based on the remaining energy of the sensor nodes in their clusters. The node with the highest remaining energy within a certain range from the cluster head node is elected to be the new cluster head node. The clustering process terminates in  $O(1)$  iterations and does not depend on the network topology or size. This algorithm is evaluated using NS2 Simulator. The proposed algorithm is intended to help in extending the life time of the network and balancing the energy consumption among different nodes by rotating the cluster head role.

**Keywords**—Clustering; energy aware clustering; multihop clustering; overlapping clustering; sensor networks

## I. INTRODUCTION

Wireless Sensor Networks (WSNs) are no longer a new technology and there are lots of applications used in our daily lives, in military and civilian domains. One of the challenging topics in designing an application in this field is energy-aware protocols.

Energy is a scarce resource for wireless sensor nodes. The energy consumed is divided into two categories either according to useful work or wasteful work. Energy consumed while transmitting/receiving data, querying requests and forwarding data is considered useful energy consumption. Since the communication between nodes is through a wireless medium, so there are collisions, which result in retransmissions of packets, also nodes can stay idle listening to the channel to send their data, in addition to the overhead of packet header even for small packets. These are all considered wasteful energy consumption. Energy consumption reduces network lifetime, which is defined as

the time elapsed until the first node, or a certain percentage of nodes [1], uses up its energy.

Many clustering techniques have been suggested in order to reduce the energy consumption by grouping some nodes into clusters. Clustering depends on nodes classification in which some nodes work as cluster heads being responsible for collecting data from other nodes in their clusters. Then, cluster heads can forward the data to the base station or aggregate it and send it as a single packet. Thus, aggregation reduces the overhead of data packets' headers. Clustering helps in reducing useful energy consumption by improving bandwidth utilization as well as reducing wasteful energy consumption by reducing overhead.

Among the algorithms proposed for Wireless Sensor Networks is K-hop Overlapping Clustering Algorithm for Wireless Sensor Networks (KOCA) [2]. It is a novel clustering algorithm aimed at generating overlapping multihop clusters. Overlapping clusters are useful in many network applications, such as intercluster routing [3], node localization [4][5][6][7] and time synchronization protocols [8].

Since nodes in a sensor network have limited energy, extending the network lifetime and improving scalability become essential. Cluster head nodes are supposed to consume more energy than other nodes due to the management procedure for cluster formation and also data aggregation from all nodes to be sent to the base station or gateway. After a certain time, cluster head nodes lose all their energy and then their clusters will stop working.

In this work, we present a distributed energy aware overlapping multihop clustering algorithm based on the remaining energy of the sensors. Each node first elects itself as a cluster head according to a certain probability. During the next rounds, cluster head nodes select the new cluster head nodes based on the remaining energy of the sensor nodes in the cluster. The node with the highest remaining energy within a certain range from the cluster head node is elected to be the new cluster head node.

The paper is structured as follows: Section II formulates the energy aware overlapping k-hop clustering problem. In Section III, we present the details of the EKOCA heuristic algorithm for solving the problem. Section IV provides sim-

ulation experiments for evaluating the EKOCA algorithm. In Section V, we present related work. Finally, Section VI provides the concluding remarks.

## II. EKOCA: ENERGY AWARE OVERLAPPING MULTIHOP CLUSTERING

In this section, we present the system model, definitions and notations and the problem formulation.

### A. System Model

Sensors will be randomly and uniformly deployed in the area of interest. Each node has a unique ID and is not equipped with GPS. All sensor nodes transmit packets at the same power so they all have the same transmission range. All communications are over a single shared wireless channel. A wireless link between two nodes is established if they are within the wireless transmission range of each other. Two nodes are immediate neighbors if they have a wireless link between each other and the two nodes are referred to as 1-hop. The communication environment is assumed to be contention-free and error-free, so no retransmission is required. Some of the nodes will be elected as Cluster Heads (CHs). The number of clusters is variable. Nodes are static which is a typical assumption for sensor networks. CH is a relaying element which sends the data received to the base station. There is a continuous reporting for the monitored event. The network is homogeneous which is also typical for sensor networks and commonly used in literature, thus all the nodes are alike (i.e., all having equal capacity in terms of computation, communication and power). CHs are picked from the deployed sensors. Multihop will be used for intracluster routing. The clustering process is distributed which means there are no base stations responsible for clusters formation and all nodes have to cooperate to make decisions.

### B. Definitions and Notations

- Network size ( $n$ ): The number of sensor nodes in the network. Since the deployment area is fixed so changing ( $n$ ) changes the node density.
- Cluster Radius ( $k$ ): The maximum distance between any node in the cluster and the cluster head node.
- Short Round Cluster Radius ( $m$ ): The maximum distance between any node in the cluster and the cluster head node while choosing the new cluster head node in the short round. The cluster head chooses the node with the highest residual energy to be the next cluster head from the nodes within this range.
- Cluster head probability ( $p$ ): The probability that a node can be a cluster head at the first round.
- Overlapping degree between two clusters: The number of nodes which are common between the two clusters.

### C. Problem Formulation

Energy aware overlapping clustering algorithm can be formulated as finding the set of cluster head nodes satisfying the following conditions:

- Coverage condition: This means that each node is either a cluster head or within  $k$ -hops from at least one cluster head.
- Overlapping condition: This means that for each cluster there is at least one cluster that overlaps with it and the overlapping degree between them is greater than a certain threshold.
- Connectivity condition: This means that during each short round, cluster head node is changed but no cluster formation is done and the nodes should communicate with the new cluster head smoothly without any problems.

## III. EKOCA HEURISTIC

There are three kinds of nodes in the clusters generated by EKOCA:

- Cluster Head (CH): A cluster head maintains the required information about the nodes in the cluster and the neighboring clusters and how to reach them.
- Normal Node: It is a member node in only one cluster.
- Boundary Node (BN): It is a member node in more than one cluster.

In this section, we discuss the necessary data structures maintained at each node. Then, we describe the cluster head selection process and the cluster membership.

### A. Data Structures

Each node maintains the following variables:

- 1) Node ID (NID): A unique ID assigned to each node before deploying the network.
- 2) Adjacent Clusters Table (AC\_table): A table maintained by CH nodes to store information about adjacent clusters. The table consists of tuples in the form (CHID, BNL, costL, oldCHID), where CHID is the CH node ID of adjacent cluster, BNL is a list of IDs of boundary nodes to reach this CH accompanied by the cost required to reach it and oldCHID is used to maintain the transition between the old cluster head and the new cluster head when no cluster reformation is required in short rounds. [see Table I]
- 3) Boundary Table (CH\_table): A table maintained by each node to store information about the clusters known to this node. If the table contains more than one entry, this means that the node is a boundary node; otherwise, it is a normal node. The table consists of tuples in the form (CHID, HC, prev, oldCHID), where CHID is the CH node ID, HC is the number of hops leading to this cluster head, prev is the node ID of a 1-hop neighbor node that can lead to this CH node



using minimum number of hops and oldCHID is used to maintain the transition between the old cluster head and the new cluster head when no cluster reformation is required in short rounds. [see Table II]

- 4) Nodes Table (nodes): A table maintained by CH node to store information about the nodes in its cluster.

### B. Cluster Head Selection

The main issue in the clustering process is to select a group of nodes to be cluster head nodes and these nodes then should gather other nodes to form different clusters. The cluster head node is supposed to consume more energy than the other nodes in the cluster for the management tasks it performs as well as aggregating data and sending it to the base station or the gateway. If this cluster head node lost all its energy, then the cluster would not be able to send the sensed data and this cluster will be separated from the whole network. Hence, we need to replace this cluster head node with another node so as not to let this cluster stop working. Rotating the role of CHs among nodes in the cluster can also be a means for fault-tolerance in addition to their load balancing advantage. The EKOCA clustering process is divided into three different types of rounds: First or Initial Round, Short Round and Long Round. The selection of the cluster head nodes differs in each of these rounds. We will now illustrate the differences in the selection process among these rounds.

This clustering algorithm is a distributed clustering which means that all nodes should cooperate to collectively specify the cluster heads and form overlapping clusters. There is no gateway or base station which has all information to decide how clusters will be formed.

1) *First Round:* This is the initial round that happens only once at the beginning of the clustering process. During this period each node generates a random number between zero and one and based on a certain probability, which is a parameter to the algorithm, this node will either be elected as a cluster head node or not. Then cluster head nodes will advertise themselves through an advertisement message (CH\_AD) to the sensors within their ranges. This message is then forwarded to all sensors that are no more than k-hops away from the cluster head which represents the cluster radius. This forwarding limitation is used to limit the flood of advertisement messages. A sensor that receives such advertisements joins the cluster even if it already belongs to another cluster. Since forwarding of the advertisement message is limited to k-hops, if a node did not receive any advertisement message within a reasonable time duration, it deduces it is not in the range of any cluster head and hence elects itself as a cluster head node and starts sending advertisement message. After this round, clusters are formed and nodes start sensing data and send these data to the cluster head. The cluster head then sends the data to the base station. After a certain time, the energy of the cluster head node is

consumed more than other nodes so we need to change the cluster head node through the short round.

Example for AC\_Table for CH node (n9) of cluster G:

TABLE I: EXAMPLE OF AC\_TABLE

CHID	oldCHID	BNL
n4	n7	(n1,5), (n2,4)
n6	n8	(n5,2), (n3,1)

Example for CH\_Table for node n3:

TABLE II: EXAMPLE OF CH\_TABLE

CHID	oldCHID	HC	prev
n9	n12	3	n10
n6	n8	1	n11

2) *Short Round:* In this round, each cluster head node sends an energy request (ENREQ) message to the nodes in the cluster within m-hops where m represents the short round cluster radius and it is less than or equal to k (the cluster radius). When a node receives this ENREQ message, it sends its current energy through an energy reply (ENREP) message to the cluster head node and forwards the ENREQ message to all sensors that are no more than m-hops. The current cluster head node compares the energies received from different nodes and selects the node with the highest residual energy to be the next cluster head node of the cluster. Thus, an election (ELECTSHRTRND) message is sent from the current cluster head node to the new elected node and this message contains the information maintained about the cluster to be stored in the new cluster head node. Upon receiving the election message, the new elected node updates its data structures and hence, sends a new advertisement (NewCH\_AD) message to all the nodes in the cluster in order to inform them that it is now the new cluster head node to which the sensed data shall be sent. This NewCH\_AD message is forwarded to all sensors that are no more than  $(k + \text{roundNum} * m)$  hops, as the selected node is now deviated from the center and k-hops are not enough to reach all the nodes in the cluster. During this round, no cluster reformation is performed, only the new cluster head maintains the cluster information from the old cluster head and the nodes update their data structure according to the new cluster head node. There can be one or more short rounds before the long round.

3) *Long Round:* In this round, the current cluster head node sends an energy request (ENREQ) message to all the nodes in the cluster. When a node receives this ENREQ message, it sends its current energy through an energy reply (ENREP) message to the cluster head node and forwards the ENREQ message to all sensors that are no more than  $(k + \text{maxNumberOfShortRounds} * m)$  hops. The current cluster head node compares the energies received from different nodes and selects the node with the highest residual energy

to be the next cluster head node of the cluster. An election (ELECTNDRND) message is then sent from the current cluster head node to the new elected node. Upon receiving the election message, clustering reformation is required and the new elected node starts advertising itself through an advertisement message (CH\_AD) to the sensors within its range. This message is then forwarded to all sensors that are no more than  $k$ -hops away from the cluster head which represents the cluster radius. There can be one or more short rounds before the long round. A cluster cycle is defined as one or more short rounds followed by a long round.

### C. Cluster Membership

Each node has a CH\_table which stores information about the cluster(s) it belongs to. Upon receiving a CH\_AD message, a node adds an entry to its CH\_table. The CH\_AD message's header includes SID, CHID, and HC, where SID is the sender node ID, CHID is cluster head ID, and HC is the number of hops leading to the CH node. The SID field is used to update the CH\_table.prev field for each node to know the path to the cluster head. The HC field is used to limit the flood of CH\_AD message to  $k$ -hops. Upon receiving a new CH\_AD message, a node will add an entry to its CH\_table. The SID field is used to update the CH\_Table.prev field. If another CH\_AD message was received from the same CHID, the HC is then compared to that stored in the table and if it was less than the old one, the HC and SID field are updated for the new shorter path to the CHID. Usually shorter path messages arrive faster but there may be a delay in the MAC layer. If a node has more than one entry in its CH\_Table, this means that this node is a boundary node. After the propagation of CH\_AD messages, each node then sends a join request (JREQ) message to each cluster it belongs to so that it becomes a member of this cluster. The JREQ message's header includes CHID, SID, RID, RSID,  $nc$  and  $(CHID, cost)_{1..nc}$ , where SID is the sender node ID, CHID is cluster head ID to join, RID is the receiver node ID in order to send this message to the CHID, its value is obtained from the CH\_Table.prev field of the entry of this CHID, RSID is the node ID which originally sends this JREQ to the CHID,  $nc$  is the number of clusters this node can hear from, which is equal to the size of the CH\_Table, and  $(CHID, cost)_{1..nc}$  represents the cluster(s) this node can hear from and the cost to reach the CHID of these clusters.

Each cluster head node maintains a list of all the member nodes in the cluster and also another list of all adjacent clusters. These adjacent clusters are known from the JREQ  $(CHID, cost)_{1..nc}$  field and store all the boundary nodes between the adjacent clusters and the hop count to reach the adjacent cluster head node. There can be more than one boundary node between overlapping clusters. Also the same node can be a boundary node for more than two overlapping clusters. Each CHID sets a timeout for receiving the JREQ messages which is enough for a message to be forwarded

along  $k$ -hops.

EKOCA terminates in  $O(k + c m)$  steps. Typically,  $k$ ,  $m$  and  $c$  are constants, so the clustering process terminates in a constant number of iterations regardless of the network size for all the different rounds.

## IV. PERFORMANCE EVALUATION

In this section, we describe the network settings used to evaluate the performance of EKOCA protocol using NS2 Simulator [9]. First, we define the simulation environment then the performance metrics used. This is followed by the experimental results and comparison with KOCA algorithm.

### A. Simulation Environment

Sensor nodes are deployed randomly over a flat square area of dimensions  $100 \times 100 m^2$ . Simulation parameters are shown in Table III.

The experiments consist of an initial round, followed by a 3 short rounds and then a long round. After adjusting the clusters for each round, data sensing and forwarding processes starts till the next coming round.

### B. Performance Metrics

We use the following performance metrics:

- 1) *Percentage of covered nodes*: defined in the first and long rounds as the percentage of nodes which became cluster heads or heard an advertisement message from a cluster head within  $k$ -hops in the first wave of cluster head advertisements.
- 2) *Communication overhead*: defined as the total number of messages transmitted.
- 3) *Time until the first node dies*: defined as the time elapsed until the first node in the network dies.
- 4) *Average Remaining Energy*.
- 5) *Number of nodes alive per certain time in seconds*.

### C. Results

Fig. 1 shows the effect of changing the cluster head probability  $p$  and the cluster radius  $k$  on EKOCA. Fig. 2 compares the performance of EKOCA vs. KOCA.

TABLE III: SIMULATION PARAMETERS

Parameters	Value
Cluster head probability ( $p$ )	0.05, 0.1, 0.15 and 0.2
Number of nodes	200
Deployment Area	$100 \times 100 m^2$
Cluster Radius ( $k$ )	2, 3, and 4
Short Round Cluster Radius (m)	1
Transmission range of each node	10m
Number of short and long Rounds	4
MAC Layer	TDMA
Propagation Model	Free Space model

1) *Effects of Parameters on Node Coverage:* Fig. 1a shows the percentage of nodes covered increases by increasing  $p$  as the number of cluster heads increases so there is more chance to cover all nodes. We can also notice that by increasing  $k$  the percentage of nodes covered increases as more nodes are reached and the nodes per cluster increases.

2) *Effects of Parameters on Communication Overhead:* Fig. 1b shows the communication overhead increases by increasing the cluster head probability  $p$  which leads to increasing the number of clusters which leads to the increase of the number of communication messages sent or forwarded. It also increases with increasing the cluster radius  $k$  due to increasing the number of nodes per cluster which leads to more communication overhead messages.

3) *Effects of Parameters on First Node Dies:* Fig. 1c shows the time when the first node dies decreases by increasing the cluster head probability  $p$  which leads to increasing the number of clusters and since they are overlapping clusters so the boundary nodes will be responsible for sending and forwarding more communication messages. This time also decreases with increasing the cluster radius  $k$  due to increasing the number of nodes per cluster which leads to more communication overhead messages.

4) *EKOCA vs. KOCA First Node Dies:* Fig. 2a shows that due to the rotation of the CH role, the time when the first node dies in EKOCA is more than that of the KOCA algorithm. We can also notice that this time decreases by increasing the cluster radius  $k$  due to increasing the cluster size which accordingly causes the increase of communication overhead.

5) *EKOCA vs. KOCA Average Remaining Energy:* Fig. 2b shows that the average remaining energy in EKOCA is higher than that of KOCA algorithm due to energy balance. We can also notice that the average remaining energy decreases by increasing the cluster radius  $k$  due to the increase in the cluster size which accordingly causes the increase of communication overhead.

6) *EKOCA vs. KOCA Number of Nodes Alive:* Fig. 2c shows that the number of nodes die in KOCA is higher than that of EKOCA and this gap increases as time passes due to the CH rotation and the balance in the energy consumption. Thus, we can notice that the nodes die slower in case of EKOCA algorithm. This extends the network lifetime more than KOCA.

7) *EKOCA vs. KOCA Communication Overhead:* The communication overhead of the EKOCA algorithm is higher than that of the KOCA algorithm and this is due to the energy request and reply messages sent or forwarded to determine the node with the highest remaining energy to be the new cluster head. This difference increases by increasing the cluster radius ( $k$ ) due to increasing the cluster size.

## V. RELATED WORK

There are many algorithms proposed for energy efficient clustering for wireless sensor networks either single hop or multihop for nonoverlapping clustering. Among these algorithms are [10][11][12][13]. (a) Energy Efficient Hierarchical Clustering (EEHC) [10]: Bandyopadhyay and Coyle proposed a distributed randomized clustering algorithm for WSNs with the objective of maximizing the network lifetime. Their technique depends on two stages: initial and extended. In the initial stage nodes form single hop clusters and in the extended stage they build  $h$  levels of clustering hierarchy which is recursively repeated to form additional tier. (b) Low Energy Adaptive Clustering Hierarchy (LEACH) [11]: Heinzelman et al proposed a distributed algorithm where some nodes elect themselves as CHs based on probability  $p$  and broadcast it for other nodes to join their clusters based on minimum communication energy between these nodes and CH. The role of being a CH is rotated periodically among the nodes of the cluster in order to balance the load. (c) Hybrid Energy-Efficient Distributed Clustering (HEED) [12]: is a distributed clustering algorithm where nodes are selected to be cluster heads based on energy and communication cost and other nodes go through several iterations until they find the CH they can communicate with using the least transmission power. (d) Distributed Weight-Based Energy-Efficient Hierarchical Clustering (DWEHC) [13]: Ding et al. proposed a distributed algorithm which differs from HEED in generating balanced cluster sizes and optimizing the intra-cluster topology. Nodes are selected to be CH based on their energy and proximity of their neighbors and other nodes try to join minimum cost reachable CH. Up to our knowledge KOCA was the first overlapping multihop clustering algorithm in which nodes can be members of more than one cluster at the same time, but it was not based on residual energy of sensor nodes while forming clusters.

## VI. CONCLUSION AND FUTURE WORK

In this paper, we presented EKOCA algorithm as an energy-aware overlapping multihop protocol and explained the behavior of the algorithm in short and long rounds. The performance of EKOCA was simulated using NS2 simulator. We studied the effect of different parameters on EKOCA and compared it with KOCA algorithm. The results show that EKOCA helps in extending the life time of the network and balancing the energy consumption among nodes by rotating the cluster head role. It also increases the time before the first node dies and as it may not be a CH, it will not prevent the cluster from working. This comes with an increase in total communication overhead due to control messages for re-clustering. This work can be extended by analyzing the protocol at different error rates and proposing a schedule for the time to turn different nodes ON and OFF.

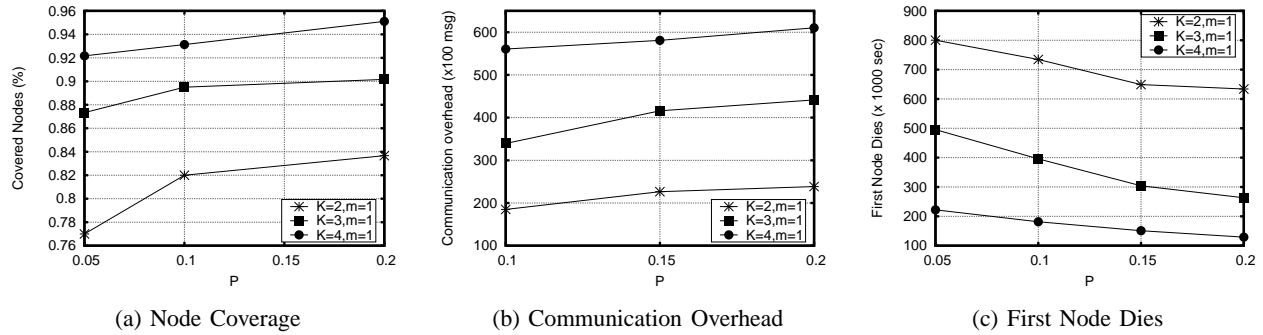


Figure 1: Effect of changing the parameters on the performance of EKOCA.

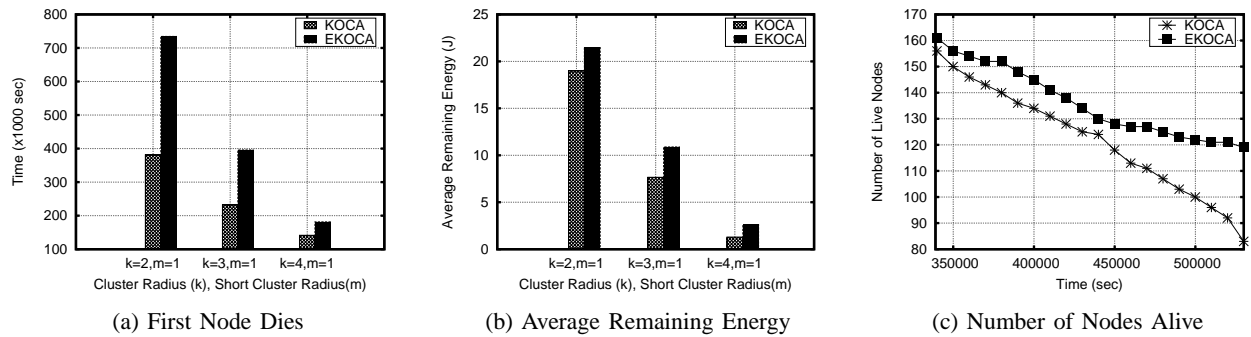


Figure 2: Comparing the performance of EKOCA and KOCA

REFERENCES

[1] D. M. Blough and P. Santi, "Investigating upper bounds on network lifetime extension for cell-based energy conservation techniques in stationary ad hoc networks," in Proceedings of the 8th annual international conference on Mobile computing and networking. ACM, 2002, pp. 183–192.

[2] M. Youssef, A. Youssef, and M. Younis, "Overlapping multihop clustering for wireless sensor networks," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 20, no. 12, Dec. 2009, pp. 1844–1856.

[3] P. Krishna, N. H. Vaidya, M. Chatterjee, and D. K. Pradhan, "A cluster-based approach for routing in dynamic networks," *ACM SIGCOMM Computer Communication Review*, vol. 27, no. 2, 1997, pp. 49–64.

[4] A. Youssef, "Salam: A scalable anchor-free localization algorithm for wireless sensor networks," PhD dissertation, Computer Science Dept., Univ. of Maryland, 2006.

[5] A. Youssef, A. Agrawala, and M. Younis, "Accurate anchor-free localization in wireless sensor networks," in Performance, Computing, and Communications Conference, 2005. IPCCC 2005. 24th IEEE International. IEEE, 2005, pp. 465–470.

[6] Y. Shang and W. Ruml, "Improved mds-based localization," in INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 4. IEEE, Mar. 2004, pp. 2640–2651.

[7] X. Ji and H. Zha, "Sensor positioning in wireless ad-hoc sensor networks with multidimensional scaling," in INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 4. IEEE, 2004, pp. 2652–2661.

[8] M. Mamun-Or-Rashid, C. Hong, and C.-H. In, "Passive clus-

ter based clock synchronization in sensor network," in Proc. Advanced Industrial Conf. Telecomm./Service Assurance with Partial and Intermittent Resources Conf./E-Learning on Telecomm. Workshop (AICT/SAPIR/ELETE '05), Jul. 2005, pp. 340–345.

[9] "The network simulator (ns2) <http://www.isi.edu/nsnam/ns/>," [accessed 17.06.2013].

[10] S. Bandyopadhyay and E. J. Coyle, "An energy efficient hierarchical clustering algorithm for wireless sensor networks," in INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies, vol. 3. IEEE, Apr. 2003, pp. 1713–1723.

[11] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on. IEEE, Jan. 2000, pp. 10–19.

[12] O. Younis and S. Fahmy, "Heed: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," vol. 3, no. 4, 2004, pp. 366–379.

[13] P. Ding, J. Holliday, and A. Celik, "Distributed energy-efficient hierarchical clustering for wireless sensor networks," in Distributed Computing in Sensor Systems. Springer, Jun. 2005, pp. 322–339.

# Quantum Key Distribution over Collective Amplitude Damping Quantum Channels

Elloá B. Guedes and Francisco M. de Assis

Institute for Studies in Quantum Computation and Information (IQuanta)  
 Post Graduation Programs in Electrical Engineering and Computer Science  
 Federal University of Campina Grande (UFCG)  
 Campina Grande – Paraíba – Brazil  
 Email: {elloaguedes, fmarassis}@gmail.com

**Abstract**—One of the most mature quantum information techniques nowadays is Quantum Key Distribution (QKD) in which two legitimate parties make use of a protocol to create a symmetric private key using a quantum channel. The quantum channel is not secure, since there may be an eavesdropper intercepting and re-sending the quantum states that are sent through it. One of the main problems in using QKD protocols is the existence of noise which can make difficult the task of eavesdropping checking. Considering these issues, this paper presents a QKD protocol over a collective amplitude damping quantum channel that makes use of decoherence-free subspaces and subsystems. The QKD protocol proposed is noiseless despite the errors existing in the quantum channel. Moreover, it makes the probability of the eavesdropper's retrieve the secret message negligible asymptotically. Besides, the probability of eavesdropper detection is stable during the whole communication which eases the eavesdropping checking procedures.

**Keywords**—Quantum Key Distribution; One-Time Pad; Decoherence-Free Subspaces and Subsystems.

## I. INTRODUCTION

The principles of Quantum Mechanics provide novel ways for quantum information transmission and processing, such as Quantum Computation and Quantum Communication. Regarding Quantum Communication, in particular, some intrinsic properties of Quantum Mechanics enable features that do not have a counterpart in Classical Communication, such as: (i) a qubit does not have not a definite value until the moment after it is read; (ii) every measurement in a qubit may disturb it; (iii) arbitrary states of qubits cannot be copied; (iv) qubits can be entangled; among others [1]. Thanks to these Quantum Mechanics principles, in certain scenarios, unconditional security can be achieved in information conveying through quantum channels.

The *Quantum Key Distribution* (QKD) [2]–[5] is one of the most mature quantum information techniques nowadays. According to QKD, two remote users can create a private key securely. This key is then used to crypt the secret message into a ciphertext through a classical cryptographic scheme such as the one-time pad, and the ciphertexts are then sent from one user to another through a classical channel. However, in a practical transmission process, the channel noise cannot be avoided completely. Noise can increase not only the error rate of the sending message, but also the difficulty of finding an eavesdropper in the process of a security check.

In order to avoid the noise, some QKD protocols [6], [7] considered the use of quantum channels which are subject to *collective decoherence*. In this scenario, all qubits which suffer

noise are affected exactly in the same way [8]. Considering this particularity, in such quantum channels it is possible to find some symmetries that protect the information from the noise. The states which remain unaffected by the decoherence compose a *decoherence-free subspace or subsystem* (DFS) [9].

Boileau et al. [6] proposed two QKD protocols using the DFS existing in the collective rotation quantum channel. The first protocol considers a subspace and the second a subsystem, both free of decoherence. Their protocol considers also the use of singlets and the encoding is based on the parity of qubits. Thanks to that, an uncertainty is inserted about the state originally sent from the perspective of the eavesdropper. However, it does not affect the legitimate parties of the protocol, enabling them to create a private key that can be later used to encrypt a classical message. It is important to emphasize that the eavesdropper is not able to affect the qubits exchanged, nor gather information about the key. Based on similar ideas, Li et al. [7] proposed two QKD protocols using DFS and considering the collective rotation and dephasing quantum channels.

The *amplitude damping* is a type of quantum noise which can make a qubit be lost. This type of error is also subject to collective decoherence, characterizing the *collective amplitude damping quantum channels*. Although these quantum channels have a DFS, no QKD protocols have been developed for them. Hence, the main objective of the present work is to characterize a QKD protocol over collective amplitude damping quantum channels, aiming at providing a secure way to create private keys between the legitimate parties despite the existence of an eavesdropper on the channel.

The present work is organized as follows. The decoherence-free subspaces and subsystems are characterized and exemplified in Section II. The collective amplitude damping quantum channels and the decoherence-free subspaces existing on their structure are shown in Section III. The model of communication considered as well as the steps that comprise the protocol proposed are shown in Section IV. An analysis of security is discussed in Section V. Lastly, final remarks and suggestions for future work are presented in Section VI.

*Notations and Conventions* – The Dirac notation [10] will be used to denote quantum states and operations over them throughout the paper. A quantum state is said to be pure if it can be represented by a unitary vector in the Hilbert space  $\mathcal{H}$ . The *Hadamard operation*, implemented by the gate  $H$ , has the following matricial representation  $H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ . The symbol  $\mathbb{1}$  denotes the *identity matrix*.

## II. DECOHERENCE-FREE SUBSPACES AND SUBSYSTEMS

Due to decoherence, a quantum system may begin to lose energy into the environment and decay to a ground state, its relative phase may be erased and, thus, the information it carries may be lost [11]. In this section, we will show how to avoid these undesired effects despite the existence of decoherence.

Let a closed quantum system be composed of the *system of interest*  $S$  defined on a Hilbert space  $\mathcal{H}$  and of the *environment*  $E$ . The Hamiltonian that describes this system is defined as follows:

$$\mathbb{H} = \mathbb{H}_S \otimes \mathbb{1}_E + \mathbb{1}_S \otimes \mathbb{H}_E + \mathbb{H}_{SE}, \quad (1)$$

where  $\mathbb{1}$  is the identity operator; and  $\mathbb{H}_S$ ,  $\mathbb{H}_E$  and  $\mathbb{H}_{SE}$  denote the Hamiltonians of system, environment and system-environment interaction, respectively.

In order to prevent errors, it would be ideal that  $\mathbb{H}_{SE}$  were equal to zero, indicating that system and environment are decoupled and evolve independently and unitarily under their respective Hamiltonians  $\mathbb{H}_S$  and  $\mathbb{H}_E$  [9]. However, in practical scenarios, such an ideal situation is not possible since no system is noiseless. So, after isolating a system to the best of our ability, we should aim for the realistic goals of the identification and correction of errors when they occur and/or avoiding noises when possible and/or suppressing noise in the system [12].

If some symmetries exist in the interaction between the system and the environment, it is possible to find a “quiet corner” in the system Hilbert space not experiencing decoherence. Let  $\{E_i(t)\}$  be a set of operators in the *operator-sum representation* (OSR) corresponding to the evolution of the system. We say that a system density matrix  $\rho_S$  is *invariant* under the OSR operators  $\{E_i(t)\}$  if  $\sum_i E_i(t)\rho_S E_i^\dagger(t) = \rho_S$ . We are now able to define the decoherence-free subspaces whose states are invariant despite a non-trivial coupling between the system and the environment.

**Definition 1** (Decoherence-Free Subspace). *A subspace  $\tilde{\mathcal{H}}$  of a Hilbert space  $\mathcal{H}$  is called decoherence-free with respect to a system-environment coupling if every pure state from this subspace is invariant under the corresponding OSR evolution for any possible environment initial condition:*

$$\sum_i E_i(t)|\tilde{k}\rangle\langle\tilde{k}|E_i^\dagger(t) = |\tilde{k}\rangle\langle\tilde{k}|, \forall |\tilde{k}\rangle\langle\tilde{k}| \in \tilde{\mathcal{H}}, \forall \rho_E(0). \quad (2)$$

Let the Hamiltonian of the system-environment interaction be  $\mathbb{H}_{SE} = \sum_j \mathbf{S}_j \otimes \mathbf{E}_j$ , where  $\mathbf{S}_j$  and  $\mathbf{E}_j$  are the system and environment operators, respectively. We consider that the environment operators  $\mathbf{E}_j$  are linearly independent. The symmetries required to define a decoherence-free subspace are described in the theorem below. For a detailed proof or different formulations, see [9, Section 5].

**Theorem 1** (Decoherence-Free Subspace Conditions). *A subspace  $\tilde{\mathcal{H}}$  is decoherence-free iff the system operators  $\mathbf{S}_j$  act proportional to the identity on the subspace:*

$$\mathbf{S}_j|\tilde{k}\rangle = c_j|\tilde{k}\rangle \quad \forall j, |\tilde{k}\rangle \in \tilde{\mathcal{H}}. \quad (3)$$

The notion of a subspace which remains decoherence-free throughout the evolution of a system is not, however, the most general method for providing decoherence-free encoding of information in a quantum system [9]. Knill et al. discovered a method for decoherence-free encoding into subsystems instead of into subspaces, which is presented below [13].

**Definition 2** (Decoherence-Free Subsystem). *Consider a decomposition of the whole Hilbert space  $\mathcal{H} = (\mathcal{H}^A \otimes \mathcal{H}^B) \oplus \mathcal{K}$ , where  $\dim(\mathcal{H}) = \dim(\mathcal{H}^A) \cdot \dim(\mathcal{H}^B) + \dim(\mathcal{K})$ . A subspace  $\mathcal{H}^B$  of the full Hilbert space is a decoherence-free subsystem if, for a quantum channel  $\mathcal{E}$ :*

$$\forall \rho^A, \forall \rho^B, \exists \tau^A : \mathcal{E}(\rho^A \otimes \rho^B) = \tau^A \otimes \rho^B, \quad (4)$$

where  $\rho^A, \tau^A \in \mathcal{B}(\mathcal{H}^A)$ , and  $\rho^B \in \mathcal{B}(\mathcal{H}^B)$ .

In fact,  $\mathcal{H}^B$  is said to encode a decoherence-free subsystem if (4) is satisfied. In particular, when  $\dim(\mathcal{H}^A) = 1$ ,  $\mathcal{H}^B$  is a decoherence-free subspace.

To make explicit the difference between decoherence-free subspaces and subsystems, consider the encoding of a generic qubit  $\alpha|0\rangle + \beta|1\rangle$  into  $\alpha|01\rangle + \beta|10\rangle$ . In this case, the information has been encoded into a *subspace* of the two qubit Hilbert space. Suppose now that the information is encoded only into the first qubit of the two qubits available, i.e.,  $\alpha|0\rangle + \beta|1\rangle \mapsto (\alpha|0\rangle + \beta|1\rangle) \otimes |\psi\rangle$ . Since this second encoding is a one-to-many mapping from the quantum information in one qubit to a two qubit Hilbert space, then it is said that the information has been encoded into a *subsystem*.

### A. Example

The *collective rotation quantum channel* acts on the input as follows:

$$|0\rangle \mapsto \cos\theta|0\rangle + \sin\theta|1\rangle, \quad (5)$$

$$|1\rangle \mapsto -\sin\theta|0\rangle + \cos\theta|1\rangle, \quad (6)$$

where  $\theta$  is the collective rotation parameter which fluctuates over time  $t$ . Two states that are immune to the decoherence caused by this quantum noisy channel are the following Bell states

$$|\beta_{00}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \quad (7)$$

$$|\beta_{11}\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle). \quad (8)$$

Despite being entangled, these states are distinguishable and can be properly obtained at the channel’s end using Bell measurements.

If one encodes a generic quantum state  $|\psi\rangle = a|0\rangle + b|1\rangle$  using the mentioned Bell states as logic qubits, i.e.,  $|\psi_L\rangle = a|\beta_{00}\rangle + b|\beta_{11}\rangle$ , we have that the resulting encoded state is protected from decoherence since the logic states are immune to the decoherence caused by the collective rotation quantum channel  $\mathcal{E}$  as follows:

$$\begin{aligned} \mathcal{E}(|\beta_{00}\rangle) &= \frac{1}{\sqrt{2}} [(\cos \theta|0\rangle + \sin \theta|1\rangle) \otimes (\cos \theta|0\rangle + \sin \theta|1\rangle) \\ &\quad + (-\sin \theta|0\rangle + \cos \theta|1\rangle) \otimes (-\sin \theta|0\rangle + \cos \theta|1\rangle)] \\ &= |\beta_{00}\rangle, \end{aligned} \quad (9)$$

and

$$\begin{aligned} \mathcal{E}(|\beta_{11}\rangle) &= \frac{1}{\sqrt{2}} [(\cos \theta|0\rangle + \sin \theta|1\rangle) \otimes (-\sin \theta|0\rangle + \cos \theta|1\rangle) \\ &\quad + (-\sin \theta|0\rangle + \cos \theta|1\rangle) \otimes (\cos \theta|0\rangle + \sin \theta|1\rangle)] \\ &= |\beta_{11}\rangle. \end{aligned} \quad (11)$$

Besides the collective rotation quantum channel, the collective amplitude damping and the collective dephasing quantum channels are also examples of noisy quantum channels that have subspaces and subsystems that are immune to the existing decoherence.

### III. COLLECTIVE AMPLITUDE DAMPING QUANTUM CHANNEL

The phenomenon of energy dissipation when conveying a quantum state is modeled by the *collective amplitude damping quantum channel*. This channel has the following OSR:

$$\mathcal{E}(\rho) = A_0 \rho A_0^\dagger + A_1 \rho A_1^\dagger, \quad (13)$$

where the operation elements  $A_0$  and  $A_1$  are as follows:

$$A_0 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{bmatrix}, \quad A_1 = \begin{bmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{bmatrix}, \quad (14)$$

where  $\gamma$  is the damping rate which can be thought of as the probability of losing a photon [1, p. 380].

In this channel, due to its collectiveness behaviour, all qubits which suffer amplitude damping are subject to the same damping rate. Thanks to that, it is possible to find a “quiet corner” in the Hilbert space of this channel whose states do not suffer from the effects caused by this type of decoherence. Such states are said to belong to a DFS  $\tilde{\mathcal{H}}$  of the input Hilbert space  $\mathcal{H}$  of this quantum channel [9]. If a state  $\rho \in \tilde{\mathcal{H}}$ , where  $\tilde{\mathcal{H}} \subset \mathcal{H}$ , then it is not affected by the existing decoherence on the collective amplitude damping quantum channel  $\mathcal{E}$ , i.e.,  $\mathcal{E}(\rho) = \rho$ .

In this quantum channel, there are three different DFS, with dimensions 1, 2 and 3, respectively, as shown below:

$$\tilde{\mathcal{H}}_1 = \{|1\rangle\}, \quad (15)$$

$$\tilde{\mathcal{H}}_2 = \left\{ |00\rangle, \frac{|01\rangle - |10\rangle}{\sqrt{2}} \right\}, \quad (16)$$

$$\tilde{\mathcal{H}}_3 = \left\{ \frac{1}{\sqrt{6}}(-2|001\rangle + |010\rangle + |100\rangle), \frac{1}{\sqrt{2}}(|011\rangle - |101\rangle), |000\rangle \right\}. \quad (17)$$

In particular, the DFS  $\tilde{\mathcal{H}}_2$  will be used in the quantum key distribution protocol that will be described in the next section.

## IV. PROPOSED PROTOCOL

Our protocol considers the scheme of communications showed in Figure 1. The legitimate parties (Alice and Bob) are connected through a classical channel and also through a collective amplitude damping quantum channel. Both channels are considered insecure. Despite of that, the objective of Alice and Bob is to create a private key to perform a secure classical message exchange.

The eavesdropper Eve has access to the quantum channel between Alice and Bob. She makes use of a device, which measures the quantum states sent through the channel and stores the basis used for measurement as well as the classical result obtained.

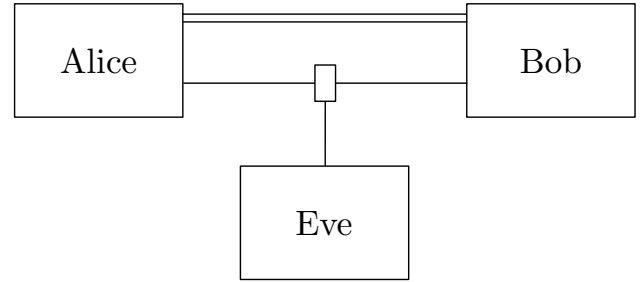


Fig. 1. Communication model considered. The single line wire represented is used for quantum communications while the double line wire is used for classical communications.

The idea of this protocol is very similar to the BB84 QKD protocol [2], but with the advantage of the noise avoidance due to the use of the DFS existing. The description of the protocol will be presented in the sections below.

### A. Protocol Description

The legitimate parties Alice and Bob makes use of the following quantum states:

$$|\rightarrow\rangle = |00\rangle, \quad (18)$$

$$|\uparrow\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}, \quad (19)$$

$$|\nearrow\rangle = |++\rangle, \quad (20)$$

$$|\searrow\rangle = \frac{|+-\rangle - |-+\rangle}{\sqrt{2}}, \quad (21)$$

where  $|-\rangle = \frac{|0\rangle - |1\rangle}{\sqrt{2}}$  and  $|+\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}}$ . Notice that the quantum states  $|\nearrow\rangle$  and  $|\searrow\rangle$  are obtained from  $|\rightarrow\rangle$  and  $|\uparrow\rangle$  by a Hadamard operation. Thanks to the DFS properties, none of the quantum states presented in Eqs. (18)-(21) are affected by the collective amplitude damping. The quantum circuits illustrated on Figure 2 show how to obtain such quantum states.

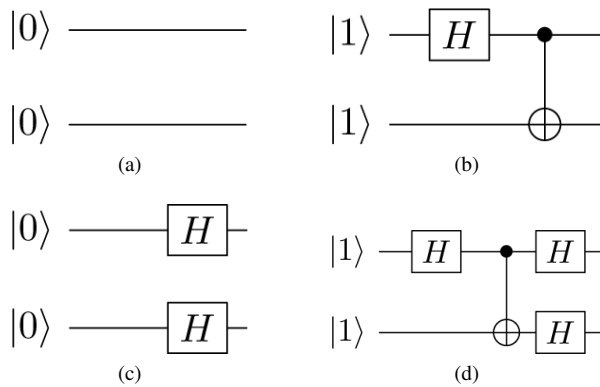


Fig. 2. Quantum circuits that implement the states  $| \rightarrow \rangle$ ,  $| \uparrow \rangle$ ,  $| \nearrow \rangle$ , and  $| \searrow \rangle$ , respectively.

Alice starts the protocol sending states randomly chosen from the set  $\{ | \rightarrow \rangle, | \uparrow \rangle, | \nearrow \rangle, | \searrow \rangle \}$ . Bob and Eve measure the states received using the bases horizontal-vertical  $+ = \{ | \rightarrow \rangle, | \uparrow \rangle \}$  or diagonal  $\times = \{ | \nearrow \rangle, | \searrow \rangle \}$  also randomly chosen.

Let's first consider that Eve is not affecting the communication between Alice and Bob. Table I shows some examples of the results obtained by Alice and Bob in order to create their private symmetric key. If Alice sends  $| \rightarrow \rangle$  or  $| \uparrow \rangle$  and Bob measures with  $+$ , he will obtain bits 0 and 1, respectively, with 100% of certainty. The same is true when Alice sends  $| \nearrow \rangle$  and  $| \searrow \rangle$  and Bob measures with  $\times$ . However, for instance, if Alice sends  $| \rightarrow \rangle$  and Bob measures with  $\times$ , then there is a probability of 0.5 that he will receive the bit 0 and of 0.5 regarding the bit 1.

TABLE I. RESULTS OBTAINED BY BOB AFTER MEASURING THE QUANTUM STATES SENT BY ALICE WITH THEIR RESPECTIVE PROBABILITY.

Alice sends	$  \rightarrow \rangle$	$  \uparrow \rangle$	$  \nearrow \rangle$	$  \searrow \rangle$
Bob measurement	$+$	$+$	$\times$	$\times$
Bit obtained	0	1	0	1
Probability	1	1	1	1

Alice sends	$  \rightarrow \rangle$	$  \searrow \rangle$	$  \nearrow \rangle$	$  \uparrow \rangle$
Bob measurement	$\times$	$+$	$+$	$\times$
Bit obtained	0 or 1	0 or 1	0 or 1	0 or 1
Probability	0.5	0.5	0.5	0.5

In order to avoid uncertainties regarding the bits obtained by Bob, he will communicate to Alice the sequence of bases he used to measure the qubits that she sent. Alice will return to Bob a string of 0's and 1's, where 0 indicates that the respective measurement must be discarded because it leads to uncertainty. After this process, even without communicating the results of the measurements, Alice and Bob agree on the results obtained after the measurement. The bits resultant will compose the private symmetric key that they will use in an one-time pad encryption of the secret classical message sent through the classical channel.

To illustrate the protocol proposed, let's suppose that Alice sends to Bob the following sequence of qubits:  $| \nearrow \rangle, | \uparrow \rangle, | \uparrow \rangle, | \rightarrow \rangle, | \nearrow \rangle, | \searrow \rangle, | \uparrow \rangle$ . Bob uses the sequence of bases  $+, +, \times, +, \times, +, +$  and obtains the sequence of bits given by 0100011. Bob sends the sequence of bases he used

through the classical channel and Alice returns him the sequence 0011101. The sequence of bits sent by Alice indicates that the first, second and sixth bits obtained by Bob must be discarded. So, the private symmetric key between Alice and Bob will have length 5 and will be equal to 00001. With this key, Alice can send Bob a classical message in secrecy by using the one-time pad scheme.

The *one-time pad* encryption scheme requires that the message and the secret key must be of equal length. Let  $m$  be the message and  $k$  be the key, both with  $n$  bits. The encrypted version of the message  $e$  is obtained by  $e_i = m_i \oplus k_i$ , for  $i = 1, \dots, n$ , where  $\oplus$  denotes the addition modulo 2. If the key is used only a single time and if it is kept in secret, then the conditions for *perfect secrecy* in the communication are guaranteed [14].

In the considered example, let's suppose that Alice wants to send a message  $m = 10101$  to Bob. She will follow the one-time pad steps, considering the key  $k = 00001$ , and will obtain  $e = 10100$  that will be sent through the classical channel to Bob. Upon receiving  $e = 10100$ , Bob will use the key  $k = 00001$ , and will retrieve the message sent by Alice by also using the  $\oplus$  operation, which results in  $m = 10101$ . This way, the quantum key distribution protocol and the secret classical message exchange conclude successfully.

In the characterization of the protocol presented, the eavesdropper Eve makes no action during the key creation process. However, it is very unrealistic and her action on the quantum channel must be considered. The next section shows how she can gather information from the private key created by Alice and Bob and how they can use strategies in order to detect her presence and to avoid her success.

### B. Eavesdropping Checking

According to the model of communications considered, Eve can perform measurements in the state sent by Alice, recover a bit from it, and resend the resulting state to Bob. During this process, Eve can not only recover bits from the private key, but also change the quantum state originally sent to Bob.

Eve performs measurements in the state sent by Alice using the bases  $+$  and  $\times$  randomly chosen, i.e, using the same strategy than Bob. To do so, she uses a device which gets the input on the quantum channel, measures it, and resend the resulting quantum state to the channel's output. The effects on the measurements performed by her may degrade the information received by Bob. Table II synthesizes the effects of Eve on the quantum channel.

If by random choice Eve chooses the same basis that Alice used to prepare the quantum states, as shown in the first part of Table II, she will measure the same bit than Bob and will cause no disturbance on the system. However, if she measure the states using the wrong basis, as shown in the second part of Table II, she will have no certainty about the state sent by Alice and will also modify the state received by Bob. When this second situation happens, Alice and Bob can perform successfully the eavesdropper detection.

To perform the eavesdrop detection, besides the bases exchange between Bob and Alice, Bob will also reveal to Alice some bits that he obtained after the measurement. Those bits revealed are important to detect the eavesdropper but they must be discarded after that in order to not compromise the security



TABLE II. STATES SENT BY ALICE AND MEASURED BY THE EAVESDROPPER EVE.

Alice sends	$ \rightarrow\rangle$	$ \uparrow\rangle$	$ \nearrow\rangle$	$ \searrow\rangle$
Eve measurement	+	+	×	×
Eve's resulting bit	0	1	0	1
Probability	1	1	1	1
Bob received state	$ \rightarrow\rangle$	$ \uparrow\rangle$	$ \nearrow\rangle$	$ \searrow\rangle$

Alice sends	$ \rightarrow\rangle$	$ \uparrow\rangle$	$ \nearrow\rangle$	$ \searrow\rangle$
Eve measurement	×	×	+	+
Eve's resulting bit	0 or 1	0 or 1	0 or 1	0 or 1
Probability	0.5	0.5	0.5	0.5
Bob received state	$ \nearrow\rangle$ or $ \searrow\rangle$	$ \nearrow\rangle$ or $ \searrow\rangle$	$ \rightarrow\rangle$ or $ \uparrow\rangle$	$ \rightarrow\rangle$ or $ \uparrow\rangle$

of private key. To illustrate such situation, let's suppose that Alice sent Bob the state  $|\nearrow\rangle$ , Eve measured it with the basis  $+$  and obtained the bit 1, and Bob measured with the basis  $\times$  and received the bit 1. When Bob tells Alice that he used the basis  $\times$  and obtained the bit 1, she can notice that something is wrong and can conclude that there exists an eavesdropper in the quantum channel, because the scenario considered is noiseless.

So, in order to create a private key in secrecy, they must communicate not only the bases used for measurement, but also some of the results obtained. It is essential to ensure the security in the protocol proposed as it is going to be shown in the next section.

## V. SECURITY ANALYSIS

The goal of an ideal key distribution is to allow Alice and Bob, who share no information initially, to share a secret key (a string of bits) at the end. Eve, the eavesdropper, should not obtain information about the key. Also, whatever Eve does, Alice's and Bob's key should be identical. It is assumed that all quantum and communication between Alice and Bob passes through Eve, and similarly for classical communication [5].

No quantum key distribution protocol can succeed if Eve has the power to impersonate Alice while communicating with Bob and to impersonate Bob while communicating with Alice. If Alice and Bob meet previously, there are authentication techniques which can be used to ensure unconditional security [15]. However, in a scenario where Alice and Bob have never exchanged a secret key before, one must assume that Alice and Bob have access to a faithful (classical) public channel so a third part cannot accomplish the impersonation attack without being detected.

Different from classic communication, the security of quantum communication is based on the laws of physics rather than the difficulty of computation. The eavesdropper Eve is so powerful that her ability is only limited by the principles in quantum mechanics. However, the *No-Cloning Theorem* [16] forbids Eve to eavesdrop the quantum signals freely and fully as her action will inevitably disturb the unknown states and leave a trace in the outcomes obtained by the two legitimate users [7]. These facts will help us in the characterization of the security in the proposed protocol.

The eavesdropping strategy that we will consider in the security analysis of the proposed protocol is the *intercept and resend attack* [17] in which Eve measures the quantum state sent by Alice, obtains a bit, and re-sends the resulting quantum

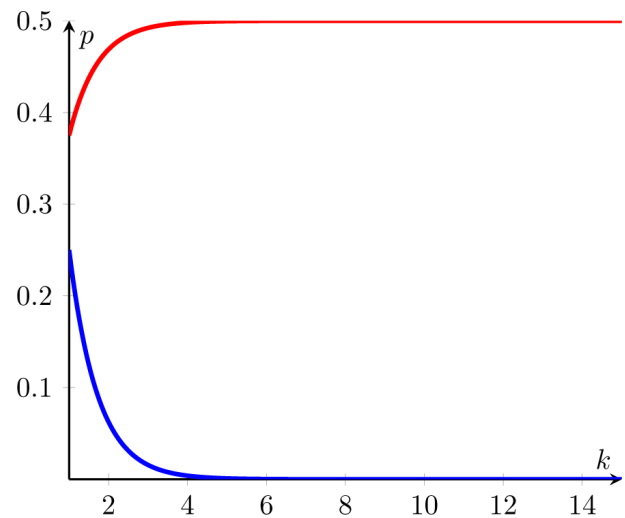


Fig. 3. Graphic showing the exponential decrease in Eve's success probability in recovering all key bits (blue line) and the probability of Alice and Bob detecting the eavesdropper per bit exchanged to create the secret key (red line).

state to Bob. This kind of attack was already depicted in Section IV, but the probability of eavesdropping detection and of Eve's success is going to be described from now on. It is important to emphasize that since this protocol is based on the BB84 QKD protocol [2], the same proofs of security are adequate to our proposition. We strongly suggest the work of Mayers [5] as a source of a more formal approach to these proofs.

If Alice wants to send a bit  $b$  to Bob, she can encode it in two different ways. Eve, upon intercepting it, can also use two options of measurement bases. Her chance of guessing the correct bit is equal to 50%. But if she uses an incorrect basis and it leads to Bob receiving a different bit than was originally sent by Alice, the chance of eavesdropping detection is also of 50% per bit sent.

Let us suppose that Alice and Bob want to create a private key of size  $k$ . Given that Eve may perform intercept and resend attacks, they will reinforce the eavesdropping checking procedure by using  $k$  additional bits. The probability of Eve measuring correctly the  $2 \cdot k$  bits exchanged between the legitimate parties is of  $p(2 \cdot k) = 0.5^{2 \cdot k}$  which decreases exponentially as the size of the key increases as shown in the blue line in Figure 3.

However, if Eve mistakes a single bit in a  $2 \cdot k$  bits sequences (probability equal to  $p_{\text{error}}(1) = 0.5$  per bit), it may result in a bit error detection by Bob and Alice (probability equal to  $p_{\text{detection}}(1) = 0.5$  per bit missed by Eve). Considering these probabilities, the chance of detecting Eve at the  $n$ -th bit goes asymptotically to 0.5, as shown in the red line of the graphic in Figure 3, i.e., it is strongly related with the probability of error detection. Differently from the probability of Eve success, the probability of eavesdropping detection is independent per bit exchanged. Since Eve can change a qubit and this alteration may not be detected, as reported in Tables I and II, there always a probability of not detecting the eavesdropper in the communication.

As it can be seen, while the success of Eve depends on guessing all bits without disturbing the communication between Alice and Bob, her detection depends on one mistake

on her measurements which disturbs the bit received by Bob used in the eavesdrop checking process. Thus, we can conclude that the ability of the protocol to detect eavesdropping is high, ensuring enough security for practical scenarios of its use. This concludes the analysis of security of the quantum key distribution protocol proposed.

## VI. CONCLUSION

The first practical demonstration of a QKD protocol took place in the early 1990s using photons over a distance of 30 cm through air. After that, the next implementation over the atmosphere guaranteed a secure communication with quantum bits over a distance of 2 km. After that, QKD protocols could be implemented in distances up to 250 km [18]. Nowadays, even commercial devices are being developed and sold to perform secure quantum key distribution [19].

However, one of the main problems in practical QKD is the noise, which can not only affect the communication between the legitimate parties, but can also favor an existing eavesdropper. In the attempt to minimize such problems, we proposed a QKD protocol over a collective amplitude damping quantum channel where an eavesdropper performs intercept and resend attacks.

This protocol is mainly based on BB84 QKD protocol [2], but since it considers the existing DFS on the quantum channel taken into account, the communication is noiseless. However, we consider the existence of an eavesdropper which aims at discovering the private key in the attempt to make a breach of security in the message exchange between Alice and Bob. In order to avoid it, the legitimate parties must use extra bits, randomness and also certain procedures for eavesdropping checking. As shown in Section V, it causes a very low probability of Eve's success while the probability of eavesdropping detection is high.

This work contributes to the use of the DFS in secure communications. If the eavesdropper is passive, following the model of quantum wiretap channels [20], [21], it is possible to reach unconditional security in the communications [22], [23]. This scenario not always occurs, so it is essential to consider other protocols and techniques. Since DFS arise where collective decoherence takes place [24], other works developing QKD protocols for collective dephasing and rotation quantum channels were already considered [6], [7]. However, so far, no protocol for QKD on collective amplitude damping quantum channels were known.

In practical scenarios, some works already report the implementation of quantum channels with DFS [25]–[27] even in long distance [28]. With this already existing technology, the proposed protocol can be adopted in realistic scenarios to provide secure communications.

In future work, we aim at proposing other protocols and techniques for secure communications over eavesdropped quantum noisy channels.

## ACKNOWLEDGEMENTS

The authors acknowledge the financial support rendered by the Brazilian funding agencies CAPES and CNPq, by the Post Graduation Program in Electrical Engineering at the Federal University of Campina Grande, and by the project QUANTA/RENASIC/FINEP.

## REFERENCES

- [1] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 2nd ed., Bookman: Cambridge University Press, 2010.
- [2] C. H. Bennett and G. Brassard, "Quantum Cryptography: Public Key Distribution and Coin Tossing," *ICSSP Press*, 1984, pp. 175-179.
- [3] A. K. Ekert, "Quantum Cryptography Based on Bell's Theorem," *Phys. Rev. Lett.*, vol. 67, 1991, pp. 661-663.
- [4] C. H. Bennett, "Quantum Cryptography Using any Two Nonorthogonal States," *Phys. Rev. Lett.*, vol. 68, 1992, pp. 3121-3124.
- [5] D. Mayers, "Unconditional Security in Quantum Cryptography," *Journal of the ACM*, vol. 48, 2001, pp. 351-406.
- [6] J.-C. Boileau, D. Gottesman, R. Laflamme, D. Poulin and R. W. Spekkens, "Robust Polarization-Based Quantum Key Distribution over a Collective-Noise Channel," *Phys. Rev. Lett.*, vol. 92, 2004, pp. 017901.
- [7] X.-H. Li, F.-G. Deng and H.-Y. Zhou, "Efficient Quantum Key Distribution Over a Collective Noise Channel," *Phys. Rev. A*, vol. 78, 2008, pp. 022321.
- [8] J. Stolze and D. Suter, *Quantum Computing – A short course from theory to experiment*, Wiley: VCH Verlag, 2004.
- [9] D. A. Lidar and K. B. Whaley, "Decoherence-Free Subspaces and Subsystems", arxiv: quantum-ph/0301032v1, 2003. (Retrieved: May, 2013).
- [10] P. Dirac, *The principles of Quantum Mechanics*, 4th ed., Oxford University Press: Oxford, 1982.
- [11] A. Shabani and D. A. Lidar, "Theory of initialization-free decoherence-free subspaces and subsystems," *Phys. Rev. A*, vol. 72, 2005, pp. 042303.
- [12] M. S. Byrd, L.-A. Wu and D. A. Lidar, "Overview of quantum error prevention and leakage elimination," *Journal of Modern Optics*, vol. 51, 2004, pp. 2449-2460.
- [13] E. Knill, R. Laflamme and L. Viola, "Theory of quantum error correction for general noise," *Phys. Rev. Lett.*, vol. 84, 2000, pp. 2525.
- [14] C. Paar and J. Pelzl, *Understanding Cryptography*, Springer: Springer, 2010.
- [15] M. N. Wegman and J. L. Carter, "New Hash Function and their Use in Authentication and Set Equality," *J. Comput. Syst. Sci.*, vol. 22, 1981, pp. 265-279.
- [16] W. K. Wootters and W. H. Zurek, "A Single Quantum Cannot be Cloned," *J. Comput. Syst. Sci.*, vol. 299, 1982, pp. 802-803.
- [17] D. Kalamidas, "Single-Photon Quantum Error Rejection and Correction With Linear Optics," *Phys. Lett. A*, vol. 343, 2005, pp. 331-335.
- [18] J. Mullins, "Making Unbreakable Code," *IEEE Spectrum*, vol. May, 2002, pp. 40-45.
- [19] ID Quantique, "Quantum Key Distribution", <http://www.idquantique.com>, 2013. (Retrieved: May, 2013).
- [20] N. Cai, A. Winter and R. W. Yeung, "Quantum privacy and quantum wiretap channels," *Problems of Information Transmission*, vol. 40, 2004, pp. 318-336.
- [21] I. Devetak, "The private classical capacity and quantum capacity of a quantum channel," *IEEE Transactions on Information Theory*, vol. 51, 2005, pp. 44 -55.
- [22] E. B. Guedes and F. M. Assis, "Utilização de Subespaços Livres de Descoerência em Comunicações Quânticas Incondicionalmente Seguras," *Proc. Simpósio Brasileiro de Telecomunicações (SBrT'12)*, SBrT Press, 2012, pp. 1-5.
- [23] E. B. Guedes and F. M. Assis, "Unconditional Security with Decoherence-Free Subspaces", arXiv:quant-ph/1204.3000, 2012, pp. 1-6. (Retrieved: May, 2013).
- [24] P. Zanardi and M. Rasetti, "Noiseless quantum codes," *Phys. Rev. Lett.*, vol. 79, 1997, pp. 3306.
- [25] G. Jaeger and A. Sergienko, "Constructing four-photon states for quantum communication and information processing," *Int. J. Theor. Phys.*, vol. 47, 2008, pp. 2120.
- [26] U. Dorner, A. Klein and D. Jaksch, "A quantum repeater based on decoherence free subspaces," *Quant. Inf. Comp.*, vol. 8, 2008, pp. 468-490.
- [27] Y. Xia, J. Song, Z.-B. Yang and S.-B. Zheng, "Generation of four-photon polarization-entangled decoherence-free states within a network," *Appl. Phys. B*, vol. 99, 2010, pp. 651-656.
- [28] P. Xue, "Long-distance quantum communication in a decoherence-free subspace," *Phys. Lett. A*, vol. 372, 2008, pp. 6859-6866.

## Live Replication of Virtualized VoIP Servers

Jiri Hlavacek, Robert Bestak

Department of Telecommunications  
Faculty of Electrical Engineering, CVUT Prague  
Prague, Czech Republic  
e-mail: hlavaji1@fel.cvut.cz, robert.bestak@fel.cvut.cz

**Abstract**— Virtualization technology enables resource sharing and it can also improve system availability thanks to the continuous real-time migration mechanism. Highly available Internet Protocol communication systems are still expensive as there is no standard solution and customized software development is lengthy and expensive. In this paper, we analyze the continuous live replication mechanism of Xen hypervisor in the context of virtual machine hosting a Voice over Internet Protocol server processing both signalization and user data. We demonstrate that, nowadays, replication mechanism is unsuitable for soft real-time applications. The main drawbacks are jitter and packet bursts generated by the replication mechanism output buffering. To eliminate this, we propose to classify output packets and to let packets with real-time data bypass the buffering. The results obtained confirm that the soft real-time application's availability can be significantly improved by using virtual machine's live replication.

**Keywords**-VoIP; system availability; virtualization; live replication.

### I. INTRODUCTION

Voice over Internet Protocol (VoIP) technology [1] is becoming omnipresent. Its rich set of services, low cost for operators and ease of integration with other technologies are the main reasons for its success. However, one of the drawbacks is lower availability compared to legacy systems. IP networks are designed to be fault tolerant for general use cases, for example client-server applications. Telephony is a soft real-time application requiring availability equal to 99.999%, i.e., 5.26 minutes downtime per year. In order to achieve such requirements, specific network configuration and application design are needed. Replication enabled applications are complex and, therefore, expensive; thus, other solutions are under research.

Virtualization is widely used for its advantages such as better resource usage, flexibility and scalability. Moreover, it can also be employed to improve system availability using live migration techniques. The main advantage of virtualization consists in the fact that the migration is completely transparent for applications. Live migration mechanisms are still in development. The main challenges include scheduling of virtual machines [2] and network bandwidth optimization [3]. Other issues are latency, jitter and packet bursts introduced by the replication process itself.

In this paper, we propose a modification of the network buffer used by the replication mechanism. This buffer enables saving of all outbound data between two successive replication steps to ensure data consistency in case of failure. Our proposition consists in network packet inspection and classification, where only Transmission Control Protocol (TCP) packets [4] are buffered. User Datagram Protocol (UDP) and Real-time Transport Protocol (RTP) packets [5][6] are forwarded, bypassing the live replication buffering. This way, the latency and jitter for RTP packets are minimized and the voice quality of calls is improved. Our modification minimizes virtualization's impacts on real-time data flows.

The rest of this paper is organized as follows. Section II provides an overview of related work. Section III presents the main virtualization principles followed by an analysis of virtualization's impact on voice quality. Section IV describes our proposition and outlines a potential implementation. In Section V, we present the testbed used and the results obtained. Finally, Section VI concludes our work.

### II. RELATED WORK

The problem of VoIP system's availability is addressed in many proprietary or standardized approaches but none of them really fits all real scenarios. In this work, we focus on the VoIP server availability in the standard server/client architecture; peer-to-peer networks are not taken into account in our study.

A common point of the solutions described below is the use of the Internet Protocol (IP) failover mechanism [7]; the active server uses a virtual IP address that is taken by the backup server when the active one fails. Thanks to that, the process is entirely transparent to the clients.

#### A. Application Dependent Solutions

Application level replication is described by A. Gorti [8]. The proposed solution requires a specific software development, which is expensive and long. Replication enabled applications based on software frameworks for high availability (e.g., Terracota) are hard to configure and maintain. Furthermore, there are several requirements on the software architecture such as thread safety [9]. G. Kambourakis et al. propose a database-based state sharing mechanism [10]. Contexts are saved in a database and the replication is done by the database engine. This solution is relatively easy to implement as only a database connector

needs to be developed instead of a complex replicated system preserving data consistency. Nevertheless, the architecture remains complex and hard to integrate with existing applications.

A common problem to all application dependent solutions is the breakdown of TCP and Transport Layer Security (TLS) connections when taking over the IP address. These connections can't be migrated without specialized operating systems or replication aware clients.

### B. Application-Transparent Solutions

Application transparent solutions usually require more resources, but it is compensated by easier development and maintenance of applications. One of the recommended methods to ensure service continuity in case of server failure is the use of the Domain Name System Service Record (DNS SRV) mechanism [11][12]. In this method, a single point of failure is the DNS server itself. A secondary DNS server is usually present but the timeout and takeover adds some seconds to the call establishment. Furthermore, the context replication issue is not addressed and thus this solution is suitable only for stateless servers.

Virtualization techniques are nowadays very popular, mainly because of their better hardware resources usage. These techniques are also employed to facilitate hardware maintenance and fault tolerance. A special use case is continuous live migration of running Virtual Machine (VM) used for high available systems. The principle of the continuous live migration is to replicate the primary machine state to the backup machine continuously and to start the backup machine if the primary one fails. However, this technique is resource demanding. It also introduces a periodic short interruption necessary for synchronization to the primary machine execution. Two ways of machine's replication exist. The first way is based on replay of non-deterministic events received by the primary machine on the backup one [13]. This technique is used for example by VMware [14]. The deterministic replay is not adapted to the Symmetric Multi Processing (SMP) environments since ordered memory access is needed [15]. The second machine's replication approach is based on periodic replication of checkpoints, i.e., processor and memory states, with a high frequency [16]. This replication method is appropriate for SMP environments, but requires more bandwidth for the replication process.

Virtualization also reduces virtual machine performances. A detailed analysis of XEN hypervisor's scheduler identifying bottlenecks for media applications is presented by M. Lee et al. [17]. Modifications aiming at a better adaptation of the XEN scheduler for media applications are presented by M. Lee et al. [2]. These two works show that virtualization supports performance requirements of VoIP servers. Adaptation of virtualized environment for high available VoIP servers is proposed by D. Patnaik et al. [18], where authors show that real-time replication without network buffering performs well enough to run a VoIP server. Nevertheless, data consistency isn't preserved without network buffering; primary and backup machine states can become desynchronized. The main

virtualization's advantage is its transparency at the application level, i.e., any VoIP server implementation can be used.

In order to support complex VoIP services, servers are composed of many different modules. Therefore, the application level replication approach becomes too complex and security requirements are hard to meet [9]. On the other hand, the migration of virtualized machines is transparent at the application level and preserves TCP/TLS connections.

## III. PROBLEM ANALYSIS

Virtualization [19] is a new challenging mechanism and its use for real-time and multimedia applications is being actively investigated by researchers. In our study, we focused on network metrics such as jitter and burst generation, as these aspects are strengthened by the replication mechanism itself.

### A. Virtualization Basics

Virtualization allows one or more virtual machines to run simultaneously on one hardware platform. Hardware sharing is ensured by a specific layer between the hardware and virtualized machines called the hypervisor. Each machine disposes of several resources that are attributed by the hypervisor and the machine can be unaware of being virtualized. Hypervisors often allow migration of virtual machines between different hardware servers in a shutdown state. Current hypervisor implementations improved the migration mechanisms and enable a real-time migration without virtual machine shutdown. This approach is called live migration. The live migration is useful to change the server hosting the VM to obtain more resources or to enable hardware maintenance. Another usage is to build high availability systems. The live migration process can be described as follows. The virtual machine execution is paused by the hypervisor; the virtual machine's state (including memory content and processor state) is transferred to the backup server where the virtual machine is resumed.

### B. Live Migration

To overcome hardware faults, live migration has been improved to support continuous live migration. This type of migration allows immediate resuming of a backup virtual machine when the primary server goes down or becomes unreachable. The execution of the primary virtual machine is divided into small periods of time and the state of the primary machine is entirely replicated on the secondary server at the end of each period. These periods are called epochs. The backup machine on the secondary server waits in a paused state. The replication period depends on the machine purpose and typically ranges from 40ms to 200ms [16]. In case of failure, the backup machine on the secondary server is resumed at the last completed checkpoint. This means that a packet loss will be observed from the user's perspective.

### C. Data Consistency

The challenge of live migration mechanism is to ensure data consistency in case of primary server failure. To

minimize the interruption and to preserve performances, the replication is being done continuously. A synchronization checkpoint ensuring the consistency is introduced at the end of each replication cycle. Synchronization checkpoint consists of stopping the execution, synchronization and consistency verification and resuming the execution [16].

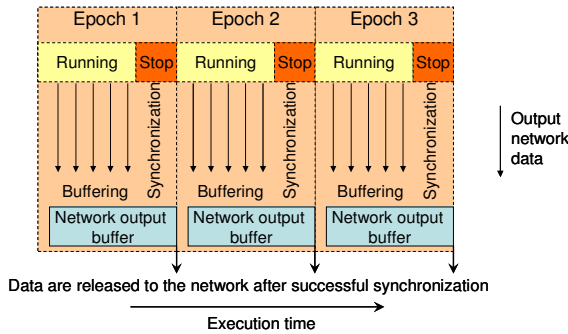


Figure 1. Network buffering during continuous live migration.

To integrate such a system into the real environment, network outputs need to be saved in a buffer between each checkpoint and released once the state of virtual machine is successfully replicated on the backup server. This buffering process introduces jitter and packets burst. This mechanism is depicted in Fig. 1. Although these events can be tolerated in interactive client/server applications, it is a significant issue for real-time applications such as VoIP services. In certain cases, the buffering impact can be accepted for signalization, but needs to be definitively eliminated for user data like RTP flows.

#### IV. PROPOSED SOLUTION

This section details the proposed solution to the problem described above, including a possible way of its implementation.

##### A. Principle

Current replication implementation buffers all outgoing packets (signaling, data). The buffering of real-time transport protocol packets has a negative impact on real-time applications. Thus, the replication mechanism needs to be enhanced by inspecting packets and classifying them in two categories: i) packets that need to be buffered and ii) packets that can be sent without being buffered. As we focus on the VoIP service, we simplify this classification as follows. All TCP packets are buffered in order to ensure TCP state consistency and to avoid signaling packet loss when using TCP or TLS protocol. Non-TCP packets are sent to the network without any processing by hypervisor. This approach minimizes a negative impact on network forwarding such as latency increase, jitter introduction or burst generation. The classification needs to be refined to match only packets with real-time data.

##### B. Implementation

In our testbed, we use the latest XEN hypervisor version 4.2-unstable [20]. A server with XEN hypervisor includes the following components:

- XEN hypervisor.
- Domain 0 – the XEN’s terminology for privileged domain with direct hardware access run by the hypervisor to manage the server and control other virtual machines.
- Domain U – the XEN’s terminology for unprivileged domains hosting virtualized machines.

The XEN hypervisor live migration implementation is provided by Remus project [16]. Remus provides fault detection, virtual machine state (memory and processor state) replication and network buffering. The network buffer implementation is based on Linux traffic control together with Intermediate Functional Block [21]. This component allows ingress traffic queuing. The traffic sent by virtual machines is buffered at the Domain 0. Standard Linux traffic control tools are used to redirect all incoming traffic from the replicated machine to the buffer. We analyze all passing packets and redirect to the buffer only TCP packets. The traffic classification can be done directly by the Linux kernel. Thus, the kernel code doesn’t need to be modified. Our implementation proposal consists of the following 3 steps:

- Packet classification using Linux packet inspection.
- Introduction of redirection rules to buffer only non-real time packets.
- Real-time packets are transmitted without any buffering.

#### V. OBTAINED RESULTS

This section presents the results of the measurements performed.

##### A. Testbed Description

The testbed platform with low performance machines and 100 Mbit/s network switch is considered. One server includes a double core Intel Pentium-M processor running at 1.8 GHz (3 GB RAM), whereas the second server is built on a single core AMD Duron processor running at 900 MHz (1 GB RAM). Both servers are running XEN hypervisor, version 4.2-unstable and Linux Ubuntu 12.04 [22] with 3.2.0-24-pae kernel in Domain 0 and the same kernel in Domain U. Fig. 2 illustrates the testbed network configuration. Additionally, Freeswitch [23] version 1.0.7 as a SIP (Session Initiation Protocol) server is used together with one SIP VoIP phone and one SIP softphone. The voice signal processing is done using G.711 A-law codec with 20 ms packetization time (ptime).

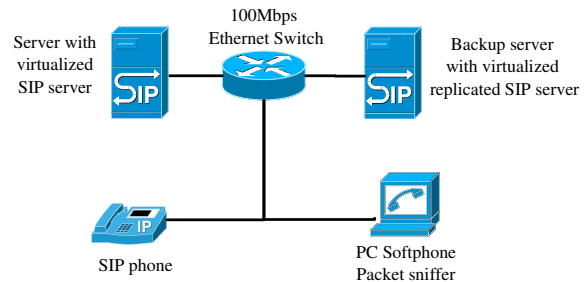


Figure 2. Testbed configuration.

**B. Jitter calculation**

As there is no packet loss except during the effective migration, we consider jitter values to evaluate the impact of virtualization. Jitter calculation is implemented as described in [24]. It is a first-order estimator with noise reduction using a gain parameter.

**C. Virtualization Impact**

Figs. 2 to 6 show the jitter observed under different conditions. The horizontal axes represent time and vertical axes jitter. As shown in Fig. 3, when running a SIP server without virtualization, the jitter is about 100µs with a minimal variation. Performances of Freeswitch running on a virtualized machine (Fig. 4) are not as good as without the virtualization, but still lower than 500µs, which is usually the limit in the Service Level Agreement among operators.

**D. Live Replication Impact**

Fig. 5 illustrates the impact of continuous live migration process on the machine performances. Without any modifications, network buffering introduces an important jitter and packet bursts at each checkpoint. A peak can be observed approximately every 400 ms, which equals to the checkpoint interval.

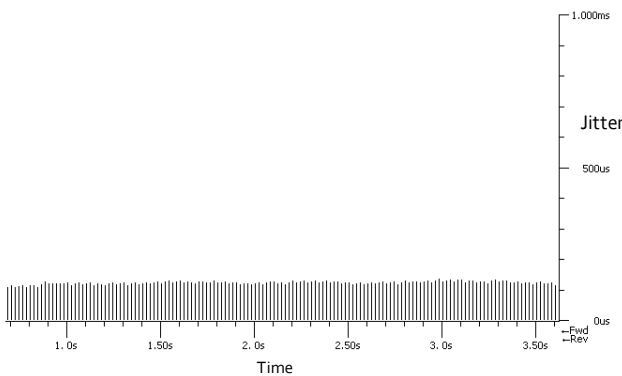


Figure 3. Jitter without virtualization.

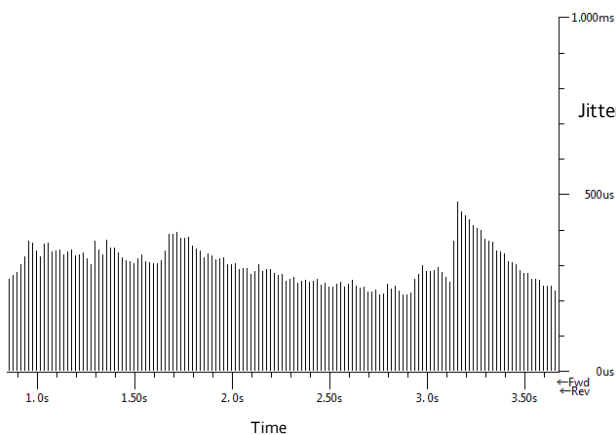


Figure 4. Jitter measured calling via a virtualized Freeswitch server.

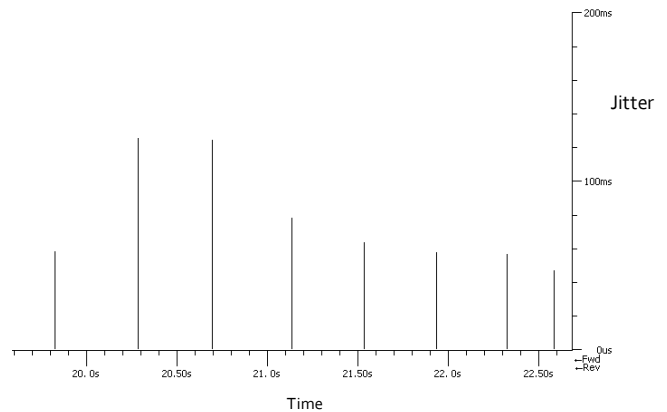


Figure 5. Jitter using unmodified replication mechanism.

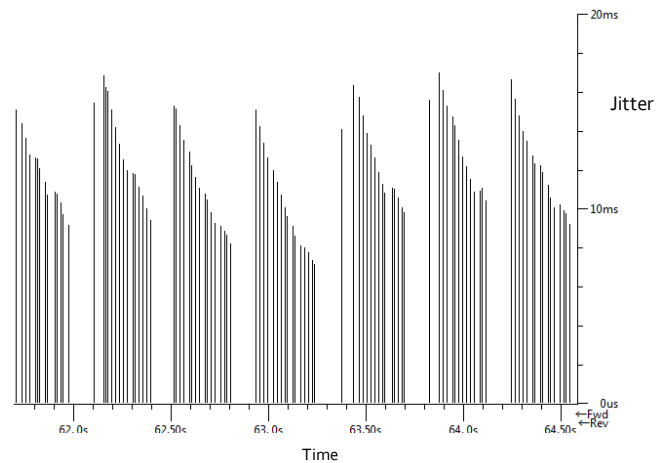


Figure 6. Jitter using modified replication mechanism.

These peaks correspond to a burst of data released from the buffer. With a 20 ms packetization time, each peak represents about 20 RTP packets by call and direction. The observed degradation of network characteristics deteriorates quality of VoIP calls. Such packet bursts can overload network equipments and cause packet loss. The impact of the continuous live replication is studied in more detail in [25].

Our proposal's impact is shown in Fig. 6. Contrary to the unmodified replication mechanism, RTP packets are forwarded as long as the virtual machine runs. The observed jitter is more important with continuous live replication because of performance impact generated by the continuous VM state replication. This effect is emphasized by the low-performance CPU. Execution interruptions can be observed every 300ms, which corresponds to the configured checkpointing interval.

The considered testbed is composed of relatively low performance machines and a single 100 Mbit/s network, while a 1 Gbit/s network dedicated to replication is recommended. Used low performance testbed allows us to verify the behavior of the proposed modification with limited resources.

Measured jitter and interruption length are therefore quite high, but fully rectified by jitter buffer and the impact on the call quality is unnoticeable from the user's point of view. Note that the calculation method defined in [24] considers the previous packet jitter; therefore we can observe a descending trend following each interruption.

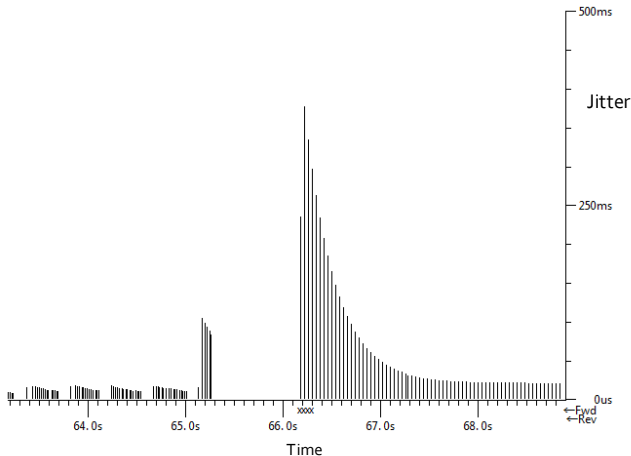


Figure 7. Jitter and packet loss during the migration following a failure.

The objective of the continuous live migration is to maintain established calls and connections in case of hardware or network failure. Once a fault is detected, the replicated machine resumes on the backup server without call interruption. To announce the new location of the virtual IP address, a gratuitous Address Resolution Protocol (ARP) request is used. The jitter observed during the migration is depicted in Fig. 7. On the left side we observe the system in a stable state, continuous VM replication running. The failure of the primary server is represented by a period without packets, which lasts about 1s. The detection of the primary server's failure, generation of the gratuitous ARP and its processing by network components is time consuming. In our configuration it's almost 1s, but the time strongly depends on network hardware. Users of the VoIP service perceive the failover as a short interruption. As the second virtual machine runs without continuous live replication, the jitter observed after the migration is stable without interruption. The descending trend is due to the used calculation method discussed above.

## VI. CONCLUSION AND FUTURE WORK

Recent VoIP systems are complex and often interconnected with external services. The possibility of implementing a high availability system with no impact on the application is therefore a very challenging task. In this article, we study the jitter as a major drawback of virtual machine's live migration. We propose a modification to improve networking properties of this mechanism without impact on data consistency. The proposition is easy to implement and its impact on the system performance is negligible compared to the unmodified system. Our measurements demonstrate that the modification is beneficial

for VoIP and other soft real-time applications. Contrary to the conventional implementation, our implementation does not introduce any jitter to the real-time packet flow except the jitter caused by the interruption required to synchronize replicated virtual machines.

The remaining jitter is caused by the interruption necessary for checkpointing. In our future work, we will focus on reducing the necessary interruption time. The optimal length of checkpointing interval is another point that is worth to be investigated as well as better packet classification.

## ACKNOWLEDGMENT

This research work was supported by the Grant Agency of the Czech Technical University in Prague, grant no. SGS13/199/OHK3/3T/13.

## REFERENCES

- [1] P. Mehta and S. Udani, "Voice over IP", IEEE Potentials, vol. 20, issue 4, November 2001, pp. 36-40.
- [2] M. Lee, A. S. Krishnakumar, P. Krishnan, N. Singh, and S. Yajnik, "Supporting soft real-time tasks in the Xen hypervisor," VEE 2010, Pittsburgh, PA, March, 2010, pp. 97-108.
- [3] Ch. Clark et al., "Live migration of virtual machines," In Proceedings of the 2nd conference on Symposium on Networked Systems Design & Implementation, vol. 2, USENIX Association, Berkeley, CA, USA, 2005, pp. 273-286.
- [4] J. Postel, "Transmission Control Protocol", RFC 793, Internet Engineering Task Force, September 1981.
- [5] J. Postel, "User Datagram Protocol", RFC 768, Internet Engineering Task Force, August 1980.
- [6] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", RFC 3550, Internet Engineering Task Force, July 2003.
- [7] K. Singh and H. Schulzrinne, "Failover, load sharing and server architecture in SIP telephony," Comput. Commun. 30, 5, March 2007, pp. 927-942.
- [8] A. Gorti, "A fault tolerant VoIP implementation based on open standards," In Proceedings of the IEEE Dependable Computing Conference, Dependable Computing Conference, 2006, pp. 35-38.
- [9] J. Hlavacek and R. Bestak, "Improvements in the availability of SIP networks," In Proceedings of the 2010 Networking and Electronic Commerce Research Conference, Dallas, TX: American Telecommunications Systems Management Association Inc., 2010, pp. 109-117.
- [10] G. Kambourakis et al., "High availability for SIP: Solutions and real-time measurement performance evaluation," International Journal of Disaster Recovery and Business Continuity, vol. 1, no. 1, 2010, pp. 11-30.
- [11] J. Rosenberg and H. Schulzrinne, "Session Initiation Protocol (SIP): Locating SIP servers," RFC3263, Internet Engineering Task Force, June 2002.
- [12] A. Gulbrandsen, P. Vixie, and L. Esibov, "A DNS RR for specifying the location of services (DNS SRV)," RFC2782, Internet Engineering Task Force, February 2000.
- [13] D. J. Scales, M. Nelson, and G. Venkitachalam, "The design of a practical system for fault-tolerant virtual machines," SIGOPS Operating Systems Review, vol. 44, issue 4, December 2010, pp. 30-39.
- [14] VMware, VMware Incorporation, <http://www.vmware.com>, [retrieved: March, 2013].

- [15] G. Kambourakis et al., "Hardware and Software Approaches for Deterministic Multi-Processor Replay of Concurrent Programs," In Intel Technology Journal, vol. 13, issue 4, 2009, pp. 20–41.
- [16] B. Cully et al., "Remus: High availability via asynchronous virtual machine replication," In Proc. Symp. Network Systems Design and Implementation (NSDI), 2008, pp. 161–174.
- [17] M. Lee, A. S. Krishnakumar, P. Krishnan, N. Singh, and S. Yajnik, "XenTune: Detecting Xen Scheduling Bottlenecks for Media Applications," IEEE Globecom 2010 (Communications Software, Services and Multimedia Applications Symposium), Miami, FL, Dec 6-10, 2010, pp. 1-6.
- [18] D. Patnaik, A. Bijlani, and V. K. Singh, "Towards high-availability for IP telephony using virtual machines," IEEE 4th International Conference on Internet Multimedia Services Architecture and Application (IMSAA), December 2010, pp. 1-6.
- [19] S. Crosby and D. Brown, "The Virtualization Reality", Queue, vol.4, issue 10, December 2006, pp. 34-41.
- [20] XEN, Xen hypervisor, <http://xen.org>, [retrieved: March, 2013].
- [21] IFB, The Intermediate Functional Block device, <http://www.linuxfoundation.org/collaborate/workgroups/networking/ifb>, [retrieved: March, 2013].
- [22] Ubuntu, Open Source Linux operating system, <http://www.ubuntu.com>, [retrieved: April, 2013].
- [23] Freeswitch, Open Source Multi-Protocol Soft Switch, <http://www.freeswitch.org>, [retrieved: April, 2013].
- [24] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A transport protocol for real-time applications," RFC 3550, Internet Engineering Task Force, July 2003.
- [25] J. Hlavacek and R. Bestak, "Configuration of Live Migration for VoIP Applications," In Proceedings of 15th Mechatronika 2012, Praha, Czech Technical University in Prague, 2012, pp. 187-190.



# Reasoning About Consistency of Relational Knowledge Bases

Tadeusz Pankowski

Institute of Control and Information Engineering

Poznań University of Technology

Poznań, Poland

Email: tadeusz.pankowski@put.poznan.pl

**Abstract**—We study a translation of relational database into an extended DL knowledge base that is both sound and complete with respect to semantics preservation. The issue is of special importance while creating Web of Data, where relational databases are exposed in the Web to be searched, integrated and exchanged without loss of information and semantics. Our approach considers all fundamental requirements of DL/OWL knowledge bases such as OWA and rejecting the UNA. We propose a system of axioms for representing database integrity constraints, which guarantees information and semantics preserving data-to-knowledge exchange.

**Keywords**—knowledge bases; data integration; integrity constraints; data exchange; ontology-based data management.

## I. INTRODUCTION

In this paper, we study the problem of transforming a relational databases to *Description Logic* (DL) knowledge bases. We propose a method to realize such a transformation and study its correctness. The translation is correct if in the target knowledge base both information and semantics are preserved. The problem of representing a relational database in a DL knowledge base has gained in importance in recent years, mainly due to the emergence of the Semantic Web. It is expected that the Semantic Web enables web-wide integration of data coming from various sources, also referred to as Web of Data. The Web Of Data can be formally perceived as a giant DL knowledge base (or DL ontology)  $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ , where  $\mathcal{T}$  is a set of axioms modeling the intensional knowledge (the TBox axioms), and  $\mathcal{A}$  is a set of assertions forming the extensional knowledge (the ABox assertions) [1]. It is expected that in such a knowledge base the underlying source repositories are represented as precisely as possible and that distributed execution of queries over these repositories is possible allowing for so called *ontology-based data access* [2]. Most of the data exposed in the Web are originally stored in relational databases, and have to be mapped or translated to *Resource Description Framework* (RDF)/*Ontology Web Language* (OWL) representations. Sometimes we face an inverse problem – Web data are to be stored in a relational database, so the relational structures and constraints have to be discovered. Thus, the similarities and differences between databases and knowledge bases has been an important and attractive field of research since many years [3], [4], [5], [6].

**Related work.** Some recent results of representing relational databases in the Semantic Web are surveyed in [7]. There are two approaches to integrating relational databases with the Semantic Web: (1) *direct mapping* [8], [9] – consists in

translating the database contents and its schema directly to representations in Semantic Web languages, i.e., in RDF and OWL, in result an ontology is created from the database; (2) *mapping to ontology* [2] – assumes the existence of a shared domain ontology and a language for mapping database schema to this ontology. Then, the *extensional* layer of the database is represented by RDF graph in a form of a set of semantic triples (*subject, predicate, object*) [10]. The *intensional* layer is captured by means of axioms specified in OWL [11] or RDFS [10]. The direct mapping of relational database schema to OWL DL is the subject of the current draft of the W3C Group [12]. A formal basis for RDF(S) and OWL provides *Description Logic* (DL), that is a family of knowledge representation languages [13] being usually a decidable subset of first-order logic. A relationship between relational databases and DL knowledge bases are studied in [6]. There are three main differences between databases and knowledge bases making the translation between them difficult [5], [4]: (a) databases are based on the *Closed Word Assumption* (CWA) while knowledge bases follow usually the *Open World Assumption* (OWA); (b) databases make the *Unique Name Assumption* (UNA) while knowledge bases usually do not accept it; (c) integrity constraints in databases are interpreted as checks while in knowledge bases all rules are interpreted as deductive rules. It turns out that incorporating integrity constraints into knowledge bases is the most challenging issue. To deal with the problem, the following three approaches have been proposed: (a) introduction of epistemic operators **K** and **A** into the knowledge base, to capture the semantics of integrity constraints by restricting interpretation of some formulas to the "state" of the knowledge base, i.e., to "what the knowledge base knows" [14]; (b) knowledge bases are restricted only to those, which accept UNA [15], (c) defining an *extended DL knowledge base* (EKB), where the set  $\mathcal{T}$  of TBox axioms is divided into *standard* TBox axioms,  $\mathcal{S}$ , and *integrity constraint* TBox axioms,  $\mathcal{C}$ ; semantics of  $\mathcal{S}$  and  $\mathcal{C}$  are defined in different ways.

**Contribution.** In this paper we will use the notion of EKB to represent a relational database in DL. We define a *data-to-knowledge exchange* (*dk-exchange*) system that defines translation of relational database schema, its integrity constraints and instances into an extended DL knowledge base. The resulting extended DL knowledge base is referred to as a *relational knowledge base* (RKB). This approach provides an automatic translation and a relational schema is mechanically transformed into TBox axioms. However, for the approach to be meaningful, some assumptions must be made about the way

in which the underlying relational database has been created. To this order we assume that the relational database under consideration implements a conceptual model developed using ER (*Entity-Relationship*) or its extension EER (*Expanded Entity-Relationship*) approach [16], [17]. Additionally, we assume that all relationships in ER schema are binary and one-to-many. It means that any many-to-many relationship is replaced with two one-to-many relationships. Such the replacement is always possible and commonly applied to form UML class diagrams [18].

In Figure 1, the EER diagram describes students, courses and exams taken by students [9]. In particular, a student is a specialization of a person. Every entity set in Figure 1 (b), i.e., Person, Student, Exam, and Course, has a key (identifier). Other attributes will be specified below in SQL definition.

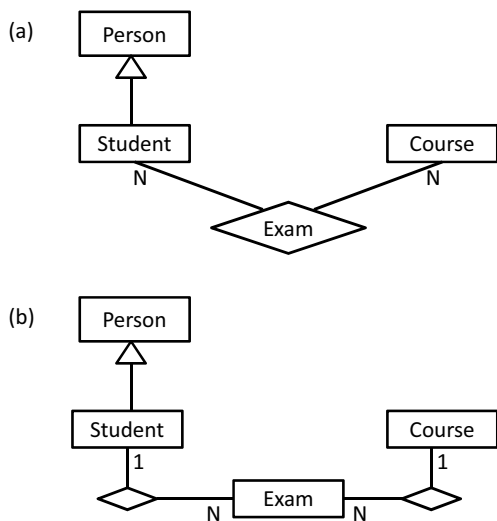


Fig. 1. EER diagrams of a university domain. (a) Exam is a many-to-many relationship, (b) Exam is a member in two one-to-many relationships [9].

A specification of the EER diagram in Figure 1 (b) using SQL notation is given in Figure 2.

```
create table Person(
  PId int primary key,
  Name char(20));

create table Course(
  CId char(5) primary key);

create table Student(
  SId int primary key foreign key references Person(PId),
  Faculty char(20) not null);

create table Exam(
  EId int primary key,
  ESId int foreign key references Student(SId),
  Course char(5) foreign key references Course(CId),
  Grade char(1));
```

Fig. 2. SQL specification of the EER diagram in Figure 1 (b) [9].

We assume that a relational database is created with the following assumptions: (1) The ER methodology is followed, and all relationships are binary and one-to-many. Any primary key and any foreign key consists of exactly one column. (2) A

single and partial specialization (inheritance) between entity sets is allowed. (3) The final database schema is defined in SQL with the following integrity constraints: unique, not-null, primary key, and foreign key. Some combinations of primary keys and foreign keys will be referred to as *inheritance constraints* (see the definition of SId column in Student table).

Analysis of relational database schema suggests a possible translation of this schema into an ontology. Some natural semantic interpretations of schema elements and integrity constraints and their translation into extended DL knowledge base (or an OWL ontology), are:

- Both relation names and attribute type names become atomic concepts (class names).
- Attributes become atomic roles (object property names) connecting individuals in classes corresponding to relation names, with individuals in classes corresponding to attribute type names.
- Integrity constraints become TBox axioms, divided properly between standard TBox axioms and TBox ic-axioms.

The main novelty of this paper is that we propose a system of standard TBox axioms and TBox ic-axioms, such that the final dk-exchange system is both sound and complete with respect to semantics preservation. We show how consistency of the resulting relational knowledge base can be verified.

The structure of the paper is as follows. In Section II, we review some definitions concerning relational databases (RDB). In Section III, we discuss a relational knowledge base and translation of RDB to RKB. Dk-exchange system and its fundamental property concerning semantics preservation is discussed in Section IV. Checking consistency of RKB is discussed in Section V. Section VI concludes the paper.

## II. RELATIONAL DATABASES

### A. Schemas and instances

A (relational) *database schema* (*db-schema*) is a pair  $(\mathbf{R}, \mathbf{IC})$ , where  $\mathbf{R} = \{R_1, \dots, R_n\}$  is a *relational schema* consisting of a set of *relation symbols*, and  $\mathbf{IC}$  is a set of *integrity constraints* over  $\mathbf{R}$ . Each *relation symbol*  $R \in \mathbf{R}$  has a *type*, which is a nonempty finite set  $U$  of *attributes*, denoted  $att(R) = U$ . Without loss of generality, we can assume that types of relation symbols are pairwise disjoint (this can be achieved, for example, by prefixing attribute names with relation symbols),  $arity(R)$  denotes the cardinality of  $att(R)$ .

Let  $\mathbf{Const}$  be a countable infinite set of *constants*, and  $\mathbf{NULL}$  be a reserved symbol not in  $\mathbf{Const}$ .

*Definition 2.1:* An *instance*  $I$  of a relational schema  $\mathbf{R}$  is a finite multiset of *facts* (or *ground atoms*) of the form  $R(A_1 : c_1, \dots, A_m : c_m)$ , where  $R \in \mathbf{R}$ ,  $att(R) = \{A_1, \dots, A_m\}$ , and  $c_i \in \mathbf{Const} \cup \{\mathbf{NULL}\}$ ,  $1 \leq i \leq m$ . By  $Inst(\mathbf{R})$  is denoted the set of all instances of  $\mathbf{R}$ . The set of all values from  $\mathbf{Const} \cup \{\mathbf{NULL}\}$  occurring in  $I$  is referred to as the *active domain* of  $I$  and is denoted  $adom(I)$ .  $\square$

An instance of  $R$  corresponds to the notion of Codd table [4]. Since an instance is a multiset, thus the concept of duplicates

within a table is captured (as in SQL databases). Because the NULL value can occur in instances, the three-valued logic is used in SQL databases, where besides TRUE and FALSE, the UNKNOWN logical value is also needed [4].

*Definition 2.2:* Let  $\mathcal{DB} = (\mathbf{R}, \mathbf{IC}, I)$  be a (relational) database with a db-schema  $(\mathbf{R}, \mathbf{IC})$  and an instance  $I$  of  $\mathbf{R}$ .  $\mathcal{DB}$  is consistent, if  $I$  satisfies (is a model of) all integrity constraints, denoted  $I \models \mathbf{IC}$ . Otherwise we say that  $\mathcal{DB}$  is inconsistent.  $\square$

### B. Integrity constraints

Integrity constraints in databases play a dual role. On the one hand, they describe all possible worlds, and, on the other hand, they describe the allowed states of the database [4]. Integrity constraints can be used in data reasoning tasks, such as checking the correctness of database data, as well as in schema reasoning tasks, such as computing query containment (or subsumption).

We assume that  $\mathbf{IC} = \text{Unique} \cup \text{NotNull} \cup \text{PKey} \cup \text{FKey} \cup \text{Inherit}$ , where: *Unique*, *NotNull*, *PKey*, *FKey*, and *Inherit* are sets of, respectively, *unique*, *not-null*, *primary key*, *foreign key*, and *inheritance* integrity constraints.

Let  $\mathbf{R}$  be a relational schema,  $R \in \mathbf{R}$  be a relation symbol of type  $\text{att}(R) = \{A_1, \dots, A_m\}$ , and  $A_k \in \text{att}(R)$ .

1) **Unique integrity constraint:** A *unique integrity constraint* is an expression of the form  $\text{unique}(R, A_k)$ . An instance  $I$  of  $\mathbf{R}$  is consistent with  $\text{unique}(R, A_k)$ , iff for every  $i$ ,  $0 \leq i \leq m$ ,  $I$  satisfies the formula

$$R(t_1) \wedge R(t_2) \wedge t_1.A_k = t_2.A_k \wedge t_1.A_i \neq \text{NULL} \wedge t_2.A_i \neq \text{NULL} \Rightarrow t_1.A_i = t_2.A_i. \quad (1)$$

Note that the occurrence of NULLs affects the interpretation of the satisfaction of  $\text{unique}(R, A)$ . An instance  $I$  with NULLs is consistent with  $\text{unique}(R, A)$ , if there is such a replacement of NULLs with constants (every occurrence of NULL can be replaced with different constant) that the result instance is consistent with  $\text{unique}(R, A)$ .

*Example 2.3:* Let  $R(A, B)$  denote a relation symbol  $R$  of type  $\{A, B\}$  and  $\text{unique}(R, A)$  be a unique integrity constraint. Let  $I_1, I_2$  (Figure 3) be instances of  $R$ . Then,  $I_1$  is consistent with  $\text{unique}(R, A)$ , while  $I_2$  is not.

A	B
a	b
a	b
a	NULL
NULL	c
NULL	NULL

A	B
a	b
a	c
NULL	b
NULL	c
NULL	NULL

Fig. 3. Consistent,  $I_1$ , and inconsistent,  $I_2$ , instances of  $R(A, B)$  with respect to  $\text{unique}(R, A)$

2) **Not-null integrity constraint:** A *not-null integrity constraint* is an expression of the form  $\text{nonnull}(R, A_k)$ . An instance  $I$  of  $\mathbf{R}$  is consistent with  $\text{nonnull}(R, A_k)$ , if for any fact  $R(t) \in I$ ,  $t.A_k$  is a constant, i.e., iff  $I$  satisfies the formula

$$R(t) \Rightarrow t.A_k \neq \text{NULL}.$$

3) **Primary key integrity constraint:** A *primary key integrity constraint* is an expression of the form  $\text{pkey}(R, A_k)$ . An instance  $I$  of  $\mathbf{R}$  is consistent with  $\text{pkey}(R, A_k)$  iff it is consistent with  $\text{unique}(R, A_k)$ , and  $\text{nonnull}(R, A_k)$ .

4) **Foreign key integrity constraint:** Let  $R, R' \in \mathbf{R}$ ,  $A \in \text{att}(R)$ , and  $A' \in \text{att}(R')$ . A *foreign key integrity constraint* is an expression of the form  $\text{fkey}(R, A, R', A')$ . An instance  $I$  of  $\mathbf{R}$  is consistent with  $\text{fkey}(R, A, R', A')$  iff  $I$  satisfies  $\text{unique}(R', A')$ , and  $I$  satisfies the following existence constraint

$$R(t) \wedge t.A \neq \text{NULL} \Rightarrow \exists t'. (R'(t') \wedge t'.A' = t.A). \quad (2)$$

5) **Inheritance constraint:** A pair  $(\text{pkey}(R, A), \text{fkey}(R, A, R', A'))$  is referred to as an *inheritance constraint*. It specifies that some entities in a class  $C_{R'}$  corresponding to the table  $R'$ , can be *specialized* in a class  $C_R$  corresponding to the table  $R$  (or that each entity in  $C_R$  *inherits* from, exactly one, entity in  $C_{R'}$ ). An instance  $I$  of  $\mathbf{R}$  is consistent with  $(\text{pkey}(R, A), \text{fkey}(R, A, R', A'))$  iff it is consistent with  $\text{pkey}(R, A)$  and  $\text{fkey}(R, A, R', A')$ .

For example, (see Figure 1), the inheritance constraint

$$(\text{fkey}(\text{Student}, \text{Sid}), \text{fkey}(\text{Student}, \text{Sid}, \text{Person}, \text{Pid}))$$

specifies that a student is uniquely identified by  $\text{Sid}$ , and simultaneously  $\text{Sid}$  refers to a person identified by  $\text{Pid}$ . It means, that description of a person is contained in part in the *Person* table, and in part in the *Student* table (we say that *Person* objects are *specialized* as *Student* objects). Then, the composition  $(s, p)$  of a tuple  $s$  describing a student and a corresponding tuple  $p$  describing a person (such that  $s.\text{Sid} = p.\text{Pid}$ ) fully describes a person under consideration.

## III. RELATIONAL KNOWLEDGE BASES

To carry out reasoning procedures about properties of relational databases (such as preservation of semantics during data exchange) it is useful to represent the relational database by means of a knowledge base. However, a traditional DL knowledge base understood as a pair  $(\mathcal{T}, \mathcal{A})$  is unable to model integrity constraints [6]. The reason is two-fold: firstly, axioms in  $\mathcal{T}$  are interpreted under the standard first-order semantics and are treated as deductive rules and not as checks, and secondly, the UNA (*Unique Name Assumption*) is not accepted in general in DL knowledge bases, it means that two different individual names can denote the same individual.

### A. Syntax of relational knowledge bases

In this section, we define a *relational knowledge base* (RKB) that is a DL knowledge base adequately representing a relational database. RKB is based on the concept of EKB [6]. We propose and discuss a system of TBox axioms, which properly represents a relational database defined in the previous section.

*Definition 3.1:* A *relational knowledge base* is a tuple  $\mathcal{R} = (\mathcal{N}, \mathcal{S}, \mathcal{C}, \mathcal{A})$ , where:

- 1)  $\mathcal{N}$  is the *vocabulary* (or *signature*) of  $\mathcal{R}$ , consisting of a set  $\mathcal{N}_{\text{Ind}}$  of *individual names*, a set  $\mathcal{N}_{\text{Cl}}$  of *class*

names (or atomic concepts), a set  $\mathcal{N}_{OP}$  of object property names (or atomic roles).

- 2)  $\mathcal{S}$  is a finite set of standard TBox axioms, which are treated as deductive rules and can infer new assertions (see Table I).
- 3)  $\mathcal{C}$  is a finite set of integrity constraint TBox axioms. These axioms act as checks and are used to check whether a set of assertions implied by  $\mathcal{A}$  and  $\mathcal{S}$ , satisfies conditions specified by  $\mathcal{C}$  (see Table II). Axioms in  $\mathcal{C}$  cannot imply new assertions.
- 4)  $\mathcal{A}$  is a set of ABox assertions, which consists of factual knowledge about universe of discourse, i.e., properties of individual objects.

**Creating vocabulary.** Let  $(\mathbf{R}, \mathbf{IC})$  be a db-schema,  $\mathbf{Const}$  be a countably infinite set of constants, and  $\Delta_{\text{Var}}$  be a countable infinite set of labeled nulls disjoint from the set of constants; labeled nulls, denoted  $X, X_1, X_2, \dots$ , are used as "fresh" Skolem terms, which are placeholders for unknown values, and can thus be seen as variables [19]. Then the vocabulary  $\mathcal{N} = \mathcal{N}_{Cl} \cup \mathcal{N}_{OP} \cup \mathcal{N}_{Ind}$ , is created as follows:

- 1) Distinguished class names Tuple, and Val, are inserted into  $\mathcal{N}_{Cl}$ . These names are independent of a concrete db-schema: Tuple is a class of individual names called *tuple names* (or *tuples*), and Val is a class of individual names called *attribute value names* (or *attribute values*).
- 2) For each relational symbol  $R \in \mathbf{R}$ , there is a class name  $C_R \in \mathcal{N}_{Cl}$ ; this class is a subclass of Tuple,  $C_R \sqsubseteq \text{Tuple}$ .
- 3) For each attribute  $A \in \text{att}(R)$ ,  $R \in \mathbf{R}$ , there is a class name  $C_A \in \mathcal{N}_{Cl}$  being a subclass of Val ( $C_A \sqsubseteq \text{Val}$ ), and an object property name  $P_A \in \mathcal{N}_{OP}$ ; the object property  $P_A$  connects a tuple in  $C_R$  with an attribute value in  $C_A$ .
- 4) A set  $\mathcal{N}_{Ind}$  of individual names consists of the union of  $\mathbf{Const}$  and  $\Delta_{\text{Var}}$ . We assume that the UNA does not hold.

**Creating standard TBox axioms.** The set  $\mathcal{S}$  of standard TBox axioms is given in Table I. Since each axiom in  $\mathcal{S}$  is a deductive rule then the assertion on the right hand side of any FOL rule is inserted into the set  $\mathcal{A} \cup \mathcal{S}$ , i.e., into the closure of  $\mathcal{A}$  with respect to  $\mathcal{S}$ .

TABLE I. STANDARD TBOX AXIOMS OF RELATIONAL KNOWLEDGE BASE

	Constraints of relational db	DL	FOL
S1	$R \in \mathbf{R}$	$C_R \sqsubseteq \text{Tuple}$	$C_R(x) \Rightarrow \text{Tuple}(x)$
S2	$A \in \text{att}(R), R \in \mathbf{R}$	$C_A \sqsubseteq \text{Val}$	$C_A(v) \Rightarrow \text{Val}(v)$
S3	range of $P_A$	$\exists P_A^- \sqsubseteq C_A$	$P_A(x, v) \Rightarrow C_A(v)$
S4	domain of $P_A$	$\exists P_A^- \sqsubseteq C_R$	$P_A(x, v) \Rightarrow C_R(x)$
S5	$\text{unique}(R, A)$ unique constraint	(func $P_A^-$ )	$P_A(x_1, v_1) \wedge P_A(x_2, v_2)$ $\wedge v_1 = v_2 \Rightarrow x_1 = x_2$
S6	$(pkey(R, A),$ $fkey(R, A, R', A'))$ inheritance constraint	$P_A \sqsubseteq P_{A'}$	$P_A(x, v) \Rightarrow P_{A'}(x, v)$

(S1) and (S2) belong to translation of facts that  $R \in \mathbf{R}$  and  $A \in \text{att}(R)$ ; they say that appropriate tuple names and attribute value names are to be inserted into, respectively, classes Tuple and Val. (S3) and (S4) belong to translation of the fact that  $A \in \text{att}(R)$ , where: (S3) defines the range

of the object property  $P_A$  corresponding to an attribute  $A$  of the relational symbol  $R$ , and (S4) defines the domain of this object property. (S5) is the result of translation of a unique constraint  $\text{unique}(R, A)$ , and enforces equality between  $x_1$  and  $x_2$ , if hold all  $P_A(x_1, v_1)$ ,  $P_A(x_2, v_2)$ , and  $v_1 = v_2$  ( $A$  in  $R$  has the key property for not null values). (S6) results of the translation of an inheritance constraint of the form  $\{pkey(R, A), fkey(R, A, R', A')\}$ , and says that extension of  $P_A$  is to be inserted into the extension of  $P_{A'}$ .

**Creating integrity constraint TBox axioms.** The set  $\mathcal{C}$  of TBox ic-axioms is given in Table II. Note that ic-axioms are checks, so we expect that the value of such an axiom is either TRUE or FALSE.

TABLE II. INTEGRITY CONSTRAINT TBOX AXIOMS OF RELATIONAL KNOWLEDGE BASE

	Constraints of relational db	DL	FOL
C1	disjointness	$\text{Tuple} \sqsubseteq \neg \text{Val}$	$\text{Tuple}(x) \Rightarrow \neg \text{Val}(x)$
C2	$A \in \text{att}(R), R \in \mathbf{R}$ functional property	(func $P_A$ )	$P_A(x_1, v_1)$ $\wedge P_A(x_2, v_2)$ $\wedge x_1 = x_2 \Rightarrow v_1 = v_2$
C3	$\text{nonnull}(R, A)$ not-null property	$C_R \sqsubseteq \exists P_A$	$C_R(x) \Rightarrow \exists v. P_A(x, v)$
C4	$fkey(R, A, R', A')$ foreign key	$\exists P_A^- \sqsubseteq \exists P_{A'}$	$P_A(x_1, v)$ $\Rightarrow \exists x_2. P_{A'}(x_2, v)$
C5	$(pkey(R, A),$ $fkey(R, A, R', A'))$ inheritance constraint	$C_R \sqsubseteq C_{R'}$	$C_R(x) \Rightarrow C_{R'}(x)$

(C1) tests disjointness of Tuple and Val. (C2) belongs to translation of  $A \in \text{att}(R)$  and checks if  $P_A$  has the functional property. (C3) is the result of translation of  $\text{nonnull}(R, A)$  constraint and tests if any tuple name in  $C_R$  is connected via  $P_A$  with some value name. (C4) is the result of translation of a foreign key constraint,  $fkey(R, A, R', A')$ , and tests the inclusion of the set of value names in the "second column" of  $P_A$  in the set value names in the "second column" of  $P_{A'}$ . (C5) belongs to the result of the translation of  $\{pkey(R, A), fkey(R, A, R', A')\}$ , and tests the inclusion of  $C_R$  in  $C_{R'}$ .

**Creating ABox assertions** ABox assertions are expressions of the form:  $C(a), P(a_1, a_2), a_1 = a_2$ , where  $C \in \mathcal{N}_{Cl}$ ,  $P \in \mathcal{N}_{OP}$ , and  $a, a_1, a_2 \in \mathcal{N}_{Ind}$ . Translation of an instance  $I$  of  $\mathbf{R}$  can be performed using Algorithm 1

#### Algorithm 1 Creating ABox assertions

**Input:** Instance  $I$  of  $\mathbf{R}$ , and an empty ABox  $\mathcal{A}$ .  
**Output:** ABox assertions in  $\mathcal{A}$  representing  $I$ .  
**for each**  $R(t) \in I$   
      $U_{R,t} := \{A \in \text{att}(R) \mid t.A \neq \text{NULL}\}$   
      $X :=$  a fresh labeled null in  $\Delta_{\text{Var}}$   
     **if**  $U_{R,t} = \emptyset$  **then**  
          $\mathcal{A} := \mathcal{A} \cup C_R(X)$   
     **else**  
         **for each**  $A \in U_{R,t}$   
              $\mathcal{A} := \mathcal{A} \cup P_A(X, t.A)$   
     **end**

#### B. Semantics of relational knowledge bases

The semantics of  $\mathcal{R}$  is defined in the standard way [13] by means of the interpretation  $\mathcal{I} = \langle \Delta^{\mathcal{I}}, \mathcal{I}^{\mathcal{I}} \rangle$ , where  $\Delta^{\mathcal{I}}$  is a non-empty set of individuals, and  $\mathcal{I}^{\mathcal{I}}$  is an interpretation function

assigning to every vocabulary element (i.e., to every individual name, class name or property name) elements in  $\Delta^{\mathcal{I}}$ , subsets of  $\Delta^{\mathcal{I}}$ , or binary relations over  $\Delta^{\mathcal{I}}$ , respectively. Semantics of complex expressions is defined inductively in a set-theoretical manner based on semantics of simpler expressions.

The difference between standard TBox axioms in  $\mathcal{S}$  and TBox ic-axioms in  $\mathcal{C}$ , appears mainly in definition of consistency [6]. A relational knowledge base  $\mathcal{R} = (\mathcal{N}, \mathcal{S}, \mathcal{C}, \mathcal{A})$  is *consistent*, if

- 1)  $\mathcal{A} \cup \mathcal{S}$  has a model, i.e., there is at least one interpretation  $\mathcal{I}$  in which all formulas belonging to  $\mathcal{A} \cup \mathcal{S}$  are satisfied, written  $\mathcal{I} \models \mathcal{A} \cup \mathcal{S}$ ; and
- 2) any minimal Herbrand model of  $\mathcal{A} \cup \mathcal{S}$  is also a model of  $\mathcal{C}$ , i.e., the entailment  $\mathcal{A} \cup \mathcal{S} \models_{mHm} \mathcal{C}$  holds.

Intuitively, (1) says that there is no any contradiction between assertions in  $\mathcal{A}$  and the general knowledge in  $\mathcal{S}$ . In (2), a Herbrand model of  $\mathcal{A} \cup \mathcal{S}$  contains in fact *all the knowledge base knows*. This model is a finite set of assertions since  $\mathcal{A}$  is finite (see Algorithm 1), and axioms in  $\mathcal{S}$  are existential-free and without functional symbols (see Table I).

#### IV. DATA-TO-KNOWLEDGE EXCHANGE SYSTEMS

A *data-to-knowledge exchange (dk-exchange system)* systems we define as a mapping  $\mathcal{M} = (\tau, \Sigma)$ , from relational databases to relational knowledge bases, such that for each db-schema  $(\mathbf{R}, \mathbf{IC})$  and every instance  $I$  of  $\mathbf{R}$ ,

$$\mathcal{M}(\mathbf{R}, \mathbf{IC}, I) = (\tau(\mathbf{R}, \mathbf{IC}), \Sigma(I)) = (\mathcal{N}, \mathcal{S}, \mathcal{C}, \mathcal{A}),$$

where  $\tau(\mathbf{R}, \mathbf{IC}) = (\mathcal{N}, \mathcal{S}, \mathcal{C})$ , and  $\Sigma(I) = \mathcal{A}$ .

One of the most challenging issues in dk-exchange is to show that the semantics of the source data is not lost by the transformation into a knowledge base. The preservation of semantics of a dk-exchange system  $\mathcal{M} = (\tau, \Sigma)$  can be understood in two ways:

1. *Soundness of dk-exchange systems with respect to semantics preservation.*  $\mathcal{M} = (\tau, \Sigma)$  is *sound* w.r.t. *semantics preservation* if every consistent database  $(\mathbf{R}, \mathbf{IC}, I)$  is transformed into a consistent relational knowledge base  $(\mathcal{N}, \mathcal{S}, \mathcal{C}, \mathcal{A})$ , i.e.,

$$I \models \mathbf{IC} \wedge \tau(\mathbf{R}, \mathbf{IC}) = (\mathcal{N}, \mathcal{S}, \mathcal{C}) \wedge \Sigma(I) = \mathcal{A} \Rightarrow \mathcal{A} \cup \mathcal{S} \models_{mHm} \mathcal{C}.$$

2. *Completeness of dk-exchange systems with respect to semantics preservation.*  $\mathcal{M} = (\tau, \Sigma)$  is *complete* w.r.t. *semantics preservation* if every inconsistent database  $(\mathbf{R}, \mathbf{IC}, I)$  is transformed into an inconsistent relational knowledge base  $(\mathcal{N}, \mathcal{S}, \mathcal{C}, \mathcal{A})$ , i.e.,

$$I \not\models \mathbf{IC} \wedge \tau(\mathbf{R}, \mathbf{IC}) = (\mathcal{N}, \mathcal{S}, \mathcal{C}) \wedge \Sigma(I) = \mathcal{A} \Rightarrow \mathcal{A} \cup \mathcal{S} \not\models_{mHm} \mathcal{C}.$$

We can show that a dk-exchange system  $\mathcal{M} = (\tau, \Sigma)$ , where: (1)  $\tau$  creates a vocabulary  $\mathcal{N}$ , a set of standard TBox axioms  $\mathcal{S}$ , a TBox ic-axioms  $\mathcal{C}$ ; and (2)  $\Sigma$  creates ABox  $\mathcal{A}$ , in the way described in Section III-A, is both sound and complete w.r.t. semantics preservation ([9]).

Let  $I$  be an instance of  $\mathbf{R}$  (satisfying or not integrity constraints in  $\mathbf{IC}$ ). Let  $\mathcal{M}(\mathbf{R}, \mathbf{IC}, I) = (\mathcal{N}, \mathcal{S}, \mathcal{C}, \mathcal{A})$ . Let us denote  $\mathcal{S}_0 = \{S1, S2, S3\}$ , and  $\mathcal{C}_0 = \{C1\}$ . Then the following properties can be shown:

- 1)  $\mathcal{M}$  such that  $\mathcal{S} = \mathcal{S}_0 \cup \{S5\}$  and  $\mathcal{C} = \mathcal{C}_0 \cup \{C2\}$  is sound and complete w.r.t. unique integrity constraint preservation.
- 2)  $\mathcal{M}$  such that  $\mathcal{S} = \mathcal{S}_0 \cup \{S4\}$  and  $\mathcal{C} = \mathcal{C}_0 \cup \{C3\}$  is sound and complete w.r.t. not-null integrity constraint preservation.
- 3)  $\mathcal{M}$  such that  $\mathcal{S} = \mathcal{S}_0 \cup \{S4, S5\}$  and  $\mathcal{C} = \mathcal{C}_0 \cup \{C2, C3\}$  is sound and complete w.r.t. primary key integrity constraint preservation.
- 4)  $\mathcal{M}$  such that  $\mathcal{S} = \mathcal{S}_0 \cup \{S5\}$  and  $\mathcal{C} = \mathcal{C}_0 \cup \{C2, C4\}$  is sound and complete w.r.t. foreign key integrity constraint preservation.
- 5)  $\mathcal{M}$  such that  $\mathcal{S} = \mathcal{S}_0 \cup \{S4, S5, S6\}$  and  $\mathcal{C} = \mathcal{C}_0 \cup \{C2, C3, C4, C5\}$  is sound and complete w.r.t. inheritance integrity constraint preservation.

#### V. CHECKING CONSISTENCY OF RELATIONAL KNOWLEDGE BASES

Now, we will consider the problem of verifying whether a given relational knowledge base is consistent. The problem is significant if data from many databases is transformed into a repository in a form of DL assertions, and we are interested if this repository poses a "relational nature", i.e., whether satisfies a set of standard- and ic-axioms being result of some dk-exchange. We will use the *chase procedure* that was developed as a method for investigating data dependencies [4] and data exchange [19]. This will be done in two steps: (1) We chase a minimal Herbrand model  $\mathcal{H}$  of  $\mathcal{A} \cup \mathcal{S}$ . Since  $\mathcal{S}$  is existential-free (thus it is also *weakly acyclic*), and with no function symbols – we can obtain a finite solution in a polynomial time. (2) We will test satisfaction of  $\mathcal{C}$  in  $\mathcal{H}$ .

The chase producing  $\mathcal{H}$  from  $\mathcal{A}$  is defined by Algorithm 2. By  $h$  we denote a *homomorphism* that maps constants to themselves and variables into the set  $\mathbf{Const} \cup \Delta_{\mathbf{Var}}$ .

#### Algorithm 2 Chasing a minimal Herbrand model of $\mathcal{A} \cup \mathcal{S}$

---

*Input:*  $\mathcal{A}$  – a set of assertions;  
 $\mathcal{S}$  – a set of standard TBox axioms.  
*Output:*  $\mathcal{H}$  – a minimal Herbrand model of  $\mathcal{A} \cup \mathcal{S}$ .

- (1)  $\mathcal{H}_0 := \mathcal{A}; i := 0;$
- (2) **for each**  $\xi := (\varphi(\mathbf{x}, \mathbf{v}) \Rightarrow \psi(\mathbf{x}, \mathbf{v})) \in \mathcal{S}$
- (3)     **while**  $\mathcal{H}_i \not\models \xi$  **do**
- (4)         **for each** homomorphism  $h$  over  $(\mathbf{x}, \mathbf{v})$
- (5)             **if**  $\mathcal{H}_i \models_h \varphi(\mathbf{x}, \mathbf{v})$
- (6)                 **do**  $\mathcal{H}_i \xrightarrow{\xi, h} \mathcal{H}_{i+1};$   
 $i := i + 1;$
- end**
- end**
- end**
- (7)  $\mathcal{H} := \mathcal{H}_{i+1}$

---

*Chase steps* – comments to Algorithm 2

- 1)  $\mathcal{H}_0$  is initiated with  $\mathcal{A}$  – line (1).
- 2) Let  $\xi \in \mathcal{S}$  be an axiom of the form  $\varphi(\mathbf{x}, \mathbf{v}) \Rightarrow \psi(\mathbf{x}, \mathbf{v})$  – line (2).
- 3) If  $\xi$  is not satisfied by  $\mathcal{H}_i$ , then there is a homomorphism  $h$  over  $(\mathbf{x}, \mathbf{v})$  mapping variables in  $\mathbf{x}$  into  $\Delta_{\mathbf{Var}}$  and variables in  $\mathbf{v}$  into  $\mathbf{Const}$ , such that

$\mathcal{H}_i \models h(\varphi(\mathbf{x}, \mathbf{v}))$  and  $\mathcal{H}_i \not\models h(\psi(\mathbf{x}, \mathbf{v}))$ . Then, we say that  $\xi$  is *applicable* to  $\mathcal{H}_i$  with  $h$  – lines (3–5).

- 4) The new state of  $\mathcal{H}$  is equal to  $\mathcal{H}_i \cup h(\psi(\mathbf{x}, \mathbf{v}))$ , and the result of *applying*  $\xi$  to  $\mathcal{H}_i$  with  $h$  is then denoted as  $\mathcal{H}_i \xrightarrow{\xi, h} \mathcal{H}_{i+1}$  – line (6).
- 5) The final result is denoted by  $\mathcal{H}$  – line (7).

The dk-chasing described by Algorithm 2 produces  $\mathcal{A} \cup \mathcal{S}$  as a closure of  $\mathcal{A}$  w.r.t.  $\mathcal{S}$ , and is denoted by  $\mathcal{A} \xrightarrow{\mathcal{S}} \mathcal{H}$ , meaning that all applicable axioms in  $\mathcal{S}$  have been applied to  $\mathcal{H}$ .

*Example 5.1:* Given a set of assertions

$$\mathcal{A} = \{Pid(X_1, 1), Name(X_1, john), Pid(X_2, 2), Name(X_2, ann), \\ Pid(X_3, 3), Name(X_3, eva), Course(X_4, db), Sid(X_5, 1), \\ Faculty(X_5, math), Sid(X_6, 3), EId(X_7, 1), ESid(X_7, 3), \\ Course(X_7, db), Grade(X_7, A)\},$$

verify if it is a representation of an instance of a university database in Figure 1. For this database schema we have

$$\mathbf{R} = \{P(erson), C(ourse), S(tudent), E(xam)\}, \\ \mathbf{IC} = \{pkey(P, Pid), pkey(C, CId), pkey(S, Sid), pkey(E, EId), \\ notnull(S, Faculty), fkey(S, Sid, P, Pid), \\ fkey(E, ESid, S, Sid), fkey(E, Course, C, CId)\},$$

and implicitly in  $\mathbf{IC}$  are integrity constraints implied by primary keys and foreign keys (i.e., *uniques* and *not-nulls*).

In result of translation, we have  $\tau(\mathbf{R}, \mathbf{IC}) = (\mathcal{N}, \mathcal{S}, \mathcal{C})$ , where  $\mathcal{S}$  is a set of standard TBox axioms of the form given in Table I, and  $\mathcal{C}$  is a set of TBox ic-axioms of the form given in Table II. We will use names, respectively,  $R$  and  $A$  to denote a class  $C_R$  and an object property  $P_A$ . In particular, (S4), (S5) and (S6) generate the following axioms:

- (a)  $PId(x, v) \Rightarrow P(x)$ ,
- (b)  $PId(x_1, v_1) \wedge PId(x_2, v_2) \wedge v_1 = v_2 \Rightarrow x_1 = x_2$ ,
- (c)  $SId(x, v) \Rightarrow PId(x, v)$ .

Then, in  $\mathcal{H}$ , being the result of the chase procedure, the following set of derived assertions is included:

$$\mathcal{A}' = \{P(X_1), P(X_2), P(X_3), C(X_4), S(X_5), S(X_6), E(X_7), \\ Pid(X_5, 1), Pid(X_6, 3), X_1 = X_5, X_3 = X_6)\}.$$

Now, it is easily seen that all ic-axioms generated by (C4) and (C5) are satisfied in  $\mathcal{H}$ . However, the ic-axiom

- (d)  $S(x) \Rightarrow \exists v. Faculty(x, v)$ ,

corresponding to  $notnull(S, Faculty) \in \mathbf{IC}$ , is not satisfied. Thus,  $(\mathcal{N}, \mathcal{S}, \mathcal{C}, \mathcal{A})$  do not correspond to any consistent instance of the database in Figure 1.  $\square$

## VI. CONCLUSION

In this paper, we studied the problem of representing relational databases in extended DL knowledge bases in a way that satisfies both information and semantics preservation. We mainly focused on semantics preservation meaning that a consistent database is mapped into consistent knowledge base (the soundness), and any inconsistent database is mapped into inconsistent knowledge base (the completeness). The importance of the problem is met while creating Web of Data integrating data coming from many distributed and heterogeneous sources

[20], [21], [22]. Then, the knowledge base (ontological) layer is used to reason about the data, the metadata and queries in data integration and data exchange procedures [1].

## REFERENCES

- [1] T. Pankowski, "Data Exchange Between Relational Knowledge Bases In The Web Of Data," in 15th International Conference on Enterprise Information Systems (ICEIS), July 04 - 07, 2013 - Angers, France. INSTICC Press, 2013, pp. 1–6.
- [2] A. Poggi, D. Lembo, D. Calvanese, G. De Giacomo, M. Lenzerini, and R. Rosati, "Linking data to ontologies," in Journal on Data Semantics X. Springer-Verlag, 2008, pp. 133–173.
- [3] J. D. Ullman, Principles of Database and Knowledge-Base Systems, Vol. I/II. Computer Science Press, 1988/1989.
- [4] S. Abiteboul, R. Hull, and V. Vianu, Foundations of Databases. Reading, Massachusetts: Addison-Wesley, 1995.
- [5] R. Reiter, "Towards a logical reconstruction of relational database theory," in On Conceptual Modelling. Perspectives from Artificial Intelligence, Databases, and programming Languages, 1982, pp. 191–233.
- [6] B. Motik, I. Horrocks, and U. Sattler, "Bridging the gap between OWL and relational databases," Journal of Web Semantics, vol. 7, no. 2, 2009, pp. 74–89.
- [7] J. Sequeda, S. H. Tirmizi, Ó. Corcho, and D. P. Miranker, "Survey of directly mapping SQL databases to the Semantic Web," Knowledge Eng. Review, vol. 26, no. 4, 2011, pp. 445–486.
- [8] J. Sequeda, M. Arenas, and D. P. Miranker, "On Directly Mapping Relational Databases to RDF and OWL (Extended Version)," CoRR, vol. abs/1202.3667, 2012, pp. 1–17.
- [9] T. Pankowski, "Semantics Preservation In Data-To-Knowledge Exchange From Relational Databases To Knowledge Bases." (submitted), 2013.
- [10] RDF Vocabulary Description Language 1.0: RDF Schema, 2004, www.w3.org/TR/rdf-schema/.
- [11] OWL 2 Web Ontology Language. Structural Specification and Functional-Style Syntax, 2009, www.w3.org/TR/owl-syntax.
- [12] M. Arenas, A. Bertails, E. Prud'hommeaux, and J. Sequeda, "A Direct Mapping of Relational Data to RDF," 2012, http://www.w3.org/TR/rdb-direct-mapping.
- [13] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. Petel-Schneider, Eds., The Description Logic Handbook: Theory, Implementation and Applications. Cambridge University Press, 2003.
- [14] F. M. Donini, M. Lenzerini, D. Nardi, W. Nutt, and A. Schaerf, "An epistemic operator for description logics," Artif. Intell., vol. 100, no. 1-2, 1998, pp. 225–274.
- [15] D. Calvanese, G. D. Giacomo, D. Lembo, M. Lenzerini, and R. Rosati, "Data complexity of query answering in description logics," Artif. Intell., vol. 195, 2013, pp. 335–360.
- [16] P. P. Chen, "The entity-relationship model - toward a unified view of data," ACM Transactions on Database Systems, vol. 1, no. 1, 1976, pp. 9–36.
- [17] R. Elmasri and S. B. Navathe, Fundamentals of Database Systems. Redwood City: The Benjamin/Cummings, 1994.
- [18] Unified Modeling Language. Resource Page, 2005, www.uml.org/.
- [19] R. Fagin, P. G. Kolaitis, R. J. Miller, and L. Popa, "Data exchange: semantics and query answering," Theor. Comput. Sci, vol. 336, no. 1, 2005, pp. 89–124.
- [20] T. Pankowski, "Using Data-to-Knowledge Exchange for Transforming Relational Databases to Knowledge Bases," in RuleML 2012, LNCS 7438. Springer, 2012, pp. 256–263.
- [21] G. Brzykcy, J. Bartoszek, and T. Pankowski, "Schema Mappings and Agents' Actions in P2P Data Integration System," Journal of Universal Computer Science, vol. 14, no. 7, 2008, pp. 1048–1060.
- [22] T. Pankowski, "Semantics preservation in schema mappings within data exchange systems," in Knowledge Engineering, Machine Learning and Lattice Computing with Applications, Revised Selected Papers, KES 2012, LNAI 7828. Springer, 2013, pp. 88–97.

# Predicting Early Students with High Risk to Drop Out of University using a Neural Network-Based Approach

Miguel Gil, Norma Reyes, María Juárez, Emmanuel Espitia, Julio Mosqueda and Myriam Soria

Information Technologies  
Universidad Tecnológica de León  
León, Guanajuato, México

e-mail: mgil@utleon.edu.mx, navila@utleon.edu.mx, mdjuarez@utleon.edu.mx, eespitia.celula@utleon.edu.mx, jmosqueda@utleon.edu.mx, msoria@utleon.edu.mx

**Abstract**—This research is focused on the use of an Artificial Neural Network-based prototype in order to measure and predict the probability that students drop out of university. This probability is calculated in an early stage when the students get enrolled in some of the university study programs. Once we obtain the results they are analyzed and compared in order to know how the factors affect to the model behavior and the predicted result. Finally, we describe how this research can assist student advisors by using a support tool that helps them to identify in a fast way those students with a high risk of dropping out of university, and help them before they quit school.

**Keywords**—artificial neural networks; students drop out; early prediction; university drop out

## I. INTRODUCTION

Dropping out of university is a common problem with students in México [1]. For governments, this is a worrisome situation since the state must provide all support to students in order to allow them to graduate with a professional college degree and advanced technical skills.

Providing higher education to all sectors of a nation's population means confronting social inequalities deeply rooted in history, culture and economic structure that influence an individual's ability to compete. Quality assurance in higher education has raised to the top of the policy agenda in many nations [2][3].

Some studies show that students' previous behavior is a good predictor of future behavior [4][5][6]. This way, the Universidad Tecnológica de León (UTL) applies a survey to know some aspects about the students' life. This survey involves questions about students' family and their relationship which each member, vices, health and frequent diseases, skills and assignments that are relevant to them finishing the program successfully. However, this survey many times is not used correctly, since it collects too much information and it is hard to understand for most of the teachers and advisors, since they are not trained as psychologists and by consequence they do not have this profile.

In the scholar model of the UTL, students are grouped and each group has a teacher that acts as an advisor for the group. Advisors are responsible to follow the academic performance of their groups. They do some activities related to the students' behavior in order to detect those students

with a high risk to drop out of university. The most valuable instrument that advisors have in order to detect those students with a high risk is the information provided by a personal interview with each student of the group. This is a highly time-consuming task, and it requires a lot of days even weeks since some groups have up to 30 students. Many times, students drop out of university before attending an interview with the advisor and the advisor is unable to detect students' problems. Detection of student's problems is a hard task since most of the teachers have not been trained as a psychologist. This way, the proposed solution is very important, since it can be used as an advice tool in order to help advisors to detect those students with a high risk to drop out of university, since the very first course.

In this research, a Back Propagation Artificial Neural Network-based system is proposed in order to measure the university students dropping out probability once they are enrolled in some university academic study program. This research was developed at the UTL and it was applied to students enrolled in the Information and Communication Technologies academic programs. In the second section, an introduction to educational data mining is described since this research is related with educational purposes but also, it involves data analysis and pattern discovery to estimate a future behavior. In the third section, the dataset used for prediction is described. In the fourth section, the neural network architecture is analyzed but also, the reason why that technique was used is described. In the fifth section, the prototype development is described. In the sixth section, the obtained results are discussed. Finally, in the seventh section, conclusions and future research are described.

## II. EDUCATIONAL DATA MINING

Into the educational field, there is a wide variety of problems concerning students' behavior and some kind of development using intelligent computational techniques and algorithms are well documented. For example, some researches are focused to discover if a student is cheating on an online assessment when he/she is taking it [7][8]. Others are focused on studying how students learn [9].

In all cases mentioned above, data mining plays a very important role in order to provide with solutions in each described scenario and it is called "educational data mining".

Educational data mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from the educational context [10].

Close works with this paper are the research works related with the students' successful academic performance. Nearly all works try to identify the main factors that could be used as the best predictors for successful students more than measure or build a computational predictive model [7][8][9][11][12]. However, that approach is complex to be applied for most of teachers in their advisory sessions, and they must follow the educational model including procedures, formats and surveys and its implementation may require a lot of effort and spent time since in the UTL, the number of factors to be considered to predict are too many. The closest work to this research that was found was [3]. In that research, a Bayesian classification model is proposed in order to classify students in one of two groups: performance or underperformance. This classification is made using a dataset that includes some students' features. However, in this research, since the number of features is considerably large and the features relationship is unknown, a Bayesian based-model was not used. Instead, an Artificial Neural Network-based (ANN) [13] approach was used since it is a more soft-classification technique and they are widely applied in pattern recognition and prediction problems when data nature, mean or relationships between them are unknown [13].

### III. THE TRAINING DATASET

The dataset used in this research includes many features from students. Those features are collected from a survey to all new students when they get enrolled in the UTL. The survey features include information about high school name and type (government or private), high school final grade, vices and addictions (alcohol, smoke, drugs, etc.), common diseases, relationship with parents, brothers and family, different kinds of problems: economics, transport, delinquency, etc. The dataset had 302 total historic records with 34 features. The dataset was split in two parts: 219 records were used to train the neural network and the remaining 83 records were used in prediction. The used records correspond to historical data in order to allow measuring the neural network accuracy. Fig. 1 shows the factors taken from the initial survey, which students previously answered when they get enrolled into the university.

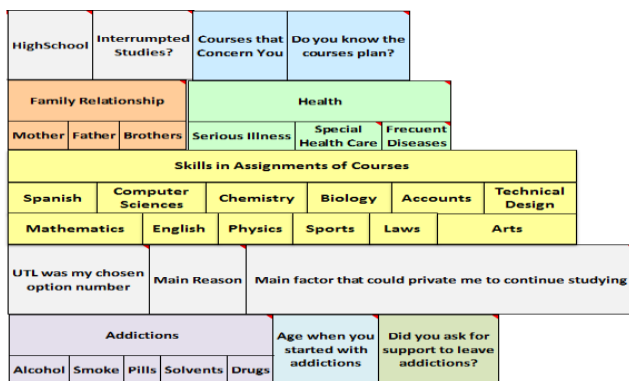


Figure 1. Survey factors that are used in order to predict graduate students' probability.

The survey is answered by new students when they get enrolled in the UTL. It has been applied since year 2000 and it was applied on paper. Only until year 2010, the survey began to be applied on a web-based system. This means that a large number of surveys remained on paper. However, many of them no longer exist because teachers and advisors got rid of them every three years. For this reason, only 302 surveys could be found for the Information Technologies area. This scenario is similar for the other areas in the UTL.

### IV. NEURAL NETWORKS

#### A. Neural Network Architecture

The main ANNs' advantage is learning capability. ANNs are very important in many technological fields; for example, biometric recognition sensors (fingerprints, retinal scanners, etc.), pattern recognition (handwriting recognition, face recognition, etc.) [14][15][16][17][18]. Neural networks have demonstrated they are functional because they have achieved good results. For example, in diagnosis of breast cancer, neural networks performance was in 89% to detect positive cases and the average for the same task performed by specialists was 84% [14]. Artificial neural networks are often compared with logistic regression, and it can be seen that both models perform on about the same level more often than not, with the more flexible neural networks generally outperforming logistic regression in the remaining cases [23].

A wide range of types of neural networks has been developed to date; for example, Radial Base [27][28][29], Hopfield [30][31][32][33] and Back Propagation (with many variants) [13][34]. Many of them are based on the basic Perceptron ANN [19][20][21][22]. In this research, a multilayer neural network with error back propagation and momentum was used [22][34].

#### B. Neural Network Design

The neural network input layer was designed using all dataset features and one additional threshold neuron (33 + 1 neurons). The final number of nodes for the hidden layer was 69. One layer of hidden neurons is generally sufficient for classifying most data sets. The number of neurons in the hidden layer needs to be set empirically, e.g., by cross-validation or bootstrapping [23]. The output layer has only one neuron with two limit values: 0 and 1, since only two possibilities are considered to measure the final status for the students when they leave the university. The two possible scenarios for any student that leaves the university are: 1: if the student successfully finished his studies and gets graduated; 0: if the student drops out the university. However, an interesting approach with ANN's is that their output values may have intermediate values between 0 and 1, considering that output value as the probability that any student have to finish his studies. A higher value in the output neuron means that the student has major probability to successfully finish his studies. For this research, if the neuron is activated (with a value major or equals to 0.50) is considered that the student will have a good probability to



finish his studies successfully and advisors could be focused firstly on those students for whom its output value was less than 0.5. Fig. 2 shows the neural network design architecture.

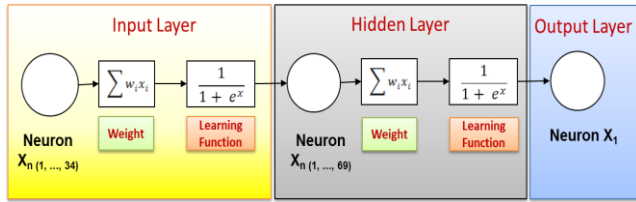


Figure 2. Neural network design architecture.

Fig. 2 shows the architecture design for the used neural network. In the model of an ANN, the most important part is the neuron. The neuron receives a set of input signals from the environment (each survey factor) or from another neuron.

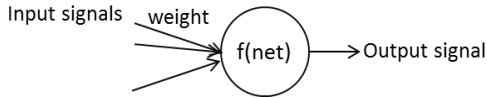


Figure 3. Illustration of an artificial neuron.

As shown in Fig. 3, an artificial neuron receives input signals, and each input signal has a weight. Also, each neuron computes a net input signal using the input signals and the weight of each input signal. Net input signal was calculated using Equation (1):

$$net = \omega_1 \chi_1 + \omega_2 \chi_2 + \dots + \omega_i \chi_i \tag{1}$$

where *net* is the net input signal,  $W_i$  is the current weight for each neuron and  $X_i$  is each input value. Once the net input signal was calculated, the output (also called activation signal) must be calculated. The output signal was calculated using a sigmoid function:

$$1 / (1 + e^{-\lambda \cdot (net - \theta)}) \tag{2}$$

where  $\lambda$  is a parameter to control the steepness of the function and usually equals to 1,  $q$  is a bias or threshold value. When simple neurons implement, for example, Boolean functions, it is easy to calculate them. But, when prior knowledge about the function is missing - except for data -, the  $\theta$  and  $\omega_i$  values are adjusted through training [13].

The neural network training was performed using the Neuroph Studio software [25]. The optimal neural network parameters were: *learning rate* = 0.1428; *momentum* = 0.8192; *training error* = 0.01. Fig. 4 shows the training performance.

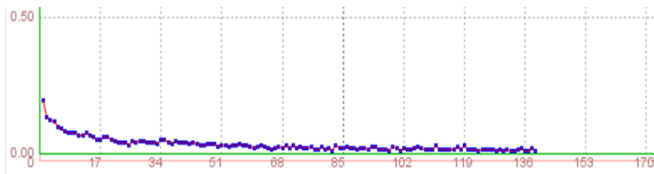


Figure 4. Neural network raining performance.

As it is shown in Fig. 4, approximately 135 iterations were needed in order to achieve the training error. After the neural network was trained, it was used with the test records in order to calculate and evaluate the results.

### V. PROTOTYPE BUILDING

In order to use the neural network designed and trained previously, a software prototype was developed. It was developed in the Java programming language [24]. Fig. 5 shows the main user interface.

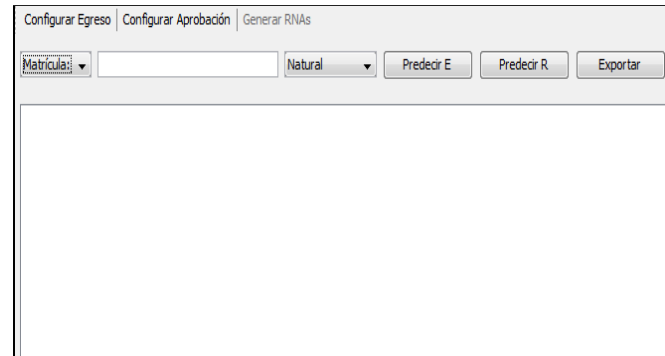


Figure 5. Prototype main user interface.

Into the main interface, the user must select and load the neural network file (created with Neuroph Studio in the neural network design stage). The prototype uses the Encog framework in order to manipulate and use the neural network [26].

Once the neural network is loaded, the next step is load a file containing the students' enrollment key field. Then, the system queries for those students' data into a database, in order to compute their success probability, and finally it will display a table containing all students' data including their success probability. Fig. 6 shows an example of a table results displayed by the program.

Mostrar Filtro Dinámico		Mostrar Detalles						
Información Tutorial de Alumnos								
	[29] Adicción Tabaco	[30] Adicción Pastillas	[31] Adicción Solventes	[32] Adicción Drogas	[33] Edad inicio consumo	[34] Apoyo dejar Adicción	[35] Probabilidad Egreso	
1	1	0	0	0	15	0	0.3581	
2	0	0	0	0	0	0	01.000	
3	1	0	0	0	17	0	01.000	
4	0	0	0	0	0	0	0.0000	
5	0	0	0	0	17	0	0.1177	
6	2	4	0	0	0	0	0.9996	
7	1	0	0	0	0	0	0.1175	
8	0	0	0	0	0	0	01.000	
9	2	0	0	0	0	0	0.0763	
10	0	0	0	0	18	0	0.3945	
11	0	2	0	0	12	0	0.0044	

Figure 6. Results table example.

Fig. 6 shows how the system calculates and presents the results. Each part of the survey is marked with a distinct

color and the last column represents the probability that students successfully finish their studies.

### VI. RESULTS

To measure the results, sample records were taken from the original dataset. Table 1 shows the sample is distributed.

TABLE I. SAMPLE DATASET SUMMARY

Gender	Finished Successfully	Dropout
Female	12	29
Male	18	24
<b>Total</b>	<b>30</b>	<b>53</b>

As it is shown in Table 1, the sample contains 41 records for women (12 finished successfully, 29 dropped out) and 42 records for men (18 finished successfully, 24 dropped out).

In a first-view, major drop outs are for women (29) and a woman is ranking less on successfully finishing her studies. In the sample, only 29% of women finished successfully their studies and the other 71% dropped out. For male data, 43% finished successfully their studies and the other 57% dropped out.

Unfortunately, the university does not know the factors that cause those students' drop out. The only evidence, in most cases is that the student fails to assist to his courses, but underlying causes remain unknown for nearly all cases. Teachers including the advisor know the causes why students that dropped out but, this only happens when students tell them why they are leaving the school. However, even in these cases, the cause or factor is not recorded. This is important to be mentioned because the students' data that we have in order to predict their probability to finish their studies successfully is the survey that they must answer once they are enrolled into the university programs.

Fig. 7 shows the results produced by the system using the dataset mentioned in Table 1. It is divided in two sections. On the left section, the results for students that drop out of university are shown. In the right section, the results of the students that finished their studies successfully are shown. That records was subtracted randomly from the original dataset and it was not used on the training stage. Record Id column is a unique id assigned to each record. Desired output column means the student real result: 0 if the student drops out, 1 if the student finished successfully. The output column shows the result produced by the system. System error column is calculated as:

$$\text{Desired output} - \text{System output} \tag{3}$$

Based on the obtained results by the program, reliability was in a 76%. That means that program failed to predict 20 cases and asserts to predict the remaining 63 cases.

Record Id	Desired Output	System Output	System Error	Gender	Record Id	Desired Output	System Output	System Error	Gender
1876	0	0.1085	-0.1085	M	1890	1	0.9994	0.0006	F
1826	0	1	-1	F	1857	1	0.1552	0.8448	F
1884	0	0.9994	-0.9994	F	1821	1	1	0	M
1855	0	0.9989	-0.9989	F	1901	1	0.9916	0.0084	M
1832	0	0.1175	-0.1175	F	1868	1	1	0	F
1825	0	0.1175	-0.1175	M	1899	1	1	0	M
1834	0	0	0	F	1858	1	0.9995	0.0005	M
1851	0	0	0	F	1894	1	0.1668	0.8332	F
1864	0	0.1174	-0.1174	F	1887	1	1	0	F
1867	0	0.1275	-0.1275	F	1829	1	0.0044	0.9956	M
1830	0	0.7853	-0.7853	F	1896	1	1	0	M
1846	0	0.9994	-0.9994	M	1822	1	0	1	M
1866	0	1	-1	M	1892	1	0.1199	0.8801	M
1838	0	0.3756	-0.3756	F	1895	1	0.9996	0.0004	M
1845	0	0.9996	-0.9996	M	1882	1	0.0882	0.9118	F
1842	0	0.1153	-0.1153	F	1878	1	0.0169	0.9831	M
1848	0	0.8919	-0.8919	F	1885	1	0.1183	0.8817	M
1833	0	0	0	M	1820	1	1	0	F
1880	0	0.1219	-0.1219	M	1893	1	1	0	M
1871	0	1	-1	M	1890	1	0.9994	0.0006	F
1859	0	1	-1	F	1823	1	0.1177	0.8823	F
1862	0	1	-1	M	1892	1	0.1199	0.8801	M
1849	0	0	0	M	1883	1	0.9898	0.0102	M
1852	0	0.1182	-0.1182	F	1897	1	1	0	F
1850	0	0.1247	-0.1247	M	1898	1	0.0002	0.9998	F
1853	0	0.1176	-0.1176	M	1858	1	0.9995	0.0005	M
1835	0	0.1182	-0.1182	F	1819	1	0.3581	0.6419	F
1869	0	0.9991	-0.9991	F	1874	1	0.9996	0.0004	M
1872	0	0.1114	-0.1114	F	1887	1	1	0	F
1839	0	0.9988	-0.9988	F	1857	1	0.1552	0.8448	F
1825	0	0.1175	-0.1175	M	1821	1	1	0	M
1828	0	0.3945	-0.3945	M	1886	1	0.9967	0.0033	M
1865	0	0.1179	-0.1179	F	1889	1	1	0	M
1879	0	0	0	F	1868	1	1	0	F
1863	0	0.117	-0.117	F	1861	1	0.211	0.789	F
1888	0	1	-1	M	1878	1	0.0169	0.9831	M
1824	0	0.9996	-0.9996	F	1894	1	0.1668	0.8332	F
1870	0	0.9276	-0.9276	M	1885	1	0.1183	0.8817	M
1841	0	0.1174	-0.1174	M	1881	1	0.9994	0.0006	M
1840	0	0	0	F	1891	1	0.5898	0.4102	M
1900	0	1	-1	M	1822	1	0	1	M
1836	0	0.9973	-0.9973	F	1892	1	0.1199	0.8801	M
1860	0	0.1178	-0.1178	M	1882	1	0.0882	0.9118	F
1856	0	1	-1	M	1883	1	0.9898	0.0102	M
1849	0	0	0	M	1897	1	1	0	F
1847	0	0.069	-0.069	F	1898	1	0.0002	0.9998	F
1854	0	0.0661	-0.0661	M	1820	1	1	0	F
1844	0	0.9902	-0.9902	F	1893	1	1	0	M
1877	0	0.0006	-0.0006	F					
1831	0	0.1475	-0.1475	M					
1872	0	0.114	-0.114	F					
1837	0	0.1175	-0.1175	F					
1827	0	0.0763	-0.0763	F					
1843	0	0.0005	-0.0005	M					
1873	0	0.1162	-0.1162	M					
1875	0	0.3821	-0.3821	M					
1849	0	0	0	M					

Figure 7. Results obtained by the system.

A failed prediction means that desired output was 0, and program output value was major or equals to 0.5. The 0.5 value is the output neuron's threshold activator and, while the output value is greater, the probability for students finish their studies successfully increased.

An interesting aspect in the survey is the "main factor" column. In this part, the students must select their main issue that could impede them to finish the school program. Fig. 8 shows the frequency of answers for this column.

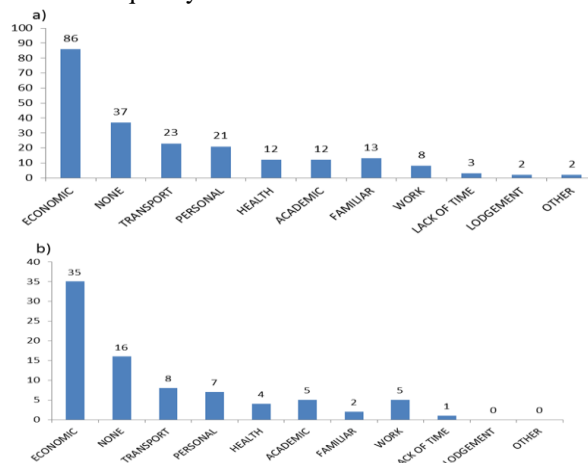


Figure 8. Frequency of responses for column "main factor".

Fig. 9a shows the frequency of answers in the training dataset, and Fig. 9b corresponds to the test dataset. In both cases, the major frequency is the economic factor. However, the second factor varies from males and females.

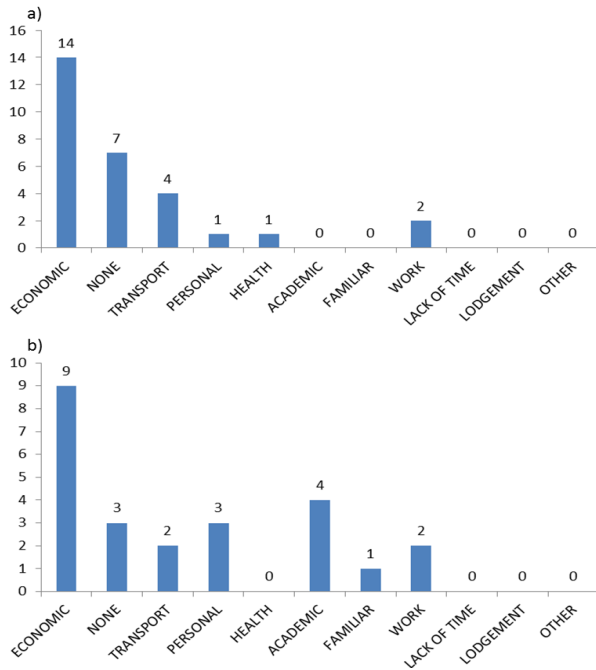


Figure 9. Drop out frequency of responses for column “main factor”.

As it is shown in Fig. 9, the second most frequent factor for drop out in female students was “none” (9a), while for males was “academic” (9b). On the skills part, frequencies are shown in Table 2.

TABLE II. SKILLS FREQUENCY FOR MEN AND WOMEN

Assignment skills frequency	Gender	
	Female	Male
Spanish Skill	97.56%	85.71%
Spanish Diifcult	2.44%	4.76%
Spanish No Answer	0.00%	4.76%
Computer Skill	73.17%	80.95%
Computer Diifcult	21.95%	9.52%
Computer No Answer	4.88%	4.76%
Mathematics Skill	41.46%	33.33%
Mathematics Diifcult	53.66%	59.52%
Mathematics No Answer	4.88%	2.38%
English Skill	34.15%	42.86%
English Diifcult	65.85%	50.00%
English No Answer	0.00%	2.38%
Technology Skill	56.10%	85.71%
Technology Diifcult	36.59%	4.76%
Technology No Answer	7.32%	4.76%

In Table 2, major frequency skills are in Spanish course, but more difficult courses are mathematics and English for both, men and women. The difficult issue for computers and technology are more frequent on women than on men. An important aspect here is that neural network cannot be trained if we skip the survey’s skills section. A deeper study is needed in order to understand how these factors alter the final result, but it requires some data mining techniques to achieve that goal since we are dealing with a considerably amount of data and fields.

VII. CONCLUSIONS AND FUTURE RESEARCH

This information is relevant in a first-stage, because it allows advisors to identify those students with a high risk to drop out of university from the beginning. For the cases where the program fails to predict the correct result, some inconsistencies in the survey answers were found. For example, some students indicated to have high ability in math, but in another survey section, they wrote comments indicating that they felt concern by math. Additionally, as seen on Fig. 8, in the results section, those women that dropped out the school and where the system predicted a high probability to finish successfully, the second major cause was “none”. This can be explained in an empirical way because teachers are noticed frequently by female students that they have unplanned pregnancies. This type of information cannot be inferred by the survey. Maybe a new question about the student’s sexual life could be useful to predict this type of drop outs. The UTL continues to support this research in order to implement this prototype as an active module on its scholar control system to be used by teachers and advisors as a complementary advice tool.

Future research will include modifications to the survey to add more information features. Is planned to add more neurons to the neural network output layer in order to have multiple outputs and, later calculate their average and use it as the result. Also, is necessary a deeper study in how the survey’s factors affect to the students behavior and their decision to leave the school.

ACKNOWLEDGMENT

Thanks to the Universidad Tecnológica de León by supporting this research.

REFERENCES

- [1] UTL, UTL statistics, Universidad Tecnológica de León, 2012, In: [http://www.utleon.edu.mx/index.php?option=com\\_content&view=article&id=229&Itemid=59](http://www.utleon.edu.mx/index.php?option=com_content&view=article&id=229&Itemid=59) [retrieved: July, 2013].
- [2] P. Altbach, L. Reisberg, and L. Rumbley, Trends in Global Higher Education: Tracking an Academic Revolution, UNESCO, 2009.
- [3] U. Kumar and S. Pal, “Data Mining : A prediction of performer or underperformer using classification,” International Journal of Computer Science and Information Technologies, vol. 2, Apr. 2011, pp. 686-690.
- [4] W. Camara and E. Kimmel, “Choosing students: Higer education admissions tools for the 21st century,” Lawrence Erlbaum Associates, Publishers, 2005, pp. 53–79.

- [5] K. Frome, W. Corbin, and M. Kruze, "Behavioral risks during the transition from high school to college. U.S.," National Institutes of Health's National Library of Medicine: Developmental Psychology, Sep. 2008, pp. 1497-1504.
- [6] C. Noyes, J. Gordon, and J. Ludlum, "Engagement Half-Life: The Impact of Incoming Student Characteristics throughout the College Career," Georgia Institute of Technology, Jun. 2006. In: <https://www.assessment.gatech.edu/documents/> [retrieved: July, 2013].
- [7] J. Hernández, A. Ochoa, J. Muñoz, and G. Burlak, "Detecting cheats in online student assessments using Data Mining," DMIN'06, Las Vegas, NV: Conference on Data Mining, 2006, pp. 204-210.
- [8] B. Whitley, "Factors associated with cheating among college students: A review," Springer, vol. 39, 1998, pp. 235-274. doi: 10.1023/A:1018724900565.
- [9] A. El-Halees, "Mining Students Data to Analyze e-learning Behavior: A Case Study," Islamic University of Gaza, 2009. In: [http://uqu.edu.sa/files2/tiny\\_mce/plugins/filemanager/files/30/papers/f158.pdf](http://uqu.edu.sa/files2/tiny_mce/plugins/filemanager/files/30/papers/f158.pdf) [retrieved: July, 2013].
- [10] C. Romero, S. Ventura, and E. García, "Data mining in course management systems: Moodle case study and tutorial," Computers & Education: Elsevier, vol. 51, Aug. 2008, pp. 368-384.
- [11] M. Pritchard and G. Wilson, "Predicting Academic Success in Undergraduates", Boise State University, vol. 11, Oct. 2007, pp. 201-206. In: [http://scholarworks.boisestate.edu/psych\\_facpubs/10/](http://scholarworks.boisestate.edu/psych_facpubs/10/) [retrieved: July, 2013].
- [12] S. Senanayaque, K. Liyanage, and P. Dadigamuwa, "Factors Affecting on Student Unsuccessfulness in Engineering Programmes in Distance Education," International Journal of Instructional Technology & Distance Learning, vol. 2 Jun. 2005. In: [http://www.itdl.org/Journal/Jun\\_05/article07.htm](http://www.itdl.org/Journal/Jun_05/article07.htm) [retrieved: July, 2013].
- [13] A. Engelbrecht, "Computational Intelligence, An Introduction," University of Pretoria South Africa; Wiley, 2007, pp. 7, 15, 17-53. 98-102.
- [14] Y. Wu, et al., "Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer." US National Library of Medicine, Apr. 1993, pp.81-87. In: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.122.5805&rep=rep1&type=pdf> [retrieved: July, 2013].
- [15] G. Gross, J. Boone, V. Greco-Hunt, and B. Greenberg, "Neural networks in radiologic diagnosis: II. Interpretation of neonatal chest radiographs", Investigative Radiology, vol. 25, Sep. 1990, pp. 1017-1023.
- [16] P. Lisboa, "A review of evidence of health benefit from artificial neural networks in medical intervention," Elsevier, vol. 15, Jan. 2002, pp. 11-39.
- [17] M. Joo, S. Wu, J. Lu, and H. Lye, "Face recognition with radial basis function (RBF) neural networks," IEEE Transactions on Neural Networks, vol. 13, Aug. 2002, pp. 697-710.
- [18] J. Lin, P. Ligomenides, M. Freedman, and S. Mun, "Application of artificial neural networks for reduction of false-positive detections in digital chest radiographs," Proceedings of annual symposium on Computers Applied to Medical Care, Nov. 1993, pp. 434-438.
- [19] L. Fausett, "Fundamentals of Neural Networks," Pearson, 1993, pp. 59-76.
- [20] C. Bishop, "Pattern Recognition and Machine Learning," Springer, 2006, pp. 225-272.
- [21] X. Basogain, "Redes neuronales artificiales y sus aplicaciones," Universidad del País Vasco, 2008, pp. 22-25.
- [22] K. Carbajal, "Introducción de la inteligencia artificial en la ingeniería civil a través de las redes neuronales artificiales," Congreso Nacional de Ingeniería Civil: ICG, 2003 In <http://www.bvsde.paho.org/bvsacd/coursouni/EM-53.pdf> [retrieved: July, 2013].
- [23] S. Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural network classification models: a methodology review," Journal of Biomedical informatics, vol. 35, Oct. 2002, pp. 352-359.
- [24] Oracle, Java platform web page, Oracle, 2013. In: <http://www.oracle.com/technetwork/es/java/javase/downloads/index.html> [retrieved: July, 2013].
- [25] Z. Sevarac, et al. Neuroph, Sourceforge, 2013. In: <http://neuroph.sourceforge.net/neuroph-2.5-released.html> [retrieved: July, 2013]
- [26] Heaton Research, Encog, Heaton Research, Inc., 2012. In: <http://www.heatonresearch.com/encog> [retrieved: July, 2013]
- [27] A. Sagahyoon and J. Abdalla, "Area and Timing Estimation in Register Files Using Neural Networks," Circuits and Systems, vol. 3, Jun. 2012, pp. 269-277, doi: 10.4236/cs.2012.33037.
- [28] A. Cai, X. Xiong, Y. Liu, W. An, J. Tan, "Artificial neural network modeling of reduced glass transition temperature of glass forming alloys," Applied Physics Letters, vol. 92, Mar. 2008, doi:10.1063/1.2899633.
- [29] G. Bingjun and Y. Jinshou, "A single-neuron PID adaptive multicontroller scheme based on RBFNN," Transactions Of The Institute Of Measurement and Control, vol. 274, pp. 243-259. doi:10.1191/0142331205tm147oa.
- [30] L. Maldonado, A. Peña, and O. Gualdrón, "Identificación Automática de Cilindros de Almacenamiento de Gas Utilizando Redes Neuronales Tipo Hopfield," Revista UIS Ingenierías, vol. 11 Jun. 2012, pp.103-111.
- [31] G. Pajares, C. López-Martínez, J. Sánchez-Lladó, and I. Molina, "Improving Wishart Classification of Polarimetric SAR Data Using the Hopfield Neural Network Optimization Approach," Remote Sensing, vol. 4, Nov. 2012, pp. 3571-3595, doi:10.3390/rs4113571.
- [32] E. Konishi, "Modeling quantum mechanical observers via neural-glia networks," International Journal Of Modern Physics B, vol. 26, Jul. 2012, pp. 1250060-1-1250060-24, doi: 10.1142/S0217979212501111.
- [33] A. Tatem, H. Lewis, P. Atkinson, and M. Nixon, "Increasing the spatial resolution of agricultural land cover maps using a Hopfield neural network," International Journal Of Geographical Information Science, vol. 17, Oct. 2003, pp. 647-672. doi:10.1080/1365881031000135519.
- [34] M. Watson, "Practical Artificial Intelligence programming with Java," 2008, pp. 110-128. In: [http://www.markwatson.com/opencontent\\_data/JavaAI3rd.pdf](http://www.markwatson.com/opencontent_data/JavaAI3rd.pdf) [retrieved: July, 2013].

# A Kinetic Light Shelf Unit as an Integrated Intelligent Control Device for Optimizing Interior Illumination

Ok-Kyun Im  
 School of Architecture  
 Chonnam National University  
 Gwangju, Korea  
 ilmeil@naver.com

Kyoung-Hee Kim  
 School of Architecture  
 University of North Carolina Charlotte  
 Charlotte, NC, USA  
 kkim33@unc.edu

Seung-Hoon Han  
 School of Architecture  
 Chonnam National University  
 Gwangju, Korea  
 hshoon@jnu.ac.kr  
 (Corresponding Author)

**Abstract** - This study suggests a Kinetic Light Shelf Unit (KLSU) which is possible to change the function intelligently according to information for the altitude of the Sun. When the Sun's altitude is high, the kinetic shelf unit works as the light shelf that can improve the interior illumination. Also, when the Sun goes down, the kinetic shelf unit operates as the vertical louver which blocks the direct rays of the Sun.

**Keywords** - Kinetic Building Facade; Light Shelf Unit; KLSU; Optimized Louver; Integrated Intelligent Control Device

## I. INTRODUCTION

Nowadays, building energy management is more important than ever. Especially, lighting energy used in the building occupies 30% of the whole consumed energy. This amount increases continuously. [1] Therefore, the solution to reduce the energy consumption is needed in these days. If appropriate method for providing of daylight in interior spaces was applied, it would help in energy saving. Light shelf is well known as the eco- friendly lighting technology. Furthermore, building design including light shelf is acceptable for the Leadership in Energy & Environmental Design (LEED).

The light shelf system has two major functions; it can reflect light deeper into interior space and block the direct light as well. When the reflected light on light shelf comes into the deeper interior space, the lighting energy will be decreased during a day time. Moreover, the solar radiation will be blocked by the light shelf as a shading device. These two functions are deeply related to the use of energy. [2]

For this reason, this proposed device is considered as a promising and practical skill. But this device is significantly affected by the altitude and motion of the sun, so it may have a limitation about the range of use. Accordingly, the purpose of this study aims at suggesting a kinetic light shelf makes the functions to be formed according to the altitude of the sun.

## II. KINETIC LIGHT SHELF UNIT

### A. State of the art

A light shelf is an architectural technology that makes light to penetrate deeper into the inner space. Figure 1 that shows light shelf is reflecting the light. This reflected light reduces the needed energy for artificial lighting to keep on acceptable indoor illumination in buildings.

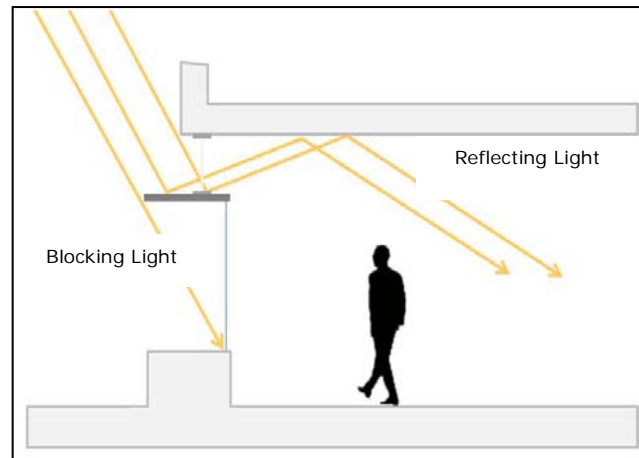


Figure 1. Concept and Function of the Light Shelf

The second function of the light shelf is to block the direct solar radiation that considered the major factor of heat gain inside indoor spaces. [3] Hence, the cost of keeping the indoor environment comfortable is going to decrease.

### B. Limitation of the light shelf associated with the altitude of the Sun

In order to the function shown by 'Figure 1', the position of the sun should be located between AM 11 and PM 1 (See Figure 2). This is because the light generated from the higher

altitude of the sun can be reflected by light shelf. [4] But when the sun set down into the ground, this function of light shelf will become dwindle. [5] And the sunlight that penetrate window will have an impact on the indoor environment.

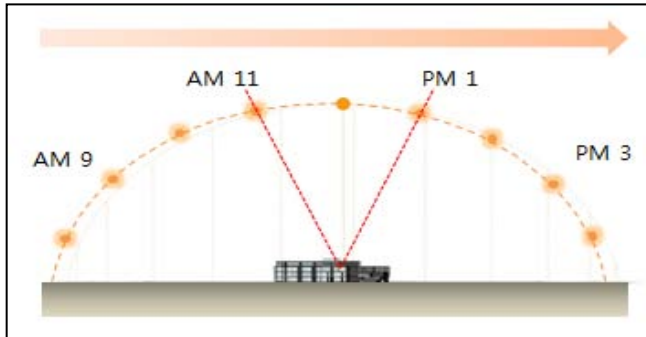


Figure 2. Performance Section of Light Shelf

C. Concept of the Kinetic light shelf which can adapt to altitude of the Sun

The most important feature of the KLSU is possible to adjust in the altitude of the sun. When the sun is located between a and b including meridian transit altitude (lower-left corner), KLSU works as a light shelf. Through this a to b course, indoor space can get the natural light reflected by surface of the light shelf.

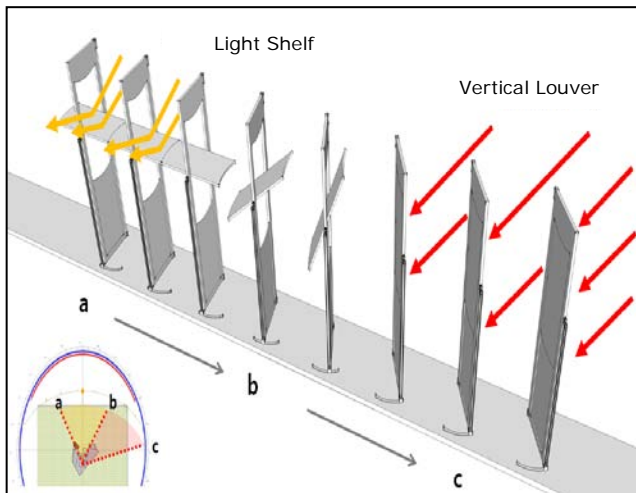


Figure 3. Conversion Process of KLSU

The surface of the shelf is covered with high reflectivity material. And when the sun is located between b and c (red part, lower-left corner), KLSU operates as a vertical louver to block out the direct light. Two types of the function are

shown by Figure 3 which represents the changing course from horizontal (light shelf) to vertical (louver). Like this, KLSU can be changed light shelf into vertical louver and it works to fit in each situation.

D. Design process of KLSU

In this part, we have looked at the first step of the KLSU design. First, a main frame that contains a light shelf is designed as a modular. So it can be installed to match the scale of the building. [6] [7] Design process II shows the course about removing conflict parts.

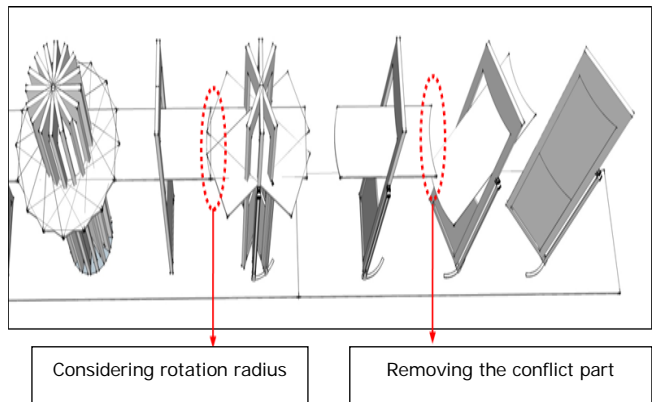
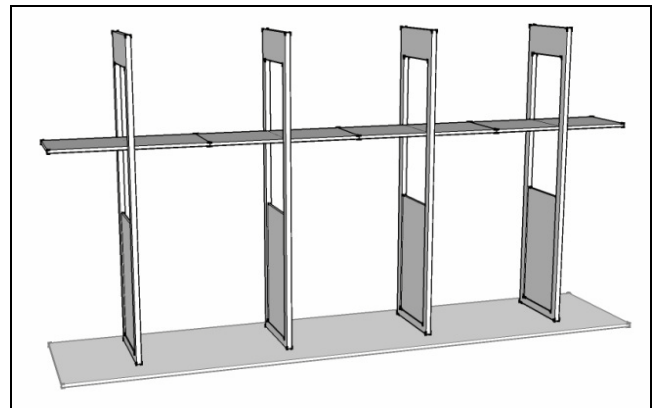


Figure 4. Design Process I (up) and II (down)

When connected KLSU elements are moved along the rotation rail, the edge of the shelf conflicts each other (Red dotted line). Thus the shelf design is decided by this process.

E. System components and operating steps

1) System components

In this part, it shows that components of the KLSU and operation steps. Consists of following elements (See Figure 5). 1<sup>st</sup> is main frame fixed to the axis of rotation. 2<sup>nd</sup> is the

light shelf reflecting the light. 3<sup>rd</sup> is the circle rod that penetrate the center of the shelf, so it makes possible to change horizontal into vertical.

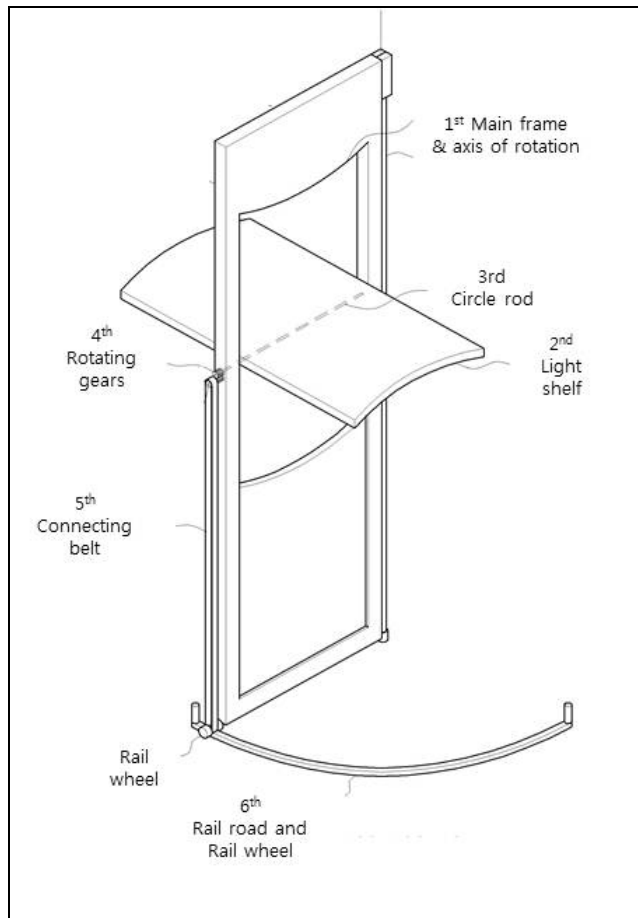


Figure 5. Components of a KLSU

4<sup>th</sup> through 6<sup>th</sup> are the key of this mechanism. 4<sup>th</sup> is rotating gears which are connected each other and one is bonded to the end of the rod. 5<sup>th</sup> is connecting belt that connect the gear with the rail wheel. 6<sup>th</sup> is the rail frame and rail wheel; the main frame is moved along this rail line around the axis.

2) Operating steps (from horizontal to vertical)

Step 1) KLSU works as the light shelf; Step 2) Rail wheel moves along the rail road; Step 3) Connecting belt is operated by the rail wheel; Step 4) Light shelf is turned into a vertical position by the movement of the gear that is affected by belt.

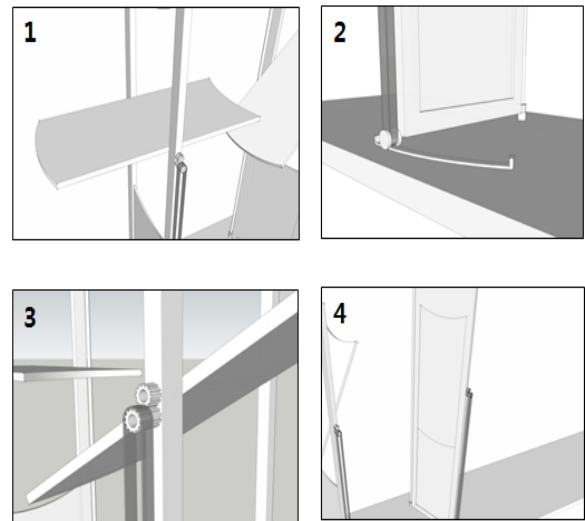


Figure 6. Operating Process

In sum, the suggested intelligent light shelf unit has been successfully responsible to the altitude of the sun, and the following is the flowchart showing a series of procedures towards its optimized functionality.

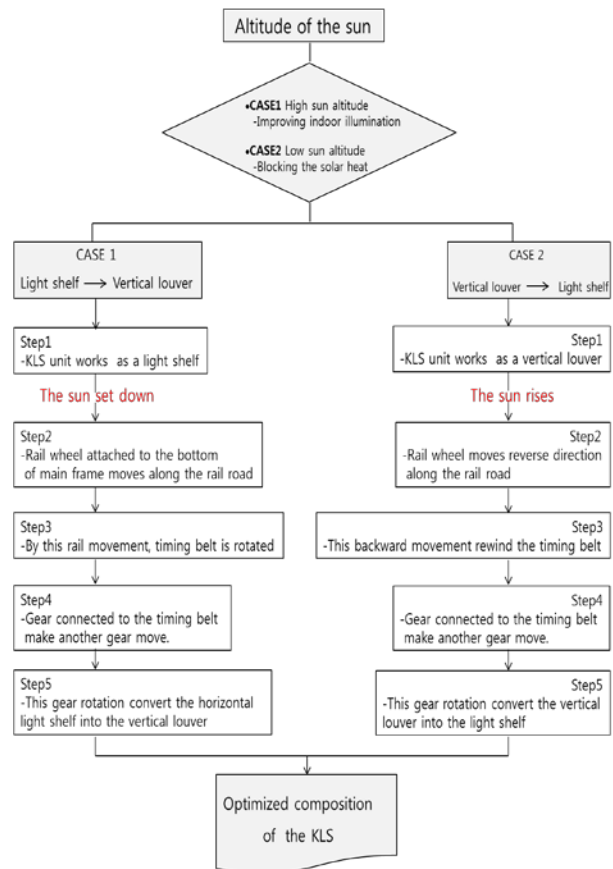


Figure 7. Flowchart of the Optimizing Process

KLSU proposed through this paper has been testified by multiple simulations such as Project Vasari and Autodesk Ecotect with Desktop Radiance and its adaptability remains as ongoing issue for future studies.

### III. CONCLUSION

Through this study, KLSU was introduced. This system is a new mechanism that is operated by interlocking main frame and light shelf. So, the kinetic shelf unit works as the light shelf and the kinetic shelf unit also operates as the vertical louver when the sun goes down. Based on this study, next research will proceed about the energy efficiency as well as the functional elements by using scale model.

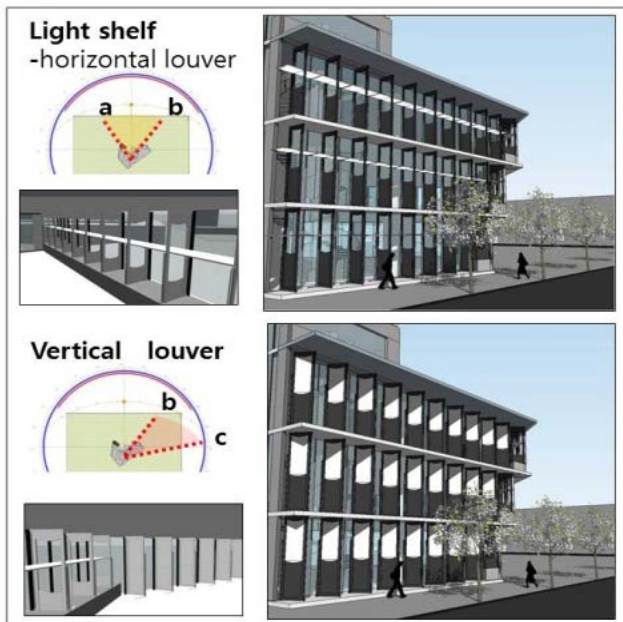


Figure 8. Application to the Building Façade

It will focus on the solar energy accumulation method associated with the altitude. The effect of variation in façade also can be offered in aspect of design. The kinetic façade changing from horizontal to vertical gives rise to the value such as building cognition. Advantages in all things will increase the possibility of eco-friendly technology.

### ACKNOWLEDGMENT

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0010519).

### REFERENCES

- [1] B. Park, I. Yang, and M. Na, "An Analysis of Daylighting Performance of Light Shelf in Office Building," *Journal of the Korean Institute of Illumination and Electrical Installation Engineers*, Vol. 22, No. 8, pp. 1-11, August 2008.
- [2] Y. Cho, B. Kim, and J. Lee, "Visual Performance Evaluation Study of a Scaled Light-Shelf Model," *Journal of the Korean Solar Energy Society*, Vol. 23, No. 3, pp. 63-71, September 2003.
- [3] H. Song, J. Lee, and K. Song, "A Study on the Application Methods Characteristics of Daylighting Systems to Building," *Korean Institute of Architectural Sustainable Environment and Building Systems*, Vol. 2, No. 1, pp. 236-241, 2008.
- [4] S. Kim, H. Kim, G. Han, R. Kwak, and J. Lee "Integrated System of Light Shelf and Venetian Blind for Improving Daylight Performance," *The Regional Association of Architectural Institute of Korea*, Vol. 2011, No. 1, pp. 505-510, October 2011.
- [5] J. Kim, K. Kim, and G. Kim, "Development and Performance Evaluation of a Sloped Lightshelf Daylighting System," *Journal of the Korean Solar Energy Society*, Vol. 24, No. 2, pp. 39-50, 2004.
- [6] T. Seo, H. Lee, and Y. Kim, "A Study on Light-Shelf System Using Context Awareness Technology for Energy Saving in Housing Space," *The Architectural Institute of Korea*, Vol. 28, No. 11, pp. 357-365, 2011.
- [7] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Daylighting Performance of a Light Shelf within Partitioned Space," *Korean Institute of Architectural Sustainable Environment and Building Systems*, Vol. 1, No. 3, pp. 30-37, 2007.