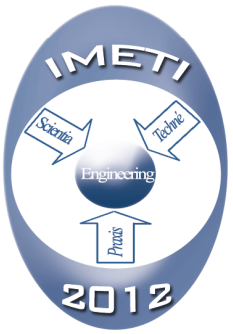


The 5th International Multi-Conference on Engineering and Technological Innovation

July 17th - 20th, 2012 – Orlando, Florida, USA

PROCEEDINGS

Post-Conference Edition

Edited by:

Nagib Callaos
Hsing-Wei Chu
Belkis Sánchez
Michael Savoie
Andrés Tremante
C. Dale Zinn

Organized by

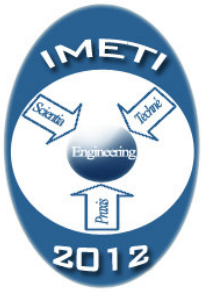
International Institute of Informatics and Systemics

Member of the International Federation for Systems Research (IFSR)

COPYRIGHT

The papers of this book comprise the proceedings of the conference mentioned on the title and the cover page. They reflect the authors' opinions and, with the purpose of timely disseminations, are published as presented and without change. Their inclusion in these proceedings does no necessarily constitute endorsement by the editors.

ISBN- 978-1-936338-74-0



PROGRAM COMMITTEE

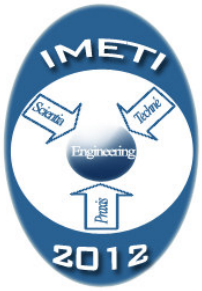
Chairs: Hsing-Wei Chu (USA) and C. Dale Zinn (USA)

Adascalitei, Adrian	Technical University	Romania
Aguilar Torres, Fernando J.	Almeria University	Spain
Ahmad, Ashraf	Princess Sumya University for Technology	Jordan
Al Obaidy, Mohaned	Gulf College	Oman
Al-Aomar, Raid	Jordan University of Science and Technology	Jordan
Alshara, Osama	Higher Colleges of Technology	Uzbekistan
Ariwa, Ezendu	London Metropolitan University	UK
Audant, A. Babette	CUNY Kingsborough Community College	USA
Ayuga Téllez, Esperanza	Universidad Politécnica de Madrid	Spain
Barberà, Elena	Universitat Oberta de Catalunya	Spain
Batos, Vedran	University of Dubrovnik	Croatia
Benbouziane, Mohamed	University of Tlemcen	Algeria
Bernardino, Jorge	Instituto Superior de Engenharia de Coimbra	Portugal
Bönke, Dietmar	Reutlingen University	Germany
Braiman, Avital	Brown University	USA
Bubnov, Alexey	Institute of Physics	Czech Republic
Calenzo, Patrick	IM2NP	France
Carbon, Margarita	The Boeing Company	USA
Castaño-Moraga, Carlos	Universidad de Las Palmas de Gran Canaria	Spain
Chang, Maiga	ELNP	Taiwan
Chang, Ruay-Shiung	National Dong Hwa University	Tokelau
Chang, Wen-Kui	Tunghai University	Taiwan
Chauhan, K. A.	S.V. National Institute of Technology-Surat	India
Chen, Yuhua	University of Houston	USA
Chen, Zhi	University of Kentucky	USA
Cheng, Xiaochun	University of Reading	UK
Chinesta, Francisco	ENSAM	France
Chiou, Yin-Wah	Nanhua University	Taiwan
Clarke, Tim	University of Wales	UK
Cone, Chris	Georgia Institute of Technology	USA
Curran, Kevin	University of Ulster	UK
Davies, B. J.	IJAMT	UK
Davim, J. Paulo	University of Aveiro	Portugal
Dierneder, Stefan	Linz Center of Mechatronics Gmbh	Austria
Dodig-Crnkovic, Gordana	Mälardalen University	Sweden
Dudek, Agnieszka	Wroclaw University of Environmental	Poland
Ehmann, Kornel F.	Northwestern University	USA
Erkollar, Alptekin	University of Applied Sciences Wiener Neustadt	Austria

Estrems, Manuel	Universidad Politécnica de Cartagena	Spain
Faura, Félix	Universidad Politécnica de Cartagena	Spain
Forbes, Alex	Georgia Institute of Technology	USA
Franco, Patricio	Universidad Politécnica de Cartagena	Spain
Fu, Yonggang	Shanghai Jiao Tong University	China
Fúster-Sabater, Amparo	CSIC	Spain
Gelbukh, Alexander	National Polytechnic Institute	Mexico
Gibson, Thomas L.	General Motors Research and Development Center	USA
Glovnea, Romeo P.	Brunel University	UK
Gomes, Samuel	UTBM	France
Goriachkin, Oleg	Volga State Academy of Telecommunication and Informatics	Russian Federation
Greenberg, William	Virginia Tech	USA
Grenmyr, Gustav	Chalmers University of Technology	Sweden
Gschweidl, Manfred	University of Applied Sciences Vorarlberg	Austria
Hepdogan, Seyhun	University of Central Florida	USA
Hetzer, Dirk	T-Systems Nova Deutsche Telekom	Germany
Higashiyama, Yoichi	Ehime University	Japan
Hirschberg, Wolfgang	Graz University of Technology	Austria
Holifield, David	University of Wales	UK
Huang, Yo-Ping	Tatung University	Taiwan
Ioannides, Stathis	SFK	Netherlands
Iovan, Stefan	Romanian Railway	Romania
Jiménez Rodríguez, Lourdes	University of Alcala	Spain
Jingchao, Chen	University Donghua	China
Kalogiannis, Konstantinos	Brunel University	UK
Kaminski, Jacek	Chalmers University of Technology	Sweden
Keiski, Riitta L.	University of Oulu	Finland
Khamba, J. S.	Punjabi University	India
Kobielarz, Magdalena	Wroclaw University of Technology	Poland
Koeglmayr, Hans-Georg	Pforzheim University of Applied Sciences	Germany
Koleva, Maria	Bulgarian Academy of Sciences	Bulgaria
Komerath, Narayanan	Georgia Institute of Technology	USA
Kuropka, Piotr	Wroclaw University of Technology	Poland
Lahlouhi, Ammar	University of Biskra	Algeria
Lamo, Yngve	Bergen University College	Norway
Lappas, Georgios	TEI of Western Macedonia	Greece
Li, Jackie	The City College of New York	USA
Li, Longzhuang	Texas A & M University	USA
Li, Man-Sze	IC Focus Ltd.	UK
Litvin, Vladimir	California Institute of Technology	USA
Liu, Jun	University of Ulster	UK
Lloret Mauri, Jaime	Polytechnic University of Valencia	Spain
López Román, Leobardo	University of Sonora	Mexico
Luh, Guan-Chun	Tatung University	Taiwan
Macke, Janaina	Universidade de Caxias do Sul	Brazil
Mansikkamäki, Pauliina	Tampere University of Technology	Finland
Mares, Cristinel	Brunel University	UK
Masoumi, Nasser	University of Tehran	Iran
Masunov, Artëm E.	University of Central Florida	USA
Mbobi, Aime Mokhoo	Red Knee Inc.	Canada

Mehrabian, Ali	University of Central Florida	USA
Mikhaylov, Ivan A.	University of Central Florida	USA
Mosley, Pauline	Pace University	USA
Mueller, Christoph Jo.	INCAP GmbH	Germany
Mura, Gianluca	Politecnico di Milano University	Italy
Murugan, Natarajan	Coimbatore Institute of Technology	India
Nagai, Yasuo	Tokyo University of Information Sciences	Japan
Nagar, Atulya K.	Liverpool Hope University	UK
Narasimhan, V. Lakshmi	Western Kentucky University	USA
Newman, Stephen	University of Bath	UK
Ngabonziza, Yves	LAGCC of the City University of New York	USA
Nicolescu, Cornel Mihai	The Royal Institute of Technology	Sweden
Nippa, Markus	Pforzheim University of Applied Sciences	Germany
Oberer, Birgit	University of Klagenfurt	Austria
Ong, Pangleen	University of Kentucky	USA
Orosco, Henry	University of Houston	USA
Ouwerkerk, David B.	General Motors Advanced Technology Center	USA
Paredes, Julia	GGC	USA
Patel, Anant	Georgia Institute of Technology	USA
Pennington, Richard	Georgia Gwinnett College	USA
Pérez, Carlos A.	Colombian Institute of Petroleum	Colombia
Petit, Frédéric	École Polytechnique de Montréal	Canada
Podaru, Vasile	Academia Tehnica Militara	Romania
Potorac, Alin Dan	University of Suceava	Romania
Praus, Petr	Charles University	Czech Republic
Pursell, David	Georgia Gwinnett College	USA
Rai, Bharatendra K.	University of Massachusetts	USA
Ramessur, Roshan	University of Mauritius	Mauritius
Remondino, Marco	University of Turin	Italy
Revetria, Roberto	Università Degli Studi di Genova	Italy
Rieder, Mathias	University of Applied Sciences Vorarlberg	Austria
Rodger, P. M.	University of Warwick	UK
Rodríguez Lozano, Gloria I.	National University of Colombia	Colombia
Rossbacher, Patrick	Graz University of Technology	Austria
Rudakov, Fedor	Oak Ridge National Laboratory	USA
Sahli, Nabil	Ministry of Technologies of Communication	Tunisia
Sala, Nicoletta	Universitá della Svizzera Italiana	Italy
Samtani, B. K.	S.V. National Institute of Technology	India
Sarate, João Alberto Rubim	Centro de Ensino Superior Cenecista de Farroupilha	Brazil
Sauder, Deborah	Georgia Gwinnett College	USA
Schaeffer, Donna M.	Marymount University	USA
Schaetter, Alfred	Pforzheim University	Germany
Schulman, Stuart	CUNY Kingsborough Community College	USA
Schumacher, Jens	University for Applied Sciences Vorarlberg	Austria
Shiraishi, Yoshiaki	Nagoya Institute of Technology	Japan
Siddique, Mohammad	Fayetteville State University	USA
Singh, Harwinder	Guru Nanak Dev Engineering College	India
Skawinska, Eulalia	Poznan University of Technology	Poland
Spearot, James A.	General Motors Research and Development Center	USA
Staretu, Ionel	Transilvania University of Brasov	Romania
Sulema, Yevgeniya	National Technical University of Ukraine	Ukraine

Suomi, Reima	Turku School of Economics	Finland
Suzuki, Junichi	University of Massachusetts	USA
Szotek, Sylwia	Wroclaw University of Technology	Poland
Tchier, Fairouz	King Saud University	USA
Thundat, Thomas	Oak Ridge National Laboratory	USA
Tousignant, Michel	Univerity of Sherbrooke	Canada
Toussaint, Luis	Universite de Technologie de Belfort Montbeliard	France
Trimble, Robert	Indiana University of Pennsylvania	USA
Troudt, Edgar E.	CUNY Kingsborough Community College	USA
Tsoi, Mai Yin	GGC	USA
Valakevicius, Eimutis	Kaunas University of Technology	Lithuania
Valdez, Pierre	Georgia Institute of Technology	USA
Vasinek, Vladimir	Technical Univerzity of Ostrava	Czech Republic
Venkataraman, Satyamurti	AIAMED	India
Vinod, D. S.	Sri Jayachamarajendra College of Engineering	India
Wallner, Daniel	Graz University of Technology	Austria
Wang, Lei	University of Houston	USA
Warwick, Jon	London South Bank University	UK
Wellons, Jonathan	Vanderbilt University	USA
Whiteley, Rick	Calabash Educational Software	Canada
Yadav, S. M.	S.V.National Institute of Technology-Surat	India
Yaghmaee, Mohammad H.	Ferdowsi University of Mashhad	Iran
Yanagisawa, Hideaki	Tokuyama College of Technology	Japan
Yingling, Yaroslava	North Carolina State University	USA
Yoon, Changwoo	ETRI	South Korea
Zalewski, Romuald I.	Poznañ University of Economics	Poland
Zaretsky, Esther	Hebrew University	Israel
Zelinka, Tomas	Czech Technical University in Prague	Czech Republic
Zhu, Hui	Soochow University	China
Zobaa, Ahmed	University of Exeter	UK

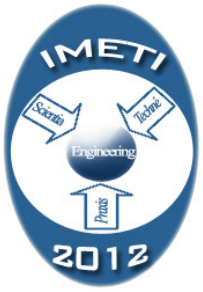


ADDITIONAL REVIEWERS

Acosta Rodríguez, José Ángel	University of Seville	Spain
Aigner, Werner	Johannes Kepler University Linz	Austria
Albayrak, Y. Esra	Galatasaray University	Turkey
Alhassan, Mohammad	Purdue University Fort Wayne	USA
Altan, Metin	Anadolu University	Turkey
Andreev, Rumen	Bulgarian Academy of Sciences	Bulgaria
Arifin, Achmad	Institute of Technology Sepuluh Nopember	Indonesia
Arpaci, İbrahim	Middle East Technical University	Turkey
Arrabales Moreno, Raúl	Carlos III University of Madrid	Spain
Badran, Omar	Al-Balqà Applied University	Jordan
Balas, Valentina	Aurel Vlaicu University of Arad	Romania
Barreiro, Joaquín	University of Leon	Spain
Barzev, Kiril	University of Rousse	Bulgaria
Bayraktar, Seyfettin	Yildiz Technical University	Turkey
Bilbao, Josu	IKERLAN Technology Research Centre	Spain
Bilich, Feruccio	University of Brasilia	Brazil
Botella, Federico	University Miguel Hernandez of Elche	Spain
Boulanger, Jean-Louis	UTC Heudiasyc	France
Branzila, Marius	Gheorghe Asachi Technical University of Iaşi	Romania
Cardoso, Pedro	University of Algarve	Portugal
Carrió Pastor, M. Luisa	Polytechnic University of Valencia	Spain
Challoo, Linda	Texas A&M University	USA
Cheggour, Mohamed	Ecole Normale Superieure	Morocco
Chen, Ping-Hei	National Taiwan University	Taiwan
Christian, Jeff	Lawson Lundell LLP	Canada
Chu, Hsing-Wei	Lamar University	USA
Cisneros-Cohernour, Edith	Autonomous University of Yucatan	Mexico
Cotet, Costel Emil	Polytechnic University of Bucharest	Romania
Dijkstra, Jan	Eindhoven University of Technology	Netherlands
Dougherty, Edmond	Ablaze Development Corp	USA
El Abed, Haikal	Braunschweig Technical University	Germany
Elwany, Hamdy	Alexandria University	Egypt
Fathy, Sherif Kassem	King Fasial University	Saudi Arabia
Feldmann, Birgit	University of Hagen	Germany
Ferrando, Emanuele	Selex Galileo	Italy
Fidan, Ismail	Tennessee Technological University	USA
Fiorini, Rodolfo A.	Polytechnic Institute of Milan	Italy
Floros, Andreas	Ionian University	Greece
Garro, Alfredo	University of Calabria	Italy

Gheno, Simoni	Federal University of São Carlos	Brazil
Glowacki, Miroslaw	AGH University of Science and Technology	Poland
Goel, Arun	National Institute of Technology Kurukshetra	India
Goi, Chai Lee	Curtin University of Technology	Malaysia
Gonzalez, Emmanuel A.	De La Salle University Manila	Philippines
Granovsky, Alex	Moscow State University	Russian Federation
Gu, Fei	Zhejiang University	China
Gujarathi, Ashish M.	Birla Institute of Technology and Science	India
Hammad, Mahmoud	GNREADER	Jordan
Hardman, John	Florida Atlantic University	USA
Haring, Roland	Ars Electronica	Austria
Harris, Marilyn	Capella University	USA
Hassini, A.	Laboratory of Application and Analysis of Radiations	Algeria
Hassini, Abdelatif	University of Oran	Algeria
He, Hongyu	Louisiana State University	USA
Hirz, Mario	Graz University of Technology	Austria
Hlynka, Myron	University of Windsor	Canada
Hrebicek, Jiri	Masaryk University in Brno	Czech Republic
Jastroch, Norbert	MET Communications	Germany
Jiang, Jinlei	Tsinghua University	China
Kaewarsa, Suriya	Rajamangala University of Technology Isan	Thailand
Kakanakov, Nikolay	Technical University of Sofia	Bulgaria
Kamrani, Ehsan	Polytechnic Montreal	Canada
Kässi, Tuomo	Lappeenranta University of Technology	Finland
Kawamura, Hidenori	Hokkaido University	Japan
Ke, Bwo-Ren	Chung Chou University of Science and Technology	Taiwan
Kechadi, Tahar	University College Dublin	Ireland
Khatib, Hisham	World Energy Council	Jordan
Kim, Hyun-Jun	Samsung Electronics	South Korea
Ksenofontov, Alexandre	Moscow Engineering Physics Institute	Russian Federation
Kuang, Yu	Stanford University	USA
Kunold, Ingo	University of Applied Sciences Dortmund	Germany
Kureshi, Nadeem	Michigan State University	Pakistan
Lee, Joo Hwan	Seoul National University	South Korea
Li, Dayong	Shanghai Jiao Tong University	China
Li, Yu-Chiang	Southern Taiwan University	Taiwan
Lin, Chih-Ting	National Taiwan University	Taiwan
Lin, Chun Yuan	Chang Gung University	Taiwan
Liu, Tingyang Lewis	National Kaohsiung Normal University	Taiwan
Lopez de Lacalle, Luis	University of the Basque Country	Spain
Luqman, Chuah A.	University Putra Malaysia	Malaysia
Mahmoudi, Saïd	University of Mons	Belgium
Maldonado R., José L.	Center for Research in Optics	Mexico
Markakis, Euaggelos	Technological Educational Institution of Crete	Greece
Mattila, Anssi	Laurea University of Applied Sciences	Finland
McCormick, John	SELEX SAS	UK
Medina, Dulce Yolotzin	Metropolitan Autonomous University	Mexico
Mensah-Biney, Robert	North Carolina State University	USA
Mokhtari, A.	University of Versailles Saint-Quentin-en-Yvelines	France
Morini, Mirko	University of Ferrara	Italy
Moschim, Edson	State University of Campinas	Brazil

Motieifar, Alireza	University of British Columbia	Canada
Muensit, Nantakan	Prince of Songkla University	Thailand
Neaga, Elena Iirna	Loughborough University	UK
Nisar , Humaira	Gwangju Institute of Science and Technology	South Korea
Omidvar, Hedayat	Nacional Iranian Gas Company	Iran
Pascal, Goureau	ENSEA	France
Patel, Ankit A.	Gujarat Technological University	India
Pearson, Esther	Lasell College	USA
Pedamallu, Chandra Sekhar	Dana-Farber Cancer Institute	USA
Pfliegl , Reinhard	AustriaTech Ltd	Austria
Pimentel, David	Cornell University	USA
Pogarčić, Ivan	Polytechnic of Rijeka	Croatia
Popa, Rustem	University of Galati	Romania
Price, Howard	Equipment Design	USA
Puslecki, Zdzislaw	Adam Mickiewicz University	Poland
Raj, Piyush	Institute of Management Prabandh Nagar	India
Rana, Mukhtar Masood	Anglia Ruskin University	USA
Ribakov, Y.	Ariel University Center of Samaria	Israel
Ros, L.	Polytechnic University of Valencia	Spain
Rot, Artur	Wroclaw University of Economics	Poland
Safarik, Pavel	Czech Technical University in Prague	Czech Republic
Schenk, Peer	The University of Queensland	Germany
Sheng, Ye	University of California Berkeley	USA
Shyr, Wen-Jye	National Changhua University of Education	Taiwan
Silva, Daniel P.	Tiradentes University	Brazil
Singh, Kaushlendra	West Virginia University	USA
Skoko, Hazbo	Charles Sturt University	Australia
Snow, Richard	Embry-Riddle Aeronautical University	USA
Song, Bo	University of Illinois at Chicago	USA
Song, Yulin	Memorial Sloan-Kettering Cancer Center	USA
Šraml, Matjaž	University of Maribor	Slovenia
Stematiu, Dan	Technical University of Civil Engineering Bucharest	Romania
Süral, Haldun	Middle East Technical University	Turkey
Sutherland, Trudy	Vaal University of Technology	South Africa
Tang, Hua	University of Minnesota Duluth	USA
Taylor, Craig	University of Southern California	USA
Trifonova, Silvia	University of National and World Economy	Bulgaria
Tsoi, Mai Yin	Georgia Gwinnett College	USA
Turcu, Cristina	Stefan cel Mare University of Suceava	Romania
Vaganova, Natalia	ICM&MG SB RAS	USA
Vandeyar, Thiru	University of Pretoria	South Africa
Varga, Mladen	University of Zagreb	Croatia
Verstichel, Stijn	Ghent University	Belgium
Vimarlund, Vivian	Linköping University	Sweden
Virvilaite, Regina	Kaunas University of Technology	Lithuania
Vukadinovic, Dinko	University of Split	Croatia
Yazici, Ali	Middle East Technical University	Turkey
Yuan, Shengli	University of Houston-Downtown	USA
Zakrzewska, Danuta	Technical University of Lodz	Poland
Zeman, Klaus	Johannes Kepler University Linz	Austria
Zuccarello, Mario	University of Cincinnati	USA

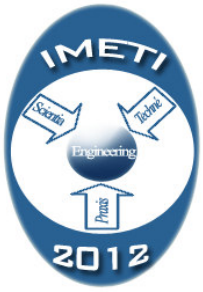


ADDITIONAL REVIEWERS FOR THE NON-BLIND REVIEWING

Allahverdi, Ali	Kuwait University	Kuwait
Andrade, Célio	Federal University of Bahia	Brazil
Andronikashvili, Teimuraz	Institute of Physical and Organic Chemistry	Georgia
Azambuja, Marcelo	University of Southern Illinios	USA
Bhuiyan, Sharif	Tuskegee University	USA
Boets, Sofie	Belgium Road Safety Institute	Belgium
Bousaba, Chafic	North Carolina A&T State University	USA
Brinksmeier, E.	University of Bremen	Germany
Buerck, John	Saint Louis University	USA
Cao, Yiding	Florida International University	USA
Cho, Woong	ETRI	South Korea
Das, Asesh	Pennsylvania College of Technology	USA
Dmitry, Kuznetsov	National University of Rwanda	India
Emam, Samir	United Arab Emirates University	UAE
Frohm, Jörgen	Vattenfall	Sweden
Gao, Liang	Huazhong University of Science and Technology	China
Gatti, Laura	University of Modena and Reggio Emilia	Italy
Gebken, Richard	Missour State Univeristy	USA
Goncalves, Juliano	IFSC	Brazil
Gong, Jie	University of Southern Illinios	USA
Gurupur, Varadraj	Texas A&M University-Commerce	USA
Haik, Yousef	United Arab Emirates University	UAE
Harger, Eliza	Federal University of Santa Catarina	Brazil
Heskova, Marie	University of Economics, Prague	Czech Republic
Hong, Jinglan	Shandong University	China
Ibhadode, Akii	University of Benin	Nigeria
Issa, Raymond	University of Florida	USA
Izui, Shozo	University of Geneva	Switzerland
Jagadish, R.	Indian Concerete Institute	India
Jean Kileshye, Onema	Waternet	Zimbabwe
Jesmin, Khan	Tuskegee University	USA
Jilek, Miroslav	Czech Technical University in Prague	Czech Republic
Kalantaridis, Christos	Bradford University	UK
Kalid, Ricardo	Federal University of Bahia	Brazil
Kastner, Kyle	Southwest Research Institute	USA
Kobayashi, Kentaro	Nagoya University	Japan
Kodacsy, Janos	Kecskemét College	Hungary

Kollman, Thomas	Texas A&M University-Commerce	USA
Kuriyagawa, Tsunemoto	Tohoku University	Japan
Kuznetsov, Dmitry	National University of Rwanda	Rwanda
Lawrence, Daniel	Technology Service Corporation	USA
Lei, Weimin	Northeastern University	China
Lim, Geonpyo	Kepco Research Institute	South Korea
Madichie, Dr. Nnamdi	University of Sharjah	UAE
Maheshwari, Bharat	University of Windsor	Canada
Manous, Joe	Institute for Water Resources	USA
Meesmann, Uta	Belgium Road Safety Institute	Belgium
Melde, Gretchyn	The MITRE Corporation	USA
Messner, John	Pennsylvania State University	USA
Michl, Markus	University of Erlangen-Nuremberg	Germany
Miller, Bob	The MITRE Corporation	USA
Montanari, Roberto	University of Modena and Reggio Emilia	Italy
Mouette, Dominique	University of Sao Paulo	Brazil
Nair, Suku	Southern Methodist University	USA
Nousala, Susu	Aalto University	Finland
Pearson, Esther	Lasell College	USA
Pessoa, Fernando	Federal University of Rio de Janeiro	Brazil
Prikrylova, Jana	Skoda Auto University	Czech Republic
Puslecki, Zdzislaw	Adam Mickiewicz University	Poland
Ramachandran, K. P.	Caledonian College of Engineering	Oman
Robles, Rosslin John	University of San Agustin	Philippines
Saeed, Babar	Pakistan Air Force Academy	Pakistan
Safarik, Pavel	Czech Technical University in Prague	Czech Republic
Sanchez, Giovanni	University of San Buenaventura	Colombia
Saraiji, Riad	United Arab Emirates University	UAE
Sarper, Huseyin	Colorado State University-Pueblo	USA
Scannelle, Kevin	Saint Louis University	USA
Shimoda, Yukio	Tokyo Womens Medical University	Japan
Smith, Brian	RDECOM	USA
Sun, Zixin	Nanjing University of Posts and Telecommunications	China
Sutcu, Hale	Zonguldak Karaelmas University	Turkey
Tang, Dunbing	Nanjian University of Aeronautics and Astronautics	China
Tang, Hua	University of Minnesota Duluth	USA
Tang, Wenyong	Shanghai Jiao Tong University	China
Tansel, Ibrahim	Florida International University	USA
Tomczyk, Andrzej	Rzeszow University of Technology	Poland
Tugrul, Nurcan	Yildiz Technical University	Turkey
Vasudevan, R.	Vellore Institute of Technology University	India
Velazquez-Araque, Luis	Czech Technical University in Prague	Czech Republic
Villaverde, Bernard	Mapúa Institute of Technology	Philippines
Vollmar, Jan	Siemens AG, Corporate Technology	Germany
Widlroither, Harald	Fraunhofer Institute for Industrial Engineering	Germany
Xu, Weichao	Guangdong University Of Technology	China
Zeman, Klaus	Johannes Kepler University Linz	Austria
Zeng, Qing-An	North Carolina A&T State University	USA

The 5th International Multi-Conference on Engineering and Technological Innovation: IMETI 2012



HONORARY CHAIR
William Lesso

PROGRAM COMMITTEE CHAIRS
Hsing-Wei Chu
C. Dale Zinn

GENERAL CHAIR
Nagib Callaos

ORGANIZING COMMITTEE CHAIRS
Belkis Sánchez
Andrés Tremante

CONFERENCE PROGRAM MANAGER
Leonisol Callaos

HARDCOPY PROCEEDINGS PRODUCTION CHAIR
Maria Sánchez

TECHNICAL CONSULTANT ON COMPUTING SYSTEM /
CD PROCEEDINGS PRODUCTION CHAIR
Juan Manuel Pineda

META REVIEWERS SUPPORT
Dalia Sánchez

SYSTEM DEVELOPMENT, MAINTENANCE AND DEPLOYMENT
Dalia Sánchez
Keyla Guédez
Nidimar Díaz
Bebzabeth García

OPERATIONAL ASSISTANTS
Marcela Briceño
Cindi Padilla

HELP DESK
Louis Barnes
Sean Barnes
José Aponte

Number of Papers Included in these Proceedings per Country

(The country of the first author was the one taken into account for these statistics)

Country	# Papers	%
TOTAL	35	100%
Brazil	1	2,86%
Chile	1	2,86%
Czech Republic	1	2,86%
France	1	2,86%
Germany	1	2,86%
Greece	4	11,43%
India	1	2,86%
Iraq	1	2,86%
Italy	1	2,86%
Japan	4	11,43%
Pakistan	1	2,86%
Poland	1	2,86%
South Africa	1	2,86%
South Korea	1	2,86%
United Arab Emirates	1	2,86%
United States	13	37,14%
Venezuela	1	2,86%
Brazil	1	2,86%
Chile	1	2,86%
Czech Republic	1	2,86%
France	1	2,86%
Germany	1	2,86%
Greece	4	11,43%
India	1	2,86%
Iraq	1	2,86%
Italy	1	2,86%
Japan	4	11,43%
Pakistan	1	2,86%
Poland	1	2,86%
South Africa	1	2,86%
South Korea	1	2,86%
United Arab Emirates	1	2,86%
United States	13	37,14%
Venezuela	1	2,86%
South Africa	1	2,86%

Foreword

Engineering activities are based on the development of new Knowledge (*Scientia*), new 'made things' (*Techné*), and/or new ways of working and doing (*Praxis*). *Scientia*, *Techné,* and *Praxis* are three important dimensions of a comprehensive conception of Engineering as a whole. Engineering, as *Scientia*, is mostly developed in academia; as *Techné,* is practiced in industry generating technological innovations; and as *Praxis,* is carried out in technical and non-technical organizations, supporting managerial activities and technical procedures, via methodical and methodological design and implementation. This is why Engineering provides one of the most solid academic and professional substrata for bridging among universities, industries and governments.

Publications and conferences related to Engineering are usually oriented to one of its three dimensions. While this is an adequate thing to do when disciplinary focus is sought, it does not represent Engineering as a whole and it misses the very important synergic relationships among the three kinds of engineering activities mentioned above. This is why a group of scholars, professionals, and consultants, in the field of engineering, considered the possibility of organizing a conference where presentations would not be reduced to one specific Engineering dimension, but would foster the participation of academics, practitioners, and managers in the three dimensions of Engineering, in the same conference, so they can synergistically interact with each other. A consequence of this purpose is the organization of *The 5th International Multi-Conference on Engineering and Technological Innovation:* IMETI 2012, where submissions were accepted for the presentation of:

- **New knowledge** (Engineering as *scientia*);
- **New products and services**, i.e. technological innovations (Engineering as *techné*);
- **New technical and managerial methods and methodologies** (Engineering as *praxis*);
- **New meta-engineering** (Engineering of Engineering activities) knowledge, innovations, and methodologies.

IMETI 2012 was organized and sponsored by the International Institute of Informatics and Systemics (IIIS, www.iiis.org), member of the International Federation of Systems Research (IFSR). The IIIS is a ***multi-disciplinary organization for inter-disciplinary communication and integration***, which includes about 4500 members. Consequently, a main purpose of the IIIS is to foster knowledge integration processes, interdisciplinary communication, and integration of academic activities. Based on 1) the transdisciplinartity of the systemic approach and its emphasis on *relationships* and *integrating* processes, and 2) the multi-disciplinary support of cybernetics' and informatics' concepts, notions, theories, technologies, and tools, the IIIS has been organizing multi-disciplinary conferences as a platform for fostering inter-disciplinary communication and knowledge integration processes.

Multi-disciplinary conferences are organized by the IIIS as support for both **intra-** and **inter-disciplinary** communication. Processes of intra-disciplinary communication are mainly achieved via traditional paper presentations in corresponding disciplines, while conversational sessions, regarding trans- and inter-disciplinary topics, are among the means used for inter-disciplinary communication. Intra- and inter-disciplinary communications might generate *co-regulative cybernetic loops*, via negative feedback, and *synergic* relationships, via positive feedback loops, in which both kinds of communications could increase their respective effectiveness. Figure 1 shows at least two cybernetic loops if intra- and inter-disciplinary are adequately related. A necessary condition for the effectiveness of Inter-disciplinary communication is an adequate level of **variety** regarding the participating disciplines. *Analogical thinking and learning processes* of disciplinarians depend on it; which in turn are potential sources of the creative tension required for cross-fertilization among disciplines and the generations of new hypothesis. An extended presentation regarding this issue can be found at www.iiis.org/MainPupose.

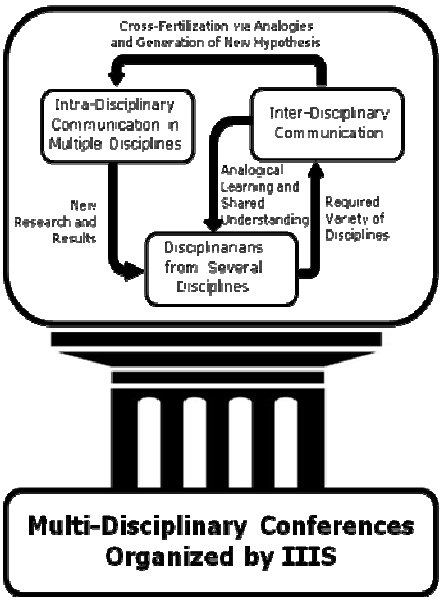


Figure 1

IMETI 2012 was organized jointly with other multi-disciplinary events with the purpose of providing a communicational forum to researchers, engineers, practitioners, developers, consultants, and end-users of computerized, communications, and/or control systems and technologies in the private and the public sectors. This multi-disciplinary forum provides the opportunity to share experience and knowledge by facilitating discussions on current and future research and innovation. Participants can explore the implications of relationships between new developments and their applications to organizations and society at-large. One of the primary objectives of IMETI 2012, and its collocated events, is to promote and encourage interdisciplinary cross-fertilization and knowledge communication. This might foster systemic thinking and practice, including the analogical thinking that characterizes the Systems Approach, which is, in most cases, the required path to logical thinking, scientific hypothesis formulation, and new design and innovation in engineering.

On behalf of the Organizing Committee, I extend our heartfelt thanks to:

1. the 199 members of the three Program Committees from 62 countries;
2. the 283 additional reviewers, from 46 countries, for their **double-blind peer reviews**; and
3. the 105 reviewers, from 38 countries, for their efforts in making the **non-blind peer reviews**. (Some reviewers supported both: non-blind and double-blind reviewing for different submissions)

A total of 705 reviews made by 388 reviewers (who made at least one review) contributed to the quality achieved in IMETI 2012. This means an average of 8.01

reviews per submission (88 submissions were received). ***Each registered author had access, via the conference web site, to the reviews that recommended the acceptance of their respective submissions***. Each registered author could also get information about: 1) the average of the reviewers evaluations according to 8 criteria, and the average of a global evaluation of his/her submission; and 2) the comments and the constructive feedback made by the reviewers, who recommended the acceptance of his/her submission, so the author would be able to improve the final version of the paper.

In the organizational process of IMETI 2012 (including the events organized in its context) about 88 papers/abstracts were submitted. These pre-conference proceedings include about 35 papers, from 17 countries, that were accepted for presentation. We extend our thanks to the invited sessions organizers for collecting, reviewing, and selecting the papers that will be presented in their respective sessions. The submissions were reviewed as carefully as time permitted; it is expected that most of them will probably appear in a more polished and complete form in scientific journals.

This information about IMETI 2012 is summarized in the following table, along with the other collocated conferences:

Conference	# of submissions received	# of reviewers that made at least one review	# of reviews made	Average of reviews per reviewer	Average of reviews per submission	# of papers included in the proceedings	% of submissions included in the proceedings
WMSCI 2012	**258**	**822**	**1348**	**1.64**	**5.22**	**98**	**37.98%**
IMSCI 2012	**79**	**341**	**573**	**1.68**	**7.25**	**45**	**56.96%**
IMETI 2012	**88**	**388**	**705**	**1.82**	**8.01**	**35**	**39.77%**
CISCI 2012	**188**	**629**	**1283**	**2.04**	**6.82**	**61**	**32.45%**
TOTAL	**613**	**2180**	**3909**	**1.79**	**6.38**	**153**	**38.99%**

We are also grateful to the co-editors of these proceedings for the hard work, energy, and eagerness they displayed preparing their respective sessions. We express our intense gratitude to Professor William Lesso for his wise and opportune tutoring, for his eternal energy, integrity, and continuous support and advice as Honorary President of WMSCI 2012 and its collocated conferences, as well as for being a very caring old friend and intellectual father to many of us. We also extend our gratitude to Professor Belkis Sánchez, who brilliantly managed the organizing process. Special thanks to Dr. C. Dale Zinn, Professors Hsing-Wei Chu, Andrés Tremante, Michael Savoie, and Belkis Sánchez for chairing, or co-chairing the Program and/or the Organizing Committees of IMETI 2012 and/or the events organized in its context.

We also extend our sincere gratefulness to Drs. Ranulph Glanville, Thomas Marlowe**,** Ruth Bergman, Karl H. Müller, Shigehiro Hashimoto, T. Grandon Gill, Marta White Szabo**,** Jeremy Horne, Mario Norbis, Susu Nousala, Richard Segall, Kathy Kovacs Burns, Rahul Bedi, Darshan Deseai, and Heidi Hahn for accepting to address the audience of the General Joint Plenary Sessions with keynote conferences.

We extend our gratitude as well to María Sánchez, Juan Manuel Pineda, Leonisol Callaos, Dalia Sánchez, Keyla Guédez, Nidimar Díaz, Bebzabeth García, Marcela Briceño, Cindi Padilla, Louis Barnes, Sean Barnes, José Aponte, Abrahan Marín, Yulaime Padilla, and Freddy Callaos for their knowledgeable effort in supporting the organizational process producing the hard copy and CD versions of the proceedings, developing and maintaining the software supporting the interactions of the authors with the reviewing process and the Organizing Committee, as well as for their support in the help desk and in the promotional process.

Professor Nagib C. Callaos,
IMETI 2012 General Chair

CONTENTS

(Post-Conference Edition)

Interdisciplinary Research, Analogy based Modeling and Problem Solving

Engineering Education

Dynamic Variability in Families of Clouds

Jorge Fox,*
Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" (ISTI-CNR)
Area della Ricerca CNR di Pisa, Italy
fox@isti.cnr.it

Abstract

Cloud computing, a promising new type of parallel and distributed system offers the opportunity to achieve almost unlimited scalability of the IT infrastructure and improvement of IT-Services. However, broad acceptance of this technology is curbed by legitimate questions regarding security, performance, and reliability, among others. A critical aspect of Cloud Computing lies in the fact that use of a cloud service requires a company to trust the service to comply with given Service-level Agreements (SLA). In order to foster broader acceptance of the Cloud, we propose a framework for preservation of Quality of Service (QoS) levels supported by product line engineering. We argue that coupling the abstraction capabilities of product line approaches and SLA's offers enhanced support for monitoring, control and enforcement of QoS parameters. In the long run, our work aims to provide a generic framework for non-functional requirements enforcement in Cloud Computing.

1 Introduction

Cloud computing offers the opportunity to improve IT -Services and to achieve almost unlimited scalability of the IT infrastructure, and to achieve all of this at a significantly reduced cost than relying on internal resources. However, the use of a cloud service requires the User to trust that the chosen cloud services comply with given Service-level Agreements (SLA) [1]. SLAs specify the conditions under which a certain service is provided by a provider to a customer as explored by Comuzzi et al. [2].

Preserving compliance to SLAs on runtime is a resource intensive task, that has to be performed continuously at the case of systems subject to fluctuating conditions in the operational environment, which is the case of cloud services. In order to facilitate this task, we conceive the development of a framework for Cloud Computing that preserves Quality of Service (QoS) standards. Supported by model-driven development techniques, the framework will facilitate development of Cloud product lines or families. This work contemplates the integration of two software engineering paradigms, Product Line Engineering (PLE) and Cloud Computing. By means of product line engineering, we introduce variability in service definitions and open the door for runtime adaptability. The objective is to design a framework that monitors, negotiates and preserves SLAs in Cloud Computing Services. Thus, supporting QoS through dynamic variability.

Moreover, we envision the development of dynamically adaptable Cloud Families, supported by lightweight reconfiguration mechanisms. The framework also considers on-the-fly incorporation of new functionalities.

In our model, adaptations are triggered when the guarantees specified in SLAs are injured, initiating a reconfiguration in the system to preserve the guarantees established in the SLAs. Adaptation triggers in our Cloud Family are defined as a collection of Service-Level Constraints based on the conceptual mode of SLAs and SLA-Templates from the EU FP7 program SLA@SOI [1]. By defining a family of Clouds as a product line, we allow the service provider and the customer to select the required service tailored to its needs, with the added benefit of reducing the level of complexity in develop-

*This work was done during the tenure of an ERCIM "Alain Bensoussan" Fellowship Programme

ment. Moreover, defining variability points in the cloud contributes to an enhanced reusability and evolution in service definitions.

The rest of the paper is organized as follows. Sect. 2 presents related work. Sect. 3 identifies key challenges that must be addressed for the broader adoption of Cloud Computing. Sect. 4 discusses our concept of a cloud family and presents an example. Sect. 5 outlines the SLA preservation mechanism. Sect. 6 presents future lines of research and Sect. 7 concludes.

2 Related Work

Jin et al. [3] propose a model that captures composition relationships between providers, consumers, and SLAs. The model is enhanced with simulation and sensitivity analysis. Still, the work comes short of relating to mechanisms for the enforcement of SLAs. Similarly, Hauck and Reiser [4] develop a low level architecture for monitoring quality of service parameters. Their work focuses on demonstrating the usefulness of the Java Dynamic Management Kit (JDMK) to achieve this architecture, rather than attempting to offer a more generic framework for monitoring quality of service parameters.

In Sahai et al. [5], the authors develop an automated SLA monitoring engine, which allows to aggregate SLAs between multiple web services. This engine represents a valuable step forward towards the enforcement of SLAs. However, the engine is confined to web services and lacks enforcement mechanisms for SLA preservation. A more advanced work, in terms of SLA and quality of service management, can be found in Czajkowski et al. [6]. Czajkowski et al. introduce an architecture supporting coordination and scheduling of resources from different policy domains, this being closer to a more generic approach toward quality of service management and preservation. Grit [7] contributes definitions to formulate resource contracts, necessary for SLA negotiation and definition. Nonetheless, this work stays short of introducing elements to cope with dynamicity.

3 Research Challenges

A Cloud is commonly described as an infrastructure composed of three layers, the first one providing computational resources on demand defined "Infrastructure as a Service" (IaaS), the second one providing the development and service payload environment named "Platform as a Service" (PaaS) and the last one providing on demand software known as "Software as a Service". The particular characteristics of these interacting layers imply the need for distinct engineering approaches. Moreover, in the Cloud this composed infrastructure is shared among a variety of diverse users with varying Service Level requirements. This poses conceptual questions that need to be addressed. Such as, harmonizing possibly contradictory demands on reliability, performance, throughput, as well as, security. Relating distinct Service Level requirements and specifications to End-Users while tracing these to the infrastructure represents a complex task. For instance, prioritizing SLAs and defining algorithms that reconcile them. Another challenge lies in the likelihood of changes in the operating environment that alter the behavior of components and ultimately services. Consequently, we target to define generic QoS specifications that correlate SLAs to common and variable elements in the Cloud composite.

4 Variability in Families of Clouds

The Cloud as a virtualised infrastructure allows for new service and business solutions. Some recent examples are: Hewlett Packard's (HP) *client virtualisation* or Virtual Desktop Infrastructure (VDI) which allows workers to go mobile. In this environment, exploiting commonalities while managing variations for specific customers is of central importance. In Desktop Virtualization [8], for instance, Desktops access the cloud and execute in the data centers provided using remote desktop protocols, moving Operating Systems and application execution from local devices to remote data and processing centers [9]. We find an ideal example for Cloud Families in Desktop Virtualiza-

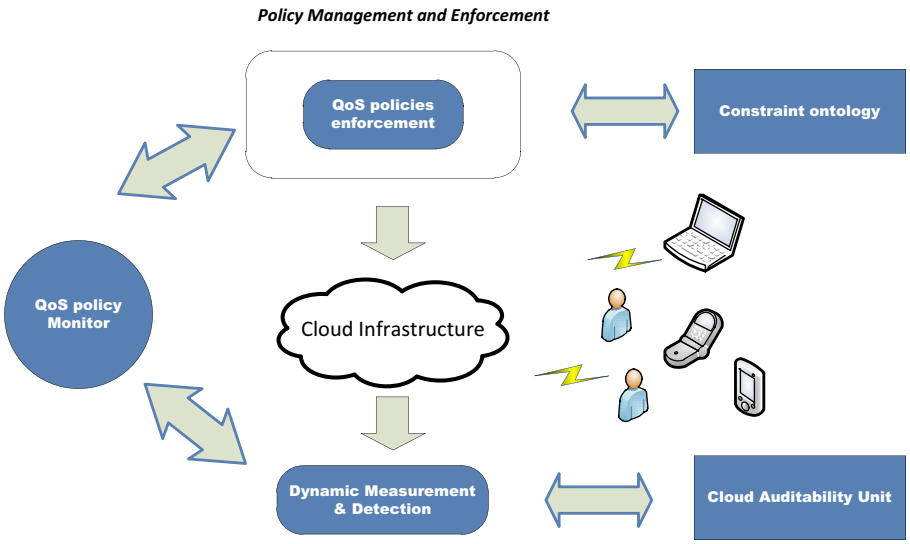


Figure 1: SLA Management and Enforcement Framework

Table 1: Common and variable services of the VDI Cloud

Common	Static variable	Dynamically variable
Application delivery layer	MS Office / Open Office	Load balancing
Resource management	Payroll, HR Sfw	Service Discovery
VM Monitor	Speech recognition	Traffic management
SLA Monitor & Negotiation	Machine Learning	User defined SLAs

tion. In this context, the convergence between the processing power of Cloud computing and the flexibility provided through dynamic variability represents the key to support trends like the gradual displacement of laptops by smartphones in the business and work environment.

Software Product Lines help identifying common and variable features of product line applications [10]. Identification of the variable features enables the development of customized applications by reusing predefined, adjustable artifacts. Diverse kinds of users, regulatory frameworks, available technologies, among other issues, demand static variations within a product line, while changes in the operating environment such as load variability, service providing failures, availability of new services, as well as changing user needs, generate other type of variability, that is, dynamic variability. We focus our proposal on the latter type of variability and the need to enforce compliance of SLAs in the face of adaptations.

An example can be drawn on a VDI where services to deliver business tasks like customer service, word processing, payroll, human resources management, need to cater for different requirements according to the end user giving rise to static variability in the product line. Moreover, changes on the operating environment or changes in user needs trigger the need to perform adaptations on runtime.

Common elements identified through PLE are allocated in the Cloud infrastructure, which in our model comprehends services pertaining to each of the three Cloud layers IaaS, PaaS, and SaaS (Fig. 1, middle). Moreover, variable elements are steered by corresponding definitions in the *QoS policy enforcement* unit (Fig. 1 upper part).

At the case of our proposed Cloud VDI, the common elements are: an application delivery layer which provides the remote applications to clients in a platform-independent manner, and the resource management layer rendering the

services provided through the cloud. The application delivery layer can be provided by protocols like SunRay, [11], Xen and PowerVM from IBM's Blue Cloud. [12]. The resource management layer and cloud infrastructure can be provisioned by Amazon Elastic Compute Cloud (Amazon EC2) [13]. Above this layer, variable assortments of desktop services like Ms Office, OpenOffice, or Human Resources Management and Payroll, can be offered to the end-user. Table 1 presents a mix of services for an hypothetical VDI, illustrating the common platform services in the product line, as well as, static and dynamic variable services.

5 Enforcing Service Level Specifications

In this section, we present the design of the SLA enforcement function related to dynamic variability in our Cloud Family Model. We acknowledge that enforcing static variability requires a different treatment and will be addressed in the future. Following the lines of [14, 15] we consider variability as an enabler for adaptability, whereby adaptable services can be modelled and designed based on each type of service variability.

The system proposed is divided in three main interacting layers: Monitoring, Delivery, and Back-end.

Three main activities are defined to keep the system conform to desired SLA's. In a first phase, metrics from the service provider and user are received and processed by the SLA Monitor, that is, a program execution analyzer [16] for SLAs. The metrics range from network traffic, load balancing, to user-defined operational performance and resource utilization, as provided by Amazon's EC2 CloudWatch [13], for instance. To illustrate this, Monitoring can be implemented as the collection of timestamps of the individual response times effecting the operation and computing the average response time on the basis of these timestamps. In case the metric approaches the tolerance range, preventive action is required, which triggers an SLA negotiation service. This is the first phase of SLA Monitoring and Negotiation. The Monitor starts the Adaptation Manager (AM), which in turn proceeds to find an alternative resource provider. If the AM fails to find an alternative provider, a reconfiguration process is started to substitute the application or service. SLA Negotiation may face conflict between multiple SLAs, this will be harmonized associating priorities. We express the Monitoring layer by means of collaborations and activities specified in UML 2.0 diagrams and substantiated with temporal logic using the methodology and tool suite Arctis [17, 18].

Arctis represents an engineering approach for reactive systems that uses collaborative building blocks in the form of reusable specification units to create more complex services through composition. These building blocks are framed as UML models. Structural aspects are expressed by means of collaborations, while detailed behavior is defined by activity diagrams. In addition, blocks built on Arctis are easily reusable. Service specifications can be analyzed with model checking, a facility incorporated in the tool suite. The specifications can be further transformed from the more abstract specifications towards implementation via refinement steps, ensuring consistency as state machines are generated automatically. The tool suite allows us to specify the SLA Monitor and AM in a streamlined fashion, therefore facilitating their implementation.

In a second phase, in case corrective action is required, a lightweight dynamic service composition process is triggered. The AM is built extending previous work further explained in [19]. In this model, reconfigurations are scheduled as non pre-emptive real-time tasks with highest priority to assure that once a reconfiguration runs, it is not disturbed by other tasks. The underlying component model is based on the UML 2.0 architecture metamodel, assuring compatibility between the components of the Monitoring layer. The AM is developed with Arctis, which facilitates its reuse as a building block. In the same way, reconfiguration is facilitated by the identification of predefined variation points in the Delivery layer.

In order to achieve reliable dynamic adaptable services we need to evaluate service modifications against a number of underlying Quality of Service properties, in addition to those

defined in the specifications that the system strives to preserve.

Running Services must comply as a minimum with the following three properties. The first one, Responsiveness, means that the service always guarantees an answer to every received service request, unless the user cancels. The second property, Availability, requires that the service is always capable to accept a request. Finally, the third property, Reliability, means that the service request can always succeed. In order to evaluate a system against these properties, we need to formulate them in a such a way that they can be analyzed, preferably in an automated manner. Therefore, the AM is tested against these properties with the model checker in Arctis.

In a third phase, an SLA Negotiation service that associates priorities to Service Level specifications is developed. This service is intended to substantiate dynamic negotiation of QoS parameters and the decision-making process effected by the AM.

6 Outlook

Cloud computing has the power to provide the environment for new business models and expertise to ensure economic value creation while enhancing productivity. It also represents novel ways to deliver services. Desirable features for the future of cloud computing include among others: improved accessibility and availability, scalability according to needs, enhanced computational power, production of customizable software products, and composition or aggregation of higher-value products. Some of these features are already within reach. However, some research avenues require further elaboration through dedicated research and development to fully achieve these features. Particularly, the definition and identification of QoS parameters that permit identification of adaptation triggers and their subsequent enforcement through dynamic reconfiguration, as well as, an enhanced preservation of security concerns.

To achieve this vision, we foresee the emergence of families of services backed by the power and flexibility of cloud computing for an ever growing mobile working force, which demands services tailored according to specific business needs, but also for a considerable number of service consumers. First, we propose to advance the use of PLE to develop lines of services in the form of cloud families. Second, we consider extending our framework to consider identification of security threats and deviations in service provision with respect to established SLAs. On the whole, we propose the development of a comprehensive framework for dependable cloud families. The definition of cloud families will ease development and reuse of common services while allowing for tailoring of services for specific users. This in turn, promotes monitoring and enforcement of quality of service properties. A valuable characteristic for environments such as mobile computing, therefore, helping promote its use in every day life.

7 Conclusions

In this paper, we present a novel approach toward Cloud Computing and Product Line Engineering that rests on these promising lines of research. That is, the use of Product Line Engineering and Service Level Agreements to ensure QoS levels. PLE facilitates identification and exploitation of commonalities among sets of related services, allowing to define tailored cloud infrastructures. Determination of variation points within the line family exhibits a twofold objective, tailoring services for particular groups of users and obtaining systems capable of coping with changing conditions in the operating environment, inclusion of new services, or altering user requirements by means of dynamic adaptation.

References

[1] C. Kotsokalis, P. Chronz, K. Lu, M. Rojas, E. Yaqub, K. Kearney, F. Torelli, B. Fuentes, A. C. Escudero, T. Tsigritis, H. Foster, G. S. F. Zulkernine, and U. Winkler, "SLA foundations and management," SLA@SOI, Deliverable, September 2010.

[2] M. Comuzzi, C. Kotsokalis, G. Spanoudakis, and R. Yahyapour,

"Establishing and monitoring SLAs in complex service based systems," in *Proceedings of the 2009 IEEE International Conference on Web Services*, ser. ICWS '09. Washington, DC, USA: IEEE Computer Society, 2009, pp. 783–790.

[3] L. Jin, V. Machiraju, and A. Sahai, "Analysis on service level agreement of web services," Software Technology Laboratory HP Laboratories Palo Alto, HP Laboratories Palo Alto HPL-2002-180, Tech. Rep., June 2002.

[4] R. Hauck and H. Reiser, "Monitoring of service level agreements with flexible and extensible agents," in *HP OpenView University Association (HP-OVUA) Plenary worhsop*, 1999.

[5] A. Sahai, V. Machiraju, M. Sayal, L. J. Jin, and F. Casati, "Automated SLA monitoring for web services," in *IEEE/IFIP DSOM*. Springer-Verlag, 2002, pp. 28–41.

[6] K. Czajkowski, I. Foster, C. Kesselman, V. Sander, and S. Tuecke, "SNAP: A protocol for negotiating service level agreements and coordinating resource management in distributed systems," in *In 8th Workshop on Job Scheduling Strategies for Parallel Processing*, 2002, pp. 153–183.

[7] L. E. Grit, "Extensible resource management for networked virtual computing," Ph.D. dissertation, Duke University, 2007.

[8] K. Beaty, A. Kochut, and H. Shaikh, "Desktop to cloud transformation planning," in *IEEE International Symposium on Parallel Distributed Processing, 2009. IPDPS 2009.*, May 2009, pp. 1 –8.

[9] G. Lai, H. Song, and X. Lin, "A service based lightweight desktop virtualization system," in *Proceedings of the 2010 International Conference on Service Sciences*, ser. ICSS '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 277–282. [Online]. Available: http://dx.doi.org/10.1109/ICSS.2010.44

[10] K. Pohl, G. Böckle, and F. J. v. d. Linden, *Software Product Line Engineering: Foundations, Principles and Techniques*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2005.

[11] "Sun ray white papers," http://www.sun.com/sunray.

[12] "Ibm introduces ready-to-use cloud computing," http://www-03.ibm.com/press/us/en/presskit, November 2007.

[13] "Amazon elastic compute cloud (Amazon EC2)," http://aws.amazon.com/ec2/.

[14] S. H. Chang and S. D. Kim, "A variability modeling method for adaptable services in service-oriented computing," in *Proceedings of the 11th International Software Product Line Conference*. Washington, DC, USA: IEEE Computer Society, 2007, pp. 261–268.

[15] Y. Kim and K.-G. Doh, "Adaptable web services modeling using variability analysis," *International Conference on Convergence Information Technology*, vol. 1, pp. 700–705, 2008.

[16] R. Jain, *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling*. New York, NY: Wiley- Interscience, April 1991.

[17] F. A. Kraemer, "Engineering reactive systems: A compositional and model-driven method based on collaborative building blocks," Ph.D. dissertation, Norwegian University of Science and Technology, 2008.

[18] F. A. Kraemer, V. Slatten, and P. Herrmann, "Tool support for the rapid composition, analysis and implementation of reactive services," *Journal of Systems and Software*, vol. 82, no. 12, pp. 2068 – 2080, 2009.

[19] S. Brennan, S. Fritsch, Y. Liu, A. Sterritt, J. Fox, E. Linehan, C. Driver, R. Meier, V. Cahill, W. Harrison, and S. Clarke, *Architecting Dependable Systems VII (ADS VII)*. Springer-Verlag Berlin Heidelberg, 2010, vol. LNCS 6420, ch. A Framework for Flexible and Dependable Service-oriented Embedded Systems, pp. 123–145.

Network Intrusion Detection System – A Novel Approach

Krish Pillai, Ph.D
Department of Computer Science
Lock Haven University of Pennsylvania
Lock Haven, PA 17745, U.S.A
kpillai@lhup.edu

ABSTRACT

Network intrusion starts off with a series of unsuccessful break-in attempts and results eventually with the permanent or transient failure of an authentication or authorization system. Due to the current complexity of authentication systems, clandestine attempts at intrusion generally take considerable time before the system gets compromised or damaging change is affected to the system giving administrators a window of opportunity to proactively detect and prevent intrusion. Therefore maintaining a high level of sensitivity to abnormal access patterns is a very effective way of preventing possible break-ins. Under normal circumstances, gross errors on the part of the user can cause authentication and authorization failures on all systems. A normal distribution of failed attempts should be tolerated while abnormal attempts should be recognized as such and flagged. But one cannot manage what one cannot measure. This paper proposes a method that can efficiently quantify the behaviour of users on a network so that transient changes in usage can be detected, categorized based on severity, and closely investigated for possible intrusion. The author proposes the identification of patterns in protocol usage within a network to categorize it for surveillance. Statistical anomaly detection, under which category this approach falls, generally uses simple statistical tests such as mean and standard deviation to detect behavioural changes. The author proposes a novel approach using spectral density as opposed to using time domain data, allowing a clear separation or access patterns based on periodicity. Once a spectral profile has been identified for network, deviations from this profile can be used as an indication of a destabilized or compromised network. Spectral analysis of access patterns is done using the Fast Fourier Transform (FFT), which can be computed in $\Theta(N \log N)$ operations. The paper justifies the use of this approach and presents preliminary results of studies the author has conducted on a restricted campus network. The paper also discusses how profile deviations of the network can be used to trigger a more exhaustive diagnostic setup that can be a very effective first-line of defense for any network.

Keywords: Fast Fourier Transforms, statistical anomaly detection, network security, signal processing, surveillance, traffic analysis, intrusion detection

1. INTRODUCTION

Case studies have long indicated that legitimate users of network resources have a specific behavioural profile. It follows that changes in usage profiles can be used to detect and differentiate a masquerader from a regular user [1] , on the basis of reference signature profiles constructed from logged information. In the simplest approach, audit trails can be used to build a signature profile for normal users. More sophisticated Intrusion Detection Systems (IDS) may generate metrics from audit records to build profiles. The use of interval timing between accesses, session counters, and resource utilization tracking has been used successfully to generate reference profiles.

Such Intrusion Detection Systems generally collect and analyze time domain volume data such as number of logins per hour, program usage, and group resource or file system usage statistics. These are compared to thresholds which if exceeded could indicate a possible intrusion. Denning's classic paper classifies Statistical Models as follows [2] :

- Operational Model
- Mean and Standard Deviation Model
- Multivariate Model
- Markov Process Model
- Time Series Model

The Operational Model tracks abnormality by comparing metrics generated by event counters to an operational reference. On the other hand, the Time Series Model maintains event counters that are used to keep track of periodicity, and thereby the frequency of events. These two models, though cognizant of the periodicity of various events that occur in the system, fall short of providing a global view of the operational model of the system.

To illustrate this exposure, consider a simple counter that maintains a count of Secure Shell (ssh) logins over time (C_{ssh}), with a view to comparing it with a reference as in an Operational Model. Establishing a tolerance limit for C_{ssh} will not be easy since the aggregate logins being measured is the sum of multiple login cycles across various users. Clients log onto the network at individual times of the day based on cycles that are different for each one of them. As a result these cycles of differing periodicity may interfere constructively or get evened out causing the standard deviation of the C_{ssh} value to fluctuate over a wide range when viewed entirely in the time domain. In other words, the metric C_{ssh} being measured tends to have cycles within cycles as a result of the convolution of various usage patterns. Setting tight tolerance limits on instantaneous C_{ssh} may cause false alarms while setting wide tolerance limits to accommodate occasional spikes may result in a weaker detection system.

On the other hand, generating a spectrum of C_{ssh} will yield a frequency domain view of the metric. The samples contributing to each cycle would be available for individual inspection since spectral content is represented along the frequency axis. As a result, interference between high frequency and low frequency cycles would be easily discernable. A reference spectral profile can now be established for the spectrum of the metric C_{ssh}, where tolerance limits for each cycle could be set independent of the other. For example, a network may have a regular metric that cycles on a daily basis from nine-to-five biased by office hours.

For the sake of illustration, let us assume that superimposed on this regular traffic, there is a sharp increase in the measured metric twice every week on Tuesdays and Thursdays, due to the development team working on code compilation related issues. The traffic from the development team adds considerably to the regular traffic on these two days, but not on other days. Setting a threshold that changes dynamically on specific calendar dates is not feasible. Additionally, setting a single limit to represent the highly variant peak is not safe since there is the same likelihood of an intrusion on a busy day as on other days. Representing the metric in frequency domain and maintaining limits on the spectral profile allows the IDS to keep track of references to specific cycles that occur for the metric being monitored.

There are several algorithms for generating discrete frequency domain data from raw time domain data, of which the algorithm most widely used in signal processing is the Fast Fourier Transform (FFT). Spectral analysis is also done using Wavelet Transforms (DWT), especially in the area of pattern recognition or extraction. DWT though computationally less complex, is unsuitable for this specific purpose since we are not looking for time domain patterns but rather its stability. Furthermore, the advantage of FFT over DWT is that in addition to providing a Dirac comb (series of Dirac-Delta impulses) that is modulated by a series of discrete-valued coefficients that are also complex-valued, the basis functions are simple sinusoids and cosines. This is particularly useful since we only have knowledge of the periodicity of the pattern we are attempting to identify and not its nature. DWT on the other hand presumes knowledge for this pattern in time domain, so that a basis function can be generated. Moreover, FFT can be easily computed in real time using the Cooley-Tukey algorithm with very low computational complexity, $\Theta(N \log N)$, equivalent to that of a simple sorting algorithm.

The preliminary step to being able to detect variations in network usage is to establish a reference frequency-domain reference profile. Once a reference pattern has been established, the near-real-time profile can be checked against this reference for deviations from the normal and tolerance limits can be established. A method for defining and capturing a reference profile is detailed in the following section.

2. PROFILE DEFINITION AND METRICS

A behavioural profile can be based on various metrics that may be defined at any layer of the TCP/IP protocol stack. A metric such as SSH logins would count shell sessions and would be an *application* layer metric. Counting the number of TCP or UDP sessions would be a *transport* layer metric. Generating audit information for Internet Control and Messaging Protocol (ICMP) [4] messages would constitute a *network* layer metric. Metrics collected at lower layers of the TCP/IP stack would encapsulate metrics collected at upper layers. For example, a count of TCP sessions would include Email, SSH, as well as any other application layer protocol that is transported on TCP.

The choice of a metric is biased by the nature of the network under surveillance. For example, in a network that deploys an email server and a secure shell server, stronger cycles may be observable at the transport layer than at the application layer if most users use SSH redirection to get to their email service. Metrics for monitoring have to be chosen on the basis of the most vulnerable protocols that needs to be protected.

2.1 Defining a profile for a network

On a typical campus LAN (Local Area Network), there are servers and there are clients. Typically the services deployed on a network, such as Network File Service (NFS) or the Light Weight Directory Access Protocol (LDAP) would run periodically, driven by client access. Using LDAP access or NFS mount requests from clients as metrics would generate profiles that correlate very strongly with user access. For example repeated attempts to gain authentication would translate to perturbations in LDAP profiles, and attempts to use a compromised LAN port to gain access to distributed file systems would translate to failed mount requests, which in turn would show up in NFS profiles.

The simplest profile may consist of aggregate data or packet count collected periodically from the network for specific metrics. But metrics can be combined to create a palette of metrics for which spectral data can be arrived at independently. A palette may be limited to key protocols that are of special interest to the network administrator.

A typical metrics palette would be-

- SSH sessions and email sessions
- Java and PHP scripts
- NFS/LDAP

The key protocols in the aforementioned palette are SSH, SMTP (Simple Mail Transfer Protocol), Hyper Text Transfer protocol (HTTP), NFS and LDAP. The process of generating the palette spectrum involves collecting raw time domain date for each key protocol and then running the FFT algorithm on each time record. The following section explains the steps involved and the pre-processing that needs to be done before FFT can be applied.

2.2 Generating time domain statistics

Pegging metrics on the network can be done with minimal engineering impact through port duplication on the default gateway or ingress point into the network. All traffic ingressing and egressing the subnet should ideally be duplicated off to a port that connects to a machine dedicated to metrics collection, so that processing overheads do not impact network bandwidth. Alternately the data can be dumped to a file for near-real time or off-line analysis. Processing a file with time-domain metrics information involves the following steps-

- Filtering out instantaneous packet counts for each metric
- Aggregating them into counts over clearly defined time intervals (packets per minute is recommended) to extract time records for the metrics palette
- Preparing each time record for spectral analysis by averaging several time records and assigning weights them
- Application of Fast Fourier Transforms to the time records

A time record represents the window of time that is of interest to the observer. The FFT algorithm assumes that the time record repeats endlessly. Therefore, it is important that the time record captures the operational behaviour of the network and can be

used as a reference for comparisons. The time record is dependent on the periodicity of the metric being measured, and should last long enough to capture at least one cycle of the metric of interest. Each time record is composed of samples of the metric that are taken periodically. The sampling rate of any metric should be at least twice the maximum cycling frequency one expects to observe for a metric, as stipulated by the Nyquist-Shannon sampling theorem[3] . The more samples in a time record, the higher the resolution of the analysis would be. Choosing the time record for FFT analysis is critical and should be done on the basis of prior knowledge of the metric being sampled. As an example, a calendar week is a good candidate for a facility that is sparingly used on weekends, but has a usage pattern that repeats every week.

Once a time record is captured, the next step in processing the profile is the translation of these records to frequency domain using Fast Fourier Transforms (FFT). If 'f' is the frequency of a metric cycle detected by the transform, and '2πf' is denoted by 'ω', then a value K(ω) can be generated for each K(t) acquired. Viewing data in frequency domain has the effect of highlighting cycles in metric variations and separating them for each frequency. In other words, FFT takes a time record containing samples along the time axis and transforms them to amplitudes along the frequency axis.

If 'N' denotes the number of samples in a time record, then FFT generates 'N' points of the spectrum of which only N/2 represent positive frequencies since the transform yields complex valued coefficients. As a result, the spectrum generated by FFT is basically reflected about the Y-axis. Each frequency point generated by the FFT algorithm is referred to as a '*bin*', and for this purpose only the positive bins are of importance. Since phase information is of little value the negative valued bins can be discarded. The resolution of the spectrum generated by the FFT algorithm depends on how well the time record lines up with variation in the metric. Since this is not controllable easily, additional processing should be done on the time record before FFT can be applied. A time record may start capturing a metric at any point within its cycle and may end at any arbitrary point, as data acquisition is not synchronized with metric variations. Since FFT assumes that time records repeat, truncated cycles can leads to spurious results. Large variations in metric values at the beginning and end of time records can cause frequency values in one bin to leak into adjacent bins, the effect being a blurring of the frequency profile. To avoid this, the beginning and end of each time record should be smoothed out with a process called *windowing*. Windowing is the multiplication of each time-record of data collected with a function that is zero at the beginning and end of the time record. This has the affect of weighing down the values of samples taken at the beginning and end of each record so that their contribution to the algorithm is diluted, reducing frequency domain blurring. The literature describes many such windowing functions. The Hanning, Rectangular, Blackman-Harris, Kaiser, and the Tukey window are extensively used in conventional signal processing.

The Tukey window [5] gives fairly good amplitude accuracy and is simple enough for this application domain. This window gave consistent results when applied to weekly time records in this case study. The windowing function should always be chosen on the basis on what needs to be observed. For example, if transient load variations are of importance, then no window or a Uniform Window Function (UWF) should be used (Figure 1.), where each value measured in the time domain is equally weighted within the time record.

To summarize, the steps involved in capturing a metric profile are as follows:

1. Definition of the metric palette
2. Identification of time record for each metric
3. Data acquisition and generation of time domain data
4. Choice of windowing function and multiplication of each time record with the windowing function
5. FFT analysis of the time records using the Cooley-Tukey algorithm

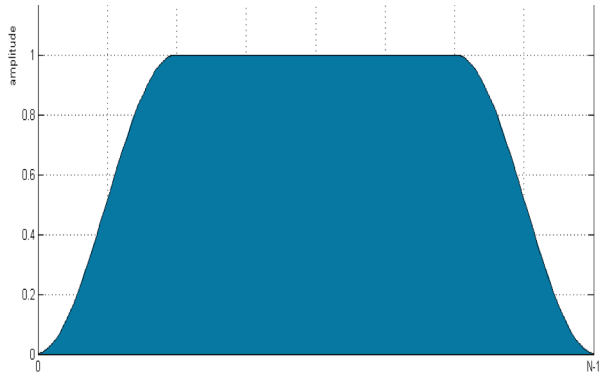


Figure 1. Tukey window to highlight data collected over week days

3. TIME DOMAIN VERSUS FREQUENCY DOMAIN

Time domain metrics collection gives a simple two-dimensional view of the behaviour of the network. The presence of cycles in the metric under observation is largely undetectable, especially if cycles of different periodicity coexist within the time record. For example, consider a Master-Slave file server system where the slave machine transfers files from the master every ten minutes. Assuming a secure setup, this would involve the slave using a protocol such as ssh to remotely copy files over 144 times a day from the master as super user. If ssh is the metric being measured, then a stable system should consistently show this base value, over which other attempts to login as super user onto the master machine would be superimposed. A system administrator may login regularly during office hours to monitor the system, and this would translate to a slower cycle with office-hours periodicity, super imposed on the base profile. An attempt to break into the master system on a daily basis by a potential intruder would affect the profile and would introduce more cycles or perturbations. Though these attempts might not affect the aggregate value of ssh login attempts substantially, it will definitely have a different periodicity from the normal behaviour of the ssh metric being collected.

To obtain a theoretical view of the application of FFT, one may visualize the network as a "black box". The network system takes an input and responds by producing an output. In the most simplistic model, the network may be categorized as a Single Input Single Output or a SISO system [6] . The input could be in reality an aggregation of several metrics or a metrics palette. Likewise, the output may be a composite attribute built up of several metrics. The discussion in this paper is limited to a single metric acquisition system for the sake of brevity.

There are certain interesting relationships between time domain and frequency domain metric variations. An impulse spike in the time domain for a metric will affect all bins in the frequency domain. A perfect cycle in the time domain on the other hand will show up as a spike in the frequency domain.

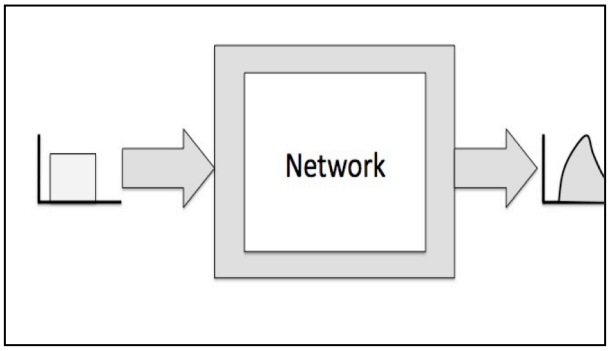


Figure 2. The network as a SISO block

Observation of metric patterns in frequency domain yields information on how the network responds to service requests. Standard control theory suggests that there is no non-determinism in how the network affects various traffic flows within the constraints imposed by the model. The objective of this paper is to use FFT to identify a signature pattern in the metric of interest, so that a reference can be established.

3.1 Capturing Metrics in Time Domain

In its simplest form the metric being observed may simply be the total number of packets flowing into the system. In the following sections, a simple traffic metric is used to illustrate the process of identifying reference patterns. Aggregate traffic can easily be captured using software such as MRTG to poll counters on a gateway or ingress point into the network. If specialized metrics are to be pegged, then more advanced software would have to be used. IPtables is an ideal user space application that allows for traffic accounting. IPtables is configured on the basis of rules that match a wide range of criteria. Rules can be implemented to match the criterion for a metric and a count of the number of times the rule gets matched can be generated. The IPtables device can then be periodically polled to generate a fairly accurate count of the metric under consideration. In the following section, we take a look at a simple traffic metric that establishes a reference profile and monitors for deviations from the reference.

3.2 A Case Study – Traffic volume as a Metric

A conventional Simple Network Management Protocol (SNMP) Management Information Base (MIB) was used to gather time domain information for the traffic volume metric. This metric is simply a bit-rate count in its simplest sense. An open source application (Multi Router Traffic Graphing Application - MRTG) was used to query the MIB [7] on the router to generate the necessary information. The average of aggregate packet flow with a five-minute sampling rate collected with MRTG is shown below. Each hour yields twelve samples and this translates to 288 samples a day. The data shown in Figure 3. shows a clear periodicity for every 288 samples or on a daily basis. This profile includes user activity superimposed over the no-load profile of the network under observation.

The network being observed has a Master-Slave file service and a centralized Network Information System (NIS) for authentication and authorization. Client machines have their file systems supplied from the file server using the "*automount*" protocol. Automount is a protocol that periodically mounts files on demand and dismounts folders from clients during moments of inactivity to optimize network traffic.

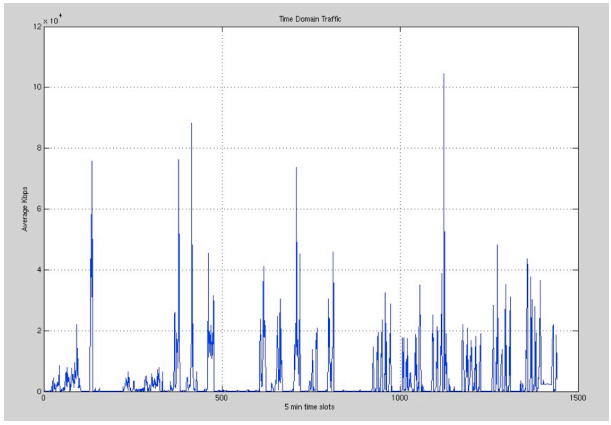


Figure 3. Five minute sample of traffic showing repetitive pattern

Distinct patterns are observable during periods of inactivity or no load, when the only traffic on the network is that generated from applications that run periodically to maintain network connectivity and the functionality of the distributed system. The figure shows traffic profile captured over a Sunday at the university UNIX Lab. The cyclic traffic from control software is evident in this graph.

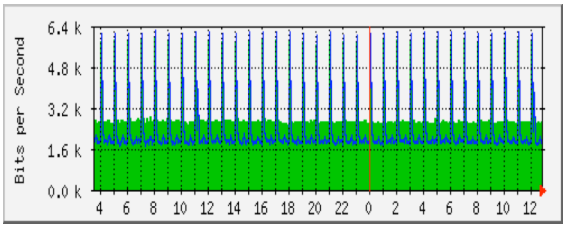


Figure 4. No-Load traffic pattern on a Sunday

A frequency spectrum would show a clear cycle associated with this record. A weekly profile, magnified here to accommodate the peaks, indicates a distinct deviation from the daily patterns, particularly on Tuesdays and Thursdays. This deviation from the pattern can be correlated to the fact that the database design class is scheduled on Tuesdays and Thursdays when client-server traffic is considerably elevated.

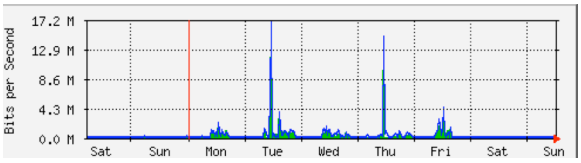


Figure 5. Weekly profile showing high activity on Tue and Thu

Converting from time to frequency domain using transient data yields unreliable results. The resulting spectrum can be stabilized considerably through *ensemble averaging*. This is basically the process of creating multiple spectra and combining them to provide a stable spectrum. As data is gathered over five minute intervals, a running average using a sliding window of a fixed number of profiles can used to track the near real-time spectrum and compare it with the signature for deviations.

For the empirical study conducted, the signature itself was built up of data captured on a typical day. FFT computation gives more stable profiles if more points are used. Since this data represents a typical day, larger calendar records were artificially generated through extrapolation and transformed to get a clearly defined signature profile.

3.3 Identifying a Signature Profile for the Metric

A unique signature would have peaks that are invariant in the profile as multiple ensembles are acquired. During periods of abnormal activity the cycles generated by usage traffic will be superimposed on the reference profile. Permanent changes in the reference profile could indicate the no-load behaviour of the network has changed, or the network steady state has been affected in some way. This could be indicative of misconfiguration or that some permanent alteration to the system has taken place. Selection of tolerance limits would require knowledge of load variations and deviations on a network. A few weeks of historical data would therefore be necessary to define a profile, and is not something that can be arrived at apriori. Once tolerance limits are established around a reference profile for the metric, deviations beyond tolerance limits can be used to trigger a Network Management Service (NMS) alarm. Multiple such profiles may be generated for a palette of metrics so that deviations along a range of values could be monitored for instability.

```
% Load file containing typical day run
load dayrecord.dat;

% Number of points to use for FFT computation
NPOINT = 256;

% Spectrum is scaled for easy graphing
% We are not interested in Power Spectral density
% Only the resonances matter
SPEC_D = abs(fft(dayrecord,NPOINT))/1000;

%
% Now generate a frequency scale that runs
% from 0 - (1 - 1/N)
% FFT yields real and imaginary components
% The whole spectrum is a reflection about the
centre
% Graph only half the screen
FREQ = [0:NPOINT-1]/NPOINT;
FREQ = FREQ *2;

plot(FREQ,SPEC_D),grid on, title('Daily Profile'),
axis([0 1 0 1000]);
```

Figure 6. Matlab code for generating FFT

A signature derived for the network under study is shown in Figure 7. The peaks circled identify the invariant part of the profile. FFT on the data set was computed using Matlab® [8] as a proof of concept, but will be implemented eventually using functionality that is part of an in-house application to facilitate this analysis.

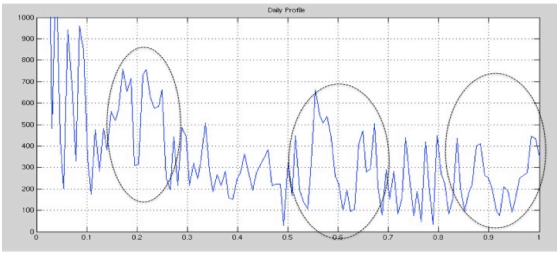


Figure 7. Signature for the lab under observation

The Matlab code snippet is fairly straightforward and can easily be implemented in C or Java using the Cooley-Tukey algorithm [9] . Packet capture can be implemented using libpcap [10] and supplemented with scripting and extraction from IPTables audits.

4. FUTURE WORK AND CONCLUSIONS

More data is being collected and various metrics are being investigated as part of this research. The author is in the processes of categorizing various use-case scenarios and mapping each to a metrics palette that would react to intrusion attempts in a timely manner. Validation would require extensive testing across various scenarios and will be done as more audit records and logs become available from various network service administrators. The functionality of the proposed detection system can be further enhanced using Artificial Intelligence (AI) [11] . The spectral analysis method described in this paper would add considerable value to a detection system that is driven by an AI Inference Engine, making it cognizant of variations in usage patterns. Identifying a signature profile for a metric and being able to represent this important attribute in quantitative terms makes it possible to use an automated scheme to track and monitor the network, and react in more sophisticated ways than just generate alarms.

As part of future work, the author plans to use this technique of monitoring networks as a backend to an expert system. The advantage of using Artificial Intelligence in network surveillance opens up various possibilities in the field of automated Intrusion Detection Systems.

References:

[1] James P. Anderson, “Computer Security Threat Monitoring and Surveillance”, **Contract 79F296400, National Institute of Standards and Technology**, April 15, 1980.

[2] Dorothy E. Denning, “An Intrusion Detection Model,” **IEEE Transactions on Software Engineering**, Vol. SE-13, No. 2, pp. 222-232, February 1987.

[3] C. E. Shannon, "Communication In The Presence of Noise", reprint as classic paper, **Proc. IEEE**, vol.86, no.2, (Feb. 1998)

[4] J. Postel, "Internet Control Message Protocol," **Internet Engineering Task Force (IETF) Request for Comments (RFC) 792**, Standard, September 1981.

[5] V. Oppenheim, R.W. Schafer and J.R. Buck, **Discrete-Time Signal Processing**, Prentice Hall, ISBN: 0-13-754920-2

[6] B.C.Kuo, and F. Golnaraghi. **Automatic Control Systems**, 8th Edition, Wiley, ISBN-13: 978-0471134763, September 2002

[7] D. Harrington, R. Presuhn, and B. Wijnen, "An Architecture for Describing Simple Network Management Protocol management Framework," **IETF, RFC 1157**, Standard, December 2002.

[8] David M. Smith, **Engineering Computation with MATLAB®,** ISBN-13: 978-0-321-48108-5, Addison-Wesley Computing, pp.369–373.

[9] Cooley, James W., and Tukey, John W., "An Algorithm for the Machine Calculation of Complex Fourier Series" in **Magnetism, Math. Comput**. vol: 19, pp.297-301, 1965.

[10] S. McCanne, and V. Jacobson, "The BSD Packet Filter: A New Architecture for User-level Packet Capture," **USENIX Conference**, January 25-29, 1993, San Diego, CA, U.S.A.

[11] S. Russell, and P. Norvig, **Artificial Intelligence: A Modern Approach,** Prentice Hall, 3rd Edition, ISBN-13: 978-0136042594, December 11, 2009.

Performance Analysis and Optimal Utilization of Inter-Process Communications on Commodity Clusters

Yili TSENG
Department of Computer Systems Technology
North Carolina A & T State University
Greensboro, NC 27411, U.S.A.

ABSTRACT

Classical science is based on theory, observation and physical experimentation. Contemporary science is characterized by theory, observation, experimentation and numerical simulation. With the use of hardware and software we can simulate lots of phenomenon. This saves time, money and physical resources. Simulation of a certain phenomenon requires lots of computing power. Answer to these computational power needs is high performance computers (HPC). HPCs consist of numerous processors working on the same task in parallel. In the past, high performance computers were very expensive and affordable by few institutions. After Message Passing Interface (MPI) library is ported to the PC platform, commodity clusters can be built of inexpensive PCs and afforded by any researcher. Lots of performance analyses have been conducted on high-end supercomputers. None has been done on commodity clusters. In this paper, experiments for six major MPI communication functions were performed on eight different configurations of clusters. Performance analyses were then conducted on the results. Based on the results, methods for optimal utilization of inter-process communications on commodity clusters are proposed

Keywords: Computational Science, High Performance Computing, Computing Education, Commodity Cluster, Parallel Programming

1. INTRODUCTION

The Need for High Performance Computing

Classical science is based on observation, theory, and physical experimentation. In contrast, with the utilization of computer hardware and software, contemporary science is characterized by observation, theory, experimentation, and numerical simulation [1].Numerical simulation is a mathematical modeling on a discrete model. It may represent a discrete approximation of the continuum partial differential equations, or it may represent a statistical representation of the microscopic model [2]. .Numerical simulation is an increasingly important tool for scientists and engineers, who often cannot use physical experimentations to test because they may be too expensive or time-consuming, because they may be unethical, or because they are impossible to perform [1]. .Numerical simulation requires computers to carry out calculations. Many important scientific and engineering problems are so complex that solving them via numerical simulation requires extraordinary powerful computers. Those complex problems are often called grand challenges for science and require high performance computers to perform numerical simulations [2]. High Performance Computers (HPCs) refer to parallel computers or multiprocessor computers. HPCs gain high throughput by having plenty of processors working in parallel. Categorized by communication models, parallel computers are classified into two major categories, namely shared-memory platform and message-passing platform [3] - [5].

Dilemma of High Performance Computing and Resolution

Although numerical simulation becomes an important tool for many science and engineering disciplines, it relies on the availability of HPCs. In the past, HPCs are very expensive and cannot be afforded by most researchers and institutions. The high costs were the barrier to wide adoption of HPCs. Nonetheless, due to the simplicity of message-passing platform's architecture and the porting of MPI library to personal computer (PC) platform by open-source software developers, now researchers and institutions can connect inexpensive PCs with generic networking and install MPI library on them to build a commodity cluster [9] - [12]. Even retired PCs can be used to build a commodity cluster [13]. That enables researchers who are interested in numerical simulations to own an effective and affordable tool to perform research through numerical simulations. The commodity clusters still have decent computing power to solve smaller-scale numerical simulations.

Related Work

Although lots of performance analyses have been conducted for HPCs and high performance networking systems in the past, they were all performed on high-end HPCs. No performance analysis has been done on commodity clusters. As they use generic parts instead of high performance parts employed by high-end supercomputers, for example Gigabit Ethernet networking versus InfiniBand networking, commodity cluster could exhibit different characteristics than those of supercomputers. Because they are more popular and adopted than before, performance analysis should be conducted on commodity clusters to investigate their characteristics in executing MPI communication functions in order to carry out optimal utilization of them.

2. INTER-PROCESS COMMUNICATION EXPERIEMNTS

MPI Library

Since commodity clusters are based on the message-passing platform, they rely on passing messages to other processes to coordinate parallel tasks. The most popular communication protocol for the message-passing platform is Message Passing Interface (MPI) while the most popular MPI library is Open MPI [6]. The Open MPI library used in the experiments is version 1.3.

Tested MPI Functions

The MPI communication functions consist of two major categories: point-to-point communication and collective communication. Point-to-point communication functions are used to send messages from one process to another. Collective communication functions involve all processes participating in a parallel program [2], [7], [8].

Point-to-Point Communication Functions: The point-to-point communication functions tested in the experiments were MPI_Send and MPI_Recv. MPI_Send is called by a process to send a message to another process. It must be paired by a MPI_Recv function called by the receiving process. Both functions are blocking functions which do not return until the message is successfully passed from sending process to receiving process. Therefore, the performance of both functions is essential to MPI programs.

Collective Communication Functions: The five collective communication functions tested in the experiments were MPI_Bcast, MPI_Reduce, MPI_Gather, MPI_Scatter, and MPI_Alltoall. A communicator is defined as all processes participating in a parallel program. MPI_Bcast is called by a process to broadcast the same message to all processes in the same communicator, including itself. MPI_Reduce collects a message from each process in the same communicator, including itself, and reduce them with a specified operation to a variable in a specified process. MPI_Gather collects a message from each process in the communicator, including itself, and stores them in an array in the order of the rank of each process. MPI_Scatter splits an array in a process and distributes one segment to one process in the communicator in the order of the rank of each process. MPI_Alltoall is equivalent to all processes in the communicator calling MPI_Gather or MPI_Scatter [3], [7], [8].The five collective communication functions are most frequently used in and hence are fundamental to MPI programs.

Methodologies

In the experiments, different MPI communication functions were executed with different sizes of messages ranging from 4 bytes, the length of an integer variable, to 1 Mbytes on different cluster configurations. The message length is multiplied by a factor of 2 for the next sample. For the point-to-point communication functions, the well-known ping-pong method was utilized [13], [15], [16].In the ping-pong method, process 0 calls MPI_Send to send a message to process 1 while process 1 uses MPI_Recv to receive the message. Process 1 immediately sends the same message back to process 0. The latency of the ping-pong operation is measured and then divided by two to determine the one-way point-to-point communication time. The procedure is repeated for 1000 times. In reality, MPI_Wtime is called before the first iteration and after the 1000^{th} iteration to dilute the overhead of calling MPI_Wtime. The average latency is calculated and used as the result. As for collective communication functions, the function is called and followed by calling MPI_Barrier to ensure the collective communication is synchronized on all processes before marking ending time [15], [17]. The procedure is repeated for 1000 iterations. Likewise, MPI_Wtime is called before and after the 1000 iterations and the average time is calculated as the result. For all functions, three iterations are executed to "warm up" the communication channels before the beginning time is marked. The number of iterations used was 20 in the methodology of Xu et al.[13] while it was 1000 in the paper of Miguel et al. [14] and OSU Benchmarks [17]. The results from Luecke et al. [15] show that 15000 iterations do not improve the accuracy significantly. Hence, 1000 iterations were used in the experiments.

Configurations of the Tested Clusters

Three physical clusters were used in the experiments: Four nodes equipped with AMD Phenom X4 2.5GHz CPU and 1GB RAM each, four nodes equipped with Intel Pentium 4 2.4GHz CPU and 256MB RAM each, and eight nodes equipped with Intel Pentium III CPU and 256MB RAM each. Eight different configurations were created by changing network interface cards, adjusting number of nodes participating the testing program, and dispatching different number of processes to be executed by each node. The detailed information of each configuration is listed in Table 2.1.

Table 2.1.Configurations of Tested Commodity Clusters

Configuration	CPU	Networking	Number of Node
1	AMD Quadcore 2.5GHz	Internal Channels	4 cores
2	AMD Quadcore 2.5GHz	Gigabit Ethernet	4
3	Intel Pentium 4 2.4GHz	Gigabit Ethernet	4
4	Intel Pentium 4 2.4GHz	100Base-T Ethernet	4
5	Intel PIII 700MHz	Gigabit Ethernet	4
6	Intel PIII 700MHz	100Base-T Ethernet	4
7	Intel PIII 700MHz	Gigabit Ethernet	8
8	Intel PIII 700MHz	100Base-T Ethernet	8

3. PEFORMANCE ANALYSIS ON RESULTS

The performance of all configurations of clusters was compared for each MPI communication function first. Then each collective communication function was compared against MPI_Send for each configuration of cluster individually to determine whether the specific function performs better than equivalent number of MPI_Send on a specific configuration of cluster. Lastly, methods for optimal utilization of inter-process communications were proposed based on the performance analyses. As the size of message length is multiplied by a factor of two for next sample, it grows exponentially and is difficult to be displayed and distinguished on the same chart due to non-uniform scale. Therefore, all results for each test are divided into four groups and are charted as four graphs so that the scale can be properly shown and all samples can be distinctly read.

Comparison of Performance of All Clusters

In this section, results from configurations with 4 nodes for the same MPI communication function are plotted on the same chart and compared. The latency of results from the configuration of PIII/100Base-T/4 nodes is too long when compared with readings from the other five configurations. Hence, they are removed so that other results can be displayed with the proper scale and are distinguishable. Charts for MPI_Send, MPI_Bcast, MPI_Reduce, MPI_Gather, MPI_Scatter, and MPI_Alltoall were plotted and displayed from Figure 3.1.to Figure 3.6. In Figure 3.1., which displays the performance of MPI_Send, it shows that, from message length of 16 Kbytes up, the latency for different cluster from low to high falls in the following order:

1) Quadcore/internal channels
2) Quadcore/Gigabit
3) P4/Gigabit
4) PIII/Gigabit
5) P4/100Base-T
6) PIII/100Base-T

Although the ranking of latency for message length under 16 Kbytes is not in the same order, the latency ranking for all other MPI communication functions tested for all message lengths are quite in this order except for MPI_Scatter and MPI_Alltoall with large messages over 32 Kbytes. The only difference is that latency of PIII/Gigabit is better than that of P4/Gigabit for those large file sizes. All other four clusters are still in the same order. That shows that generally the performance of those configurations is in this ranking order. It demonstrates that Gigabit NICs perform significantly better 100Base-T NICs. In turn, internal channels outperform Gigabit networking.

Performance Analyses on Collective Communication Functions

As MPI collective communication functions send out messages to all processes in the same communicator, the following performance analyses compare the latency of each MPI collective communication function to the aggregate latency of equivalent number of MPI_Send. For example, calling MPI_Bcast in a parallel program which is executed with four processes is equivalent to calling MPI_Send three times to send the same message to the other three processes because the overhead of sending the same message to itself is negligible.

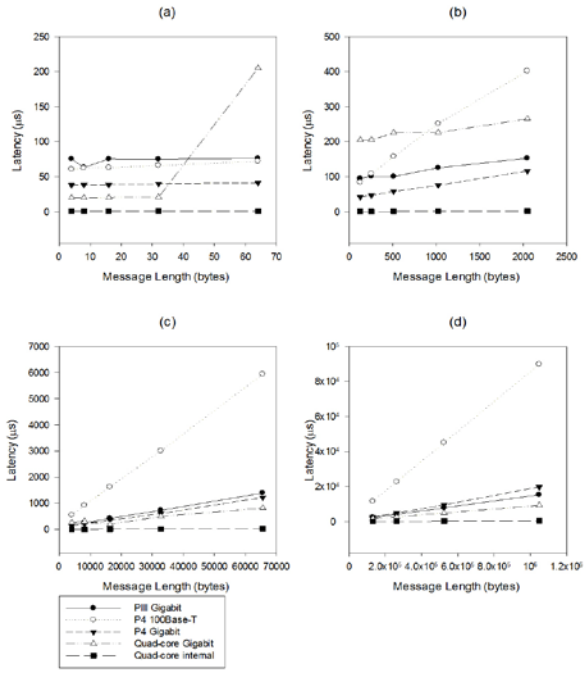


Figure 3.1.Comparison of MPI_Send Latency on All Clusters

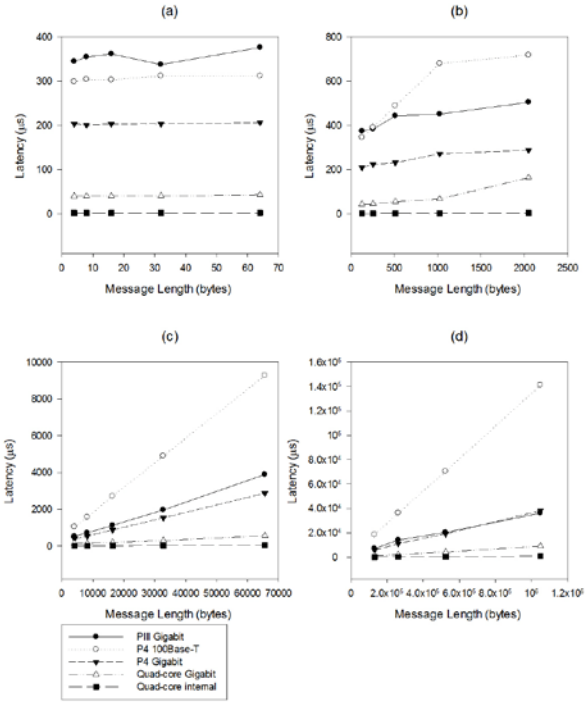


Figure 3.2.Comparison of MPI_Bcast Latency on All Clusters

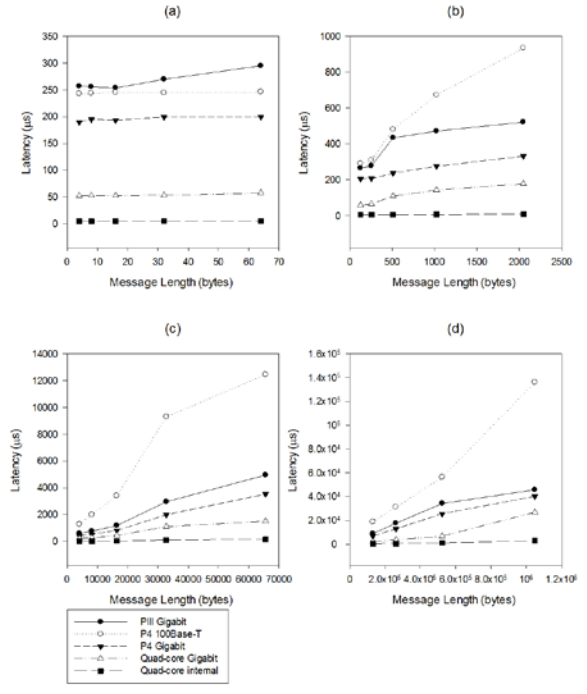


Figure 3.3.Comparison of MPI_Reduce Latency on All Clusters

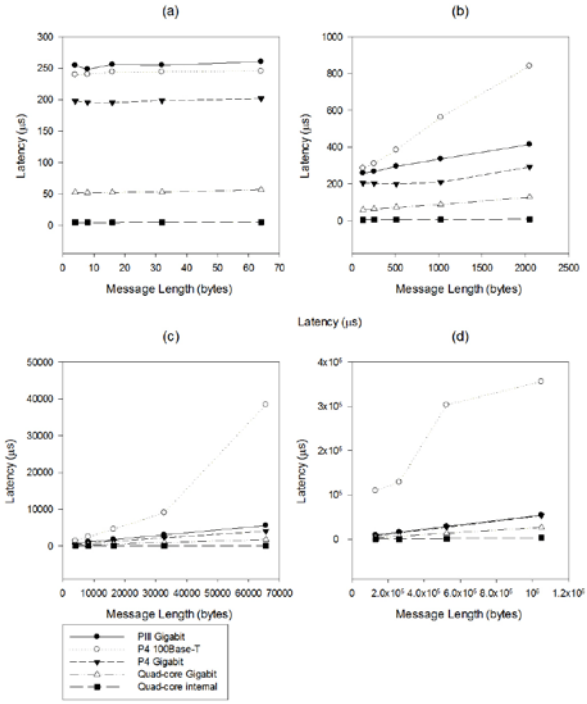


Figure 3.4.Comparison of MPI_Gather Latency on All Clusters

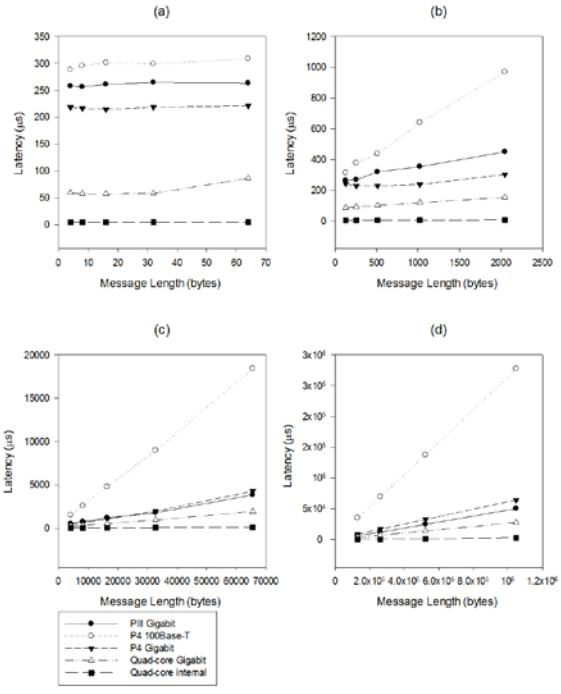


Figure 3.5.Comparison of MPI_Scatter Latency on All Clusters

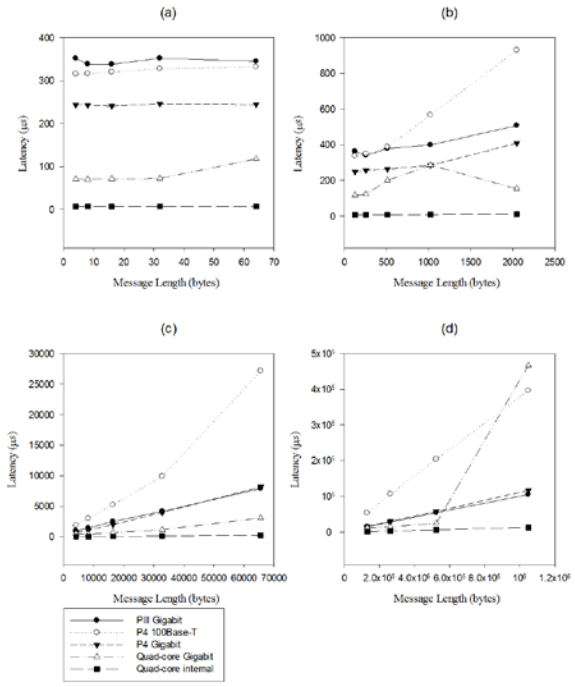


Figure 3.6.Comparison of MPI_Alltoall Latency on All Clusters

The latency of executing MPI_Bcast with four processes is compared with the aggregate latency of executing MPI_Send three times. Similarly, the latency of executing MPI_Alltoall with four processes is compared with the aggregate latency of executing MPI_Send twelve times because each process of the four processes need to collect one message from the other three processes. The comparison results for each tested collective communication function are listed in Table 3.1 through Table 3.5. The notation L means that the latency of the specific function is lower than that of the corresponding number of MPI_Send while H means higher. The changing length refers to the length of the message where the characteristic of the latency comparison changes.

Table 3.1 Observations on MPI_Bcast versus MPI_Send

Configuration	Short Message Latency	Long Message Latency	Changing Length (bytes)
Quadcore 2.5GHz/Internal Channels/4 nodes	L	L	
Quadcore 2.5GHz/Gigabit Ethernet/4 nodes	L	L	
P4 2.4GHz/Gigabit Ethernet/4 nodes	H	L	2048
P4 2.4GHz/100Base-T Ethernet /4 nodes	L	H	1024
PIII 700MHz/Gigabit Ethernet/4 nodes	H	L	262144
PIII 700MHz/100Base-T Ethernet/4 nodes	H	H	
PIII 700MHz/Gigabit Ethernet/8 nodes	H	L	128
PIII 700MHz/100Base-T Ethernet /8 nodes	L	H	16384

Table 3.2 Observation on MPI_Reduce versus MPI_Send

Configuration	Short Message Latency	Long Message Latency	Changing Length (bytes)
Quadcore 2.5GHz/Internal Channels/4 nodes	H	H	
Quadcore 2.5GHz/Gigabit Ethernet/4 nodes	L	L	
P4 2.4GHz/Gigabit Ethernet/4 nodes	H	L	2048
P4 2.4GHz/100Base-T Ethernet /4 nodes	H	L	1024
PIII 700MHz/Gigabit Ethernet/4 nodes	H	H	
PIII 700MHz/100Base-T Ethernet/4 nodes	H	H	
PIII 700MHz/Gigabit Ethernet/8 nodes	H	L	128
PIII 700MHz/100Base-T Ethernet /8 nodes	L	H	16384

Table 3.3 Observation on MPI_Gather versus MPI_Send

Configuration	Short Message Latency	Long Message Latency	Changing Length (bytes)
Quadcore 2.5GHz/Internal Channels/4 nodes	H	H	
Quadcore 2.5GHz/Gigabit Ethernet/4 nodes	L	L	
P4 2.4GHz/Gigabit Ethernet/4 nodes	H	L	262144
P4 2.4GHz/100Base-T Ethernet /4 nodes	L	H	65536
PIII 700MHz/Gigabit Ethernet/4 nodes	H	H	
PIII 700MHz/100Base-T Ethernet/4 nodes	H	H	
PIII 700MHz/Gigabit Ethernet/8 nodes	L	H	8192
PIII 700MHz/100Base-T Ethernet /8 nodes	L	H	8192

Table 3.4 Observation on MPI_Scatter versus MPI_Send

Configuration	Short Message Latency	Long Message Latency	Changing Length (bytes)
Quadcore 2.5GHz/Internal Channels/4 nodes	H	H	
Quadcore 2.5GHz/Gigabit Ethernet/4 nodes	L	L	
P4 2.4GHz/Gigabit Ethernet/4 nodes	H	H	
P4 2.4GHz/100Base-T Ethernet /4 nodes	H	H	
PIII 700MHz/Gigabit Ethernet/4 nodes	H	H	
PIII 700MHz/100Base-T Ethernet/4 nodes	L	H	1024
PIII 700MHz/Gigabit Ethernet/8 nodes	L	L	

PIII 700MHz/100Base-T Ethernet /8 nodes	L	H	16384

Table 3.5 Observation on MPI_Alltoall versus MPI_Send

Configuration	Short Message Latency	Long Message Latency	Changing Length (bytes)
Quadcore 2.5GHz/Internal Channels/4 nodes	L	H	131072
Quadcore 2.5GHz/Gigabit Ethernet/4 nodes	L	H	1048576
P4 2.4GHz/Gigabit Ethernet/4 nodes	L	L	
P4 2.4GHz/100Base-T Ethernet /4 nodes	L	L	
PIII 700MHz/Gigabit Ethernet/4 nodes	L	L	
PIII 700MHz/100Base-T Ethernet/4 nodes	L	H	8192
PIII 700MHz/Gigabit Ethernet/8 nodes	L	L	
PIII 700MHz/100Base-T Ethernet /8 nodes	L	H	16384

Methods for Optimal Utilization of Inter-Process Communications

Based on the previous observations and performance analysis, the following methods are recommended for optimal utilization of inter-process communications on commodity clusters: 1) Use faster CPUs 2) Adopt multicore CPUs to utilize the internal communication channels 3) Employ Gigabit networking 4) Run tests of different MPI communication command against MPI_Send

4. CONCLUSION

With the advancement of computer hardware and software, numerical simulation becomes an important tool for researchers. Nonetheless, complicated numerical simulations rely on HPC. In the past, HPCs were very expensive and not affordable by most researchers and institutions. Due to the simplicity of message-passing platform and the porting of MPI library to PC, commodity clusters can be built out of inexpensive PCs. They provide an effective tool for all researchers to take advantage of numerical simulations. Although a lot of performance analyses have been conducted HPCs, they are all on high-end HPCs. In order to take the most out of commodity clusters, performance evaluations should be performed on them. Experiments which executed six major MPI communication functions were performed on eight different configurations of clusters. Performance analyses were conducted on results from those experiments. It was observed that performance of networking system still plays the key role. A cluster equipped with PIII CPUs and Gigabit NICs outperforms a cluster consisting of Pentium 4 CPUs and 100Base-T NICs. Gigabit NICs are very cheap to acquire nowadays. The cost/performance ratio of upgrading to Gigabit networking is well worthy. It was also observed that collective communication functions may not perform better than utilizing multiple MPI_Send for certain configuration of cluster with certain size of message. Recommendations for optimal utilization of inter-process communications based on observations on experiment results were proposed: 1) Use faster CPUs 2) Adopt multicore CPUs to utilize the internal communication channels 3) Employ Gigabit networking 4) Run tests of different MPI communication command against MPI_Send

5. REFERENCES

[1] M. Quinn, *Parallel Programming in C with MPI and OpenMP*, McGraw-Hill, 2004.

[2] G. Karniadakis and R. Kirby, *Parallel Scientific Computing in C++ and MPI*, Cambridge University Press, 2003.

[3] P. Pacheo ,*Parallel Programming with MPI*, Morgan-Kaufmann, 1996.

[4] A. Grama, A. Gupta, G. Karypis and V. Kumar, *Introduction to Parallel Computing* ,Addison-Wesley, 2003.

[5] K. Hwang, *Advanced Computer Architecture: Parallelism, Scalability, and Programmability*, McGraw Hill, New York, 1993.

[6] Open MPI: Frequently Asked Questions. http://www.openmpi.org/faq.

[7] W. Gropp, E. Lusk and A. Skjellum, *Using MPI*, MIT Press, 1999.

[8] W. Gropp,E. Lusk, and R. Thakur, *Using MPI-2*, MIT Press, 1999.

[9] J. Sloan, *High Performance Linux Clusters*, O'Reilly, 2005.

[10] C. Bookman, *Linux Clustering*, New Riders, 2003.

[11] A. Vrenios, *Linux Cluster Architecture*, Sams, 2002.

[12] R. Lucke, *Building Clustered Linu Systems*, Prentice Hall, 2005.

[13] Z. Xu and K Hwang,"Modeling Communication Overhead: MPI and MPL Performance on the IBM SP2,"*IEEE parallel & distributed Technology*, Vol. 4,No.1, Spring 1996, pp. 9-23.

[14] J. Miguel and A. Arruabarrena, "Assessing the Performance of the New IBM SP2 Communication Subsystem," *IEEE Parallel & Distributed Technology*.1996.

[15] G. Luecke, J. Coyle, J. Hoekstra, and M. Kraeva, "Measuring MPI Latency and Communication Rates for Small Messages," Shanghai Supercomputing Center,2009.

[16] B. Barrett, J. M. Squyres, A. Lumsdaine, R. L. Graham and G. Bosilca,"Analysis of the Component Architecture Overhead in Open MPI," Euro PVM/MPI2005, 2005.

[17] The Ohio State University Benchmarks: http://nowlab.cse.ohio-state.edu/projects/mpi-iba/benchmarks.html.

Evolution of the New Construction Classroom

Dr. Allan CHASEY, Heather CISZCZON
Arundhati GHOSH, Lisa HOGLE
Del E Webb School of Construction, Arizona State University
Tempe, AZ 85281, USA

ABSTRACT

The Del E. Webb School of Construction (DEWSC) at Arizona State University has taken the initiative to transform the curriculum by evolving a method of instruction that includes more collaborative hands-on instructional methods. Courses taught in a traditional classroom setting are useful but not always the best learning environment. Transitioning to a collaborative method of teaching has transformed the physical layout of the classroom; in lieu of rows of desks and chairs, a Virtual Construction and Collaboration Lab (VC^2L) was developed. The space planning program for the VC^2L called for features that would stimulate group work and conversations with supporting tools and technology.

Based on a literature review, VC^2L was developed as a pod-based scenario, with movable furniture allowing for changing the environment for the particular task. In the VC^2L classroom, collaboration takes the form of group assignments centered on virtual construction activities. Utilizing Building Information Modeling (BIM), current leading edge technology within the construction industry, a digital representation of a project is provided to the students allows input and extraction of project information by the group. This paper summarizes the process of the methodical change in the undergraduate construction curriculum while addressing the design of the new construction classroom.

Keywords: BIM, Collaboration, Construction, Education, Virtual Construction.

1. INTRODUCTION

Construction education has evolved over the years to include components such as guest lectures, internships, professional development, site visits, and technology labs, making the student knowledgeable and experienced. While the curriculum has progressed, the user experience has not received much attention. Collaboration is a critical aspect of design, engineering, construction and management and is a requirement for establishing a higher standard of building. In academia, students can no longer be merely passive recipients of information, but must proactively participate, putting into practice what they have learned, and learning via putting theory into practice. A collaborative method of teaching encourages students to acquire teamwork skills as they mimic scenarios that could be encountered on a typical construction job-site. Understanding and implementing this concept of collaboration requires a cultural change in already established methods of instruction and learning. While the curriculum is transformed to encourage team activities, the physical environment must also be conducive for collaboration to occur.

As the industry grows and adapts to the latest trends, it is essential that students follow suit and lead the evolution. To further integrate the changing patterns of industry growth, the curriculum must now focus on the aspect of collaboration as kindled by Building Information Modeling. Building Information Modeling (BIM) is the process which aids this cultural change of collaboration. Academia must allow the permeability of BIM in all levels of courses offered in a Construction Management program. To mirror this paradigm, a Virtual Construction and Collaboration Lab (VC^2L) becomes the environment where students apply their education and skills to solve real world problems. It also becomes the platform for experimentation and research that can ultimately be brought to the industry.

2. COLLABORATION AND BIM

Collaborative or constructivist form of learning is a theory that replaces the passive instructional approach and concentrates on active learning through dialogue, constructing, inventing and creating.[2] In the field, collaboration is important for reducing design and coordination errors, cost effective and expedited delivery, and for increasing the overall value of architecture and construction. Russian psychologist Lev Vygotsky (1896-1934) conceived the idea of ‘Social Development Theory’ which forms the foundation of constructivism. [3] It states that ‘while social interaction plays a fundamental role in the process of cognitive development, most of the learning occurs in the Zone of Proximal Development (ZPD), which is the distance between the students’ ability to perform a task with

collaboration and independently'. [3] Keeping with this notion, students in the present day classroom are encouraged to actively demonstrate their learning and understanding by summarizing their ideas and developing critical questions. [4] Translating this theory into practice, while keeping up with the changes in social behavior and technology requires both a space which can allow for such events to occur supported by a curriculum that encourages independence and collaboration.

In the construction industry, BIM is emerging as at the language for communication between the parties involved. At DEWSC, BIM is viewed as a cyclical process which aids in the integration, collaboration and management for the various components of the architecture, engineering, and construction community, informing each phase of a facility's lifecycle. To fulfill the need of the industry related to the desired skills of new hires, BIM has been embraced as a core component of undergraduate Construction Management programs across the U.S. and beyond. Two approaches have been taken for the curriculum - integrating BIM in one or two courses in a Construction Management curriculum or implementing BIM through several courses, such as in a 5 year Architectural Engineering curriculum. [5] These efforts of incorporating BIM into an academic curriculum are supported by industry professionals who dedicate their time and share data or sample projects for training purposes.

3. DESIGN OF VC^{2}L

Collaboration can be achieved through various exercises such as team building, group assignments, time-based problem solving and collective conversations. Along with software programs and technology, which are potent tools to stimulate this notion, co-location, proximity and visual access are the prime considerations for designing a space for collaboration. [1] The specific features of the space have to be designed considering the behavioral patterns of different groups, ways of knowledge sharing between individuals, task structure, cognitive complexity of the group task, on-going communication needs and practice methods. Construction management is a profession that requires an individual to have the ability to clearly communicate and develop skills to identify how people behave, react and think. Most of the time as a Project Manager is spent in meetings in conference rooms and site offices. Construction trailers are often modeled on an open office concept, with maybe one closed conference room for owner-architect meetings and co-ordination sessions. With these ideas in mind, the practicality of the profession was brought closer to academia with VC^{2}L.

This experimental space has the flexibility of transforming between 'pod-based', lecture hall and conference room settings by simple re-arrangement of furniture. The primary layout for collaboration is the 'pod' - which is a workstation seating six with a single computer, two LCD monitors and power outlets for connecting personal laptops. The lab is equipped with sophisticated computing systems to handle modeling and simulation exercises, LCD screens, whiteboards and projectors with the capability of transforming any flat surface to an interactive work surface. The main controls are handled by an instructor station which functions as the centralized server, creating a local network for the lab. Sophisticated switching devices allow the function of replicating the screen content of any of the pods onto the front projection screen or any combination of screens. This gives the instructor the ability to focus the attention of the class when and where required.

Given the size of the space available, and the course intake, the maximum capacity of the lab was set at 30. After several iterations, it was concluded that the space would be apt for 5 pods with 6 students each. This also recognized previous research studies which indicate that smaller group sizes of less than 4 lack the diversity and varied expertise for collective decision making, and larger groups more than 6 do not ensure participation of all group members. [2]

Plan1.0 - Pod

The primary idea behind a 'pod' configuration is to stimulate conversation by bringing individuals in close proximity to each other while also providing 'high visual and aural access to group members'. [6] Usually the pods have a high screen around them to provide privacy to each group. It was determined for the VC^{2}L layout that the pods be open for greater visual access and interaction. Each of the workstations has two 30"x72" tables joined together at the longer edge to make up a large work surface which is 60" x 72". Students use this surface to open up large paper drawings, work on their own laptops, and have sufficient room for writing instruments and other documents. Two rollable chairs are placed on each of the three sides. The fourth side butts against a wall or a perimeter vertical surface on which two LCD monitors are mounted. Every pod has a resident computer, with the CPU and other peripherals located under the tables, keeping the work surface free from clutter. The pod is also equipped with a power strip and switcher for connecting laptops and displaying content on the screens. The central aisle is left open for ease of movement. There is also the provision of joining together some extra tables (24" x 60") to form a long drawing table on which paper drawings can be rolled out. On certain occasions, students huddle around the table for discussions pertaining to construction drawings. These tables are also used for expanding the size of the workstation to accommodate larger groups.

Plan 2.0 - Lecture Hall
The flexibility of having light rollable furniture makes it easier to alter the layout of the space to accommodate various other situations. If the instructor requires students to simulate a seminar, the students can roll their chairs and tables to arrange them in a more traditional classroom setting.

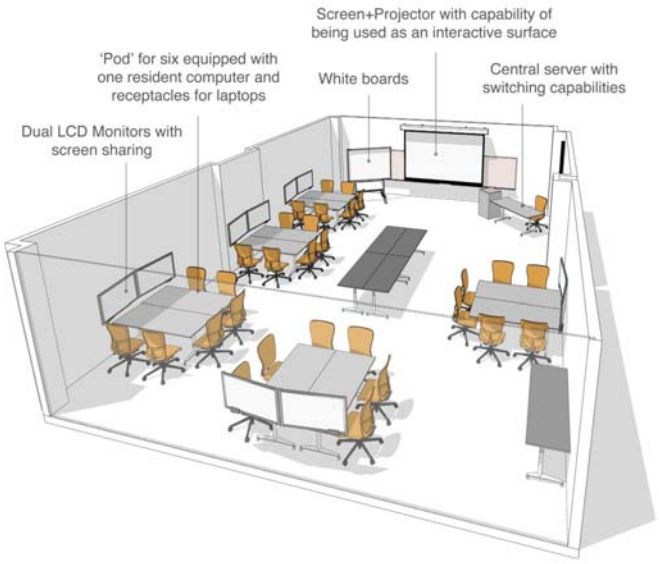


Figure 1 – Perspective view of VC²L

Figure 2 – Group work in session

Plan 3.0 – Conference Room
Added flexibility of the room allows for the configuration of a conference room or large working station option. This makes it possible to have a larger group discuss plans, hold meetings, or a variety of other uses that typically require a separate room. With the technology hosting a web based meeting is possible, or participating in a skype type call. This option makes it ideal to collaborate.

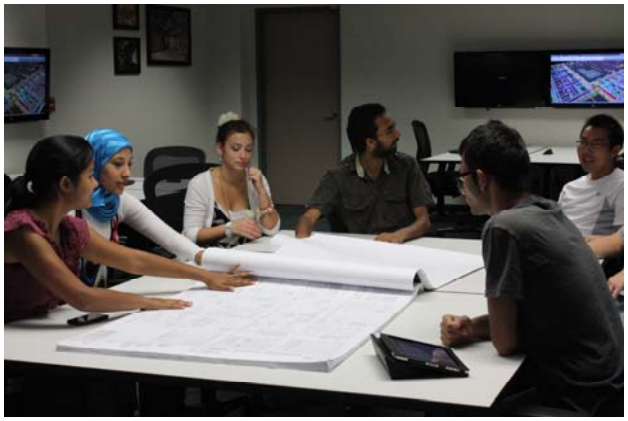


Figure 3 – Conference room setting

4. CHALLENGES AND THEIR MITIGATION

The transition to a world of collaboration from the traditional ways of learning can be challenging, and is often received with mixed feelings. Every student has a different personality and ability to adapt in a group situation. In the first semester of the VC²L, a variety of different challenging situations arose from reasons such as varying skill levels, different knowledge levels, allegiance to the team, personality traits, enthusiasm towards the course content and the difficulty level of the task. Oftentimes, the group leader would be the one with the strongest technical and practical skills, which may be beneficial for other members to organize themselves and contribute in other areas. At times, the same situation would lead the rest of the team to procrastinate as the leader single-handedly completed the entire project. There were other cases, when members of the team were absent, non-responsive, lacked interest and sometimes overconfident in their abilities. These are all real world challenges which every individual often faces in the practical world of construction. Collaborative projects test technical and academic knowhow as well as leadership skills, personality traits and performance in group situations. To mitigate these situations, the VC²L courses followed three simple rules: i) change the team and assign a new leader for every new assignment; ii) each team member would conduct performance evaluation for the rest of the group; and iii) have a member of the team, other than the leader, present the project and their results to the rest of the class. These steps resulted in a sense of competition amongst the students, challenging them to perform better.

Assessment of individuals in a group environment is always a challenge faced by the instructors. It is easier for the escapist to blend in with a group and avoid any responsibility. For the VC²L lab, a 1:10 teacher-student ratio was utilized, such that for a group of 30 students, there was 1 instructor and 2 Teaching Assistants. In other words, each instructor or TA was responsible for

overlooking 2 or less pods. These closely-knit groups made interaction and instruction more personal and tailored to account for the skill level of each individual. Each student was graded for an assignment based on four factors: i) the actual deliverable, ii) presentation score, iii) performance evaluation, and iv) observations of the instructor.

5. CURRICULUM

BIM was introduced as a lab component accompanying the senior level Project Management lecture at DEWSC in 2008. Each year the curriculum for the lab was modified to accommodate the changing software and industry improvements. Instructors for the lab are BIM experts invited from the industry. This arrangement ensures an advocate-based, best practices delivery while simultaneously exposing students to current industry trends and job opportunities. The 3-credit course meets for a 1-hour lecture twice a week, followed by a 2-hour lab meeting once a week. The lecture develops an understanding of the project management process, including defining, planning, executing, controlling and closing a project and the use of BIM to support that process. The objective of the lab is to create an overall knowledge of the software that is part of a typical BIM toolbox (Revit, Navisworks, SketchUp, Synchro) and use those tools to tackle a management issue. For many of the students, it is the first experience using any of the programs, and therefore the lab schedule was divided into individual and group project weeks. During individual weeks, the students received a tutorial outlining the basic functionalities of the program, followed by group sessions wherein they were asked to solve a challenging problem using the skills developed during the lab.

One example of a lab is Quantity Take-off and Estimates. A group of 5 was assigned the role of a trade – Architecture, Structural, Plumbing, HVAC and Electrical subcontractor. The groups used Revit Architecture to derive quantities of materials and equipment. While two members of the group worked on extracting quantities, two worked on obtaining cost data from RS Means and one member was assigned the role of compiling the data to develop an estimate for their trade. The same scenario was repeated for the second lab period and the values were compared. This exercise enabled discussions, questions and enthusiasm because each group wanted to win the bid, all within a timeframe of two hours.

6. CONCLUSION

As an overall improvement the students were more enthusiastic towards their projects, however a survey conducted showed that a majority, 60%, still felt that a traditional computer lab setting was more beneficial for them. The reason for this response, amongst many, could be heavily influenced by the fact that since this is a graduating class, in general the students are overloaded with classes and job search. As we changed the curriculum to include group assignments, the course load did increase a bit. 32% of the class felt the pace was too fast for them. While we saw a slight increase of 4% in the number of students with previous knowledge of BIM, the acceptance of the fact that it was integral part of the construction industry and a first job requirement went up by 15%. This trend reinforced the need to have BIM knowledge spread throughout the 4 years including a stand alone class which teaches the software program as shown in figure 4, a BIM curriculum should be developed throughout the entire program. This integration helps students understand the tools and realize the full benefits of BIM.

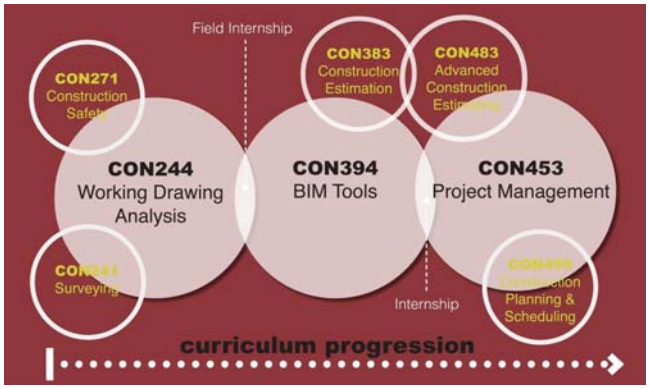


Figure 4 – BIM integration curriculum plan

7. REFERENCES

[1] "Making Room for Collaboration." *Herman Miller*. Web. 2 Jan 2012. <http://www.hermanmiller.com/MarketFacingTech/hmc/research_summaries/pdfs/wp_Collaboration.pdf>.

[2] Gokhale, Anuradha A. "Collaborative Learning Enhances Critical Thinking." *Journal of Technology Education*. 7.1 (1995): n. page. Web. 2 Jan. 2012. <http://scholar.lib.vt.edu/ejournals/JTE/v7n1/gokhale.jte-v7n1.html>.

[3] "Social development theory (Vygotsky)." *Learningtheories.com*. Learning Theories, 2008. Web. 2 Jan 2012. <http://www.learning-theories.com/vygotskys-social-learning-theory.html>.

[4] Liu, Chu Chih, and I. J. Chen. "Evolution of Constructivism." *Contemporary Issues in Education Research* 3.4 (2010): 63-6. ABI/INFORM Complete. Web. 2 Jan. 2012.

[5] Ciszczon, Heather, and Allan Chasey. "Curriculum Development for Building Information Modeling." *Proceedings EcoBuild*, Washington, D.C., 6-10 Dec, 2011

[6] Heerwagen, Judith H., , et al. "Collaborative Knowledge Work Environments." *Building Research and Information*. 32.6 (2004): 510-528. Web. 2 Jan. 2012. <http://www.interruptions.net/literature/Heerwagen-BRI04.pdf>.

Transitioning to a Global Classroom

Dr. Allan CHASEY, Lisa G. HOGLE
Del E Webb School of Construction, Arizona State University
Tempe, AZ 85281, USA

ABSTRACT

The world is no longer comprised of isolated pockets of communities, but now has become an interdependent web that spans the entire globe. Training must be able to reach across the globe to provide a worldwide training solution. As companies span the globe, education for their employees must do the same.

This is a case study combining design, construction and facility management concepts for advanced technology facilities using web-based technologies. This method provides student's access to receive the necessary training while being available at the work site for their companies. It incorporates Project Based Learning (PBL) to provide students a safe environment to experiment, create and sometimes fail, while allowing the opportunity to think, discover, and most importantly, learn.

Technology allows combining online and face to face education for optimal learning. Current research indicates students retain 75% of the information presented when they are engaged and learning by doing. Web-based modules allow students to learn at their own pace while interacting with peers, encouraging collaboration and sharing of lessons learned. Additionally a face to face learning environment enables review and clarification of topic information as well as practical, real-life application of scenarios that are common to facility and/or project managers.

Keywords: Web based learning, project based learning, facility management, project management, technology for education, training, global training.

1. INTRODUCTION

We as educators have a great opportunity to utilize technology to transform our educational practices to accommodate the global marketplace. Not only can education and training be delivered around the world, but unique knowledge and expertise can be brought into a classroom from virtually anywhere. This becomes more important as companies span the globe. Education for employees must change, allowing them to continue their employment regardless of where the job takes them.

Research indicates students retain 75% of the information presented when they are engaged and learn by doing [10]. Technology allows us to combine online and face to face education methods establishing an optimal learning opportunity for working individuals. The Project Managers Development Program (PMDP) and the Facility Managers Development Programs (FMDP) at Arizona State University show a high degree of success in utilizing this method to train individuals around the world. The PMDP covers the major concepts in the design, construction and renovation of advanced technology facilities. These facilities have additional complexities due to the various cleanliness and certification requirements, which require a higher skill set for construction. The FMDP is focused at Facility Managers of an advanced technology facility. While most Facility Managers have extensive experience operating various general purpose facilities, the complexities surrounding the operations and maintenance of a high technology facility presents unique challenges, an order of magnitude different from general purpose buildings.

The PMDP and FMDP programs are a blended learning experience that combines both on-line and residency training for the students. The activities that are part of the overall program are specifically designed to increase the retention of the information presented. On-line lectures provide the basic course knowledge. Virtual conferences allow the students to discuss the material as well as provide information about a specific topic. Finally the face to face residency is designed to have the students apply course material in a real-life scenario.

One key to the success of these programs is that students are put in a cohort providing an opportunity for students to interact and share knowledge and experiences with fellow classmates from different locations and backgrounds. To increase success students are put in groups and encouraged to collaborate and share experiences and lessons learned throughout the program.

2. USE OF TECHNOLOGY

Technology has changed how we communicate and now is able to change how we educate employees. "Technology is not another subject, another class. It represents a pervasive set of changing tools for learning and teaching" [1]. As the presence of technology increases, online education can be used to continue the tradition of passing quality information on to more individuals.

Online education enables students to access course information as many times and as often as necessary, it gives them the ability to pause and further research or clarify any point of the material [11]. Connectivity provides the potential to connect everyone to everything all of the time [1]. Online environments provide the opportunity to self pace the learning that would not otherwise be available in traditional classroom instruction methods.

Students are able to access course material even if they are not on campus because of distance or other responsibilities. This availability allows a large population the ability to receive an education or participate in an educational endeavor. [4] Technology provides the opportunity to deploy the training information across the globe, making it available 24 hours a day, seven days a week, only requiring an internet connection for access.

Instructors can engage students by providing increased opportunities to extend a student's learning experience to a global level. Instructors do so by engaging students from other locations, or diverse cohorts, in practical learning environments [5]. This provides an opportunity for students to interact and share knowledge and lessons learned with fellow classmates from all over the world. Online learning allows for asynchronous learning which is being able to see the material and information anywhere at any time in a self-paced format accommodating different time zones and schedules.

The internet has opened up many new options for communicating and collaborating. In the past it was very costly to connect the students with industry experts. Technology now enables students to have access to a wide variety of industry experts affordably. Online education continues the legacy of providing quality access to more people, providing multiple opportunities to expand the influence of education beyond a physical footprint, moving towards an "anytime anywhere (educational) model." [6, 3] Online learning is an optimal way to disperse training and information to a global audience affordably. However, there must be some additional components to provide a best case scenario.

Forming a cohort, a group of students moving through common courses together, utilizes relationships to increase learning. Students who have gone through a program as part of a cohort receive certain benefits from instruction including increases in integrating the course content [5]. Being able to share and learn from others reinforces the material and provides a support and encouragement system. "Students benefit from peer teaching – explanation, comments, and instruction from their course-mates" [6].

Engaging the students in hands on teaching methods increases the amount of information they retain. Studies have shown that students only retain 5% of the information conveyed during a lecture, increasing to 20% when audiovisual presentations are included with the lecture. Information retention increases to 30% when a skill is demonstrated and reaches 75% when participants are engaged and learn by doing [9, 10].

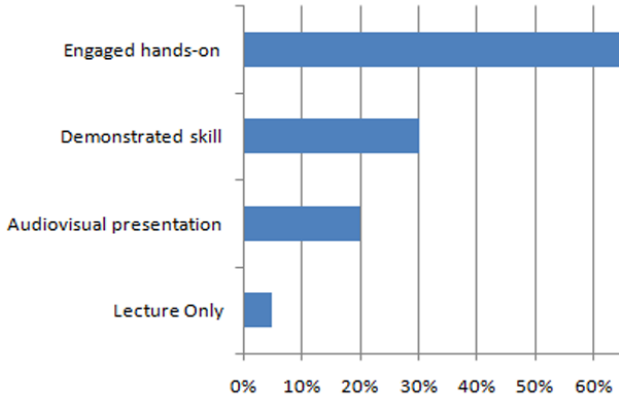


Figure 1 – Retention percentages

3. PROJECT BASED LEARNING

Project Based Learning (PBL) is an educational model where students learn through problem solving. This process includes connecting previous knowledge with new knowledge to bolster learning and retention. PBL provides an interactive hands-on approach that engages students in learning by doing, applying the information to a relevant problem. PBL encourages students to take all the individual pieces they have learned and blend them into a viable solution. It stimulates teamwork, critical thinking, problem solving skills, and reinforces the information they have learned [10].

Learning with projects provides the platform to connect all the pieces. Learning typically consists of several units or topics being presented. Concluding those units with a project drives students to link all of the topics together for a fuller more in depth understanding of the material that has been presented.

PBL also provides the opportunity to customize the learning incorporating company or industry specific challenges to work through.

During the online portion students gather all the information regarding an advanced technology facility providing them the individual pieces to the overall picture. During the face to face the students are given case studies and projects that help them put the pieces of the puzzle together and not only have an understanding of the information, but also have practiced how to implement it and how each aspect inter-relates to the other. This is not always a part of the education process.

4. PROJECT AND FACILITY MANAGERS PROGRAMS

One solution to this challenge is a program developed at Arizona State University in the Del E. Webb School of Construction. This program entitled Project Managers Development Program (PMDP) allows students to learn and receive necessary training while still fulfilling their current work assignments. The PMDP not only presents the material to the student, but provides opportunities for them to implement the learning. Students engage with others in the program to discuss topics, share any personal experience or to clarify. Students who have been involved with the program represent twelve countries.

Training and education is vital to keeping a company productive. Improving job skills through access to specialized technical information while retaining the services of employees during intensive training is a benefit for companies. The PMDP was developed because of a need to increase technical skills to the existing workforce. The recognized need was a lack of knowledge regarding the support systems and the methodologies required to design and construct complex manufacturing facilities. These complexities are apparent when the manufacturing environment must be controlled to ensure product viability.

In the case of the PMDP the complexities of the semiconductor manufacturing processes are affected by the facility systems because of the controlled environment necessary for the wafers. Controlled environments relate not only to the ambient air (HVAC), but the facility itself (structure and vibration). The process complexity due to both hazardous materials and number of different gasses and chemicals presents challenges for even the most seasoned professionals. With the need established and a graduate level course available several program level changes were instituted to improve the learning outcomes. The revamped program is available through an internet portal providing the student access 24 hours a day, 7 days a week.

5. EXAMPLES

Each week students watch a video lecture presentation from an industry recognized expert, typically about two hours in length. These lectures are web based and available anytime, anywhere an internet connection is available, enabling students to receive the training at their convenience. The students complete a study guide based on the presentation materials and meet within their group to discuss the study guide and materials in the presentation. This group remains the same throughout the online portion, approximately 9 months, forming a strong sense of community within the group. These group discussions are a time to also discuss applications or experiences with the topic that could benefit the group. The students submit the completed study guides to the program coordinator to be reviewed by the faculty, and address any gaps or misunderstandings in the responses of the students. Students are also able to send emails with any questions for further information or clarification on a topic.

Within each of the groups, a weekly team leader is assigned to coordinate the team meeting and any logistics that are required, such as phone bridges or meeting rooms. Team leaders are also responsible for tracking attendance and sending the information along with any questions to the program coordinator. Students are given the opportunity to be the team leader to practice and enhance their leadership skills.

At the end of each section, approximately 5 topics, a conference call with the entire class is set up. Typically each student is assigned outside articles to read and present a brief overview of the information to the rest of the cohort, including how the information that has been reviewed in the section can be applied to the topics or to their position. The articles could be focused on more in-depth information on the topics or additional information that enhances the topics in the session. Reviewing the articles also allows the students to think beyond their current position and current trends and information and summarize that material to others.

The online portion of the program culminates in a packed four (4)-day face to face at Arizona State University. The face to face enables students to receive additional information and clarification on the information presented during the on-line lessons and connect them with industry experts. They continue

to build their network as they are mixed with a different group to work on a "real world" project that requires them to develop a proposal, based on current industry issues, to present to a "board of directors". The project will typically address several issues including budgeting and technical systems giving them the experience to see how to implement the knowledge they have gained.

- Currently we have a student enrolled in the PMDP who is completing the training and enhancing his skill set. Without the format of the PMDP this training would not be possible because he needs to remain at his location and take care of his wife who is pregnant.
- The past year many travel budgets have been cut, one company in particular had to increase the skill set and knowledge of their employees, but all travel had been restricted. Because of the format of the PMDP the employees were still able to get the required training even with the restricted travel.
- The cost of training is another inhibitor for companies to train their employees. When compared to the cost of a year of tuition at a public university the PMDP will provide a cost savings.

6. CONCLUSION

In conclusion, the PMDP and FMDP training programs are a benefit to employees as they are able to gain needed training while maintaining their current job responsibilities. Because they are still in their positions they are able to immediately implement the knowledge they are obtaining. Companies are able to train their employees, enhancing their workforce, without significant downtime. Training is becoming an option for individuals who could not have done it previously. Companies are able to train their employees affordably and effectively.

7. REFERENCES

[1] Hutinger,P. n.d. Technology and Education: What will the Future Bring? Retrieved July 13, 2009, from http://www.wiu.edu/thecenter/articles/teched.html.

[2] Popov, (2009) Teachers' and students' experiences of simultaneous teaching in an international distance and on-campus master's programe in engineering. International Review of Research in Open and Distance Learning.10(3). Retrieved September 1, 2009 from http://www.irrodl.org/index.php/irrodl/article/view/669/1272.

[3] Katz, R. (2008) The Gathering Cloud: Is This the End of the Middle? The Tower and The Cloud, EDUCAUSE. Boulder, CO, pg. 2 – 42.

[4] Nelson, (2009) The Conversion of Traditional Course Material to the Online Learning Format. Unpublished manuscript, Capella University.

[5] Davis, G., O'Brien, L. & McLean, P. (2008) Growing in Esteem: Positioning the University of Melbourne in the Global Knowledge Economy. The Tower and The Cloud, EDUCAUSE. Boulder, CO, 64-80.

[6] Noble,D.F. (1998, January 5) Digital Diploma Mills, Part 1: The Automation of Higher Education. First Monday, Vol.3, Number 1. Retrieved September 1, 2009 from http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/569/490.

[7] Davis, B.G. (2001) Tools for Teaching. San Francisco, CA: Jossey-Bass

[8] Duderstadt, J., Atkins, D., & Houweling, D. (2002). Higher Education in the Digital Age. Westport, CT: Praeger Publishing."Educationonline" (2009) Education: Online versus Traditional. Retrieved 18 September 2009, from http://www.educationonline.net/home/education-online-versus-traditional/.

[9] Johnstone, A. H., & Su, W. Y. (1994). Lectures: A learning experience? Education in Chemistry, 31. 75 – 79.

[10] Flint, W.J. (2007) Problem Based Learning: Welcome to the"Real World". Word Unlimited

[11] Friedman, T.L. (2006) The world is flat:abrief history of the twenty-first century, New York,NY; Farrar, Straus, and Giroux

Teaching Introductory Software Engineering Course Using Software Process Simulation

Jacob SUKHODOLSKY
Department of Mathematics and Computer Science
Saint Louis University
St. Louis, Missouri 63103, USA
sukhodpj@slu.edu

ABSTRACT

Teaching an introductory Software Engineering course to undergraduate students is a major challenge mainly because of their lack of software development experience at an industrial scale. The traditional approaches, such as explanation of the Software Engineering concepts accompanied by discussing cases and a semester-long project, do not sufficiently prepare students for real-life projects. A recently introduced approach to teaching Software Engineering by using software project simulation looks very promising. The problem now is to find the most effective way of utilizing it. The article suggests an original method of teaching Software Engineering by simulating software process workflows in the classroom.

Keywords: Software Engineering, Project, Plan, Workflow, Simulation, Tools.

1. INTRODUCTION

Undergraduate Software Engineering education in most cases fails to provide students with the knowledge and skills they need to successfully start their software development careers. That happens for a number of reasons, one of which is a very limited amount of information students get from a one-semester Software Engineering course, even if the course includes a team project. A reason is that no real project can be started and finished within a one-semester time span. Another serious problem is low students' motivation. It is practically impossible to make most of the students work hard on their project tasks, unless they themselves realize that the skills they are acquiring in the process are vital for their future careers. The instructor has practically only one way to enforce appropriate course work ethics - grading. But then, not every student is motivated even by a possibility of getting a 'C-' for the course. A lowly motivated student not only harms his/her future. Often such a student affects other students' attitude. The instructor then faces a hard problem – countering the student's negative influence on the peers.

The most popular method of teaching Software Engineering today is by providing students with knowledge of various theoretical concepts they are supposed to be familiar with when working on a real software project as part of a software development team. Typically, the complexity of a real project is much higher and its size is considerably larger than a one-semester project the students participate in at school. Besides that, in industry they are required to plan their job and report their results on a regular basis. An instructor can teach them these skills but it is almost impossible to make them update the project plan on a regular basis. The most difficult problem for the instructor is how to teach students the specifics of software process workflows, such as Requirements elicitation, Analysis, Design, Testing, Installation, and Maintenance. These workflows, so important in a real software process, are practically unknown to the students taking a Software Engineering course, since the majority of them have dealt in the previously taken Computer Science courses with problems that required no serious analysis or design.

One of the newest approaches to teaching Software Engineering is by presenting a software process as a game. The Problems and Programmers [2] method uses the card game format to simulate a software process. In that game, each participant competes with the rest of the group by performing various project management tasks pertinent to the situation described on the cards they randomly pull from the stack. The game is designed to provide students with the opportunity to play the role of a project leader and compete in this role with the rest of the class. While such a game is important from the perspective of developing students' project management skills, it eliminates discussion, where each student can voice his/her own opinion. Furthermore, there is a possibility that the majority of the students will take a wrong route, while only a few "wise" ones will succeed. The teaching value of the method in the latter case is doubtful.

The article demonstrates a way of simulating a real software process, with students managing a project collectively and taking decisions as a group. In the process, they inevitably encounter some unforeseen consequences of their decisions. Such unforeseen outcomes require people effort and time to deal with. That leads to additional costs associated with the decision. Estimating those costs and then comparing them to the actual costs involved helps the students better understand software project intricacies.

2. HOW IT IS SUPPOSED TO LOOK

Students work in class on a software project, which is different from their traditional semester team project. This in-class project's Requirements, Analysis, and Design workflows are at the students' center of attention. Every class, the instructor informs the students about some kind of circumstances related to the project, which is typical for real-life software development. Those may include, for instance, request from the customer to change a certain requirement or an entirely new requirement. By the time of the following class, each of the students will have decided how to proceed with the project. During a short class discussion, each of them presents his/her solution. The group then decides picks the best, and appropriate corrections are made to the plan. When the situation has been handled, the instructor and the students collectively determine the effort that was required. It is important to emphasize that handling the circumstances takes place in parallel with the scheduled project development activities. If the problem could have been avoided, the students actually pay a penalty for not having done that. At the end of the semester, the instructor evaluates the project completion status. If the project is complete and the software product satisfies all the requirements, the instructor helps the students to understand how it happened and where the critical decisions resulted in the project success were taken. If the project is incomplete or complete but some of the requirements are not satisfied, the students are explained how that happened and where the most critical mistakes were made. In any case, that should be a very instructive experience, especially if the students compare this experience with that of their semester team project.

3. HOW TO MAKE IT WORK

Project To make the described method work, the instructor should first of all choose a right in-class project. The project should be technically simpler than the team project the students work on their own. The goal of the instructor is to make students concentrate on the software development process rather than on coding and debugging. At the same time, the project should be interesting to the students.

Evaluating Cost of Decision The cost of a decision is comprised of the cost of implementing that decision and the cost of dealing with the consequences of that decision. For example, during each of the project workflows, students create software artifacts. Some of them may be defective. Each defect, depending on its nature and the project workflow during which it was introduced, entails costs that are associated with the effort required to fix it. The cost to fix a defect changes for different projects and different

organizations. Yet it is possible to evaluate how the ratio of fix costs changes during the product's lifecycle. According to Pressman [4], the expected cost to fix defects increases during the product's lifecycle. As a general rule, the more workflows separate the current workflow from the one where the defect was introduced, the larger cost is associated with fixing it. The relative cost of correcting an error, assuming that 1 is the cost of fixing a software requirement error during the Requirements elicitation activity [5], it costs:

- 3 – 6 times more during Design
- 10 times more during Coding
- 15 - 40 times more during Development Testing
- 30 – 70 times more during System Testing
- 40 – 1000 times more during Field Operation

Let us suppose the students encounter a defect in a design artifact while coding their project. They determine that the problem originate at the Requirements workflow. Fixing the defect will involve fixing the requirements as well as the design artifacts based on them. In addition to it, part of the code already written will need changes. Based on the preceding, the costs associated with the described activities can be calculated as following:

costs = 1 + number_between_3_and_6 + 10

For the teaching purposes, the number_between_3_and_6, which reflects the costs of fixing the design artifacts dependent of the defective requirement, can be generated randomly.
The costs of fixing the defects are then added to the cost of implementing the decision responsible for the unsuccessful outcome that has been taken care of.

Class Discussion This is probably the most important component of the proposed teaching method. The challenge is to spend at maximum 50% of a class time presenting to the students the project status quo, discussing with them the possible way of handling the situation, and updating the project plan. (The other 50% of time will be spent on the new material).

Tools The students will need some kind of software tools to do analysis, design, implementation, testing, and debugging of the project. The instructor cannot afford teaching them how to use the tools. The students will have to do this on their own. The instructor should only recommend them some free tools and make sure that every student uses the same kind of tools as the rest of the class. Following is a short list of the tools the students may use.

- Analysis and Design Tools (such as SQL Power Architect and the free community edition of Visual Paradigm for UML)
- Construction Tools (such as Linux-based code compiling and linking tools)
- Configuration Management Tools (such as CVS, RCS, or SCCS for Linux).

4. EXAMPLE

Presented below is an example of using the described method.

Eliciting Requirements

Let us consider, for example, a case where a customer hires software professionals to develop software he could use to plan and track his activities and expenses. The teacher plays the role of such a customer. As a typical customer, he believes that he knows what the software should be able to do to satisfy his own requirements. The students come up with a list of questions they are going to ask to figure out what these requirements are. The questions can be following:

a. Do all your activities incur expenses?
b. Will you want to specify the period of time the activities took place or the software should list all the activities that have ever taken place?
c. Should the information about your activities be accessible over the Web?
d. Should access to the information be password-protected?

When the students feel they have got answers to all their questions, the customer (teacher) and the developers (students) proceed according to one of the three possible scenarios.

1. Developers did not extract all the requirements.

   For instance, the customer may want to be able to print the retrieved information. The teacher shows to the students that the list of requirements they came up with is incomplete and asks them to think about other requirements they have possibly omitted. The project plan has to be updated to include the process of collecting the missing requirements.

2. Developers extracted all the requirements.

   Although it is a rear case, even for experienced software developers, there is a possibility that students figure out all the teacher's requirements. The teacher should then encourage them to think about other possible requirements and emphasize the fact that if even in such a simple project the customer may want so many things, then how many more requirements should be satisfied in case of a more complicated project.

3. Students came up with requirements the customer did not even think about.

   This situation is pertinent to both case 1 and case 2. The students, for instance, may ask if the customer wants to use a smart phone application that could retrieve the information about his activities and expenses. Such an original idea may get the customer excited, but it is equally possible that the customer will not be interested in such an option or simply irritated that his/her time is wasted. The students should be told that the ultimate decision is left up to the customer (the one who pays money).

After the discussion, the updated list of questions is saved, but the old one is not deleted. This is the first students' encounter with configuration control, the main principle of which is to keep all the results of your project activities. (In their semester project, which starts at the same time as the in-class one, the students are required to apply the skills they have learned when working on eliciting requirements for the latter project).

At the end of the Requirements workflow, the instructor and students together update the project plan.

Analysis

The next error-prone software process workflow is Analysis. Its purpose is to produce a conceptual model of the information that exists in the domain being analyzed. The result of object-oriented analysis is a formal or semi-formal description of what the system is functionally required to do. If the object-oriented development method is used, the analysis starts with determining the classes of objects that can be used to construct the software. It is often impossible to come up with all the necessary classes during the initial analysis activity. At the same time, the developer should be careful not to lose any obvious classes. For instance, two obvious choices of classes in the software tracking activities and expenses are Activity and Expense. Missing any of them results in an incomplete model of the information and, ultimately, in an inadequate software product. It is important to emphasize to the students that the input to Analysis is the information obtained from the customer during the Requirements workflow. Some crucial information may be hidden in the customers' heads until late in the software development process. Imbedding this knowledge into the system will require changes to the model developed during Analysis. For example, the customer may think of some important report the software should be able to generate. Failure to extract this requirement from the customer will lead to omitting an important class during the analysis activity. In the worst-case scenario, it will be overlooked during Design and Implementation. Only when the customers test the finished product, they will notice the absence of an important software feature. The

developers will then have to change all the software artifacts, starting with the requirements and finishing with the code. The changes may lead to re-writing a lot of code, which, in turn will result in new errors.

Changing the requirements, as well as correcting the analysis artifacts, will require updating the project plan. Each missed requirement discovered during Analysis entails a 2-points penalty.

Design

This workflow is error prone because of many small details to be taken care of, not mentioning such important decisions as the one about the implementation language, database, and the computer hardware configuration. The step from what to do to how to do that involves choosing data structures and the class methods. To verify correctness of the design decisions, one should go back to the analysis artifacts that include use-case diagram, and make sure that all the use cases are supported through the chosen class methods. The students should realize that the design is correct only if the analysis and the requirements it is based on are correct. At this point, the instructor may want to change or add a few use cases, as well as add a new requirement. This will help the instructor to demonstrate to the students the amount of work to be done in order to synchronize the design with updated requirements and/or analysis artifacts. It is needless to say that some updating of the project plan will be required.

Each error in an Analysis artifacts discovered during Design entails a 3-points penalty. Each missed requirement discovered during Design entails a 3 to 6 points penalty.

Implementation

This involves implementing design using the language of choice. Coding is an error-prone activity by itself. It is even more so if the design contains errors. Fixing them will require updating the plan. At this point, it is useful to compare the latest version of the plan to the original one and estimate together with the students how much more additional time and resources have been required to fix the errors made earlier in the software development cycle.

Each error in the Design artifacts discovered during Implementation testing entails a 4 -points penalty. Each error in the Analysis artifacts discovered during Implementation entails a 20-points penalty. Each missed requirement discovered during Implementation entails a 40-points penalty.

System Testing

At this point, the whole system is tested to check whether it satisfies the Requirements. If a missed requirement is found, fixing it will cost the team 70 points. An error in an Analysis artifact will cost 40 points. A Design error will cost 20 points. An error in the code will cost 10 points to fix. Since it is impossible to know how the software will behave during field operation, System Testing is the last development activity of the project.

5. CONCLUSIONS

The proposed method of teaching has a number of merits. On a simple project, students learn the software process realities. They learn that any error, no matter how small it is, may cost a lot to fix if it is found late in the software development process. They learn to be much more serious about Requirements, Analysis, and Design than they were before, when Implementation got the first priority. The students get used to reviewing their artifacts. The teacher may want to rate the students' work on the project according to the number of points they collect over the course of the project. The classroom project may positively affect the students' semester project if they try to repeat there the practices they learn in the classroom project.

The proposed method's effectiveness could be increased by using a computer program that would allow an instructor to choose from several kinds of problems, their origin, severity, and the penalty assigned for each of them

when the instructor enters the workflow during which the problem was encountered. To generate its output, the program can use a "bank" of problems the contents of which the instructor can update as necessary.

Implementing the described method of teaching may not be as successful as it can be if the teacher completely sacrifices theoretical material in favor of discussing the classroom project. Instead, the classroom project should be an illustration of how to implement in real life the theoretical knowledge acquired in class.

The proposed method will be tested in spring 2012.

6. REFERENCES

[1] D. Pfahl, O. Laitenberger, G. Ruhe, J. Dorsch, T. Krivobokova, "Evaluating the learning effectiveness of using simulations in software project management education: results from a twice replicated experiment", **Information and Software Technology**, vol. 46, 2004, pp. 127-147.

[2] A. Baker, E. O. Navarro, A. van der Hoek, "An Experimental Card Game for Teaching Software Engineering Processes", **Journal of Systems and Software**, vol. 75, 2005, pp. 3-16.

[3] D. Carrington, A. Baker, A. van der Hoek, "It's All in the Game: Teaching Software Process Concepts", **Proceedings 35th Annual Conference In Frontiers in Education FIE '05**, 2005, pp. F4G-13-F4G-18.

[4] R. S. Pressman, **Software Engineering, A Practitioner's Approach**, 3rd Edition, McGraw Hill, New York, 1992, p.559.

[5] B. Boehm, **Software Engineering Economics**, Prentice Hall, 1981.

An Improved Image Compression Technique Based on Diagonal Edge Estimation

Huda Al-Ghaib, Reza Adhami, Hai Dinh
Electrical and Computer Engineering
The University of Alabama in Huntsville
Huntsville, AL, USA
hag0002@eng.uah.edu, adhamir@uah.edu, hqd0001@uah.edu

Abstract

Image compression is the process of compacting data while maintaining the necessary information. Image compression minimizes the memory requirement for storage of data especially when dealing with multi-dimension images. This research focuses on image compression using different wavelet transforms and diagonal edge estimation. In this research, we have shown that a better compression ratio can be achieved when eliminating the diagonal edges in the wavelet transform during the compression process and estimating them during the decompression process. Our results are compared with the JPEG2000 and JPEG.

Keywords: Image Compression, Diagonal retrieval, lossy image compression, Sobel Operator, and Wavelet transform.

Introduction

Image compression is the process of reducing data in digital images and still maintaining the visual information. Image compression is either lossless or lossy. Lossless image compression is also known as reversible image compression or noiseless coding. Lossless image compression preserves all information contained within an image, i.e., the numerical differences between the pixels of the original image and the compressed one is equal to zero. The highest compression ratio that is achieved using lossless compression is 2:1 [1]. However, this can be higher since it is image depended. Lossless compression is used with applications that do not afford the circumstances of sacrificing some information to obtain a higher compression ratio. Examples of such applications are: medical imaging, telemetry, forensic, and geophysics applications [2-3]. Lossy compression sacrifices some information to obtain higher compression ratio. It is acceptable for applications that do not require an error free reproduction of the original image such as fast transmission of not so critical images through internet. Lossy compression provides "visually lossless" images, meaning that the actual data has been lost but not sensitive to visual appearance [4].

Lossy image compression consists of the following three steps: decorrelation, irrelevancy reduction, and entropy coding. Eliminating irrelevancy reduction step results in lossless image compression [3], [5]. Usually, images contain redundant information that is either spatial and/ or spectral. The main source of the spatial redundancy is the correlation between neighboring pixels. Spectral redundancy is from the correlation between different spectral bands and/or color planes. Decorrelation is for removing the spatial and spectral redundancies. There are different techniques for decorrelation such as SCAN language based methodology, transform and predictive techniques [3], [6].

I. Image Compression Algorithm

Figure 1. illustrates the general scheme for image compression [7].

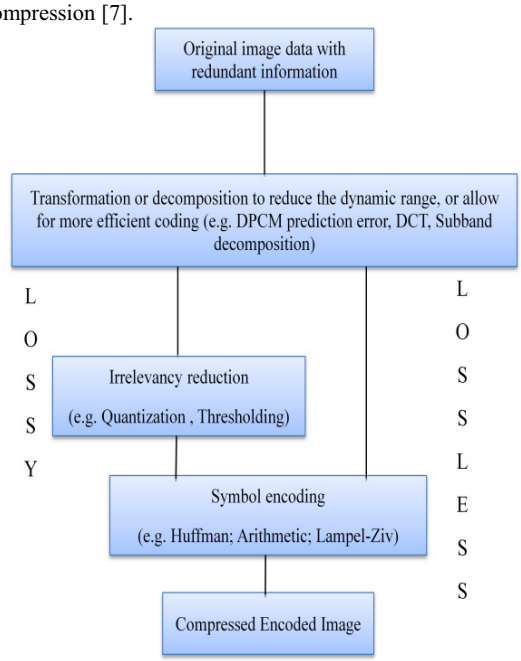


Figure 1. Generalized Image Compression Scheme

The first step in image compression is image transformation. In JPEG2000 and this research, the wavelet transform is used instead of other transformations [8-10]. Once an image is transformed/ decorrelated, symbol encoding step is performed to reduce coding redundancy. Prior to symbol encoding an optional step "quantization and thresholding" is performed. The quantization and thresholding step changes the data contained within an image. Thus, it is eliminated for lossless compression technique.

II. Generalized Image Transform

Let $f(x,y)$ denoted as an image, we define the transform of $f(x,y)$ as,

$$T(u,v) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)r(x,y,u,v) \qquad (2.1)$$

Equation (2.1) represents the general form of two dimensional transform. It is evaluated for $u = 0,1,2,\ldots,M-1$, and $v = 0,1,2,\ldots,N-1$. Where, $T(u,v)$ is the forwarded transform of $f(x,y)$, M and N represent the number of rows and columns in $f(x,y)$ respectively, x and y are the spatial variables, u and v are the transform variables, and

$r(x,y,u,v)$ is the forward transformation kernel. The inverse of $T(u,v)$ is obtained from,

$$\hat{f}(x,y) = \sum_{u=0}^{M-1}\sum_{v=0}^{N-1} T(u,v)s(x,y,u,v) \quad (2.2)$$

where $x = 0,1,2,\dots,M-1$, and $y = 0,1,2,\dots,N-1$, and $s(x,y,u,v)$ is the inverse transformation kernel [11]. If $\hat{f}(x,y) = f(x,y)$, then the transform is an invertible transformation.

Different kernels result in different transformations. In wavelet transform, the kernel of transformation consists of scaling and wavelet functions. Here, we used a separable two dimensional scaling function $\varphi(x,y)$ as follows:

$$\varphi(x,y) = \varphi(x)\varphi(y) \quad (2.3)$$

Where, $\varphi(x)$ and $\varphi(y)$ are one dimensional scaling function. The wavelet function measures intensity variation along different directions. Also, the three functions of the separable two dimensional wavelets are obtained as follows:

$$\psi^H(x,y) = \psi(x)\varphi(y) \quad (2.4)$$

$$\psi^V(x,y) = \varphi(x)\psi(y) \quad (2.5)$$

$$\psi^D(x,y) = \psi(x)\psi(y) \quad (2.6)$$

$\psi^H(x,y)$ is a directional wavelet that detects horizontal edges. Generally, horizontal edges are detected by measuring variations along columns of an image. $\psi^V(x,y)$ measures variations along rows and it detects the vertical edges, and $\psi^D(x,y)$ detects diagonal edges.

Thus, the two dimensional scaled and translated basis are defined as follows:

$$\varphi_{j,m,n}(x,y) = 2^{\frac{j}{2}}\varphi(2^j x - m, 2^j y - n) \quad (2.7)$$

$$\psi^i_{j,m,n}(x,y) = 2^{\frac{j}{2}}\psi^j\left(2^j x - m, 2^j y - n\right), \quad i = \{H,V,D\} \quad (2.8)$$

i is a subscript with values $H,V,and\ D$ that identifies the directional wavelets. The two dimensional discrete wavelet transform (DWT) is given by:

$$W_\varphi(j_o,m,n) = \frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\varphi_{jo,m,n}(x,y) \quad (2.9)$$

$$W^i_\psi(j,m,n) = \frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\psi^i_{j,m,n}(x,y), \quad i = \{H,V,D\} \quad (2.10)$$

j_o is an arbitrary starting scale, usually $j_o = 0$. Equation (2.9) calculates the approximations of $f(x,y)$ at scale j_o while the coefficients $W^i_\psi(j,m,n)$ add the horizontal, vertical, and diagonal details for scale j where$j > j_o$. Note that when $N = M = 2^J$, then $j = 0,1,2,\dots,J-1$, where J is decomposition level, and $m = n = 0,1,2,\dots,2^j - 1$. The DWT is similar to subband coding [12- 14] in which a 1-D or 2-D signal is decomposed using filter banks. Figure 2. demonstrates one level of decomposition of an image [11].

Additional decomposition can be obtained by decomposing the approximation of $f(x,y)$ using equations (2.9) and (2.10) [15-16].

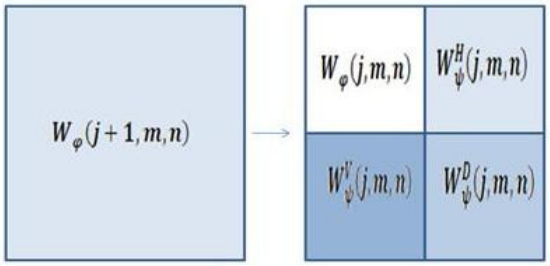


Figure 2. Level 1 decomposition

In Figure 2., $W_\varphi(j+1,m,n)$ represents the original image. $W_\varphi(j,m,n)$ is the approximation sub-image. Usually, natural images vary smoothly while fine details are represented as sharp edges. The approximation sub-image represents the smooth variations obtained from low passing the original image and decimating the results by factor two. $W^H_\psi(m,n)$, $W^V_\psi(m,n)$ and $W^D_\psi(m,n)$ are the detail images that are the results of high passing the original image and decimating the results by a factor two. These are the horizontal, vertical, and diagonal edges of the approximation image, respectively [11]. Figure 3. illustrates one level of decomposition using the wavelet transform.

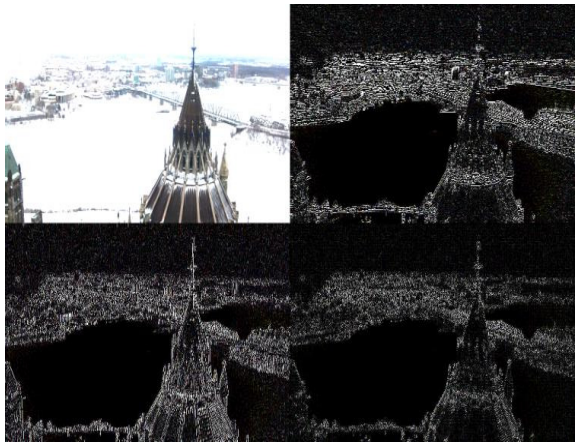


Figure 3. Parliament.jpg decomposed using the wavelet transform

The sub-image in the upper left corner of Figure 3. represents the approximation image with a size equal to $M/2xN/2$, where MxN is the size of the original image. The other three sub-images, of Figure 3. contain the horizontal, vertical, and diagonal edges, repressively.

III. Quantization and Thresholding

Once the approximation and detailed images are obtained, the quantization and thresholding step is performed. Quantization is simply rounding up the wavelet coefficients to the nearest integer. After quantization, thresholding is used to remove irrelevant information. In our algorithm iterative global thresholding was used. Thresholding is only performed on horizontal and vertical sub-images. In process of compression, we set the values in the diagonal sub-image to

zero. In the image reconstruction process, the diagonal edges are estimated mathematically.

IV. Diagonal Edges Estimation

Horizontal and vertical edges are used to estimate the diagonal edges. Sobel mask are used for this purpose. Figure 4. shows two different variations of Sobel masks that detect diagonal edges when applied to a sub-image of Figure 5.

0	1	2
-1	0	1
-2	-1	0

-2	-1	0
-1	0	1
0	1	2

Figure 4. Sobel masks for detecting diagonal edges

z_1	z_2	z_3
z_4	z_5	z_6
z_7	z_8	z_9

Figure 5. Sub-image

In our algorithm, the left mask in Figure 4. was applied to the horizontal and vertical sub-images to obtain g_x and g_y. The diagonal sub-image is computed as:

$$M(x,y) = \sqrt{g_x^2 + g_y^2} \qquad (4.1)$$

The process is not an error free but it provides edges that visually similar to the original ones. Figure 6. represents the estimated diagonal edges for the image in Figure 3. Using the Sobel operator we are able to estimate the diagonal edges that were discarded for the purpose of saving memory.

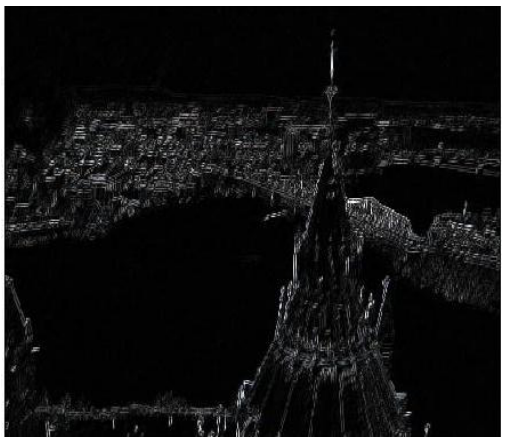


Figure 6. An estimated diagonal sub-image

V. Experimental Results and Analysis

We applied this algorithm, Mix JPEG, to two types of images, colored and grayscale. Figure 7a, and 7b, represents uncompressed colored and grayscale images respectively. Both images are of BMP extension (bitmap).

Figure 7a. Shuttle_Colored.BMP with no compression

Figure 7b. Shuttle_Gray.BMP with no compression

The Mix JPEG, is compared with two compression algorithms, JPEG (Photographs Experts Group) and JPEG2000. JPEG uses the discrete cosine transform and provides a high compression rate. JPEG2000 uses the wavelet transform and produces images with high quality and moderate compression rate. JPEG and JPEG2000 are the ISO/ITU standards [17-23]. Figure 8a, 8b, and 8c illustrate an image [24], that is compressed using different compression algorithms

Figure 8a. Shuttle_colored.jp2

Figure 8b. Shuttle_colored.jpg

Figure 8c. Shuttle_colored.mix jpg

Figure 9a, 9b, and 9c present the grayscale version of Figure 8 which are compressed using JPEG2000, JPEG, and Mix JPEG, respectively.

Figure 9a, Shuttle_grayscale.jp2

Figure 9b, Shuttle_grayscale.jpg

Figure 9c, Shuttle_grayscale.mix jpg

The compressed images in Figure 8 and 9 are visually the same compared with the original uncompressed images in Figure 7a, and 7b, respectively. Visually, there is no difference between images with extension .jp2 and .jpg but artifacts become obvious when the images are zoomed in. JPEG standard delivers high compression but the tradeoff is information lose. Figure 10a, 10b, and 10c demonstrate a zoomed region by a factor of 400% for images in Figure 8a, 8b, and 8c, respectively. Image in Figure 10a contains only edge artifacts and the image in Figure 10b contains blocking artifacts since JPEG standard uses an 8x8 blocking technique in its transformation. The overlapping nature of the wavelet transform helps alleviates the blocking artifacts that results from the DCT.

As the compression ratio increases, images compressed with the wavelet transform degrades more gracefully compared with JPEG. This is due to the functionality of the wavelet transform that supports both low and high frequencies. This is illustrated in Figure 10. The subjective evaluation of Figure 10. reveals that image in Figure 10c. is sharper than the other two images.

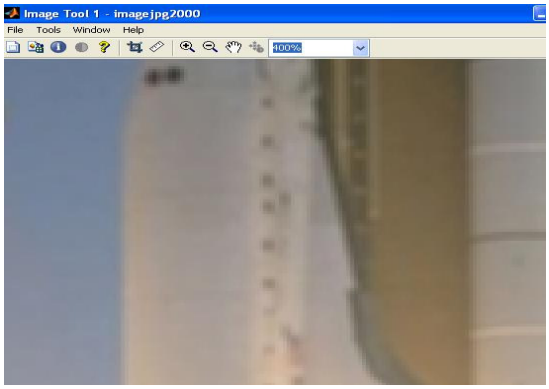


Figure 10a. Zoomed region from Shuttle_colored.jp2

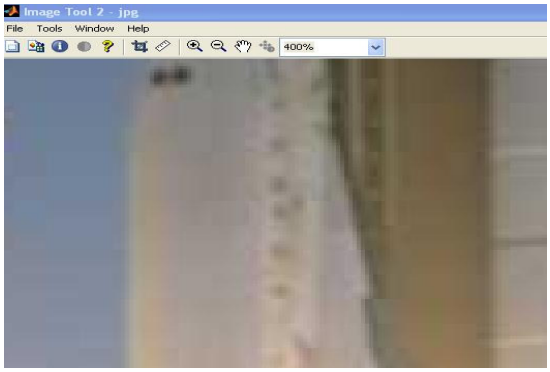


Figure 10b. Zoomed region from Shuttle_colored.jpg

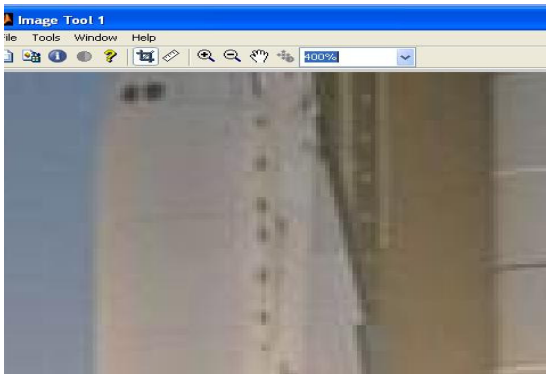


Figure 10c. Zoomed region from Shuttle_colored.mix jpg

Charts 1 and 2, illustrate the compression ratio for images in Figure 8 and 9, respectively. Note that compression ratio is based on calculating the ratio between the uncompressed image, i.e., BMP extension, and the compressed one. For colored and grayscale images we were able to obtain the highest compression ratio compared to the JPEG2000 and JEPG.

The mean square error e_{rms} and signal-to-noise ratio SNR_{rms} are used to evaluate the objective quality of images. Equations (5.1) and (5.2) illustrate the mathematical expression for e_{rms}, and SNR_{rms} ,respectively.

$$e_{rms} = \left[\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}[\hat{f}(x,y) - f(x,y)]^2\right] \quad (5.1)$$

$$SNR_{rms} = \frac{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\hat{f}(x,y)^2}{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}[\hat{f}(x,y) - f(x,y)]^2} \quad (5.2)$$

Tables 1 and 2, show e_{rms} and SNR_{rms} values for images in Figure 9, and 10, respectively. For colored images, this algorithm provides better objective quality in comparison with JPEG but slightly less than the quality of JEPG2000. This is due to quantizing and thresholding the horizontal and vertical sub-images. Eliminating quantization and thresholding part would slightly increase image size but provides better objective quality. For grayscale images, the quality of our images is still better compare to JPEG but lesser compared to JPEG2000.

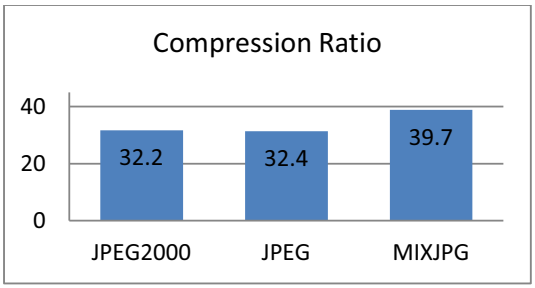


Chart 1. Compression ratio for a compressed colored image

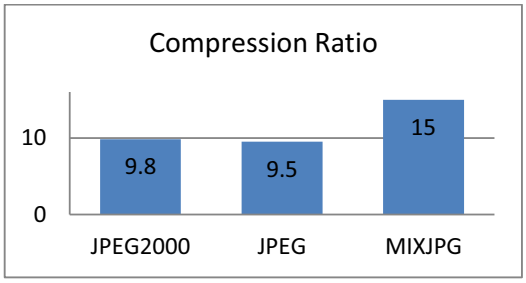


Chart 2. Compression ratio for a compressed grayscale image

	e_{rms}	SNR_{rms}	Size (KB)	Compression ratio
BMP	-	-	1361	-
JPEG2000	0.92	48.5	42.3	32.2%
JPEG	1.96	45.21	42	32.4%
MIX JPEG (quantization)	1.57	46.17	34.3	39.7%
MIXJPEG (no quantization)	0.8	49.12	35.1	38.8%

Table 1, Objective quality measurement for images in Figure 8

	e_{rms}	SNR_{rms}	Size (KB)	Compression ratio
BMP	-	-	457	-
JPEG	1.05	47.91	46.6	9.8%
JPEG 2000	4.33	41.77	48	9.5%
MIX JPEG (quantization)	5.16	41	30.5	15%
MIXJPEG (no quantization)	4.27	41.81	31.7	14.4%

Table 2, Objective quality measurement for images in Figure 9

VI. Conclusions

A novel lossy image compression is introduced in this paper. The transformation used in the compression algorithm is the wavelet transform which is the same as the one that is used in JPEG2000 but uses a different wavelet. The diagonal sub-image obtained from the wavelet transform is set to zero. This is due to the fact that the diagonal edges can be estimated using the horizontal and vertical sub-images during the decompression process. This process contributes to less memory requirement for image storage. Quantization and thresholding are performed on horizontal and vertical sub-images to increase the compression rate. This algorithm provides a higher compression rate compared to JPEG2000, and JPEG. For the images used in this research, JPEG2000, and JPEG provided a compression ratio equal to 32.2% and 32.4% for colored images and 9.8% and 9.5% for grayscale image, respectively. The same images were compressed using mix JPEG and provided 38.8% and 15% for colored and grayscale image, respectively. These compression ratios are slightly better than JPEG2000, and JPEG.

References

[1] Asraf, R.; Akbar, M.; Jafri, N.; , "Statistical Analysis of Difference image for Absolutely Lossless Compression of Medical Images," *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE* , vol., no., pp.4767-4770, Aug. 30 2006-Sept. 3 2006.

[2] R. Adhami, P.M. Meenen, III, D. Hite "Fundamental Concepts in Electrical and Computer Engineering with Practical Design Problems", 2nd edition, 2007.

[3] Ming Yang; Bourbakis, N.; , "An overview of lossless digital image compression techniques," *Circuits and Systems, 2005. 48th Midwest Symposium on* , vol., no., pp. 1099- 1102 Vol. 2, 7-10 Aug. 2005.

[4] Tzong-Jer Chen; Keh-Shih Chuang; , "A pseudo lossless image compression method," *Image and Signal Processing (CISP), 2010 3rd International Congress on* , vol.2, no., pp.610-615, 16-18 Oct. 2010.

[5] R. Sudhakar, Ms R Karthiga, S. Jayaraman, "Image Compression using Coding of Wavelet Coefficients- A Survey," ICGST-GVIP Journal, Vol. 5, Issue 6, June 2005.

[6] C.Alexopoulos and N.Bourbakis, An algebraic modeling of the SCAN language, Pattern Recognition Journal, to appear (also PhD thesis Univ. of P a m 1989).

[7] Majid Rabbani and Paul Jones, "Digital Image Compression Techniques", SPIE Press, 1991.

[8] Hua Li; Yiming Zhu; , "Lossless Image Compression Based on DPCM-IWPT," *Computing, Communication, Control, and Management, 2008. CCCM '08. ISECS International Colloquium on* , vol.1, no., pp.157-160, 3-4 Aug. 2008.

[9] S. Hamidi, R. Adhami, and L. M. Bruce, Lossless Data Compression Using Compactly Supported Wavelets Proc. Fifth International Conference on Signal Processing Applications & Technology, ICSPAT 94 Conference, Dallas, TX., 1994.

[10] S. Hamidi and R. Adhami, Wavelet Transform for Image Data Compression, Proc. 25th SSST, 1993.

[11] R. C. Gonzalez, R. E. Woods, "Digital Image Processing", 3rd edition, 2008.

[12] N. J. Fliege, "Multirate Digital Signal Processing: Multirate Systems – Filter Banks – Wavelets", Wiley &sons, Jan., 2000.

[13] Strang, G. and Nguyen, T. Wavelets and Filter Banks, Wellesley-Cambridge Press, Wellesley, MA, 1996, http://www-math.mit.edu/~gs/books/wfb.html.

[14] K. P. Soman & K.I.Ramachandran "Insight into Wavelets from theory to practice", Prentice Hall India, New Delhi, 2002.

[15] Mallat, S.G.; , "A theory for multiresolution signal decomposition: the wavelet representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , vol.11, no.7, pp.674-693, Jul 1989.

[16] Volkmer, H.; , "On the regularity of wavelets," *Information Theory, IEEE Transactions on* , vol.38, no.2, pp.872-876, Mar 1992.

[17] Penne baker, W. B.and Mitchell, J. L. JPEG - Still Image Data Compression Standards, Van Nostrand Reinhold, 1993.

[18] Wallace, G.K.,The JPEG still picture compression standard, Comm. of the ACM, 34(4), 30–44, 1991.

[19] Balster, E. J.; Fortener, B. T.; Turri, W. F.; , "Integer Computation of Lossy JPEG2000 Compression," *Image Processing, IEEE Transactions on* , vol.20, no.8, pp.2386-2391, Aug. 2011.

[20] T. Acharya and P.-S. Tsai, JPEG2000: Standard for Image Compression. Hoboken: Wiley, 2004.

[21] JPEG 2000 Part I Committee Draft Version 1.0, , Mar. 2000, ISO/IEC 1.29.15444-1.

[22] A. Samet, M. A. Ben Ayed, M. Loulou, and N. Masmoudi, "Comparison between JPEG and JPEG2000 still image compression standard," Proc. Vis., Imag., Image Process., Sep. 2002.

[23] D. Santa-Cruz, R. Grosbois, and T. Ebrahimi, "JPEG 2000 performance evaluation and assessment," Signal Process.: Image Commun., vol. 17, pp. 113–130, 2002.

[24]http://www.nasa.gov/multimedia/imagegallery/iotd.html.

EYE BLINK DETECTION USING INTENSITY VERTICAL PROJECTION

Hai DINH, Emil JOVANOV and Reza ADHAMI
Dept. Electrical and Computer Engineering, University of Alabama in Huntsville
Huntsville, Alabama, USA

ABSTRACT

Blinking is a spontaneous or voluntary action of human eyes. It has long been used for human-machine communication and human physiological detection. In these applications, detecting the eye's states, opened or closed, is an important step and affects the accuracy of detecting human intention or physiological states. In this paper, we present a robust method to detect eye blink based on the intensity vertical projection of eye's image. Unlike many previous works on eye blink detection based on edges of images, our work addresses blink detection by using the intensity of images. The method can deal with the variation of user's eye shape and distances between users and camera.

Keywords: Eye Blink, Vertical Projection, Drowsiness, Viola-Jones

1. INTRODUCTION

Eye blink and eye movement have been used for both human-machine communication to help people with disability [3, 8] and human physiological detection to improve driving safety [2]. Reports show that drowsiness is the major cause of fatal accident [1-2]. The drowsiness increase when we drive at night, in a long trip, or in uniform and monotonous roads. Many projects had been launched in Europe such as SAVE-1996 and AWAKE-2000 programs to address the drowsiness issue in drivers. There are many approaches addressing this issue such as analyzing driver behavior (i.e. steering wheel and pedal movements), mouth opening or heart rate. However, the electrode attachments make those methods intrusive and inconvenient to employ. Recent researches focus on non-intrusive approaches such as using camera to detect eye blink [1-2]. Despite promising results, they face some problems of accuracy of blink detection, variance of eyes' shape or users' movement. In this paper, we propose a new method for eye blink detection by using intensity vertical projection.

2. EXISTING APPROACHES

Applications employed eye blink detection or eye movement tracking for either communication or physiological detection are classified into two approaches: biological and optical approaches.

Biological Approaches: Biological approaches are techniques evaluating and recording the electrical activity produced by human brain, muscle, or retina. Those signals change when a person blinks or moves their eyeballs. Reference [3] proposed a method using eye blinks to activate home lighting system. This application provides an assistive communication for people with disabilities. Electrodes are placed on forehead, earlobe, and occiput to measure brain's spontaneous activity signal (Electro-encephalography, EEG). In order to accurately detect users' intent, the paper analyzed types of blink to differentiate the intentional blinks and unintentional ones. Based on that analysis, patterns of blinks are chosen so that the voluntary blinks, which we make purposely, do not interfere or are not misrecognized with other blinks. Although successfully detect eye blinks with 85% accuracy, the system fail to detect the eye movements which contain artifacts from spontaneous blinks and eyelid movement.

Reference [4] proposed a method using the resting potential of retina (Electrooculography, EOG) instead of EEG for eye mimicking in animation. The authors measure signals from biopotential electrodes placed on forehead and corners of eyes. Authors claim that eyeball orientation could be estimated by using EOG. The estimation requires signal calibration due to variety of personal characteristics and precision of electrodes placement. A detailed method for estimation has not yet been described, and the results also focus only on blink detection. Eye blinks are marked by high amplitude pulses, which extracted by subtracting original signal with signal led through median filter. Eye blinks and eyeball position are then incorporated to produce the appropriate values of morphing goals of an animated character in the computer animation.

In summary, the biological approaches can be used for blink detection but they are relative expensive and bulky. The usage of electrodes attached to human heads also makes it intrusive and inconvenient. This makes it more difficult to apply in systems that require mobility such as driver's drowsiness detection.

Optical Approaches: Optical approaches are methods using motion or pattern of objects in an image sequence. Object recognition algorithms provide accurate methods to detect eyes which have been integrated into a broad variety of applications. The optical approaches are divided into two sub methods: active and passive methods.

Active methods use infrared (IR) illumination to exploit physical properties of eyes. In [5], the red-eye effect, which occurs when using a flash very close to the camera lens, is exploited. Much of the light from the flash passes into the eye through the pupil, reflects off the fundus at the back of the eyeball and out through the pupil. Normal visible light can also create this effect, but IR light is preferred because it does not make users uncomfortable. The system includes on-axis and off-axis IR LEDs. Those LEDs alternately turn on and off while camera taking pictures. The on-axis LEDs create images with bright pupil and off-axis LEDs create images with dark pupil. Those interlaced images are subtracted and applied a threshold filter to get the difference image. Connected component algorithm labels white pixels into groups and then filters those which contain less than three pixels. Probabilistic principal component analysis (PPCA) is used to classify candidates. Each candidate's texture is projected into both eye and non-eye vector spaces which were trained by sets of images. The probability for a particular candidate belongs to either eye or non-eye space is calculated, and then Kalman filter are used to track eyes. The disadvantages are that the method requires IR sources. Furthermore, frame subtraction is sensible to light changing and motion artifact.

The passive methods are mere image processing even some of them still use IR LEDs to deal with ill-illuminated environments such as in dark room or at night. The algorithms vary from detecting eyelid movement or pupil/iris appearances. The driver vigilance diagnostic system [2] applies passive optical method for driver's drowsiness detection. The system detects eyes by initially searching the whole image to locate of face's and eye's features such as eye corners, corners of the mouth and the eyebrows. After initialization, eyes are tracked by using geometrically constrained Kalman filter. The distance between upper and lower eyelids is used to judge the state of eyes. The upper eyelid is modeled by a second degree curve and the lower eyelid is modeled by the line on edge eye region. In comparison with biological approach, this system can detect not only eye blink, but the blink amplitude, the opening degree of eye before and after blink. Detected blinks are classified into short, long, very long, and sleepy blink based on their duration. A fuzzy diagnostic is employed to deal with variability of blinking behavior. The numbers of blink in each category are counted in a time frame ΔT and multiplied by its type's degree to provide the driver vigilance level such as alert, slightly drowsy, drowsy and sleepy.

Reference [1] uses the distance between eyelids for blink detection. Eyes are located by using Viola-Jones object detection algorithm [7] which combines integral image representation and Adaboost classifier to detect objects fast. The detected eye areas are then cropped aligned and tracked by using Lucas-Kanade optical flow. Eyelids are extracted by applying 2D Gabor filter on eye regions. The Gabor filter responds strongly to edges with specific angle vector depended on the orientation angle. The connected component is used to label objects and filter out objects which contain small number of pixels. The largest arcs will be considered eyelids. The distance between those arcs is the opening of the eyes. If it is smaller than a threshold, eyes are considered closed. The improvement of the approach is that it uses Gabor filter instead of normal edge detection as in [2]. Authors claim that the system accuracy is 100%, robust to noise and variety of human eyes. The error of eye detection has not been mentioned. The speed of human head movement also has not considered and, as mentioned by author, it must guarantee that the eye image must have same size, heads rotation, and symmetry between right and left.

In summary, optical approaches are non-intrusive and more convenient. The improving of computational power makes it wearable.

3. PROPOSED METHOD

Intensity Vertical Projection Analysis: Intensity Vertical Projection (IVP) is the total intensity of object pixels in each row. As seen in Figure 1, the eyebrow and iris areas are supposed to be darker than the skin. Hence, there will be two local minima of IVP represent their centers (point A and B). The maximum between those is the center of the skin area between eyebrow and iris (point C).

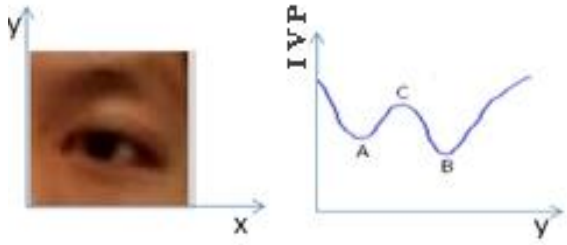


Figure 1. Eye image and IVP.

When the eye is closed, the second local minimum is now the center of the eyelids. Because the skin area is expanded when eye is closed, the IVP become wider around C and narrower around B. This is not depended on user as seen in Figure 2, note that for more convenient comparisons, eye images are rotated ninety degree in this figure.

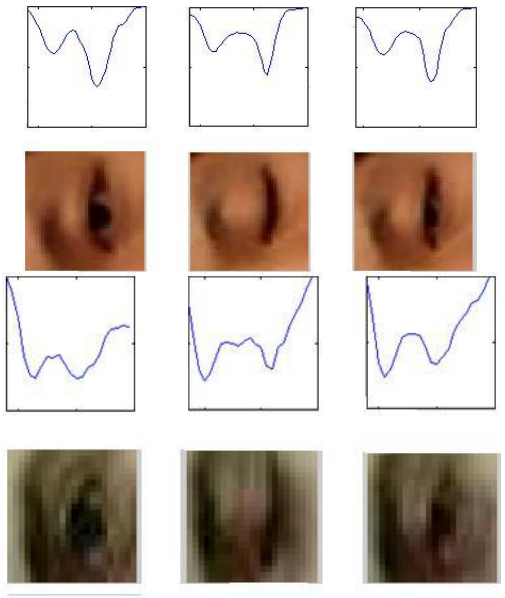


Figure 2. Opened, closed, and half-closed eyes' IPV two different users.

We define the eye opening is the ratio between iris area over the skin area. This ratio can be computed by calculate the ratio of intersection segments between IVP and the adjustment line. The line is defined as 80% of intensity at center of skin area.

$$Eye\ opening = \frac{FE}{ED} \quad (1)$$

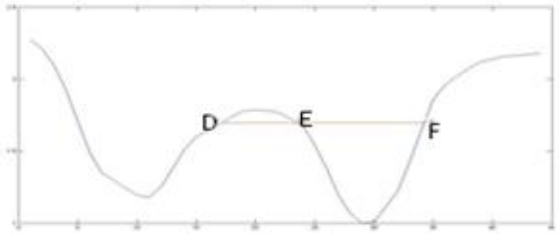


Figure 3. Eye opening parameter.

If the eye opening ratio is larger than a threshold, the eye is considered opened. Otherwise, it's considered closed.

System Overview: Our eye blink detection system includes three main steps as in Figure 4.

Figure 4. System's block diagram.

Video sources are images captured from webcam or loaded from videos. To detect user's eyes, Viola-Jones detection algorithm is applied. The system first looks for user's face and then searches for eyes within detected face. Note that the eye can be found directly in the image without the face detection step, but a two-step detection increases the confidence of detected eyes. The right eye region then will be cropped and used for eye opening calculation at each image frame. Figure 5 shows the system graphical-user-interface (GUI). The value of blink graph is one when eye is closed.

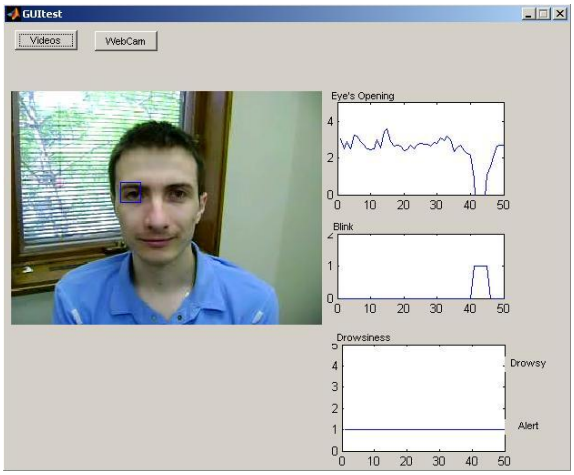


Figure 5. System GUI.

4. EXPERIMENTATION

The system runs on ACER Extensa 5420 with AMD Athlon X2 1.9 GHz CPU, 2G RAM. Videos are taken from Sony camera 10.2 mega pixels and save in avi format. The source code is developed in Matlab and can run without additional library. The trained haar classifier for Viola-Jones detection algorithm is from OpenCV website in the form of an xml file.

Result of Blink Detection: The system is tested with both static and dynamic (moving) users. The moving user makes head-movement left, right, closer and further from camera. The opening ratio is marked by red crosses for opened eyes and blue circles for closed eyes. The eye's statuses are manually checked.

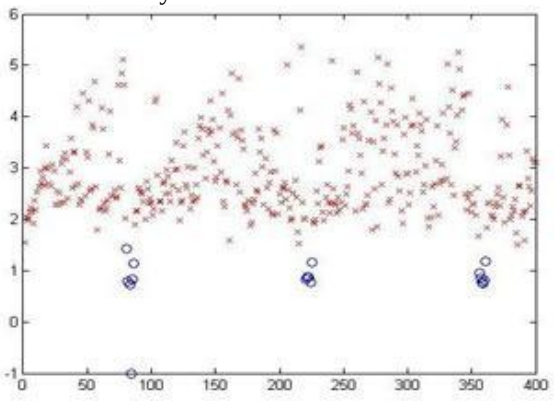


Figure 6. Eye Opening Ratio of static user.

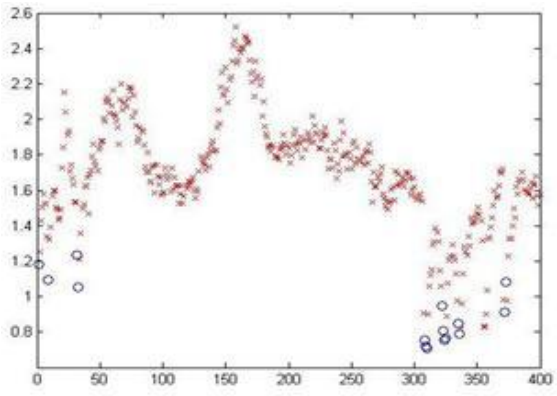


Figure 7. Eye Opening Ratio of dynamic user.

As seen in Figures 6 and 7, the opened and closed eyes can be distinguished by the opening ratio. Figures 8 and 9 show the result for blink detection from the system in comparison with manually checked result. The ratio threshold is set at 1.2 on both videos.

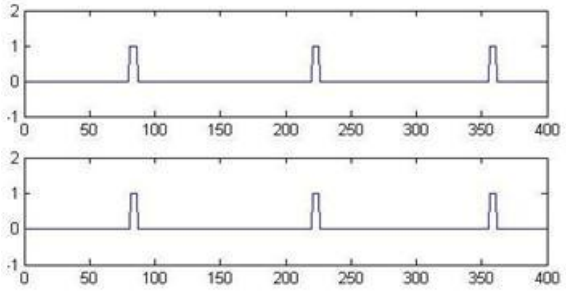


Figure 8. Blink detection result comparison between our system and manual check of static user.

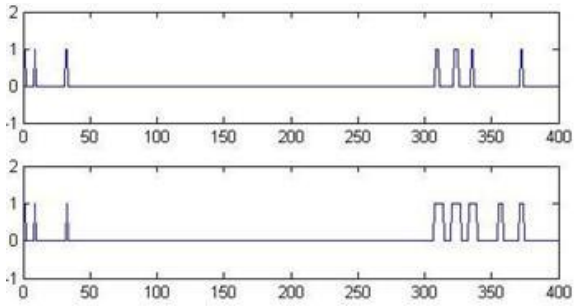


Figure 9. Blink detection result comparison between our system and manual check of dynamic user.

The accuracy is 100% when the user does not move. For the second video, there is one false positive blink, and the accuracy of eye status (opened or closed) is 94.8%.

Drowsiness Detection: The result from blink detection can be used to detect physiological state of drivers. The drowsiness signals, such as vibrates or alarm sounds, are turned on if the system detect that the driver has more than 2 long blinks (defined by the duration of closed eye) and the opening ratio is small after those blinks.

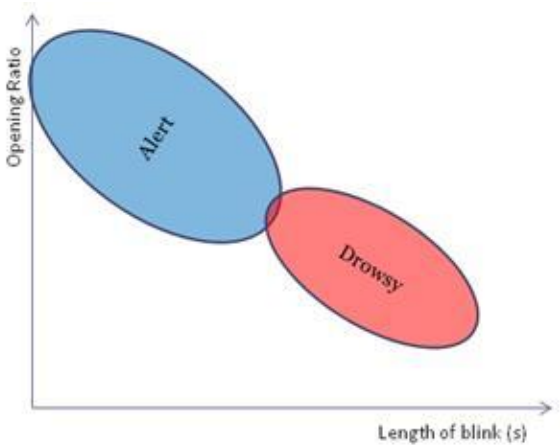


Figure 10. Drowsiness detection.

5. CONCLUSIONS

This paper proposes a new and simple method for eye blink detection by using intensity vertical projection. The proposed method calculates the opening ratio from the vertical projection of images' intensity to detect eye's states. The proposed method is validated using videos of different users. The results are very comparable to manual measurement. Another advantage of the proposed method is that it robust again the variation of eye shapes and eye movements.

6. REFERENCES

[1] Arai, K. and Mardivanto, R., "Real Time Blinking Detection Based on Gabor Filter," **International Journal of Human Computer Interaction**, 2010.

[2] Boverie, S. and Giralt, A., "Driver vigilance diagnostic based on eyelid movement observation," **The International Federation of Automatic Control, Proceedings of the 17th World Congress**. Seoul, Korea 6-11 July 2008.

[3] Rani, M.S.A. and Mansor, W., "Detection of Eye Blinks From EEG Signals for Home Lighting System Activation," **Proceeding of the 6th International Symposium on Mechatronics and its Applications (ISMA09)**. Sharjah, UAE 24-26 March 2009.

[4] Krupinski, R. and Mazurek, P., "Estimation of Eye Blinking Using Biopotentials Measurements for Computer Animation Applications," **Springer-Verlag Berlin Heidelberg**, 2009.

[5] Haro, A., Flickner, M. and Essa, I., "Detecting and Tracking Eyes by Using Their Physiological Properties, Dynamics, and Appearance," **Proceedings of Computer Vision and Pattern Recognition 2000**, pp. 163-168.

[6] Wang, P., Green, M. B., Ji, Q. and Wayman, J., "Automatic Eye Detection and Its Validation," **Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2005**.

[7] Viola, P. and Jones, M., "Robust Real Time Object Detection," **Second International Workshop on Statistical and Computational Theories of Vision-Modeling, Learning, Computing, and Sampling**. Vancouver, Canada, July 2001.

[8] Tobii Technology, ***www.tobii.com***

Mass Distribution Effect on Flutter Characteristics

Abid Ali KHAN
Aerospace Engineering Department, College of Aeronautical Engineering
National University of Sciences & Technology, Pakistan

Abdul Munem KHAN
College of Aeronautical Engineering, National University of Sciences & Technology
Risalpur, 24090 Pakistan

Nadeem Shafi KHAN
Aerospace Engineering Department, College of Aeronautical Engineering
National University of Sciences & Technology, Pakistan

ABSTRACT

Completely rigid structures do not practically exist, so every structure under the action of aerodynamic forces will normally deflect at least by a small amount. This deflection will become increasingly significant at higher air speeds. Elastic stiffness of the material will try to restore the structure to its original shape and in the process, will overshoot the static equilibrium point and the structure will be pushed into expansion. This process will cause the structure to vibrate. If the energy during the period of aerodynamic excitation is more than the natural damping of the structure, the amplitude of vibration will increase. With increase in air speed, the structural damping of the system will diminish to a point where it becomes zero and eventually, its value changes from negative to positive value. At the point, when damping becomes zero, it is said that aeroelastic phenomena commonly known as 'flutter' has occurred. This paper presents the computational work carried out using Fluid Structure Interaction module of a commercial software to describe the effect of mass distribution on flutter characteristics of structures subjected to aerodynamic forces.

Keywords: Structural dynamics, Aero-elasticity, Flutter, Vibration Modes and Resonance.

1. INTRODUCTION

Aeroelasticity is the field of science in which the emphasis is on the study of the interaction between the deformation of an elastic structure exposed to an airstream and the resultant aerodynamic force. The mutual interaction of the three disciplines namely; aerodynamics, dynamics, and elasticity can be seen in figure 1. Classical aerodynamic theories provide a mean to assist in determining the forces acting on a body of a given shape while the resultant shape of an elastic body under any load can be determined using theory of elasticity. Lastly, the effects of inertial forces in the overall system are catered by Dynamics. It is only possible with the interdisciplinary know how of these three disciplines that the designer is in a position to study and comprehend the implications of the mutual interaction of two or more of these phenomena [1]. The consequences of interaction of these phenomena can be tragic and at times result into heavy losses. One of such interactions called "Flutter" is being briefly touched upon in this paper. It has perhaps the most far reaching effects of all aeroelastic phenomena on the design of high speed aircraft.

Modal analysis is the field of study in which the dynamic response of the structures and/or fluids upon excitation by an input is measured and analysed. Using modal analysis, the dynamic response of the structure can be explicated in terms of damping, natural frequencies, mode shapes and resonance. All computational work included in this research is based upon Modal analysis.

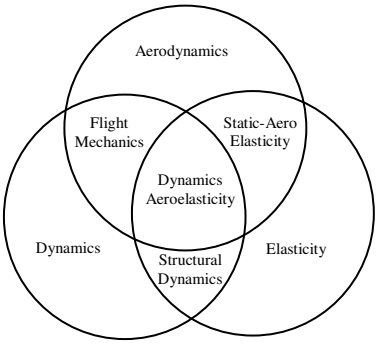


Figure 1: Schematic for the field of Aeroelasticity

2. HISTORICAL BACKGROUND

Throughout the history of powered flight, it can be witnessed that particular attention has been focused on the aeroelastic phenomenon. [2]. In 1903, significant anhedral of the wings caused lateral instability initially in the Wright flyer aircraft. However, lateral control was achieved on the Wright flyer later, through effective utilization of the controlled warping of the wings which contributed appreciably to the triumph of powered flight. Earlier in 1903, Samuel Langley twice attempted unsuccessfully to achieve powered flight from the roof top of a houseboat in the Potomac River. Aeroelastic divergence resulting from insufficient torsional stiffness ended up in catastrophic failure of the wings. The same

torsional divergence phenomenon kept the biplane design predominant until the early 1930s when adequate torsional stiffness for monoplanes was ultimately realized through "stressed skin" metallic structural configurations.

Flutter on aircraft was first recorded and documented in 1916. The Handley Page O/400 (H.P.12) bomber was World War I (WWI) British bomber which was one of the largest aircraft in the world at that time. Absence of torsion rod connection between the port and star elevators contributed excessively to violent tail oscillations. This led to dynamic fuselage twisting of almost 45 degrees which was complemented by anti symmetric elevator flapping. Airplane flutter due to its associated catastrophic failures emerged as a major design parameter during WWI. Its importance in aircraft design remains so even today.

Implications of the aeroelastic phenomenon can be as mild as onboard human discomfort to as catastrophic as failures resulting in human life loss without warning. In between the above two extreme situations, steady state and transient vibrations can be a likely cause of fatigue damage of the aviation structure on the microscopic level. It is therefore, always recommended to take this aspect of wind tunnel and flight testing thoroughly and seriously. Short cuts in this aspect can result in unwanted/unpleasant circumstances during regular flying of the designed vehicle.

3. STRUCTURES & OSCILLATIONS

Structural nonlinearities are a common concern in the design of aircraft structures, which can occur in both distributed and concentrated form. More often, distributed nonlinearities are significant for large amplitude oscillations, whereas localized nonlinearities could be important at relatively much smaller amplitude levels. Control surface nonlinearities could be attributed to manufacturing defects or wear and tear which leads to loose or worn out control surface hinges. Thus aeroelastic problems with these nonlinear effects have been an important area to study. Nonlinear systems can display sustained limit cycle oscillations at different air speeds, unlike a linear system which shows a limit cycle oscillation only at the critical speed.

In order to explain the modal analysis, a constant force is applied to one corner of a freely hung plate but the frequency of excitation is changed in a sinusoidal fashion as shown in figure 2. As the frequency of constant force changes, response at the excitation point is measured using accelerometer. It can be noticed that level of amplitude of response changes with change in frequency oscillation. It seems odd as the level of excitation is constant. A typical time response is shown in figure 3. The data is converted to frequency domain through Fast Fourier Transform and a frequency response function (FRF) plot is obtained which is shown in figure 4. The peaks in this plot correspond to the frequency of oscillation where the amplitude of response is greatest. The same can be seen by overlaying the time and frequency trace, as the amplitude of response in time trace increases, the amplitude of the FRF also increases (see figure 5). Research revealed that the points of increased amplitude occur at the natural frequency of the system, which depends on the mass and stiffness distributions [3-4].

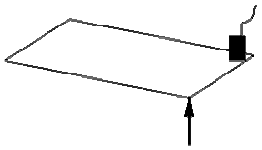


Figure 2: Simple plate excitation/response model

Figure 3: Simple plate time response plot

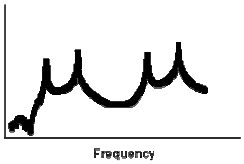


Figure 4: Simple plate frequency response plot

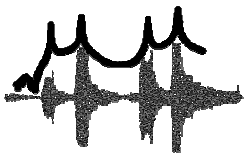


Figure 5: Overlay of time and frequency response function

4. DIFFERENT VIBRATION MODES

The excitation coincides with each of the four resonant frequencies at each of the peaks in the Frequency Response Function (FRF). This can be demonstrated by using a number of scattered accelerometers on the plate. The deformation patterns thus generated by these four modes are different from each other and can be seen in figure 6.

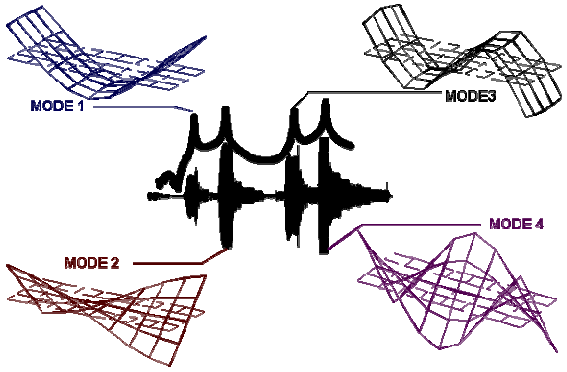


Figure 6**:** Simple plate mode shapes

(a) **Mode I :** The coinciding of excitation with first resonant frequency results in the 1st bending mode, where the shape of deformation is a bending motion.

(b) **Mode II:** The coinciding of excitation with second resonant frequency results in the 1st twisting mode, where the shape of deformation is a twisting motion.

(c) **Mode III:** The coinciding of excitation with third resonant frequency results in the 2nd bending mode, where the shape of deformation is a second bending motion.

(d) **Mode IV:** The coinciding of excitation with fourth resonant frequency results in the 2nd twisting mode, where the shape of deformation is a second twisting motion.

4. OSCILLATIONS ON A LIFTING SURFACE

There are two general classes of vibrations viz a viz excitation force - free and forced. When a mechanical system is initially excited by an input force and then the system is allowed to vibrate on its own after the removal of the exciting force, the system is said to be executing 'Free vibration'. Such a system vibrates at one or more of its natural frequency. The vibration gradually decays to zero because of the inherent structural stiffness of the elastic structure.

On the other hand, vibrations that are executed by the system during the active and continual excitation of the system is called 'Forced Vibration'. If the nature of the exciting force is oscillatory, it will compel the system to vibrate at the excitation frequency of the force. If accidently, this excitation frequency matches one of the natural frequencies of the vibrating system then amplitude of oscillation will be exponentially enhanced. This condition is known as 'resonance'. Resonance is known to have a major contribution to the destruction/ damage of airplane wings during flight.

An aerodynamic shape as shown in figure 7 that produces a typical aerodynamic force "lift", is selected for computations during research. As the structure has high stiffness and fourth natural frequency is very high therefore only first three expected vibration mode shapes for selected lifting surface are being discussed in this paper. The coinciding of excitation for the selected aerodynamic structure with its first, second and third resonant frequency results in the 1st bending mode, 1st twisting mode and 2nd bending mode respectively. The deformation patterns for the selected surface are shown in figure 8.

5. FLUTTER PHENOMENON

The Flutter or critical speed 'V_f' and frequency 'ω_f' are defined as the lowest airspeed and corresponding circular frequency at which a flying structure at a given atmospheric density and temperature will exhibit sustained simple harmonic oscillations [5]. Flutter is a self excited oscillation of an aerodynamic surface and its associated structure caused by the interaction of the aerodynamic, inertial and elastic characteristics of the components involved. At speeds below the flutter speed, oscillations are damped, but at the flutter speed, undamped oscillations persist with constant amplitude, however, at speeds above the flutter speed undamped oscillations diverge and result in damage or destruction of the structure [6-7]. The divergent behaviour can occur within a few cycles or with very little velocity increment and be catastrophic.

Flutter is said to have occurred when under the action of aerodynamic loads, the lifting surface deflects in such a way as to cause reduction of the applied load.[8] it can be described as a type of vibration which is self feeding and has the potential of destructing the structure eventually. During flutter, the frequency of the aerodynamic load couples itself to the structure's natural mode of vibration which results in rapid periodic motion. In other words, a positive feedback exists between the aerodynamic loading and structure's natural vibration characteristics. A 'divergent response' in which the amplitude tend to increase throughout the process can occur if the energy input by the air stream flow is greater than the structural damping of the structure. Practically it has been observed that flutter can be induced on an apparently unrelated aerodynamic component of the aircraft by causing a simple change of the mass distribution of an aircraft or the stiffness of a single component. Theodore Von Karman is said to have remarked that "Some men fear flutter because they do not understand it, while others fear it because they do".

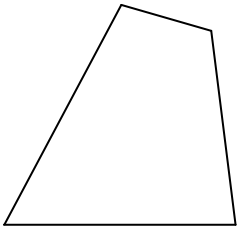


Figure 7**:** Selected Aerodynamic shape structure

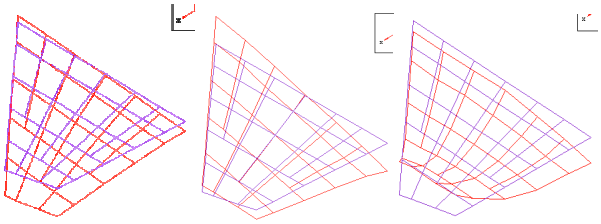


Figure 8**:** Three basic mode shapes for lifting surface

6. EFFECT ON A FLYING VEHICLE

In real life, flutter on a flying vehicle can be caused by the coupling of two or more structural modes. These coupled modes may be wing bending and torsion, wing bending-control surface hinge torsion, wing torsion-fuselage bending, or horizontal/vertical tail-fuselage modes. The phenomenon is explained in Figure 9, which shows velocity vs. frequency plot for four different structural modes. With the increase in speed, frequency of modes 3 and 4 remains stable and these modes do not interact with any other structural mode. However, frequency of modes 1 and 2 change with the increase in speed and they tend to interact with each other. This interaction is explained in Figure 10, where at velocity 'V_f' the modal damping of mode 1 has gone to zero. This is an onset condition for flutter. This is caused by the interaction of free stream energy with the elasticity of the structure and has created the effect of zero total damping (structural and aerodynamic). Increase in speed beyond this value will cause large increase in vibration amplitude, which are divergent and can cause catastrophic damage to the structure [9-10].

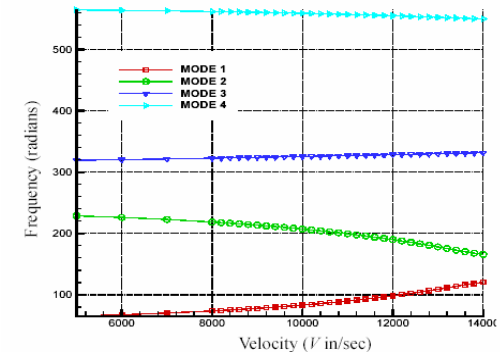


Figure 9: Flutter frequency vs velocity curve for explanation

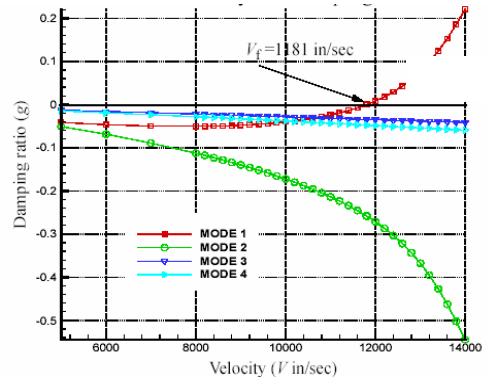


Figure 10: Flutter frequency vs velocity curve

7. BALANCE WEIGHT

Balance weights are dead weights added to different location for different purposes. In the aerodynamic structures these are generally used to achieve an optimum mass distribution. In this way, the structure although pays penalty in the shape of additional weight but achieves a required force distribution. This positive change can also result in the delay of aeroelastic phenomenon like flutter. The locations of balance weight for the selected aerodynamic shape structure can be seen in figure 11. The selected structure was modelled and the computations were done in the three configurations as given below:-

(a) Without a balance weight

(b) A balance weight of 6Kg at Leading Edge tip

(c) Different balance weights at Leading Edge root

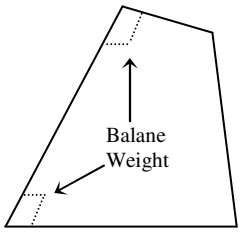


Figure 11: Balance weight locations

8. RESULTS & ANALYSIS

Results of analysis conducted without any balance weight, 6kg balance weight at the LE tip, relocating the balance weight from LE tip to LE root and then weight increments are being discussed in sub sequent paragraphs. During this analysis, no other changes to the structure model were made.

The research was aimed to reveal the affect of mass distribution while achieving the flutter speed of 400 m/s for the selected structure. The results revealed that with no balance weight and with 6 kg balance weight at the LE tip the flutter speed is much lower than the required value of 400 m/s. The results for computations using these two configurations can be seen in Table 1 whereas the results for computations using different balance weights at the Leading Edge Root are shown in Table 2. $V_{f0\%}$ and $V_{f2\%}$ are the flutter speeds for zero and 2 per cent damping respectively. Also, note that the computations for the second mode (which is the first twisting mode) have only been done for the third configuration.

Table 1: Computed Flutter for 0 & 6 kg balance wt at LE Tip

Weight (Kg)	Bending & Twist		
	$V_{f\,0\%}$ (m/s)	$V_{f\,2\%}$ (m/s)	ω_f (Hz)
0	208	242	33.8
6	368	378	27.2

Table 2: Flutter Computations with Balance Weight at LE Root

Wt (Kg)	Bending & Twist			2nd Mode		
	$V_{f\,0\%}$ (m/s)	$V_{f\,2\%}$ (m/s)	ω_f (Hz)	$V_{f\,0\%}$ (m/s)	$V_{f\,2\%}$ (m/s)	ω_f (Hz)
1	220	261	32.9	412	591	53.5
2	233	253	34	415	595	53.4
4	247	253	31.6	421	602	53.4
8	275	328	26.5	429	612	53.4
10	369	410	24.7	433	616	53.4
12	493	547	22.9	435	618	53.4
14	-	-	-	437	620	53.4
18	-	-	-	440	624	53.4

Analysis reveals that flutter speed in case of no balance weight and 6 kg weight addition at LE tip is much lower than the required value. However, the flutter speed for bending and twist modes start to improve with increase of balance weight at the LE root. It can be seen that required value of flutter speed that is more than 400 m/s is achievable with balance of 10 Kg or more. Although the improvements with higher balance weights are tremendous but weight plenty is to be paid in return for higher flutter speeds V_f. Computations for 2nd mode show that 400 m/s is not only achievable but the structure is also safe with any amount of balance weight LE root.

Figure 12 shows the trend of flutter speed for relocated different balance weights at the LE root in different modes.

Determination of the flutter boundary of aircraft or the major structural parts is an essential step to be performed during design qualification of any flying vehicle [11]. The flutter boundary can be defined either by analytical calculations or by carrying out flutter wind tunnel tests by using scaled down special flutter models of the concerned structure. The event of flutter dip is another important phenomenon to be computed for flying vehicles that are likely to go into the supersonic region or cross the transonic region. The transonic dip is found either by applying corrections from previous design experiences of similar flying structure or by carrying out the transonic speed flutter wind tunnel testing.

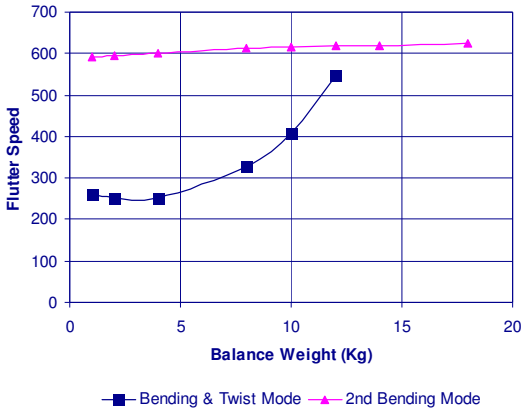


Figure 12: Computed Flutter speed Vs balance weight

Figure 13 shows the Equivalent Air Speed (Ve) and Mach Number (M) plots for different altitudes and flutter margins. Also shown is the flutter boundary and the transonic dip The V_e-M lines are drawn for different altitudes with sea level line having the highest slope and are shown by solid lines. The dotted lines show the expansion of basic V_e-M envelope expanded by using design margin of safety, i.e. 15%. This makes up the design limit speed envelope and it should still remain little below the flutter boundary curve and little right of the transonic dip. This makes up the first flutter qualification requirement for a flying vehicle.

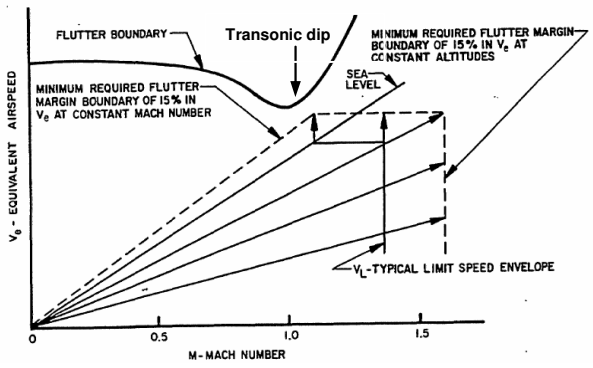


Figure 13: V_e vs M plots for different altitudes and flutter margins

9. CONCLUSION

The study and the computational work done during this research concluded that the flutter speed is a key parameter that predicts the maximum flutter free speed of flying structure. The use of balance weights can enhance the modal characteristics of the flying structures but the structures have to pay heavy plenty in terms of weight increase.

The computations revealed that addition of 10 kg at the root LE can enhance the flutter speed to a required value of 410 m/s. However increase in the supersonic region flutter speed involves more complex computations which are not included in this paper.

Also, in this research, the effect of damping in the control of flutter speed is highlighted. For the same amount of balance weight and its location, additional damping moves the flutter speed to higher values. Additionally, it can be seen that higher margins in flutter speed for the first twisting mode is available with additional damping. Critical analysis of the results further reveal that with increasing value of balance weight, lower circular flutter frequencies are achievable. However, this effect is negligible for the circular flutter frequencies associated with the first twisting mode.

It is to be kept in mind that the increase in damping ratios may increase the structural performance but cannot be kept very high as they tend to make the structures rigid. Vibrating systems are all subject to damping to some degree because energy is dissipated by friction and other resistances. If the damping is small, it has very little influence on the natural frequencies of the system, and hence the calculations for the natural frequencies are generally made on the basis of no damping. On the other hand, damping is of great importance in limiting the amplitude of oscillation at resonance.

10. FUTURE DIRECTIONS

Analysis for determining the transonic dip is also required in order to keep the flying structure safer in the transonic and supersonic region. Theoretical work in this regards will be done at a later stage and the results will be published. The experimentation to verify these computations may later be performed to verify the achieved results.

10. REFERENCES

[1] Hodges. D. H. and Pierce. G. A., **Introduction to Structural Dynamics and Aeroelasticity**. Cambridge University Press. Pub., 2002.

[2] Bisplinghoff. R. L., Ashley. H. and Halfman. R. L., **Aeroelasticity**. Addison-Wesley Publishing Company, Inc. Reading, Massachusetts, USA. Pub., 1957.

[3] Uhl. T., Lisowski. W., Mendrok. K. and Kurowski, P., New Solutions in Experimental Modal Analysis of Mechanical Structures. **Proceedings of**

International Congress on Theoretical and Applied Mechanics. Warsaw, Poland. 15 – 21 August, 2004.

[4] Avitabile. P., Experimental Modal Analysis (A Simple Non-Mathematical Presentation). **Sound and Vibration Magazine**. Vol 35, No. 1, p 20-31, 2001.

[5] Patil. J. M., Limit Cycle Oscillations of Aircraft due to Flutter Induced Drag. Paper No. 2002-1409. **Proceedings of 43rd Structures, Structural Dynamics and Materials Conference**. Denver, Colorado, USA. 22-25 April, 2002.

[6] Karpel. M., Efficient Vibration Mode Analysis of Aircraft with Multiple External Sore Configurations. **Journal of Aircraft**. Vol 25, No. 8, p 747-751, 1998.

[7] Shelton. J. D. and Tucker. P. B., Minimum Weight Design of the F-15 Empennage for Flutter. **Journal of Aircraft**. Vol 40, No. 5, p 832-842, 1975.

[8] Yang. G., Obayashi. S. and Nakamachi. J., Aileron Flutter Calculation for a Supersonic Fuselage-Wing Configuration. **Proceeding of 23rd International Council of Aerospace Sciences**. Toronto, Canada. September, 8-13, 2002.

[9] Abel. I., Research and Applications in Aeroelasticity and Structural Dynamics. **NASA Techncial Memorandum 112852**. NASA Langley Research Center. Hampton, Virginia. Pub., 1997.

[10] Bennett. R. M., and Edwards. J. W., An Overview of Recent Developments in Computational Aeroelasticity. Paper No. 98-2421 **Proceedings of 29th AIAA Fluid Dynamics Conference**. Albuquerque, NM, USA.. June 15 – 18, 1998.

[11] NASA **SP-8003**. Flutter, Buzz and Divergence. NASA Space Vehicle Design Criteria.

Fast Location Opposite Update Scheme for Minimizing Handover Latency over Wireless/Mobile Networks

Sunguk Lee
Research Institute of Industrial Science and Technology
Pohang, Gyeongbuk, 790-330, S.KOREA
sunguk@rist.re.kr

Randy S. Tolentino
Department of Multimedia Engineering
Hannam University, Daejeon, S. KOREA
daryn2004@yahoo.com

Byungjoo Park
Department of Multimedia Engineering
Hannam University, Daejeon, S. KOREA
bjpark@hnu.kr

ABSTRACT

Mobile IP (Internet Protocol) allows a mobile node to send and receive packets with its home IP address, regardless of the IP address of its current point of attachment in the Internet. Since Mobile IP induces the handover latency due to mobility management operations, fast handover algorithms have been studied to reduce the latency. In this paper, we optimize the handover procedure in Fast Handover for Mobile IPv6 (FMIPv6) scheme using reverse Binding mechanism. We will explain and discuss the scheme that supports a fast handover effectively in standard Mobile IPv6 (FMIPv6 by optimizing the associated data and the flow of the signal during handover. A new signaling messages, Fast Packet Binding Update (PBU), and Reverse PBU, are defined and utilized to hasten the handover procedure.

Keywords: MIPv6, FMIPv6, Binding Update, Fast Handover, Seamless.

1. INTRODUCTION

Mobile Internet Protocol version 6 (IPv6) allows an IPv6 node to be mobile—to arbitrarily change its location on an IPv6 network—and still maintain reachability. Connection maintenance for mobile nodes is not done by modifying Transport layer protocols, but by handling the change of addresses at the Internet layer using Mobile IPv6 messages, options, and processes that ensure the correct delivery of data regardless of the mobile node's location [1].

The MIPv6 facilitates to reduce the signaling overhead and delay concerned with the location update, in which an MN sends a local Binding Updates (BU) to the local MAP, rather than the Home Agent (HA) and Correspondent Nodes (CN). On the other hand, the Fast Handover for MIPv6 (FMIPv6) uses bi-directional tunnels between ARs and exploits various L2 triggers for supporting fast handover and further minimizing service disruption during handover.

In this paper we proposed a robust Fast handover scheme for MIPv6 using reverse binding mechanism to overcome such ineffectiveness of standard. This paper briefly describes and discusses the robust fast handover scheme in MIPv6 using reverse binding mechanism.

2. RERATED WORKS AND PROBLEMS

A. Handover Procedure in Mobile IPv6

We can define the handover procedure like as movement detection, new CoA configuration, DAD and binding update. To process Movement Detection, an MN detects that it has moved to a new subnet by analyzing the router advertisement periodically sent by

the access router (AR). The MN can also request the AR to send a router advertisement by sending a router solicitation. To initiate CoA configuration and DAD, the information contained in the router advertisement will allow the MN to create a new CoA. As specified in IPv6 [6], the MN first needs to verify the uniqueness of its link-local address on the new link. The MN performs DAD on its link-local address. Then, it may use either stateless or stateful address autoconfiguration [7] to form its new CoA.

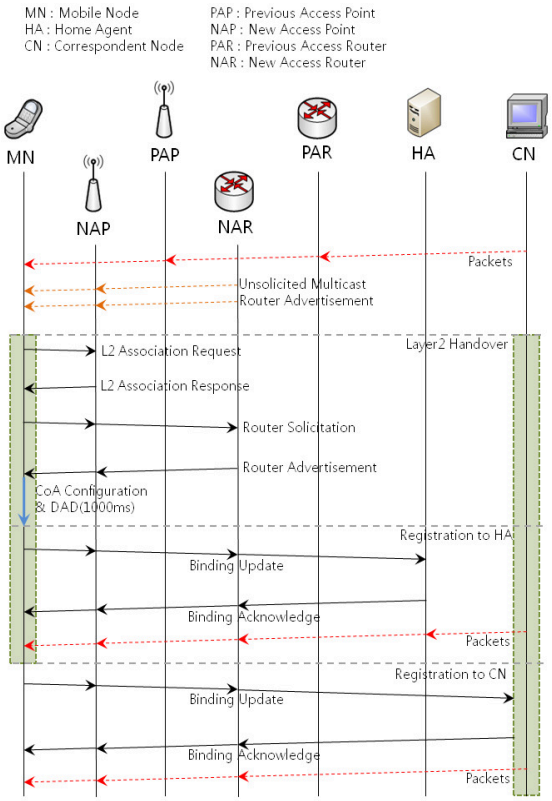


Figure 1. MIPv6 Handover Procedure

1*) Movement Detection:* The primary aim of movement detection is to identify L3 handovers. In MIPv6, movement detection generally uses Neighbor Unreachability Detection to determine when the default router is no longer bi-directionally reachable, in which case the mobile node must discover a new default router on a new link.

However, this detection only occurs when the mobile node has packets to send, and in the absence of frequent Router Advertisements or indications from the link-layer, the mobile node might become unaware of an L3 handover. After a change of Link Layer connection the MN must detect any change at the IP Layer before it can signal the change to the network. In MIPv6 this uses RS and RA to detect changes of IP network prefix. This is part of the standard Router Discovery Protocol [6]. The Router Discovery Protocol of IPv6 Neighbor Discovery contains built in timers. These timers prevent a router from sending immediate responses to RS in order to prevent multiple nodes from transmitting at exactly the same time and to avoid long-range periodic transmissions from synchronizing with each other. These are significant delays since they interfere with the MIPv6 movement detection algorithm thus preventing mobility signaling for up to 1000ms [1] [6].

2) Duplicate Address Detection (DAD): In MIPv6, after completing movement detection an MN should generate a new CoA using IPv6 stateless address auto-configuration upon moving to the new link [6] [7].

After generation of the CoA an MN should perform DAD for testing the new CoA's uniqueness within the new link. The duration required to complete DAD is up to 1 second. This delay is inherent to MIPv6.

B. Fast Handover for Mobile IPv6

FMIPv6 is designed to address issues related to IP layer movement detection, CoA configuration and BU in MIPv6. It exploits various L2 triggers to prepare an NCoA at the nAR in advance while being connected to oAR's link. Upon receiving a L2 trigger, FMIPv6 starts to 'anticipate' or prepare for the forthcoming handover before-hand. It assumes that the oAR is configured with a table containing the MAC addresses of its own and the neighboring Point-Of-Attachments (PoA) and the corresponding subnet prefixes of the neighbouring ARs. During the anticipation phase, the oAR assists in the NCoA formation by resolving subnet prefixes based on the table and the L2 identifier reported by the MN. There are three FMIPv6 signaling messages involved in the anticipation phase: Router Solicitation for Proxy Advertisement (RtSolPr), Proxy Router Advertisement (PrRtAdv) and Fast Binding Update (FBU). These messages are used for aiding IP movement detection and NCoA configuration. Through the RtsolPr and PrRtAdv messages, the MN formulates the NCoA when it is still present on oAR's link. Hence the latency due to new prefix discovery suffered by FMIPv6 is eliminated.

The MN could immediately use the prospective address after attaching to the nAR's link when the MN

has received a Fast Binding Acknowledgement (FBack) message prior to its mobility to the new link (i.e. Predictive Mode of operation).

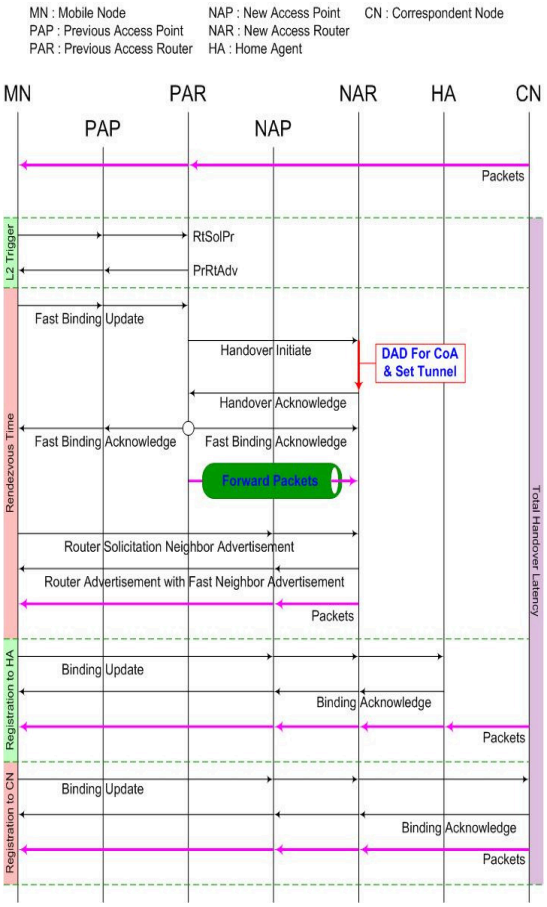


Figure 2. FMIPv6 Handover Procedure

This reduces long binding update latency suffered by FMIPv6. In the event, the MN moves upon receiving the FBack, and the NCoA would be usable only after it sends the Unsolicited Neighbor Advertisement (UNA) message and the FBU to the oAR from the nAR's link (i.e. Reactive Mode). In any case, the oAR stars tunnelling packets arriving for PCoA to NCoA. Such a tunnel is established by the Handover Initiate (HI) and Handover Acknowledgement (HAck) messages. Apart from that these message could also be used for the ARs to transfer resident contexts, such as access control, QoS, header compressions etc. There is no requirement of foreign Agents in Mobile IPv6. As mentioned previously, Neighbor Discovery and Address Auto-configuration features enable mobile nodes to function in any location without the services of any special router in that location.

There is no ingress filtering problem in Mobile IPv6 (In Mobile IPv4 this happens because the correspondent node puts its home address as the source address of the packet). In Mobile IPv6, the correspondent node puts the care-of address as the source address and having a Home Address Destination option, allows the use of the care-of address to be transparent over the IP layer. When using the FMIPv6 based mobility management, there are still rooms for improvement in reducing the handover latency and handover packet loss:

Neighboring access network discovery: The FMIPv6 doesn't address any radio access network discovery mechanism. Discovering the available PoAs by actively searching/scanning all the channels provided by the neighboring networks takes a considerable amount of time, which has a significant contribution to the overall handover latency.

Information exchange with neighboring ARs: The method by which neighbouring ARs exchange the information that enables the construction of PrRtAdv messages is not specified in the original FMIPv6. The IETF SEAMOBY WG produced the Candidate Access Router Discovery (CARD) protocol [2] to address this issue. The CARD protocol allows MNs to dynamically construct and populate their own CAR (Candidate Access Router) tables, which contain the mapping between the L2 PoA ID and corresponding IP addresses of the Candidate ARs. However, the use of CARD protocol so far is very limited for the reason that it will need additional support and upgrade to routers. Also, CARD enables a MN to gather attributes associated with target subnets so that a suitable AR could be selected for handover. But it does not provide the MN with appropriate L2 information in order to tackle the issue of radio access discovery that FMIPv6 faces. In [3], a mechanism to build (AP-ID, AR-Info) between ARs in hierarchical structure is described. However such mechanism is still very pre-mature and would require the ARs to be upgraded.

The Cost of Anticipation: In FMIPv6, the L2 handover is triggered by degrading link conditions. There is no guarantee that the MN will be connected to the oAR long enough to send and receive all FMIPv6 messages. When anticipation is used, the MN may not have sufficient time to update the oAR with the FBU. As a result, if the MN has already lost connection with oAR,

then the MN is forced to operate in the reactive mode and the handover latency will increase consequently.

The Ping Pong Movement: Time taken by the signaling exchange of the three FMIPv6 anticipation messages (RtSolPr, PrRtAdv and FBU) is long enough to increase the uncertainty of a MN's movements. For example, the handover may take place earlier than originally anticipated by the link layer. The border between overlapping cells may change dynamically due to the objects (e.g. buildings, trees etc) blocking the signals between the APs and MN. Due to the intrinsic dynamic nature of wireless channels, the MN may not move to the originally anticipated PoA. It may not move after all, or it may move somewhere else. That is to say that the MN would ping-pong between cells. Hence premature forwarding of data by the oAR (upon reception of an FBU) could be harmful because the MN may not move to the anticipated PoA. As a result there will be packet losses and long handover latencies.

C. Fast Handover and Triggers in FMIPv6

Mobile IP supports MN to maintain connectivity to the Internet during its handover from one AR to another. During handover operation Mobile IP involves movement detection, IP address configuration, and location update. This combined handover latency may preclude MN from real-time and throughput sensitive applications [3].

When MN moves from PAR to NAR, handover occurs. When packets are lost during the handover, they are retransmitted by higher-level protocols. If L2 connection between MN and previous AP (PAP) fails, then L3 connection between MN and PAR also fails. L2 and L3 handovers are started at this point. After L2 handover, MN receives Router Advertisement (RA) message or sends Router Solicitation (RS) message to NAR via new AP (NAP). Then MN requests new CoA to NAR and registers to HA. After authentication from HA, L3 handover is completed. Since router discovery delay is closely related to L2 system, an efficient L3 handover algorithm is necessary to reduce delay due to the route re-establishment.

3. Proposed Handover Scheme

In this section, we describe our proposed scheme to reduce the total latency and network load resulting from movement detection and tunneling for the fast handover in Fast Mobile IPv6 networks.

3.1. Handover procedure for FMIPv6 with reverse binding mechanism

The main goal of the proposed scheme is to incorporate the reverse binding mechanism to MIPv6 to enhance the fast handover procedure and to overcome such ineffectiveness by defining the signaling messages between NAR and HA in MIPv6 fast handover. The proposed scheme utilizes only pre-established bi-directional tunnels between NAR and HA.

The Fast MIPv6 handover using Reverse Binding Mechanism consists of the following messages.

(a). The mobile node (MN) sends a Router Solicitation for Proxy (RtSolPr) to find out about neighboring ARs.

(b). The MN receives a Proxy Router Advertisement (PrRtAdv) containing one or more [AP-ID, AR-Info] tuples.

(c). MN sends a Fast Binding Update (FBU) to the Previous Access Router (PAR).

(d). PAR sends a Handover Initiate (HI) message to the New Access Router (NAR), A duplicate address detection for Care of Address (CoA) within (a000ms) & set tunneling.

(e). NAR sends a Handover Acknowledge (HAck) message to the PAR.

(f).PAR sends a Fast Binding Acknowledgement message to the MS on the new link. The FBAck is also optionally sent on the previous link if the FBU was sent from there.

(g). MN sends Fast Neighbor Acknowledgement (FBAck) to the NAR after attaching to it.

(h). PAR starts packet forwarding message to NAR, and then the packet buffering start.

(i). MN sends a router solicitation with Fast Neighbor Advertisement (FNA) to NAR.

(j). After the NAR received the outer solicitation with Fast Neighbor Advertisement (FNA), it sends an outer Advertisement FNA to MN.

(k). NAR sends a tunneled Packet to MN.

(l). NAR sends a Enhanced PBU to the CN.

(m). CN received the Enhanced Packet and sends a Enhanced Packet Binding Acknowledgement to the CN.
(n). PAR sends a Fast Update message to HA.
(o). After HA received the Fast Update message, it sends a Fast Update message response to PAR.
(p). HA sends a Reverse Packet Binding Update to NAR.
(q). After NAR received the Reverse Packet Binding Update, it replies a Reverse Packet binding Acknowledgement to HA.
(t). HA sends the packet to MN.

3.2. Movement Detection

The primary aim of movement detection is to identify the L3 handovers. Movement detection generally uses Neighbor Unreachability Detection to determine when the default router is no longer bi-directionally reachable, in which case the mobile node must discover a new default router on a new link.

4. CONCLUSIONS

This paper has described Robust Fast handover scheme using Reverse Binding mechanism as new scheme applied in MIPv6 for supporting a fast handover effectively..

5. 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 (2010-0024401, 2011-0026286). Byungjoo Park is Correspondent author of this paper.

6. REFERENCES

[1] Davies R. W." **The Data Encryption standard in perspective,**"Computer Security and the Data Encryption Standard, pp. a29-a32.

[2] M. Liebsch, et.al, "**Candidate Access Router Discovery (CARD)**", RFC 4066, IETF, 2005.

[3] J.Zhang, H.Chen, Z. Xia, X. Duan, Z. Zhou, "**AR Information for FMIPv6**", IETF, June 2006

[4] Othman, R. 200a. **Antecedents and Outcome of IT use: How does HRM fit** in Asia Pacific Management Review 6(a) 9a-a04.

[5] Johnson, D., C. Perkins and J. Arkko. 2003. **Mobility Support in IPv6 (work-in progress)**. draft-ietfmobileip-ipv6-2a.txt.

[6] Dommety, G. 2003**. Fast Handovers for Mobile IPv6 (work-in progress)**. draft-ietf-mobileip-fastmipv6-06.txt.

Measuring the Refractive Index of Liquid and Gas by Mie Scattering

Li LIN
Electrical Engineering Department, School of Engineering and Applied Science, University of Pennsylvania
Philadelphia, PA 19104, USA

and

Li DU
College of Information Science and Engineering, Northeastern University
Shenyang, 110819, China

ABSTRACT

In this study I create a new method to measure the refractive index of liquid and gas by using Mie theory. The method is initiative though someone has used Mie scattering to measure the radius of small particles. This technique has the advantage of large range of measurement and tiny volume. I also provide two cases of the refractive index measurement devise based on this method in this paper.

Keywords: The refractive index, Mie theory, range of measurement, tiny volume and light distribution.

1. INTRODUCTION

The refractive index of liquid and gas is very important for industrial use. As the refractive index is one of the key indexes of liquid and gas, it is measured frequently in the laboratory when scientific researchers are testing or creating different kinds of solutions and gases.

Advancements in diffraction and interference technology have already made it possible for the accurate measurement of the refractive index. However, the scattering technology has not been utilized to measure this kind of index. Mie theory has been used to measure the refractive index of the particle itself [1], but it has been seldom used to find out the index of the liquid or air where the particles immersed in. So I employ the Mie theory which has been brought forward hundreds years ago to make it more useful in the liquid and gas index measurement. Besides, my design has little cost of fund and room, also the cost of the liquid and gas when they are expensive as this technology needs little sample to be test.

2. BACKGROUND

There are many other ways to measure the refractive index of the medium. Basically, they can be classified in three kinds [2]. I will introduce each of them in this section and present the basic knowledge of Mie theory later on.

Background Technology

(1)Geometric optical technology: This technology has critical angle method [3], imaging method, spectrometer method [4] and Abbe refractometer method. The main idea of the technology is to measure the angle caused by refraction or reflection and calculate the refractive index as the angle is relate to the refractive index of the medium which the light travel through. In this kind of technology, spectrometer has better resolving power but Abbe refractometer method is weak in this aspect.

(2)Wave optical technology: This technology has interference method and polarization method. The interference method based on the light length difference after the light travels through the medium, and people can judge the refractive index via counting the interference stripe. Polarization method is based on the principle that the polarization state will change when the light refract at the surface of the medium.

Nowadays many measurement systems use CCD to count interference stripe. For example, Shiqun Hua [5] used CCD to manipulate stripe based on their CCD automatic test system.

(3)Fiber optical sensors technology: This kind of technology is the most frequently used method nowadays for its high sensitiveness and the ease of operation. The most representative method to measure the refractive index is the fiber grating method [6] and the optical fiber refractive sensor based on surface plasmon resonance [7].

Fiber grating can be divided into Bragg grating and long-periodic grating. The fiber grating is very sensitive to the change of the refractive index outside. The external condition change will affect the radius and the refractive index of the fiber core and the clad. Thus, using the relations between the refractive index change outside and the amount of movement of the resonance hump, we can measure the refractive index outside accurately.

The most popular technique used to measure the refractive index is the fiber refractive sensor based on surface plasmon resonance. It can measure the refractive index of liquid with a high precision [8-9]. A plasma wave resonates effect is a physical chemistry

phenomenon happened on the interface between the metal and electrolytes, and it is very sensitive to the change of the refractive index of the medium surrounding them.

The Basic Knowledge of Mie Theory

People will draw into the dimensionless diameter index $\alpha=m_1\pi d/\lambda$, where m_1 is the refractive index of the medium surrounding the particle, λ is the wavelength of the incident light in vacuum, and d is the diameter of the particle.

Considering the condition of polarization, let the incident light go toward z direction, and electric vector goes toward x direction. The distance between the particle and the observation point of the scattering light is r. The scattering angle is represented by θ. The plan constituted by z axle and observation point is called scattering plane. The angle between the oscillating plane of the incident light and the scattering plane is φ. Now I_r represents the intensity of the light which vibrates perpendicular to the scattering plane; I_l represents the intensity of the light which vibrates parallel to the scattering plane; I_s represents the total scattering light intensity. Their expression is as follows [10]:

$$I_r = \frac{\lambda^2}{4\pi^2 r^2}|s_1(\theta)|^2 I_0 \sin^2\varphi = \frac{\lambda^2}{4\pi^2 r^2} i_1(\theta) I_0 \sin^2\varphi \quad (1)$$

$$I_l = \frac{\lambda^2}{4\pi^2 r^2}|s_2(\theta)|^2 I_0 \sin^2\varphi = \frac{\lambda^2}{4\pi^2 r^2} i_2(\theta) \mathrm{I}_0 \sin^2\varphi \quad (2)$$

$$I_0 = \frac{\lambda^2}{4\pi^2 r^2}\left[i_1(\theta)\sin^2\varphi + i_2(\theta)\cos^2\varphi\right] \quad (3)$$

For spherical particles, the scattering angle, the relative refractive index $m=m_2/m_1$ (m_2 is the refractive index of the particle itself) and the dimensionless index α which represents the particle radius can influent the intensity function $i_1(\theta)$, $i_2(\theta)$ and amplitude function $s_1(\theta)$, $s_2(\theta)$. But the azimuth angle φ has no effect on them. The infinite series constituted by Legendre function and Bessel function make up of the amplitude function.

$$s_1(\theta) = \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)}(a_n\pi_n + b_n\tau_n) \quad (4)$$

$$s_2(\theta) = \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)}(a_n\tau_n + b_n\pi_n) \quad (5)$$

a_n and b_n are called Mie scattering index, which is the function constituted by dimensionless index α and refractive index m; But τ_n and π_n is relative to scattering angle θ.

$$a_n = \frac{\psi_n(\alpha)\psi_n'(m\alpha) - m\psi_n'(\alpha)\psi_n(m\alpha)}{\xi_n(\alpha)\psi_n'(m\alpha) - m\xi_n'(\alpha)\psi_n(m\alpha)} \quad (6)$$

$$b_n = \frac{m\psi_n(\alpha)\psi_n'(m\alpha) - \psi_n'(\alpha)\psi_n(m\alpha)}{m\xi_n(\alpha)\psi_n'(m\alpha) - \xi_n'(\alpha)\psi_n(m\alpha)} \quad (7)$$

$$\pi_n = \frac{\mathrm{P}_n^{(1)}(\cos\theta)}{\sin\theta} = \frac{\mathrm{dP_n}(\cos\theta)}{d\cos\theta} \quad (8)$$

$$\tau_n = \frac{\mathrm{dP}_n^{(1)}(\cos\theta)}{d\theta} \quad (9)$$

$\xi_n(z)$ and $\Psi_n(z)$ are Ricatti-Bessel functions, and they belong to the functions of the second Hankel function and semi-integral order Bessel function.

$$\xi_n(z) = \left(\frac{\pi z}{2}\right)^{\frac{1}{2}} H^{(2)}_{n+\frac{1}{2}}(z) \quad (10)$$

$$\psi_n(z) = \left(\frac{\pi z}{2}\right)^{\frac{1}{2}} J_{n+\frac{1}{2}}(z) \quad (11)$$

$P^{(1)}_n(\cos\theta)$ and $P_n(\cos\theta)$ are the one-order associate Legendre function and Legendre function based on $\cos\theta$.

When the incident light is the natural light, the scattering light is partial polarized. Like former formulas, the scattering intensity is:

$$I_r = \frac{\lambda^2}{8\pi^2 r^2}\mathrm{i}_1(\theta) I_0 \quad (12)$$

$$I_l = \frac{\lambda^2}{8\pi^2 \mathrm{r}^2}\mathrm{i}_2(\theta) I_0 \quad (13)$$

Thus we can get the polarization degree P as Eq. (14) and the total scattering intensity I_s as Eq. (15):

$$P = \frac{i_1(\theta) - i_2(\theta)}{i_1(\theta) + i_2(\theta)} \quad (14)$$

$$I_s = \frac{\lambda^2}{8\pi^2 r^2}\left[i_1(\theta) + i_2(\theta)\right] I_0 \quad (15)$$

For polarized light and the natural light we have:

$$k_{sca} = \frac{C_{sca}}{\pi a^2} = \frac{2}{a^2}\sum_{n=1}^{\infty}(2n+1)\left(|a_n|^2 + |b_n|^2\right) \quad (16)$$

$$k_{ext} = \frac{4}{a^2}\mathrm{Re}\left[s(\theta)\right] = \frac{2}{a^2}\sum_{n=1}^{\infty}(2n+1)\,\mathrm{Re}(a_n + b_n) \quad (17)$$

$$k_{abs} = k_{ext} - k_{sca} \quad (18)$$

3. FEASIBILITY ANALYSIS

To measure the refractive index of liquid and gas, we need scatterer which is illuminated by the incident light. The scatterer may be several particles with the radius between 10nm to 10μm. I will choose the case of one small particle in this paper. The scatterer is located in a cavity with two tubes used as entrance and exit for the liquid and gas. The tubes are curving or zigzag to make sure that the light outside will not come in the cavity. There is one or more light intensity detectors distributed surround the scatterer to detect the scattering light intensity in the setting direction. The inwall of the cavity is paint with material which has high absorptivity of the incident light.

Now let's concentrate on the measurement of the refractive index of liquids, and the measurement of the refractive index of gases is similar to it. As the range of the refractive index of liquids spreads lager than the gases', the test on liquids is more representative. The refractive indexes of gases vary from 1.000132 to 1.001711, and the indexes of the liquids we have ever got are distributed between 1.31 and 2.10. The refractive indexes of the most familiar liquids in daily life are distributed between 1.33 and 1.36. We need to set the radius of the scatterer and the location of the light intensity detectors, so when the refractive index changes little, the variability of the light intensity which detectors can detect is distinguishable for detectors.

Based on the functions in Section 2.2, I compile a program to calculate the scattering light distribution after

the incident light travel through a spherical particle [11]. Ground on this, A program is compiled by Matlab in order to set the location of the detector and the radius of the particle in the case that only one small particle is used. The program is attached in the appendix.

In the program, the wavelength of the incident light is set at 1.31μm, and the incident light is same each time. I divide the whole work into several groups. In each group, I set the radius of the particle at one specific value. Then I locate detectors at the position where the scatter angle is 0.1π, 0.2π, 0.3π, 0.4π, 0.5π, 0.6π, 0.7π, 0.8π, 0.9π and 1.0π. After that, I change the refractive index of the medium surround the particle from 1.33 to 1.36 with the separation of 0.005. Finally, I collect the variance of the light intensity detected by the detector at each scatter angle. The larger the variance, the easier detectors can distinguish the change of the refractive index, which means the survey meter is more sensitive.

After collect the data of the first group, I change the radius of the particle and collect the data of the second group like the first step. The simulation data after running the program is listed as below. The variance with specific radius and the scattering angle is listed in Table 1.

From the statistical table we can summarize that generally, the variance of the scattering light at forward direction is increasing along with the increasing of the radius of the particle. This is because when the radius is increasing, the scattering light will concentrate on the forward direction. Thus, the base is larger, and the change of the refractive index will bring larger difference between the light intensity.

The conclusion that the scattering light concentrate on the forward direction along with the increasing of the radius is demonstrated by running the program calculating the light distribution after Mie scattering. For instance, Fig.1 is the statistical chart of the ratio of the backward energy of Mie scattering under the condition that the relative refractive index m ($n_{particle}/n_{surrounding\ medium}$) equal to 0.9.

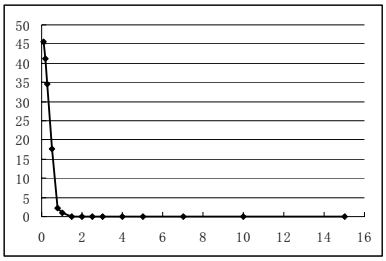


Fig.1. The ration of the backward energy changes along with diameter when m=0.9

So in general, choosing larger spherical particle will enhance the sensitivity of the refractive survey meter. Furthermore, locating the light intensity detector at the forward direction will make it detect more energy when the refractive index of the medium changes. However, if enlarge the radius without any limitation, the spherical particle will be too large and range out of the confine of Mie theory.

For example, If only one light intensity detector is used in the cavity, setting the radius at 4.0μm, and the light intensity detector is located at the scattering angle of 0.1π; if two light intensity detectors are used, setting the radius at 3.0μm, and light intensity detectors are located at the scattering angle of 0.1π and 0.2π separately.

Table 1 The variance with specific radius and the scattering angle

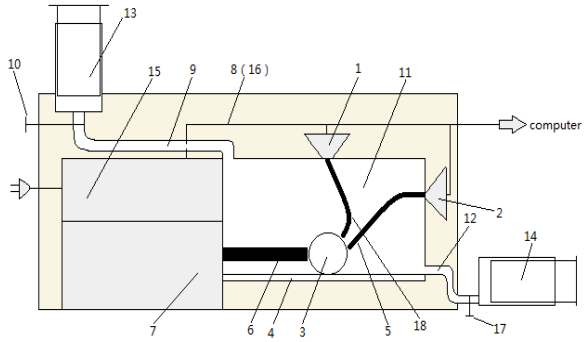
Radius/ Angle	**0.1π**	**0.2π**	**0.3π**	**0.4π**	**0.5π**	**0.6π**	**0.7π**	**0.8π**	**0.9π**	**1.0π**
0.3μm ($\times10^{-6}$)	0.1988	0.1509	0.0739	0.0130	0.0015	0.0215	0.0407	0.0477	0.0474	0.1577
0.4μm ($\times10^{-5}$)	0.1300	0.0692	0.0168	0.0000	0.0068	0.0143	0.0144	0.0105	0.0070	0.0616
0.5μm ($\times10^{-6}$)	0.0836	0.0251	0.0078	0.0806	0.0901	0.0265	0.0014	0.0507	0.1261	0.0522
0.6μm ($\times10^{-6}$)	0.0000	0.0653	0.1606	0.0742	0.0004	0.0237	0.0420	0.0388	0.0339	0.0434
0.7μm ($\times10^{-6}$)	0.9380	0.0740	0.2950	0.0072	0.0439	0.0163	0.0042	0.0261	0.0237	0.0007
0.8μm ($\times10^{-6}$)	0.5732	0.5277	0.1963	0.0405	0.0312	0.0025	0.0065	0.0028	0.0430	0.0088
0.9μm ($\times10^{-5}$)	0.0388	0.1295	0.0016	0.0095	0.0001	0.0002	0.0014	0.0021	0.0012	0.0001
1.0μm ($\times10^{-5}$)	0.0703	0.1808	0.0079	0.0053	0.0003	0.0015	0.0010	0.0004	0.0013	0.0000
1.1μm ($\times10^{-5}$)	0.0297	0.1634	0.0333	0.0000	0.0005	0.0006	0.0004	0.0000	0.0012	0.0001
1.2μm ($\times10^{-6}$)	0.0022	0.9268	0.4207	0.0051	0.0301	0.0019	0.0015	0.0201	0.0025	0.0001
1.3μm ($\times10^{-6}$)	0.0951	0.1346	0.2388	0.0038	0.0059	0.0000	0.0142	0.0000	0.0074	0.0002
1.5μm ($\times10^{-5}$)	0.3369	0.1920	0.0006	0.0072	0.0000	0.0006	0.0001	0.0019	0.0003	0.0000
1.8μm ($\times10^{-4}$)	0.1675	0.0999	0.0038	0.0000	0.0000	0.0002	0.0000	0.0001	0.0001	0.0000
2.0μm ($\times10^{-4}$)	0.1841	0.0715	0.0087	0.0021	0.0003	0.0000	0.0003	0.0001	0.0000	0.0000
3.0μm ($\times10^{-4}$)	0.7356	0.6589	0.0049	0.0003	0.0000	0.0000	0.0001	0.0001	0.0003	0.0000
4.0μm ($\times10^{-4}$)	0.4223	0.1461	0.0235	0.0026	0.0000	0.0020	0.0003	0.0000	0.0000	0.0000
5.0μm ($\times10^{-3}$)	0.5684	0.3591	0.0254	0.0000	0.0002	0.0001	0.0001	0.0002	0.0003	0.0000

4. EXAMPLES OF THE REFRACTIVE SRUVEY METER

The light source output laser at 1.31μm. The light travel through the incident fiber and the order of the core diameter of incident fiber is 1μm. There are narrow gaps between the ends of the fibers and the particle, and the order of the width of the gaps is 1nm. The particle is fixed on the fixed tray by cauterization or pasting and its radius is 3μm. The scattering light is detected by two detected fibers at the scattering angle of 0.1π and 0.2π. The test preparation is contained in the vessel and will inpour into the measurement cavity when the plunger is open. After measurement, open the plunger of the vessel containing outflow liquid (gas) and the testing material

will be extracted. The measurement cavity need to be washed by distilled water each time after measurement. In addition, each time before measurement, close the upper plunger and open the lower one, extract the air in the measurement cavity to approximative vacuum.

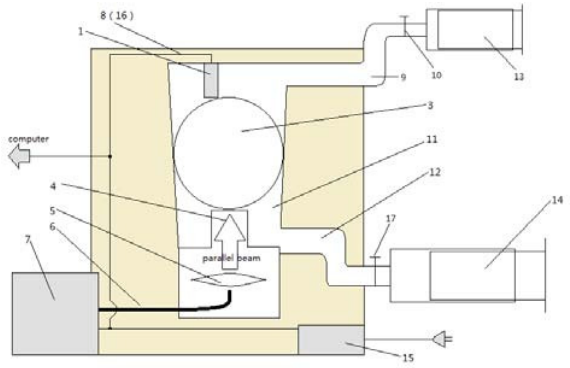
Case 1:



1, 2—light intensity detector; 3—particle; 4—fixed tray; 5, 18—detected fiber; 6—incident fiber; 7—light source; 8—data wire; 9—tube for infused liquid(gas); 10, 17—plunger; 11—measurement cavity; 13—vessel contain infused liquid(gas); 14—vessel contain outflow liquid(gas); 15—energy source; 16—power line

Fig.2. The schematic diagram of Case 1

Case 2:



1—light intensity detector; 4—incident light emitting surface; 5—lens; 6—incident fiber; 7—light source; 8—data wire; 9—tube for infused liquid(gas); 10, 17—plunger; 11—measurement cavity; 13—vessel contain infused liquid(gas); 14—vessel contain outflow liquid(gas); 15—energy source; 16—power line

Fig.3. The schematic diagram of Case 2

Before testing the unknown refractive index of medium, we need to use the liquids and gas which the refractive indexes are known to test the relation schema of the detected light intensity and the refractive index of medium. In the case of this example, a three-dimensional schema should be made. Two axises represent the light intensity detected by two detectors, and the third axis represents the refractive index of the medium. Using mediums which the refractive indexes are known, we can depict the relation schema by computer. When the unknown refractive index is measured, the energy detected by the two detectors will set the value of the first two axises of the relation schema and we can easily find out the value of the refractive index.

The difference between the Case 1 and Case 2 is that only one light intensity detector is used here. The particle has the radius of 4μm and is fixed by four slopes of the measurement cavity. The lens is used to assemble the incident light into parallel beam. The other elements act similarly as the first case.

5. CONCLUSION

The refractive index meter based on the Mie theory is initiative and has many advantages. It has a large range of measurement and can detect the refractive index of liquids and gases. Its volume is little and can be integrated on other devices. The operation is simple and the cost is low. The scattering light distribution changes with the change of the refractive index of the medium surround the scatterer, thus detect the change of light distribution will lead us to find out the refractive index of the medium. Optimize the size of the particle and the location of the light intensity detector will increase the sensitivity of the survey meter. Two examples are given to illustrate the technique.

REFERENCES

[1] P. Chylek, V. Ramaswamy and A. Ashkin, “Simultaneous determination of refractive index and size of spherical dielectric particles from light scattering data”, **Applied Optics**, vol. 22, no. 15, pp. 2302-2307.

[2] Shigang Yu. “The analysis of the method to detect the refractive index of liquid”, **Optic instrument**, vol. 29, no. 4, Aug. 2007.

[3] Guoguang Mu, Yuanling Zhan. **Optics**, Beijing: People Education Press, 1978.

[4] Zhaokui Zhang, Lianyuan Miu, Li Zhang. **University physical experiment (Second edition)**, Beijing: People Education Press, 2001, pp.91-100.

[5] Wanghua Xiang, Bing Zhang. “Cointersection spherical surface F-P interferometer measure the refractive index and density of liquid experimental study”, **Academic Journal of Tianjin University**, vol. 40, no. 1, 2007, pp. 83-87.

[6] Hong-Wei Chen, Chuen-Lin Tien, Wen Fung Liu. “The Measurement of Liquid Refractive Index by D-shaped Fiber Bragg Grating”, **2007 IEEE/LEOS International Conference,** Aug. 12, 2007, pp.119-120.

[7] Matthisa N, Andreas K. “Intrinsic fiber optical gas sensor based on surface Plasmon resonance spectroscopy”, **SPIE**, 1995, pp.303-311.

[8] Matsubarak K, Kaswata S, Minami S. “Optical chemical sensor based on surface Plasmon measurement”, **Appl Opt**, vol.27, no.6, 1988, pp.1160-1163.

[9] Jorgenson R C, Yee S S. “A fiber-optic chemical sensor based on surface Plasmon rensonance”, **Sensors and Actuators B**, vol.12, no.3, 1993, pp. 213-220.

[10] Xiaoshu Cai, Mingxu Su and Jianqi Shen. **The technology and the application of the measurement of the granularity of particles**, Beijing: Chemical Industry Press, 2010, pp. 21-40.

[11] Heyong Zhang, Weijiang Zhao and Deming Ren. "Modified Matlab algorithm of Mie scattering index of spherical particle," **The Journal of Light Scattering**, vol. 20, no. 2, Jun. 2008.

Seismic Analysis of 3-D Building Frame with Masonry Infill and Steel Bracings as Lateral Load Resisting Systems.

Venkatesh S.V
Assistant Professor, Faculty of Civil Engineering, PESIT,
Bangalore 560 085, Karnataka, India

and

Dr. Sharada Bai H.
Professor, Faculty of Engineering - Civil, UVCE, Bangalore University,
Bangalore 560 056, Karnataka, India

and

Divya S.P
Former Student, Faculty of Engineering - Civil, Bangalore University,
Bangalore 560 056, Karnataka, India

ABSTRACT

Lateral Load Resisting System (LLRS) which will supplement the behavior of moment resisting frames is important for high rise structures to resist the lateral load. Natural phenomenon like earthquake causes damage to or collapse of buildings if not designed for lateral loads resulting due to Earthquake. Shear walls, Infill frames and Steel bracings are some of the LLRS commonly used. An attempt is made in the present study, to analyze the structural behavior of 3-dimensional (3D) single-bay-four-bays, 10 storey basic moment resisting RC frames when provided with Masonry infill and Steel bracings as LLRS. The detailed investigations are carried out for zone V of Seismic zone of India, considering primary loads and their combinations. Four models are analyzed consisting of one basic moment resisting RC frame and other three include basic moment resisting RC frame with Masonry Infill (idealized as diagonal compressive strut), External and Internal Steel Bracings. The results obtained from the linear dynamic analysis are thoroughly investigated for maximum values of joint displacements, support reactions, column forces, beam forces, forces in truss elements and dynamic properties. Better resistance to lateral load of the frames in the presence of LLRS is observed from the results obtained.

Keywords: RC frame, Earthquake, Retrofit, LLRS, Masonry Infill, Steel bracings.

1. INTRODUCTION

Lack of seismic strength and detailing requirements as per Indian standard codes of practice in many existing buildings necessitate upgrading, if the structure was initially not designed and constructed to resist an earthquake i.e. designed only for gravity loads but still has not undergone failure. For structures, which have undergone failure due to earthquake, it is essential to retrofit for future use.

2. PRESENT INVESTIGATION

The present investigation is concerned with detailed 3D study of results of analysis of a ten storey Moment Resisting Frame having single bay along X and four bays along Z subjected to gravity load, seismic load and their combinations.

Frames considered are:
(a) BF: Moment Resisting Bare Frame (without any special LLRS feature) (Fig.1a & 1b)
(b) MDS: Moment Resisting Frame provided with masonry infill (idealized as diagonal compressive strut). (Fig.2a & 2b)
(c) ESB: Moment Resisting Frame provided with External Steel Bracing (Fig.3a & 3b)
(d) ISB: Moment Resisting Frame provided with Internal Steel Bracing (Fig.4a & 4b)

Only exterior frames are provided with masonry infill and two types of steel bracings.

Method of Analysis

The present study undertaken deals with Linear Dynamic analysis i.e Response Spectrum Method.

Modeling of the Structure: For the present 3D study STAAD.Pro software package is used.

Details of the problem chosen:

Plan and Height of the Bare Frame: The plan consists of Single bay of span 7.5 m along X direction, four bays of span 3.0 m each along Z direction. The typical
Ten-Storey building has each storey height of 3.0 m along Y direction.

Beam Cross-Sections:

Plinth Beam Size:
B1 (Primary Beam) (ZxY) :300mmX450mm
B2 (Secondary Beam) (XxY) :300mmX300mm

Floor Beam Size:
B1 (Primary Beam) (ZxY) :300mmX750mm
B2 (Secondary Beam) (XxY) :300mmX375mm

Column Size: Square column:
636mmX 636mm.

Masonry wall thickness **:**230mm

Steel Section **:**ISLC 225

Frames with Special Feature of LLRS:
1. Frames with External Steel Bracing (ESB)
2. Frames with Internal Steel Bracing (ISB)
3. Frames with Masonry infill (MDS) provided at end bays along X and Z directions.

Seismic Zone: Zone V as per IS code [1] for which zone factor (Z) is 0.36.

Types of Primary Loads and Load Combinations: The structural systems are subjected to three types of Primary Load Cases as per IS code [2], they are:
1. Dead Load case (Gravity load), "DL"
2. Live Load case (Gravity load), "LL"
3. Seismic (Lateral) Load along X, "ELx"
4. Seismic (Lateral) Load along Z, "ELz".

In addition, the structural systems are subjected to 13 different Load Combinations, they are:

5.1.5(DL + LL)
6.1.2(DL + LL + ELx)
7.1.2(DL + LL - ELx)
8.1.2(DL + LL + ELz)
9.1.2(DL + LL - ELz)
10.1.5(DL + ELx)
11.1.5(DL - ELx)
12. 1.5(DL + ELz)
13. 1.5(DL - ELz)
14. (0.9DL + 1.5ELx)
15. (0.9DL - 1.5ELx)
16. (0.9DL + 1.5ELz)
17. (0.9DL - 1.5ELz)

Material Properties:

Density of brick wall	: 18.85 kN/m^3
E of brick wall	: 4.2x10^6 kN/m^2
Poisson's Ratio of brick wall	: 0.15
Density of R.C.C	: 25 kN/m^3
E of concrete	: 2.17x10^7 kN/m^2
Poisson's Ratio of concrete	: 0.17

The dead load consists of self-weight of structural elements and masonry wall load of thickness 230 mm (The lateral load resistance effect of infill wall is not considered for analysis). Superimposed load on slab is 2.875kN/m^2.
The live load considered is as adopted for medium office, hospital or hostel building i.e. 4 kN/m^2 as per IS code [2]. The Response Spectrum Method of analysis is adopted for the calculation of the lateral load at each floor level as per IS code [1]. The lateral loads applied are given in Table 1.

Table 1: Joints Load at each storey in kN.					
Level	Bare Frame/Steel bracing		MDS		
	End- frame	Mid-frame	End-Frame	Mid-Frame - adjacent to corner nodes	Mid-frame
Plinth	26.25	33.75	10.826	30.795	33.75
Floor	79.92	122.34	49.308	116.482	122.34
Terrace	53.97	72.69	38.664	69.759	72.69

3. RESULTS AND DISCUSSIONS

The results obtained are observed and the maximum values obtained among all the load cases and load combinations (L/C) considered are presented in Table 2, along with the corresponding load case. The Table indicates the results of frames with LLRS considered (i.e. and ESB, ISB MDS) and for the moment resisting Bare Frame (BF). The discussions focus on the comparison between the three LLRS considered and the basic Bare Frame with respect to Maximum Joint Displacements, Maximum Support Reaction, Maximum Column Forces, Maximum Beam Forces, Maximum axial forces and axial stresses in truss elements and dynamic parameters.

Maximum Joint Displacements

For all the structural systems considered, the maximum joint displacement is observed at the top storey level (Lateral sway in X and Z directions) as expected, the Bare Frame (without any LLRS) undergoes the maximum joint displacement namely Max X, Max Z & Max Ab.

Effect of load and load combinations: For the structural systems considered, load combinations for which Max X and Max Z occur are load case 10 or 11 i.e. 1.5(DL ± ELx) and 12 or 13 i.e. 1.5(DL ± ELz) respectively, whereas Max Ab occurs for load combination 13 i.e. 1.5(DL - ELz) except in case of MDS it occurs for load case 16 i.e.0.9DL+1.5ELz.

Effect of LLRS: The value of Max X reduces by 47%, 25% and 93% in MDS, ESB and ISB respectively when compared with bare frame. It is also interesting to note that introduction of masonry infill as LLRS reduces the Max Z and Max Ab by 51%. The values of Max Z and Max Ab reduce by 32% in ESB and 93%and 88% in ISB respectively when compared with the bare frame.

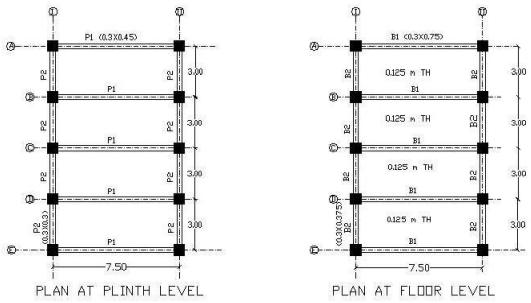


Fig 1a: BF-Moment Resisting Bare Frame

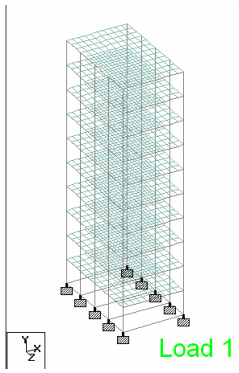


Fig 1b: Basic 3D frame

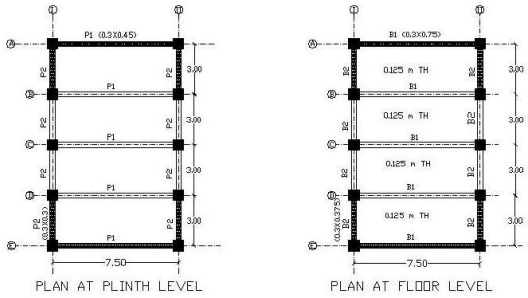


Fig 2a: MDS-Masonry infill with Moment Resisting Bare Frame

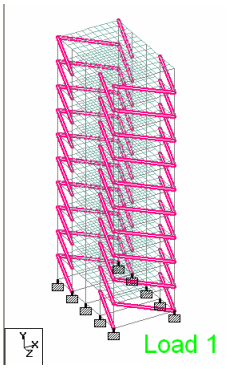


Fig 2b: 3D MDS frame

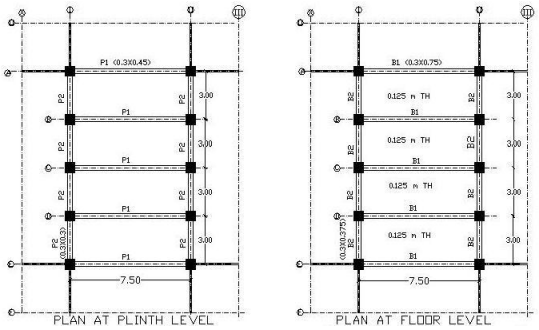


Fig 3a: ESB-External Steel Bracing with Moment Resisting Bare Frame

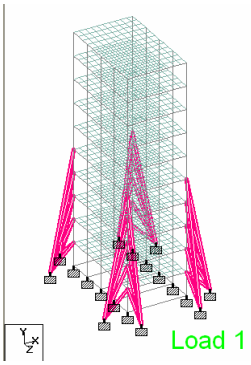


Fig 3b: 3D ESB frame

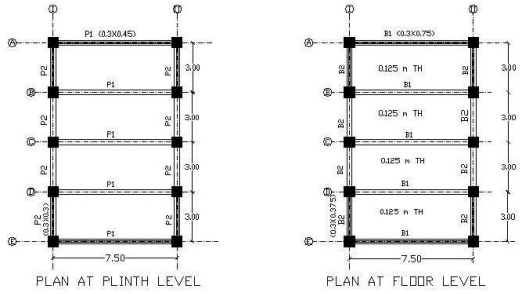


Fig 4a: ISB-Internal Steel Bracing with Moment Resisting Bare Frame

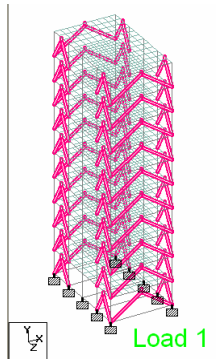


Fig 4b: 3D ISB frame

Maximum Support Reactions

Effect of load and load combinations: The maximum support reaction Fx occurs when seismic load combination 10 i.e. 1.5 (DL + Elx) is applied, except in case of ISB and MDS where it occurs when load case 5 i.e. 1.5(DL+LL) and 14 i.e. 0.9DL+1.5ELx is applied respectively. The Maximum Support reaction Fy occurs when non-seismic load case 5 i.e. 1.5(DL+LL) is applied for ESB and ISB except in case of bare frame and MDS where it occurs when load case 10 i.e.1.5 (DL + Elx) and 12 i.e. 1.5(DL+ELz) is applied.

The maximum support reaction Fz occurs when seismic load combination 12 or 13 i.e.

1.5 (DL ± Elz) is applied except in case of MDS where it occurs when load case 16 i.e.0.9DL+1.5ELz is applied. The maximum support moment Mx and Mz occur in seismic load combinations.

Effect of LLRS: The maximum support reactions Fx and Fz increase in case of ESB and

MDS whereas it reduces by 43% and 89% respectively in ISB. The maximum support reaction Mx reduces by 39%, 77% and 93% for MDS, ESB and ISB respectively. Also the maximum support reaction Mz reduces by 40%, 62% and 79% respectively for frames with LLRS considered.

Maximum Forces in Columns

Generally the maximum column forces Fx, Fy, Fz, My and Mz occur in seismic load combinations. The maximum column forces Fx, Fy, Fz, My and Mz reduces in all the LLRS considered when compared with bare frame whereas the column force Fx increase in MDS.

Table 2- Magnitude of the parameters considered and corresponding Load Case.

Parameter	Notations	Bare Frame		MDS		ESB		ISB	
		Magnitude	L/C	Magnitude	L/C	Magnitude	L/C	Magnitude	L/C
Storey Sway [mm]	Max-X	55.30	10	29.19	14	41.19	10	-3.950	11
	Max- Z	80.71	12	39.36	16	-55.25	13	-5.81	13
	Max- Ab	81.04	13	39.43	16	55.71	13	9.59	13
Support Reaction [kN, m]	Max Fx	202.88	10	394.82	14	449.38	10	-115.82	5
	Max Fy	3635.83	10	4078.08	12	3558.48	5	3510.25	5
	Max Fz	169.84	12	245.83	16	-489.79	13	-17.89	13
	Max Mx	505.72	12	307.05	12	-115.07	17	-34.62	13
	Max Mz	-473.02	11	281.94	10	178.03	10	98.56	10
Column Forces [kN, m]	Max Fx	3505.47	10	3818.06	12	3428.13	5	3379.89	5
	Max Fy	231.91	10	155.24	10	-179.23	11	133.40	10
	Max Fz	181.09	12	118.70	12	139.75	12	76.10	12
	Max My	505.72	12	253.22	16	220.11	12	115.79	12
	Max Mz	457.52	10	291.48	10	277.90	10	-216.52	11
Beam Forces [kN, m]	Axial force	66.20	5	356.98	10	68.91	10	-118.46	11
	Shear Y	268.84	10	221.43	5	241.71	10	219.46	5
	Max Mx	-70.45	13	37.34	12	56.33	12	31.86	8
	Max Mz	592.91	10	402.01	10	-492.07	11	-287.17	7
Truss Forces	Axial force [kN]	-	-	499.05	10	316.41	12	268.81	12
	Axial stress [MPa]	-	-	2.67	12	103.64	12	88.05	12

Table 3 : Dynamic Properties

Dynamic Parameter	Bare frame	MDS	ESB	ISB
Fundamental period in seconds	1.58	1.09	1.25	0.24
Base shear Vb along X in kN	658.81	904.89	751.12	1517.69
Base shear Vb along Z in kN	543.42	784.88	650.79	1576.85
Multiplying factor Vb/VB along X	1.57	1.15	1.39	-
Multiplying factor Vb/VB along Z	1.91	1.32	1.60	-

Maximum Forces in Beams

The maximum beam forces Fx, Fy, Mx and Mz occur in seismic load combinations. The greatest value of the maximum Shear force Fy, Torsion moment Mx and Bending moment Mz in beams, of all the structural systems considered, occur in bare frame. Comparing the frames with LLRS with bare frame, the maximum axial force Fx increases. The

maximum shear force Fy, moment Mx and Mz decrease for all LLRS considered.

Forces in Truss Elements

The forces considered are maximum Axial Force and Axial Stress in Masonry infill (idealized as compressive strut) and Steel Bracings. The maximum Axial force and axial stress for ESB, ISB occur when load case 12 i.e. 1.5(DL + ELz) is applied and for MDS, maximum axial force occur when load case 10 i.e. 1.5(DL + ELx) are applied respectively. The maximum axial stress values as given in Table 2 in MDS, ESB and ISB are less than the permissible stresses in masonry and steel as applicable.

Dynamic Properties

The fundamental period as obtained by software for the models considered are given in Table 3 whereas the calculated value as per IS 1893 (part I) 2002 is 0.96. It can be observed that the fundamental period obtained do not match with the values obtained from the formula given in IS code 1893 (part I) 2002. This needs to be considered and revised as the fundamental period does not depend only on the height, but also depends on the structural configuration.

The base shear along X and Z axis are presented in Table 3 for the models considered. The base shear values are different along X and Z direction for a particular type of model. The base shear (Vb) calculated using the fundamental period Ta as given in IS 1893(part I) : 2002, are compared with design base shear (VB). If VB is less than Vb then, all the response quantities for example, displacements, storey forces, member forces, storey shear and base reactions shall be multiplied by Vb/ VB. It can be observed that in case of all the models considered the value of multiplying factor (Vb/VB) is greater than 1 which indicates that the values of equivalent static force method (ESLM) are greater than Response Spectrum method (RSM). The maximum Vb/VB ratio is observed in case of bare frame along Z direction.

4. CONCLUSIONS

i. It is necessary to consider gravity and seismic loads as well as all the load combinations during analysis of the structure. Provision of MDS ,ESB and ISB effectively reduce large displacements found in bare frame.
ii. The best performing LLRS among the three LLRS considered is ISB as all the parameters considered reduce in this case when compared with bare frame.
iii. When these LLRS considered in the study are employed in field for upgrading or retrofitting a structure, it is necessary to ensure proper connections between existing structure and LLRS provided.

5. ACKNOWLEDGEMENTS

The authors wish to thank the Authorities of PESIT and Bangalore University for their support.

6. REFERENCES

[1] IS: 1893 (Part–1)–2002 - Code of Practice for Criteria for Earthquake Resistant Design of Structures, Part 1: General Provisions and Buildings, Bureau of Indian Standards, New Delhi.

[2] IS: 875 – 1987 - Code of Practice for Design Loads (other than Earthquake) for Buildings and Structures, Part 1: Dead Loads, Part 2: Imposed Loads, Part 5: Special Loads and Load Combinations, Bureau of Indian Standards, New Delhi.

[3] Ghobarah A, Abou Elfath H, "Rehabilitation of a reinforced concrete frame using eccentric steel bracing", **Engineering structures** Vol.23, 2001, pp. 745 - 755.

[4] S. V. Venkatesh, H. Sharada Bai, R. Rajeshwari , "Effect of Shear wall and Steel bracing as lateral load resisting system on a ten storey building", **The Fourth International Conference on Structural Engineering, Mechanics and Computation (SEMC 2010)**, University of Cape Town, Cape Town, South Africa.

[5].NEHRP guidelines for the seismic rehabilitation of building, **FEMA 273 /** October 1997

A Study of Thermal Performance of Contemporary Technology-Rich Educational Spaces

Sarah ELMASRY, Ahmed HASSAN and Mahmoud HAGGAG
Dept. of Architectural Engineering, United Arab Emirates University
Al-Ain, P.O.Box 17555, United Arab Emirates

ABSTRACT

One of the most dominant features of a classroom space is its high occupancy, which results in high internal heat gain (approximately 5 KW) [1]. Furthermore, installation of educational technologies, such as smart boards, projectors and computers in the spaces increases potential internal heat gain. Previous studies on office buildings indicate that with the introduction of IT equipment in spaces during the last decade, cooling load demands are increasing with an associated increase in summer electrical demand [5][6][7].

Due to the fact that educational technologies in specific correspond to pedagogical practices within the space, a lot of variations due to occupancy patterns occur [3]. Also, thermal loads caused by educational technologies are expected to be dependent on spatial configuration, for example, position with respect to the external walls, lighting equipment, mobility of devices [9].

This study explores the thermal impact of educational technologies in 2 typical educational spaces in a facility of higher education; the classroom and the computer lab. The results indicate that a heat gain ranging between 0.06 and 0.095 KWh/m^2 is generated in the rooms when educational technologies are in use.

The second phase of this study is ongoing, and investigates thermal zones within the rooms due to distribution of educational technologies. Through simulation of thermal performance of the rooms, alternative room configurations are thus recommended in response to the observed thermal zones.

Keywords: Learning Technologies, Performance-based Design, Thermal Performance, Learning Spaces, Heat Gain.

1.INTRODUCTION

Occupied with students and equipped with educational technologies, classrooms and laboratories are considered the basic unit for thermal comfort and energy consumption in educational buildings. Equipment in these spaces become significant sources of sensible and latent heat, which need to be lowered in order to reduce the energy consumption in the facilities. Existing literature and findings for low energy consumption in other building types (for example residential or office buildings) thus may not apply to educational spaces, where occupancy patterns are different, as well as the nature and patterns of use of technologies, and accordingly, internal heat gains.

Previous research has only emphasized the importance of high performance design of a school classroom disregarding the role of educational technologies as relates to thermal performance of the spaces [1][2][4][8]. Studies conclude that through enhanced design of building components such as walls and openings, school buildings might consume 180 to 80 kW h/m2 for heating, cooling, ventilation and lighting annually in hot-humid climates. This means a reduction of more than 50% of the energy consumption than typical designs[8]. These results match those found in high performance schools in Europe and US, with 55–75% savings in heating and 30–40% savings in electricity when several technologies were implemented as part of a holistic approach aiming at high energy savings and accepting medium and long payback times [8]. There is no study found in the literature which investigates the thermal impact of using learning technologies in contemporary learning environments. This study aims to bridge this gap. Two typical room configurations have been selected for investigation; a classroom, and a computer lab. The rooms are diverse in size and proportion, layout, and equipped with a variation of learning technologies.

2.EXPERIMENTAL SETUP

Two rooms were selected as a sample in a facility of higher education in the city of Al Ain, United Arab Emirates. The building typically lies in the hot arid climate. Measurements were taken in the months of November and December, where the ambient outdoor temperatures typically lie between (---). During this time of the year the facility is at its highest occupancy level and usage rates of educational technologies. However, temperature measurements were taken when the room not occupied, the results thus represent only the thermal impact of educational technologies on the space. Also, air conditioning was shut down whenever measurements were taken in the rooms. However, it should be noted that thermal conditions in both rooms vary during the day with the variation in solar radiation which they are exposed to from the east oriented facades.

The rooms represent the typical configurations of a classroom (Figure 1), and a computer lab (Figure 2). The first is equipped with a smart board system located at the front learning wall. The second is equipped with 22 desktops; 10 on two perimeter walls, as well as 12 in the center. The learning wall at the computer lab constitutes of a smart board system, and the front teaching zone is also equipped with a desktop and a printer (Figure 4). While the classroom is only equipped with a smart board mounted on the front learning wall (Figure 3).

Temperature measurements in both rooms were recorded on a central axe starting from the learning wall to the opposite external façade of the room (which in both rooms was facing east) as illustrated in Figure 3 and Figure 4. The hight of the temperatures sensors were set at the occupants' level and accordingly, were fixed on the lower surfaces of the desks.

Figure 1: Spatial Configuration of Classroom Sample

The measurement points were evenly distributed across the room on a central axe. Measurements were taken at equal intervals of 1.5 m starting from the point below the heat source (smart board projector) and along the measurement axe towards the east external walls.

Measurements in both rooms were recorded between 7:00 a.m.-7:00 p.m. daily; typically the hours in which the facility is occupied.

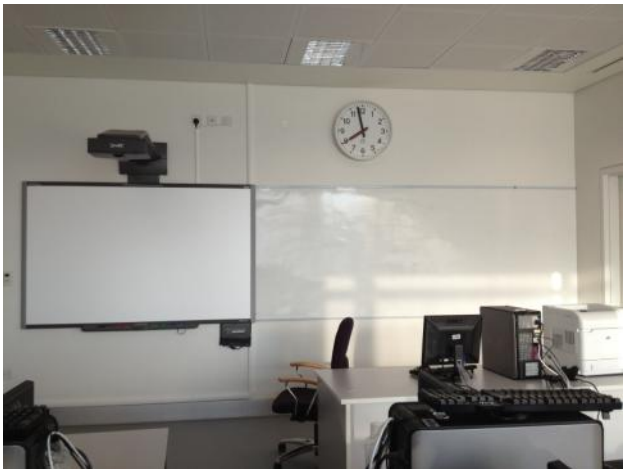


Figure 2: Spatial Configuration of Computer-Lab Sample

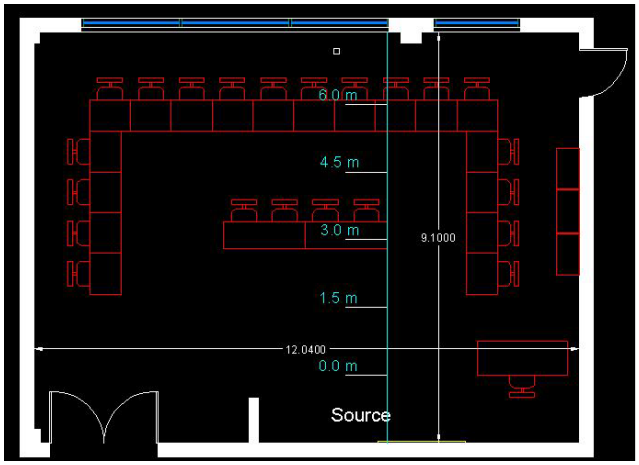


Figure 3: Measurement Points Across the Classroom

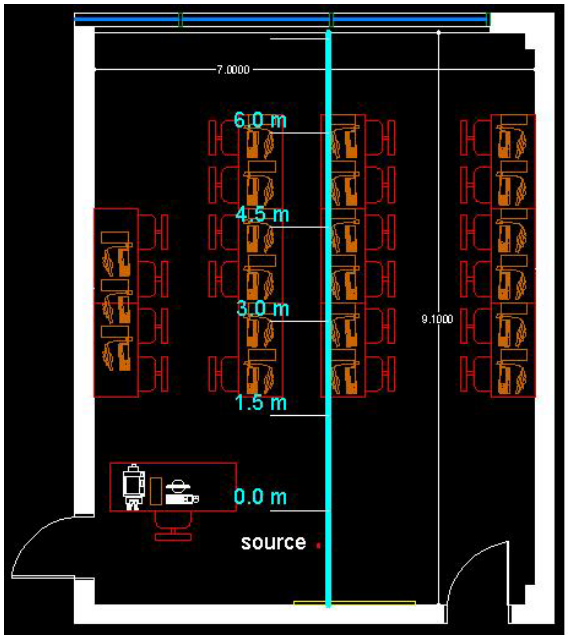


Figure 4: Measurement Points Across the Computer-Lab

3.METHODS

Temperature measurements collected are grouped in 4 data clusters for each room:

1) Morning interval temperatures with the technologies on (T_{on}): temperatures collected between 7:00 a.m. and 3:00 p.m. with all educational technologies in the room turned on.
2) Morning interval temperatures with the technologies off (T_{off}): temperatures collected between 7:00 a.m. and 3:00 p.m. with all educational technologies in the room turned off.
3) Afternoon temperatures with the technologies on (T_{on}): temperatures collected between 3:00 and 7:00 p.m. with all educational technologies in the room turned on.
4) Afternoon temperatures with the technologies off (T_{off}): temperatures collected between 3:00 and 7:00 p.m. with all educational technologies in the room turned off.

Phase one: temperature measurements are obtained to calculate the heat gain in the spaces due to full operation of educational technologies available in each room. A/C was not in use during experimentation. Also, the rooms were not occupied.

Heat gain in the room due to educational technologies was calculated based on the following equation:

$$Q = \dot{m} \cdot C_p(T_{on} - T_{off})$$

$$Q = \rho \dot{V} \cdot C_p(T_{on} - T_{off})$$

$$\dot{V} = \frac{ACH \; x \; V}{3600}$$

Where:

$\dot{m}$ = mass flow rate of air

$\dot{V}$ =Volume flow rate of air

C_p = heat capacity of air

T_{on}= Indoor air temperature while technology is on

T_{on} = Indoor air temperature while technology is off

ρ = Air density

ACH= Air changes per hour (5ACH)

Equation 1: Heat Gain Calculations

Figure 5 illustrates temperatures when the single smart board in the classroom is turned on, while Figure 6 illustrates a sample of temperatures measured in the same room with no educational technologies are in use. With an air volume of 229.3 m^3, total heat gain in the classroom of area 109.56 m^2, was found to reach 6.47 KWh, which is equivalent to 0.06 KWh/m^2. With an average rise in temperatures that reached 3.7°C due to the use of educational technologies in this room configuration.

Educational technologies in the computer lab on the other hand, with air volume of 0.32 m^3 and area 63.7 m^2, resulted in total heat gain of 6.03 KWh, which is equivalent to 0.095 KWh/m^2, with an average rise in temperatures that reached 4.5°C due to the use of educational technologies in this room configuration. Figure 7 and Figure 8 illustrate the temperature data gathered at the computer lab in both room conditions; with educational technologies turned on and off.

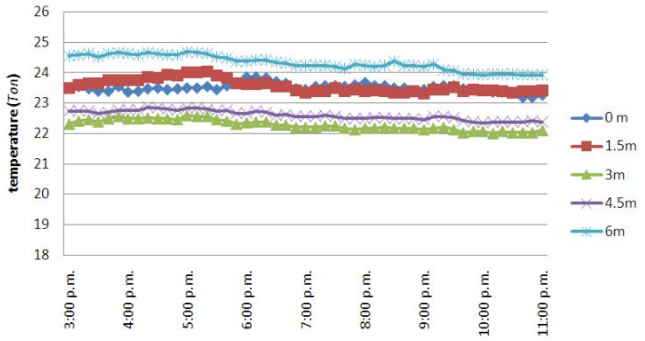


Figure 5: Sample of Temperature Measurements (T_{on}) at 5 Data Input Points in the Classroom. Afternoon Interval, Educational Technologies Turned On.

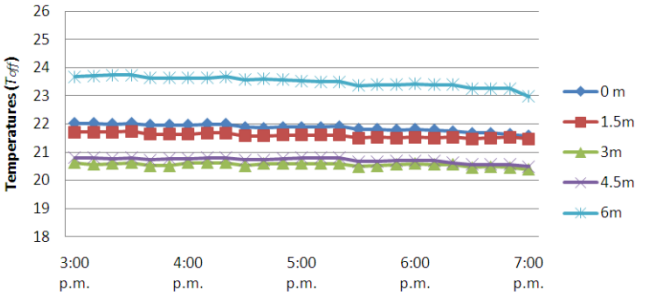


Figure 6: Sample of Temperature Measurements (T_{off}) at 5 Data Input Points in the Classroom. After noon Interval. Educational Technologies Turned Off.

It is observed from the thermal data obtained across the depth of both rooms, that the even distribution of the technologies has influenced the thermal profiles of each room. Thus a potential effect on user's thermal comfort may be

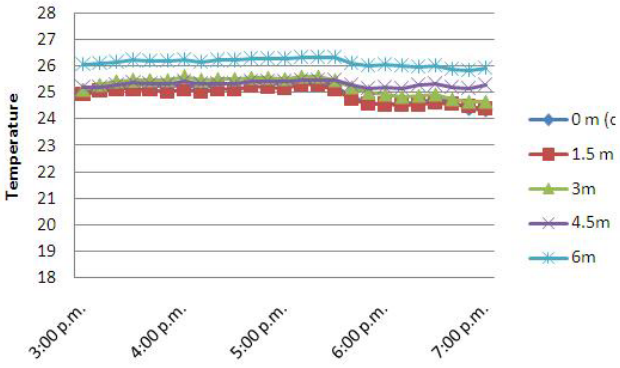


Figure 7: Sample of Temperature Measurements (Ton) at 5 Data Input Points in the Computer-Lab. Afternoon Intervals. Educational Technologies Turned Off.

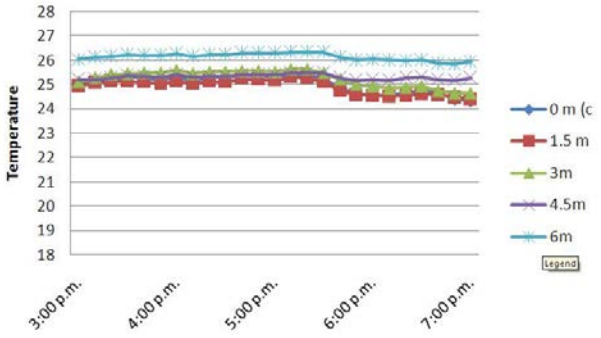


Figure 8: Sample of Temperature Measurements (Toff) at 5 Data Points in the Computer Lab. Afternoon Intervals. Educational Technologies Turned Off.

Phase two: The second phase of this study is ongoing and aims at investigating thermal profiles of the rooms as relates to the distribution of educational technologies and the influence on users' thermal comfort.

For that, each room is divided into 3 thermal zones:

1) The educational front zone: in which the typical instructional process takes place, and includes the learning wall and instructor's work station.
2) The learning zone: students occupy most of this zone, most educational technologies in the computer lab is also occupying this central area.
3) The perimeter zone: is mostly influenced by the external envelope conditions, solar radiations and wall composition. In classroom situations, the effect of learning technologies at the front zone is minimal.

In that phase, thermal zones in both rooms are simulated. And compared to alternative proposed configurations Recommendations for room configurations achieving better thermal performance and thermal balance within the spaces conclude the study.

4.CONCLUSION

In contemporary learning spaces, learning technologies play an important role in thermal performance of such spaces. The study investigates 2 typical spaces in a higher education institute; a typical classroom and a computer lab. Both rooms vary in proportion, type and number of learning technologies, as well as their distribution within the space.

During the experiment, air conditioning in the room was turned off, and the rooms were not occupied.

Thermal zones and profiles observed varied with the distribution of educational technologies in each room. Also, there has been proven rise in room temperatures and heat gain in the room.

With an area of 109.56 m^2, heat gain in the classroom resulting from a single smart board mounted on the front learning wall was found to be 0.06 KWh/m^2. With an average rise in temperatures that reached 3.7°C. Educational technologies in the computer lab on the other hand, with area of 63.7 m^2 resulted in total heat gain of 0.095 KWh/m^2, with an average rise in temperatures that reached 4.5°C due to the use of educational technologies in this room configuration.

The study also investigates thermal zones within the rooms due to distribution of educational technologies. Room configurations are thus recommended in response to the observed thermal zones.

5.REFERENCES

[1] Becker, R. et al., "Improving energy performance of school buildings while ensuring indoor air quality ventilation", **Building and Environment**, Vol. 42, 2007, pp. 3261–3276.

[2] Desiri, U. and S. Prioetti, "Analysis of energy consumption in the high schools of a province in central Italy", **Energy and Buildings**, Vol. 34, 2002, pp. 3003-1016.

[3] Elmasry, S., "Patterns for the Integration of Alternative Pedagogical Models and Learning Technologies", **Proceedings of the International Symposium on Engineering Education and Educational Technologies: EEET 2009**- Orlando, Florida, August 2009.

[4] Filippin, C., "Thermal response of solar and conventional school buildings to design-and human-driven factors", **Renewable Energy**, Vol. 30, 2005, pp. 353-376.

[5] Gartia, E. and A. De Herde, "Design of low energy office buildings", **Energy and Buildings**, Vol. 35, 2003, pp. 473-491.

[6] Gavin, D. and I. Knight, "Small power equipment loads in UK office environments", **Energy and Buildings**, Vol. 37, 2005, pp. 87-91.

[7] Jenkins et al., "Climatic and internal factors affecting future UK office heating and cooling energy consumptions", **Energy and Buildings**, Vol. 40, 2008, pp. 874–881.

[8] Perez, Y. and I. Capeluto, "Climatic considerations in school building design in hot-humid climate for reducing energy consumption", **Applied Energy**, Vol. 86, 2009, pp. 340-348.

[9] Radhi, H. et al., "Will energy regulations in the Gulf states make building more comfortable- A scoping study of residential buildings", **Applied Energy**, Vol. 86, 2009, pp. 2531-2539.

Generating Ontology Relations through Clustering for Preventive Health Care

Sang C. Suh
Lavanya Gaddam
Tarun Bheema
Department of Computer Science
Texas A&M University – Commerce
Commerce, TX 75429

Abstract- Ontology is a technique that represents data as a set of concepts within a domain and the relativity of those concepts [16]. It is used to give the rationale to the entities within the domain, and can also be used to describe about it [21]. Clustering organizes data into individual clusters. In this article we are proposing a model which uses the hierarchical relationships. We group related item sets to form clusters utilizing relational ontology. Relational ontology varies from domain to domain. As the name suggests relational ontology is more about relations. It represents the relativity of attributes and concepts. By performing clustering on the relational ontology we derived new relations through which, we can generate the new rules learnt from the database. In this article we have chosen the preventive health care domain through nutrition as an instance to show how effective is relational ontology to represent the existing data and to generate the further knowledge in the database. Furthermore, potential application of this technique on ontology relations clustering will be in the area of search engine based on semantic representation through ontology relations. While search engine process the task the link table look up the relations in the table for appropriate result.

Keywords: Clustering, Relational ontology, Ontology relation, Contributed attributes, Concepts, Combined relations.

I. INTRODUCTION

Data clustering is a technique through which we can develop cluster of objects that possess identical or similar characteristic in one way or the other. The criterion for ensuring the similarity is implementation dependent [16].

To meet the peculiar requirements like high volume of data, high dimensionality, meaningful cluster labels for clustering documents and ease for browsing we propose CBRO (clustering based on relational ontology). In which, we can ignore the number of clusters since this method is robust enough to handle different types of domains in a real-world environment. Furthermore, imprecise estimation of the number of clusters often yields poor clustering accuracy. CBRO enhances clustering accuracy by establishing related item sets.

Health care is a core component of one's life. Not to our surprise most tragic ill health conditions of our life like diabetes, cardiovascular diseases, obesity and malnutrition are preventable if taken proper measures. In several instances these problems arose out of negligence, improper care, they start as small health issues and when we ignore preventive measures, they turn into serious life threatening conditions and even fatal if they go undetected and untreated. The enduring nutritional transition expressed through increased utilization of high fat and excessive sodium containing products. It contributes to the rising burden of heart disease, stroke, obesity and diabetes. Preventive health care through proper nutrition is one of the most efficient means of triggering these issues and to stay out of most costly and disabling conditions mentioned above. We are aspiring to implement our application on domain which is useful for variety of people.

Here we are proposing a related item set based clustering, utilizing the relational ontology (has, avoids, causes and occurs-in etc). HAC is a hierarchical model tool for human cognitive concepts [8]. Our application CBRO is the reforge of HAC employing relational ontology. Related Item Set (RIS) is a set of related contributed attributes, applying RIS and relational ontology we build relationship between contributed attributes and concepts. We also implemented the concept of HAC (Hierarchy of Attributes and Concepts), but here we maintain the hierarchy based on relation. We have shown clusters in different levels using relational ontology. We are also presenting PHC through nutrition domain sample input and the output of it implementing our application.

II. RELATED WORK

The goal of clustering is to reduce the load of data by classifying it into groups depending on similarity of its objects. Such grouping is pervasive in the way humans' process information, and one of the motivations for using clustering algorithms is to provide automated tools to help in constructing categories or taxonomies. The methods may also be used to minimize the effects of human factors in the process.

Clustering methods can be divided into two basic types: hierarchical and partitional clustering. Within each of the types there exists a wealth of subtypes and different algorithms for finding the clusters. Hierarchical clustering proceeds successively by either merging smaller clusters into larger ones, or by splitting larger clusters. Partitional clustering, on the other hand, attempts to directly decompose the data set into a set of disjoint clusters [3].

HAC is both hierarchical and conceptual clustering system that organizes data so as to maximize inference ability [4]. The idea of hierarchical clustering is to begin with each point from the input as a separate cluster. We then build clusters by merging clusters that are close to each other: repeatedly merge two or more clusters that are closest to each other out of all pairs [8]. Visual Data Analytics (VDA) system represents the hierarchy of

attributes and concepts graphically. VDA is designed to represents attributes, concepts and their relationships visually [12, 13].

Figure 1 shows the example of HAC with concepts and the related attributes. Pink and green buttons shows the attributes and concepts respectively. Lines between them show the relationship between the concept and attributes. In this paper, we propose clustering based on the semantics (i.e. meaning), according to which we define ontology relation for every attribute. In a cluster every attribute is related based on that relation.

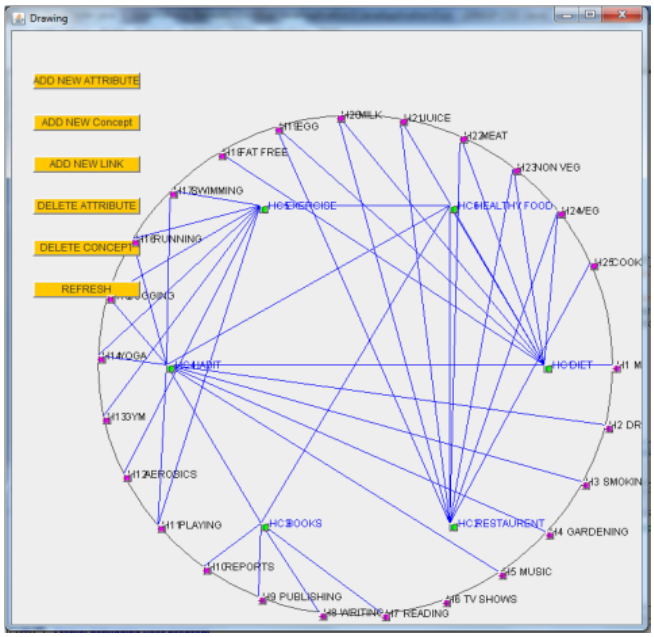


Figure 1: Example of HAC

III. RELATIONAL ONTOLOGY

Relational ontology varies from domain to domain. Ontology, a cornerstone of the semantic web, have gained wide popularity as a model of information in a given domain that can be used for many purposes, including enterprise integration, database design, information retrieval and information interchange on the World Wide Web. For our work, we describe OBO (Open Biological and Biomedical Ontologies) relational ontology [2, 6, 7].

TABLE 1
BASIC RELATIONAL ONTOLOGY RELATIONS

	Transitive	Reflexive	Anti symmetric
Is a	x	x	X
Part of	x	X	X
Integral part of	X	X	X
Proper part of	X		
Located in	X	X	
Preceded by	X		
Has	X		
Avoids			
Causes			
Occurs-In			

Transitive means "if A is related to B and B is related to C, then A is related to C"

Reflexive means the relation is bidirectional. I.e., "if A is related to B, then B is related to A".

Anti symmetric means "if A is related to B and B is related to A, then A is equal to B"

In this paper we have taken four relational ontology relations from the above-mentioned relational ontology. They are *has*, *avoids*, *causes* and *occurs in*. Using these ontology relations, contributed attributes are clustered as concepts in PHC domain. We combined two ontology relations and defined a new relation to represent a combined ontology relation.

For example:

AVOIDS ⊕OCCURS IN ⟹ GOOD FOR

Blindness come under the concept vitamin A according to ontology relation *avoids* and it also comes under the concept eyes according to ontology relation *occurs-in*. From these two ontology relations, a new ontology relation "*good for"* derived between vitamin A and eyes.

IV. CLUSTERING ON RELATIONAL ONTOLOGY

In existing clustering methods, there has been no attempt to form clusters using relations. In this paper we build a database on PHC, using word net from which we define definition of each relation for each and every contributing attribute. Word net is a lexical database for English language [14].

We build an input table which is shown in Table 6 in results section. We built contributed attribute table directly from the source table using the fields, *ca-id* and *ca-name*. We used an algorithm to build concept table. The algorithm is shown below on how the concepts and relations are generated from the source table. For example, calcium is a concept and its related attributes are cheese, molasses, vegetables according to ontology relation *has*.

INPUT: Definition of each contributed attribute of every property
OUTPUT: Concepts generated

PROCEDURE:

1. Initialize property[];
2. Repeat Until each property from the defined properties {
3. Repeat until the end of each property (column) {
4. Repeat Until(j<n) {
5. Compare the first string with all other strings if they are equal consider it as concept {
6. Add the String to the group of concepts
7. Make the property of new string as "OR" } } } }

Clustered data is shown in 2 different ways. One way is according to concept, based on ontology relation of first level as illustrated by Figure 2. Figure 2 shows the cluster of concept protein. For example eggs, milk, rice, whole grain and corn comes under the concept protein according to the ontology relation *has*. Second way is according to each relation. For example, for ontology relation *avoids* in Figure 3(a), fissured tongue and neural tube disease can be avoided by folic acid, so we cluster them as one cluster. Osteoporosis and bone disorders can be avoided by calcium, so we cluster them as another cluster under *avoids* ontology relation. Here we combined two ontology relations and by combining them we defined one new relation between two concepts.

R_1: $a_1 \xrightarrow{Avoids} a_2$

R_2: $a_3 \xrightarrow{Occurs-In} a_1$

Then,

$R_3 = R_1 \oplus R_2 \Rightarrow a_3 \xrightarrow{Good\ For} a_2$

For example, since vitamin C avoids scurvy and scurvy occurs in gums as shown below, from these relations, a new relation, *good for*, between cancer and loss of vision is derived, as shown below and also in Figure 4.

Vitamin C *avoids* → scurvy

Scurvy *Occurs in* →gums

So, Vitamin C *good for* → gums

In this way we derived new ontology relations which are shown in combined relation table (Table 2).

TABLE 2
NEWLY GENERATED RELATIONS THROUGH RELATIONAL ONTOLOGY

Combined relation	New relation
Has + Avoids	Prevents
Has + Causes	May cause
Causes + Occurs-in	May effect
Avoids + Occurs-In	Good for
Has + Causes + Occurs-In	Not good for
Has + Avoids + Occurs-In	Good for

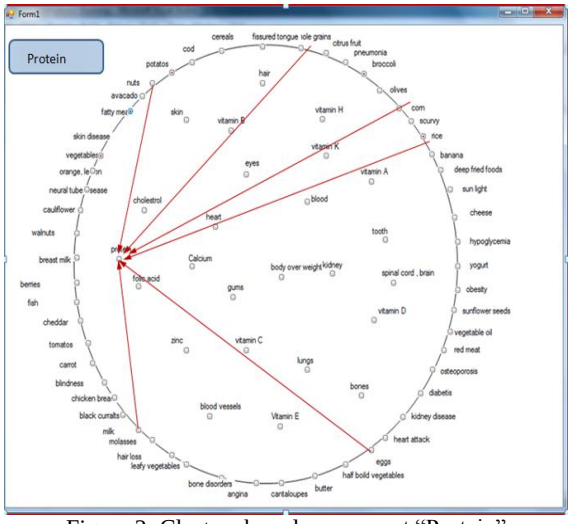
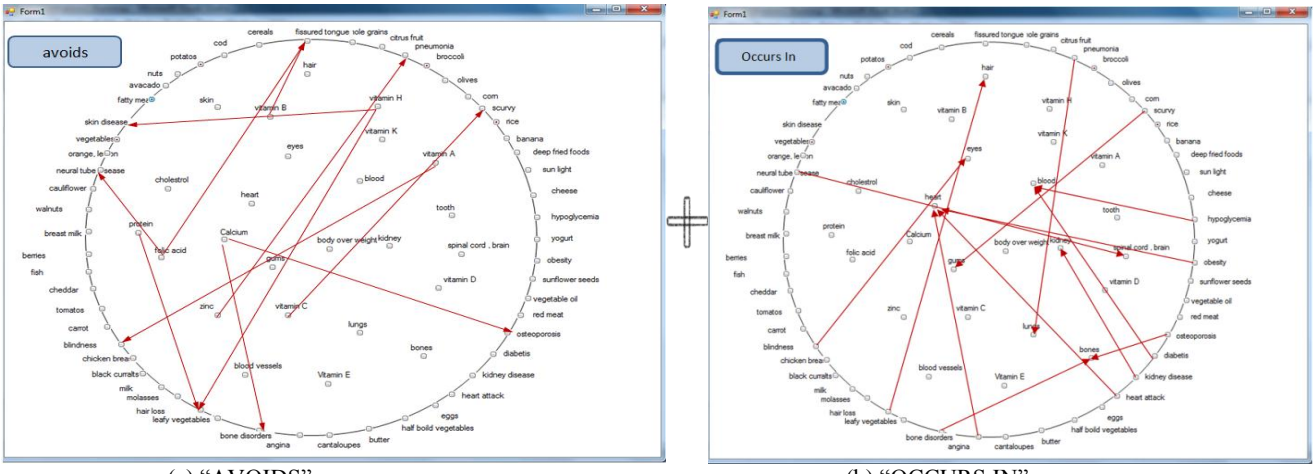


Figure 2: Clusters based on concept "Protein"

Here we combine two HACs (built based on existing ontology relations) to form a new HAC representing a new relation as shown in Figure 4.



(a) "AVOIDS" (b) "OCCURS IN"

Figure 3: Clusters based on existing relations "AVOIDS" and "OCCURS IN"

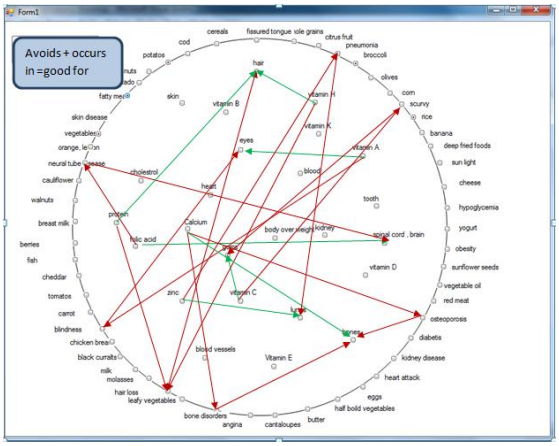


Figure 4: New relation "GOOD FOR"

TABLE 3
CONTRIUTED ATTRIBUTED TABLE

Concept_ID	CA_ID	Relation
C23	S34	Occur in
C16	S31	Occur in
C1	S4	Has
C10	S47	Has
C1	S35	Avoids
C10	S19	Avoids
C10	S32	Avoids
C13	S35	Causes
C5	S32	Causes

V. IMPLEMENTATION

We have chosen PHC through nutrition domain as the input database. Figure 6(a) shows the entire domain using circles with buttons on the circle representing the contributed attributes and buttons inside the circle representing concepts. Preventive health care takes measures to prevent diseases (or injuries) rather than curing them. From our PHC domain, we only consider *different food types, minerals, diseases caused by vitamin deficiency and diseases caused by over consumption of minerals, and the parts affected by different diseases* among many others. The most important part of preventive health care is in maintaining good healthy diet.

The database for any domain to implement our application needs four tables: contributed attribute table, concept table, link table and combined relation table. The Contributed Attribute (CA) table contains CA_ID and CA_Name as shown in Table 3, concept table contains Concept_ID and Concept_Name as shown in Table 4, link table contains Concept_ID, CA_ID and Relations as shown in Table 5, combined relation table contains combined relations and new relations as shown in Table 2.

Using these tables we represent the clusters visually in different ways. Initially it shows the circle with contributed attributes on the circle and concepts inside the circle. When we click on any concept it shows the related attributes by linking each and every related attribute with that concept. CBRO allows the user to select the relation from the dropdown list which they want, and then it shows the clusters based on that relation. CBRO combines two relations to get the new relation and forming next level clusters based on that new relation. CBRO shows the new relation with green links between two concepts like in Figure 7(a). For example when we combine *causes* and *occurs-in* we get a cluster like in Figure 7(a), here green links show the new ontology relation *may effect* in between two connected concepts, red links shows the first level ontology relations *causes and occurs-in* and all the connected concepts with their relation are shown below the circle.

TABLE 4
CONCEPT TABLE

CONCEPT_ID	CONCEPT_NAME
C1	calcium
C10	vitamin E
C11	vitamin B
C12	vitamin K
C13	protein
C14	vitamin A

TABLE 5
LOINK TABLE

CA_ID	CA_NAME
S12	vegetable oil
S13	sunflower seeds
S14	leafy vegetables
S15	orange, lemon
S16	banana
S17	scurvy
S18	blindness
S21	cheddar

When we click on the concept we will get the related contributed attributes of that concept and it will also displays the relation between the concepts and attribute. Figure 5 shows the cluster of concept vitamin H.

Example: Concept: Vitamin H
Related Contributed Attributes: skin disease, hair loss.
Relational Ontology: *avoids*

User also has an option to select the relational ontology from the drop down list which is shown in Figure 6(a).
When user selects the particular relational ontology then it shows all clusters which are built based on that relational ontology. Figure 6(b) shows the clusters based on the ontology relation *occurs-in*. Neural tube disease *occurs-in* spinal cord and brain.

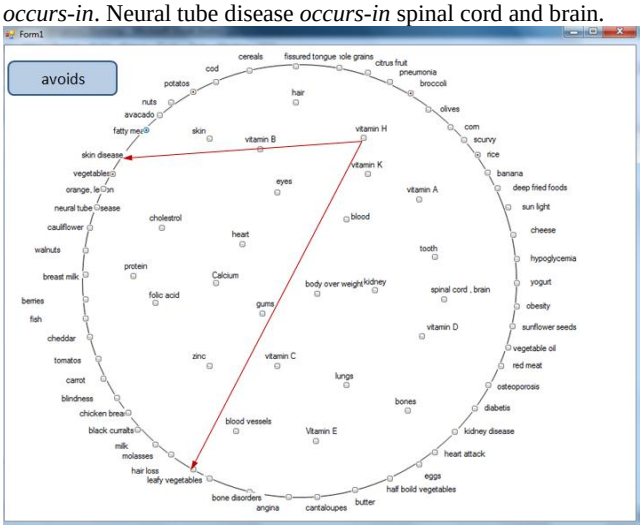
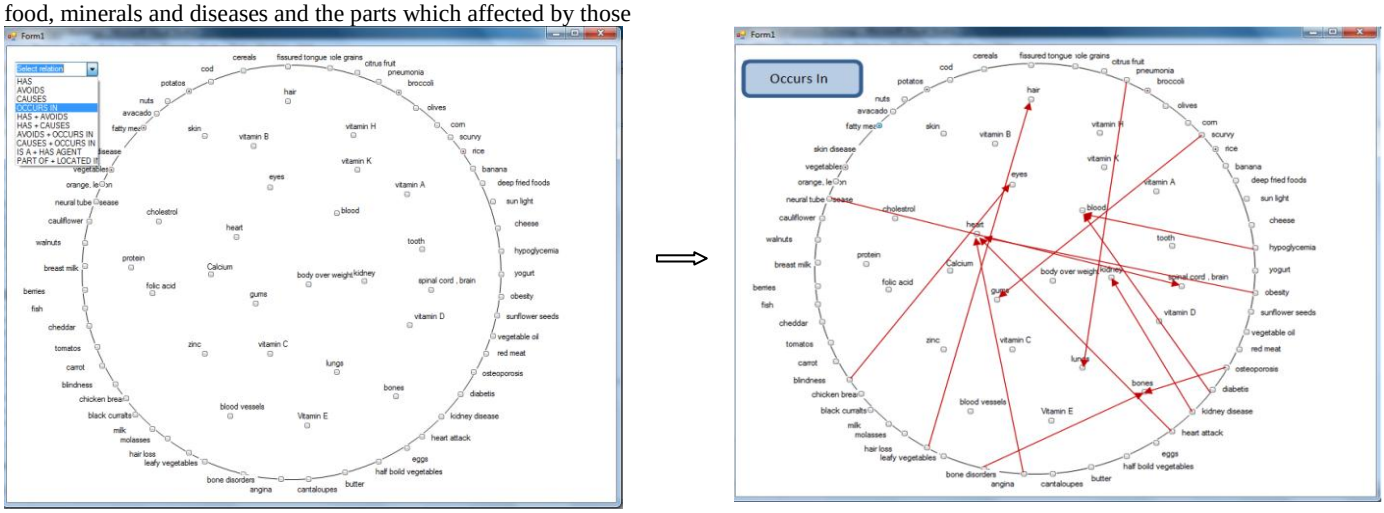


Figure 5: Cluster of Concept "VITAMIN H"

VI. RESULTS

Here is our sample input table which we used for CBRO. We built this table using information from various websites about the food, minerals and diseases and the parts which affected by those diseases. By considering food habits we built this table as shown in Table 6. To implement CBRO on PHC we derived four tables those are contributed attribute table, concept table, link table and combined relation table from the input table.



(a)View of CBRO

(b) Clusters of ontology relation "OCCURS-IN"

Figure 6

We used an algorithm to get the concept table and link table from the input table. CBRO derives the new ontology relation between concepts from the existed ontology relation between concepts and contributed attributes. Output of CBRO contains the derived new ontology relation between two concepts. Heart attack comes under concept heart based on ontology relation *occurs-in* and it also comes under concept cholesterol based on ontology relation *causes*, using these two we built a new ontology relation between two concepts cholesterol and heart that is *may effect*. Leafy vegetables *have Vitamin A* and vitamin A *avoids* Blindness, from this we derived a new ontology relation *prevents*. Leafy vegetables *prevent* blindness. Table 7 Sample output table shows some of the newly derived ontology relations.

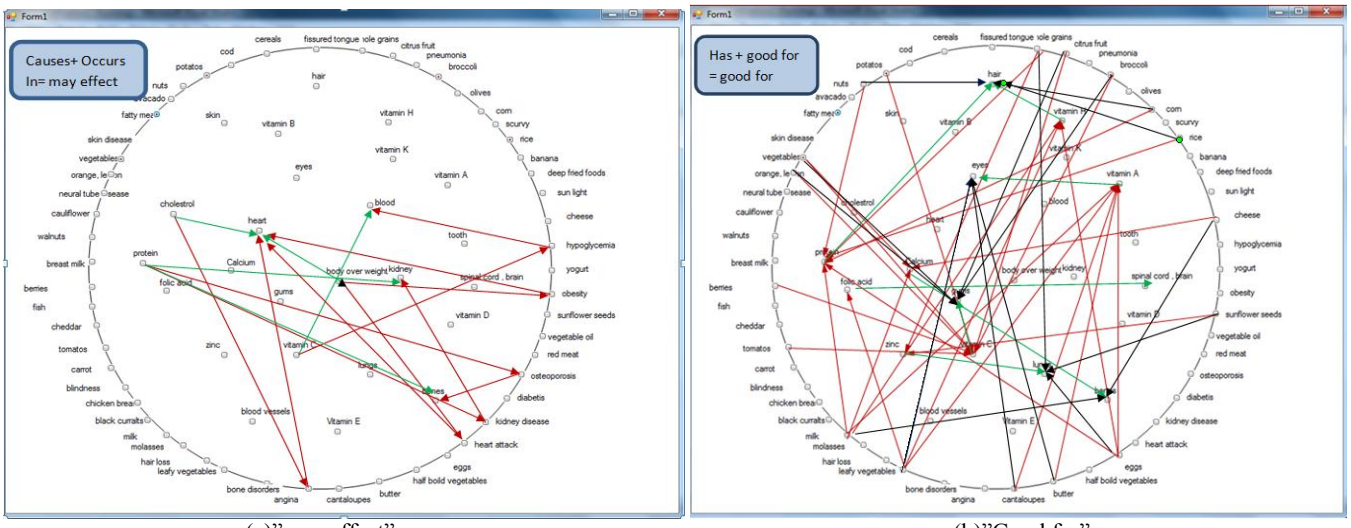
Figure 7 shows the derived ontology relations and their clusters. Figure 7(a) shows ontology relation *may effects* which is determined by combining *causes* and *occur-in*. Examples of derived relation may affects: cholesterol *may effects* heart, over consumption of vitamin C *may effects blood* etc. Figure 7(b) shows the ontology relations *good for* which is derived by adding *has+ avoids + occur-in*. In the Figure red line shows the 1st level relation has, green lines shows the 2nd level relation which is determined by adding avoids and occurs-in and the black lines shows the actual third level relations which are determined by adding has and good for. Examples of derived relations *good for*: eggs good for lungs, leafy vegetables good for eyes etc.

TABLE 6: SAMPLE INPUT TABLE

Concept_ID	Relation	Concept	Contributed Attribute
C1	Has	Calcium	S1,C4,S54,S53,S52,S44,S43,S24,S25,S26
C21	Has	Zinc	S9,S11,S12,S55,S56
C3	Avoids	Vitamin C	S8,S10
C6	Causes	Cholesterol	S1,S16
C7	Occurs In	Heart	S20,S21
C8	Occurs In	Eyes	S1,S3,S14
C10	Good for	Gums	S23,S45

TABLE 7: SET OF NEWLY GENERATED RELATIONS

Relation	Concept-Name
Has	(Broccoli ,sodium) (Pecans ,vitamin A)
Avoids	(Vitamin C, Scurvy), (Vitamin F , Fissured Tongue)
Good for	(protein, hair),(vitamin H, hair)(vitamin A, eyes),(folic acid, spinal cord),(folic acid, brain),(calcium, bones),(zinc, lungs),(Vitamin C, Gums)
Occurs-in	(Diabetes, Blood)
May effect	(Cholesterol, heart),(protein, kidney),(over body weight, heart), (Vitamin C, blood),(protein, bones)
Not good for	(Fatty food, heart), (Deep fried food, heart)
Good for	(nuts, hair),(corn, hair),(rice, hair),(citrus fruit, gums),(vegetables, gum)



(a)"may effect" (b)"Good for"

Figure 7: Clusters of newly generated relation

VII. CONCLUSION AND FUTURE WORK

In this paper, clustering of data using ontology relations in PHC domain was discussed and illustrated. Few tables like contributed-attribute table, concept table and link tables were built to implement CBRO for the PHC domain. First, data was clustered in concept level. Second based on the existing ontology relations *has, avoids, causes and occurs-in* new ontology relations are generated through this process. Ontology relations *may cause, prevents, good for and may effects* are derived between two concepts by implementing CBRO application in PHC domain (Data base). Then, we integrated HAC's of three related ontology relations. Ontology relations *good for, not good for* are derived between three concepts by implementing CBRO application in PHC domain (Data base).

In the future, further work can be done to expand this HAC by allowing more than three ontology relations to be combined as a single relation. While our application CBRO is domain dependent, it will be expanded to be suitable for any domain (domain independent) as future work.

VIII. ACKNOWLEDGMENT

This research has been partially supported by a grant from the US Department of Energy Grant (Funding #: DE-SC0001132). It has also been supported in part by an interdisciplinary research enhancement grant from the graduate school of A&M University-Commerce. This paper is the extension of role of clustering of ontology relations for preventive health care through nutrition.

IX. REFERENCES

1. Benjamin C.M. Fung, Ke Wang, "Hierarchical Document Clustering Using Frequent Item sets" Martin Ester SIMON FRASER UNIVERSITY, BC, Canada, September 2002.
2. Sudipto Guha, Rajeev Rastogi, Kyuseok Shims "ROCK: A Robust Clustering Algorithm for Categorical Attributes", Information Systems Volume 25, No. 5,pp 345-366, Elsevier Science Ltd, Britain, 2000.
3. Sang C. Suh and Kalyani Komatireddy, "Clustering of Ontology in Preventive Health Care Through Relational Ontology", IKE'11, World comp 2011, Las Vegas, Nevada, USA.
4. Ji-Rong Wen, Jian-Yun Nie, Hong-Jiang Zhang "Query Clustering Using Content Words and User Feedback", SIGIR'01, New Orleans, Louisiana, USA, 2001.
5. Sang C. Suh, Sam I. Saffer, Nikhil Goel, Young S. Kwon, "Generating Meaningful Rules Using Attribute Concept Hierarchy", Intelligent Engineering Systems through Artificial Neural Networks, Volume 16, ASME Press, 2006.
6. E. A. Freigenbaum and H. Simon, "EPAM-like Models of Recognition and Learning", Cognitive Science, Volume: 8, pp. 305 -336, 1984.
7. R.T.Ng and J.Han "Efficient and Effective Clustering Methods For Spatial Data Mining". In Proceedings of the VLDB conference, 144-155, Santiago, Chile, 1994.
8. Sang C. Suh and Gouthami Vudumula "The Role of Conceptual Hierarchies in the Diagnosis and Prevention of Diabetes", NCM 7th International Conference, 2011.
9. Sang C. Suh and Jhansi Baireddy, "Visual Representation of Hierarchical Attributes and Concepts as an Ontology for Semantic Reasoning", Intelligent Engineering Systems through Artificial Neural Networks, Volume 20, ASME Press, 2010.
10. Fellbaum, "WordNet: an electronic lexical database" Cambridge, MIT Press, 1999.
11. wordnetweb.princeton.edu/perl/webwn
12. en.wikipedia.org/wiki/Ontologies_(computer_science)
13. Jan Jantzen, "Tutorial on Fuzzy Clustering", Technical Report, DTU, August 2001.
14. H. G. Wilson, B. Boots, and A. A. Millward, "A Comparison of Hierarchical and Partitional Clustering Techniques for Multispectral Image Classification", Geoscience and Remote Sensing Symposium, IEEE International, Vol.3 pp 1624 - 1626, 2002.
15. Holmes Finch, "Comparison of Distance Measures in Cluster Analysis with Dichotomous Data", Ball State University, Journal of Data Science 3, 85-100, 2005.
16. http://members.tripod.com/asim_saeed/paper.htm

A driver-rider interaction validation study in both simulated and real environments

Katerina TOULIOU
Dimitris MARGARITIS
Evangelos BEKIARIS
Centre for Research and Technology Hellas (CERTH)
Hellenic Institute of Transport (HIT)
Thessaloniki, 57001, Greece

ABSTRACT

Motorcycle related fatalities have increased by 14% during the last decade (NHTSA, 2009). The human factor is dominant in almost 90% of accidents (ACEM, 2004) and 70% of them occur due to late detection of the PTW. Perceptual failures occur mostly in intersections. Two in-depth controlled studies were conducted in order to investigate the interaction between drivers and riders in two road conditions: *intersection* and *car following*. The drivers; perspective was the main objective of the study. Ten experienced drivers (31±4.16 years old) were included in both real (on the road) and simulated (passenger car simulator) tests. On the road tests were conducted with an instrumented vehicle and the simulator tests with a semi-dynamic Smart cabin-based driving simulator. The main vehicle parameters of interest were speed (km/h), braking activation, lateral acceleration (g) and steering wheel angle (degrees). After the end of simulator tests, general realism, simulator sickness, and driving effort questionnaires were completed. Subjective assessments were intensified by video frames data.
Intersections seem to be complex to simulate because of generic characteristics difficult for participants such as curved road segments with 90 degrees. Simulation of certain road conditions has higher inherent aspects for success (e.g. motorways). No significant mean speed differences were found in the car following scenarios between the two environments. It is important to consider that while stability was increased on the motorway environment, lateral control was still decreased compared to real setting.

Keywords: PTW, rider, simulator, intersection, car following, brake actuation, lateral acceleration.

1. INTRODUCTION

Motorcycle related fatalities have increased by 14% during the last decade [1]. The contribution of powered two-wheelers (PTWs) in motorised traffic accidents seems to be the same in UK [2]. PTWs are considered to be a small percentage in traffic but the number of riders increases dramatically as the need for flexible travelling, also, increases. Surprisingly enough most accident occur during daylight (73%) with clear weather conditions (90%) [3].

According to current research factors that may lead to an accident fall into the following three categories: a) human, b) vehicular, and c) environmental [4][5]. If the environmental conditions are safe what happens to the drivers/riders? Which human factors can be identified that play a crucial role in the accidents between passenger cars and PTWs? It is a well-researched fact that most accidents occur because of perceptual human errors (i.e. human factor is prevalent in accident analyses). The human factor is dominant in almost 90% of accidents (ACEM, 2004) and 70% of them occur due to late detection of the PTW. Human factors involve diversity of errors with the most abundant to be visual distraction and "looking-but-failing-to-see" errors. More specifically, these perceptual failures occur mostly in intersections. Drivers are trained to follow the sequence "look right, then left, then right again", therefore they tend to focus more on looking and rather less on processing incoming information.
Crundall and colleagues [6] proposed a framework to accommodate for the interaction between drivers and riders from the perspective of the driver. This framework attempts to define the driver behaviour by assuming that personal beliefs and "schemata" feed their attitudes, knowledge and driving skills and interact with their sensory and cognitive ability to "see", "recognise", and "appraise" the situation with the PTW. These influences are called top-down influences. Their counter-influences are called bottom up (e.g. movement, colour and luminance, obscurence). The interaction of these three processes could yield an accommodating model; however, research-to-date has not covered all of them in a single experimental effort.
Cognitive conspicuity depends on both sensory information processing and cognitions the driver has for driving performance and, of course, for other aspects of living. Based on "flight vs. fight" theory [7] of stress, subsequent generalisations may lead to attempts to explain human behaviour based on criticality and danger imminence. Reactions in dangerous on near dangerous situations are influenced by physiological reactions. Therefore these reactions in imminent situations could be influenced by inherent and generic aspects of PTWs. PTWs are smaller in size by default, hence they may be perceived as less dangerous. The latter is extremely important for intersection related incidents and accidents. This happens because drivers find it difficult to estimate the time it takes to collide with the intersection entering PTW

because its size is much smaller than a passenger cars'. This phenomenon is known as "size arrival illusion" [8]. Additionally, drivers that are involved in accidents with motorcycles are more likely to be unfamiliar with motorcycles [9].

Researchers have distinguished two main types of PTW related conspicuity, namely *attention* and *search* conspicuity. Attention conspicuity refers to the ability of the driver to detect the PTW when it is unexpected while search conspicuity has to do with the anticipation and the active search of the driver to "see" the PTW [5].

The current study focuses on the driver's perspective of the interaction between driver and rider on two potentially dangerous road conditions as identified by current research: a) intersections and b) car following. This study consisted of two complementary experiments, the first was conducted in real setting and the second was carried out in a simulated environment. The influence of driving environment types (intersection and car following) and PTWs speed were evaluated. In the present experiments the authors studied the possibility of relations between the recorded variables in both environments. Thus, in both experiments the same sets of variables were used as the main objective was to investigate the discrepancies (if any) between simulated and real environments.

It was hypothesized that speed will increase in the simulated environment accompanied by higher variation in related measurements. An in-depth analysis and kinematic reconstruction of 22 cases of urban accidents involving a motorcyclist and another road user showed that a motorcyclist's speed is significantly higher for motorcycle-related cases in comparison with others [10]. Speed of the motorcyclist entering the trial area was of interest as current research places emphasis on the importance of high speed of motorcycles in urban areas. In relation to this finding two different speed were chosen for the intersection condition (35 and 50 km/h).

Drivers usually fail to see riders for many reasons as discussed earlier. "Looked but failed to see" (LBFTS) accidents have been of great interest in research; however, it has been studied mostly in naturalistic observation and accident analysis efforts, therefore never controlled and measured in behavioural studies. Experiments have shown that drivers as soon as they look away they tend to remember only three or four vehicles around them even though they usually are many more in surrounding traffic. This is complimentary to the fact that usually glimpse images are incomplete and the brain steps in to fill in the gaps. Therefore, on the road "stereotypes" seem prominent to fill in the gap by giving way to similar vehicles (i.e. the majority are passenger cars). It is of no surprise that drivers who are riders show increased visual search skills compared to individuals who are just drivers. This may not only be translated to riding training for everyone but additionally to modifications and/or enhancements of existing driving training with riding and rider issues. A driving simulator could possibly be a good starting point. Past research on the interaction between PTW and a typical car driver was mostly performed in either hypothetical (video scenes) or simulated environments. It has been shown that car simulator studies can contribute to better define the problem from the perspective of the driver [11].

This is another reason why comparisons of both environments could be valuable for extrapolating findings to riders' training and not only for enhancing and refining existing driving simulators.

The focus of the study is mainly to address the possibility of using a driving simulator in order conduct such experiments which evidently would be a safer alternative. So, is the driving a reliable alternative to on road studies? Where the differences lie?

The list of questions is endless but the need to address certain issues is both a necessity; especially because this area is not studied in depth in controlled settings. Two types of accidents have been identified and selected for the experiments. Firstly, accidents in intersections are the most prevalent and have a higher statistical probability of severe injury or death when compared to accidents happening in other road areas [12]. Secondly, it is of interest to investigate the driver's reaction with regards to vehicle control when the PTW is close following and overtakes the passenger car on the motorway.

2. METHODS

Both experimental tasks aimed to utilise intersections and car following scenarios. The focus was directed towards the driver in relation to riders' presence. The decision and selection of certain scenarios were directed by research questions relevant to the study undertaken. A brief description of both the instrumented vehicle used for the on-road tests and the driving simulator is provided below.

Instrumented vehicle

The instrumented research vehicle was utilised in order to carry out the on road tests. The vehicle was equipped with a data logger, 3 different-view cameras, and a pc at the back trunk where the data files were stored. All vehicle parameters were recorded and stored in ASCII format. All vehicle instrumentation was synchronised to allow for cross-data sources' comparisons. Similarly, the driving simulator recorded data in the same format. A camera was mounted on the front windscreen in order to be able to record frames with information about the direction the driver was looking at.

Experiments required the investigation of driver's behaviour and reactions to another PTW rider; hence the instrumentation of a PTW was necessary in such a way that allowed recordings of their relative position (GPS data). Consequently, two GPS units were installed in the instrumented car and PTW, respectively. The next step was to temporally synchronise both units (global time).

Both vehicle and PTW instrumentation was verified in order to ensure recordings of reliable and accurate data. The technical verification process resembled the foreseen experiments with, of course,

limited number of repetitions. Another vital specification of the data acquisition systems is that they are totally autonomous and they do not interfere with the other functionalities of the vehicle's main PC, hence safety hazards are minimised.

Car driving simulator

The driving simulator study was conducted in the driving simulator laboratory within CERTH/HIT premises. The CERTH/HIT driving simulator is built around a Smart cabin equipped with sensors. The actuation of all control levers, windshield wipers, blinker, ignition key and light switch is electronically transmitted to the driving computer. All operational elements such as steering wheel, accelerator pedal, brake pedal, gearshift lever and handbrake lever, provide nature-true force reactions. The visual system includes five large-screens, each having a width of 2 meters. The visual system works with on-screen projection with video projectors (2500 ANSI-lumen). The sound system generates original sounds according to the situation (starter, engine noise, horn, screeching of tires, drive wind, rain, etc.). The vibration device creates natural vibrations of the car according to the revolution of the simulated engine.

The driving scenarios were developed in order to emulate the road specifics in an urban and motorway environment.

Video data

Three cameras were mounted on the instrumented vehicle for the on road tests. For the intersection scenarios the cameras were facing left, right and the face of the driver. For the car following scenarios the cameras were facing the face of the driver, the front and the rear of the vehicle. One USB camera facing the driver was mounted on the dashboard of the smart cabin for the simulator tests.

Subjective assessments

All participants filled in a background questionnaire with basic demographics. The same individuals participated in both experiments, hence all filled in the simulator sickness index (Kennedy et al., 1994). The SSQ is based on three components: *nausea, oculomotor problems,* and *disorientation.* These can be combined to produce a total SSQ score. Participants were asked after each driven scenarios in the simulator "How they felt" if any signs of sickness were suspected (e.g. sweating, uneasiness, etc.). Simulator sickness is not such a rare phenomenon (20-30%), especially with experienced drivers and more importantly especially in simulated urban environments. It was important to make sure that susceptibility was low prior participants continued with driving scenarios to limit drop-outs and have a more time-efficient testing conduction.

In addition, participants rated how much they had to try in order to complete the driving tasks [Rating Scale of Mental Effort (0-150); RSME] (Zijlstra, 1993) and how well they thought they performed in the driving tasks compared to their "on the road" driving experience (Driving Quality Scale, Brookhuis et al., 1985). In the present study, the original subjective scales were translated into Greek and back-translated by an independent professional and verified by a native English speaker.

In addition, the simulator realism questionnaire was filled in in order to investigate the subjective realism of the experience. Mental effort scores and subjective realism provide a representative account of the experience for the participants.

A focus group was conducted in the end of the study in order to discuss in depth their driving experience for both environments. The discussion included topics/themes about their reactions when they saw the rider and the level of realism in their reactions compared to everyday driving experience.

Participants

10 drivers (31±4.163 years old) participated in both the real and simulated experiments. Participants were screened in order to meet the inclusion/exclusion criteria. All participants had good or corrected (i.e. glasses/contact lenses) eyesight. The gender ratio was 7 (male) to 3 (female), although it was aimed to be more balanced.

Procedure

The experiment was divided into two successive phases. Participants gave written consent prior their participation in the tests. First, the participants filled out a demographic and driving experience questionnaire. Next they were familiarised with the instrumented vehicle and with the driving simulator. Familiarisation sessions lasted approximately 15-20 minutes and the experimental session did not start before the participant informed the study leader about them being comfortable to start. The same drivers participated in both real and simulated studies in order to ensure comparability and control for confounded variables due to personal differences. Participants were encouraged to drive as realistic as possible ("Drive as usual for this type of the road"). No other workload was given as it was decided that it would be risky for the on road tests. No other interaction or coaching or discussion was given or encouraged during the tests.

Both on the road and simulated environment experiments were conducted the same period of time in order to control for performance related and idiosyncratic attributes. A professional rider participated as a demonstrator rider in both intersection and car following scenarios for the on.

Table 1: Group statistics for speed (km/h), lateral acceleration (m/sec^2) and steering wheel angle (deg) in real and simulated environment

			Brake activation (on/off)		
			Off	On	Total
Environment	**On the road**	Count	2383	127	2510
		% within Environment	94.9%	5.1%	100%
	Simulated	Count	6098	1310	7408
		% within Environment	82.3%	17.7%	100%
Total		Count	8481	1437	9918
		% within Environment	85.5%	14.5%	100%

Table 2: Cross-tabulation counts for braking activation in real and simulated environment

	Environment	Mean	Std. Deviation
Speed (km/h)	*On the road*	58.42	34.782
	Simulated	61.97	38.828
	Total	61.07	37.874
Lateral acceleration m/sec^2	*On the road*	0.10	0.655
	Simulated	-0.02	0.511
	Total	0.009	0.553
Steering wheel angle (deg)	*On the road*	-30.83	61.168
	Simulated	-0.04	12.262
	Total	-7.84	35.187

road trials. Participants were briefed about the nature of the study without revealing the exact purposes and hypotheses. However, due to the nature and the probability of related safety issues (especially for the intersection) participants were informed about the presence of a rider in order to avoid accidents. As the main interest was to compare vehicle parameters in both environments, it was assumed that this was not seriously affecting driving behaviour.

Participants were asked to signal when they actually saw the rider. The order of the conditions (i.e. instrumented vehicle and simulator) and scenarios were counterbalanced. Therefore extra care was taken to note the sequence of scenarios each time for each participant. The whole procedure lasted approximately two hours for each participant.

Participants filled in a feedback form after the end of the real tests about their experience during the on the road tests and asked to describe what they did when they saw the rider (e.g. braked, decelerated). At the end of the simulator tests they filled three subjective questionnaires as described in the respective methods section. At the end of the study participants were invited in a focus group to discuss about the overall experience moderated by the test leader. At the end of the discussion participants were debriefed about the nature of the study.

All data were collected right after the experiments, checked for errors and/or omissions, were coded and stored in a safe place.

3. RESULTS

Paired comparisons were conducted in order to investigate differences between the real and virtual environments based on the vehicle collected data. Braking was treated as a dichotomous variable (on/off); hence cross-tabulations were carried out for the comparisons performed. More elaborate statistics could be applied with increased number of participants.

The first level of analysis was based on comparisons between the real and simulated vehicle parameters. Higher speeds were overall recorded in the simulated when compared to the real driving environment. On the contrary, steering angle was turned significantly more to the left in the real when compared to the virtual driving environment (Table 2). Lateral acceleration was almost the same for both environments.

Participants increased their speed when driving in the simulator (mean difference: 3.55 km/h). Despite the increase in speed, steering was more stable in simulated tasks when compared to real driving (mean difference: 30.79). This difference may be the result of greater variability in the driving context when compared to the simulated environment. The latter is supported also by the standard deviation values (SD difference: 0.144).

In addition, participants braked more in the simulated tasks (Table 1). Participants showed an increase of 12.6% in braking behaviour in the simulator. Overall, participants slightly increased their speed in the simulator but steering behaviour was increased and more variant in real context. These parameters may be more affected by other experimental task dependent characteristics such as speed, and path curvature. Hence, it is necessary to move forward to specific analysis per condition driven (i.e. intersection and car following).

Participants showed no difference in required mental effort for both environments (Figure 1). However, the participants had to try more and concentrate more in order to complete the intersection scenarios when compared to the car following scenarios.

Four participants reported that they believed they have driven worse than usual in the simulator for the intersection scenarios compared to no participants in the car following scenarios for both environments (Figure 2).

4. CONCLUSION

Overall comparisons in different parts of the analysis did not seem to present great discrepancies but as the level of analysis was becoming more specific towards scenario level the differences increased.

Two things are important to keep as main differences for the general picture of findings; differences in variations (i.e. standard deviations)

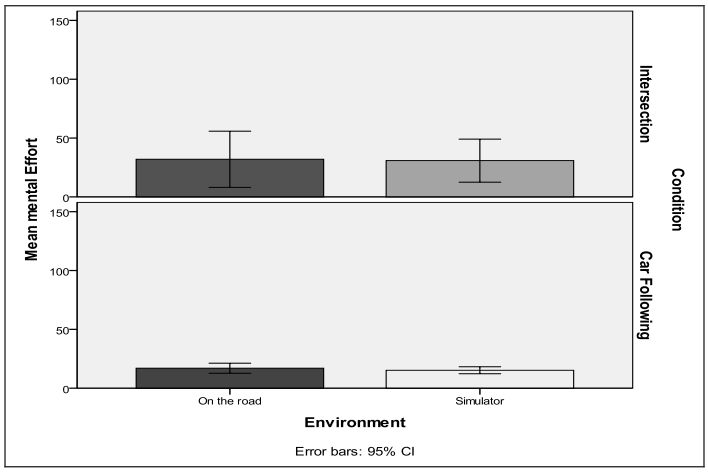


Figure 1: Mean mental effort scores (0-150)

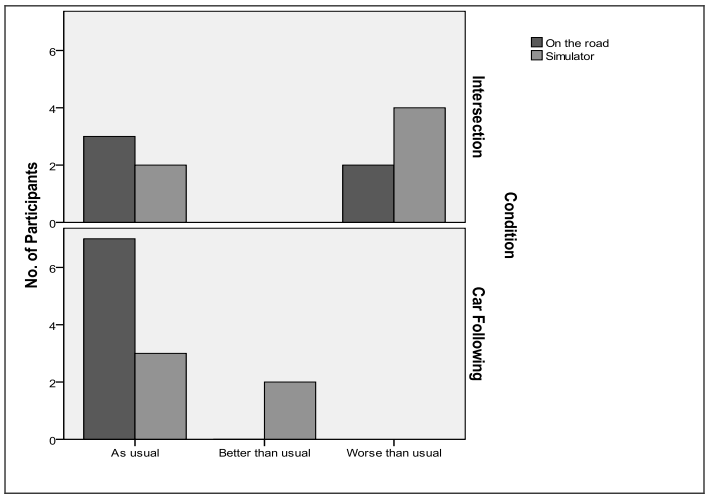


Figure 2: Subjective evaluation of driving quality per environment and per condition

with large variations found in the driving simulator and increased braking behaviour in the simulator regardless condition or scenario applied.

The speed increased in the driving simulator but the difference was really small (3.55 km/h). This finding supports the validity of the driving simulator with regards speed. Simulated environment affects the ability of drivers and potentially interferes with their learnt skills with regards speed adjustment (ie. large variations).

Despite the slight increase in speed, steering was more stable in simulated tasks when compared to real driving (mean difference: 30.79). This difference may be the result of greater variability in the driving context when compared to the simulated environment. The latter is supported also by the standard deviation values (SD difference: 0.144).

Lessons learnt

Intersections seem to be difficult to simulate because of generic characteristics difficult for participants such as curved road segments with 90 degrees. All participants in this study have driven in a static or semi-dynamic driving simulator before, therefore sickness was not high, but disorientation could not be avoided. Of course to none of the participants the level of sickness arose to points that had to stop the experiment but discomfort was mentioned and recorded (i.e. subjective scales). The research team had taken into consideration this possibility and deliberately added a small rural segment at the beginning of the all intersection scenarios in order to avoid to suddenly and rapidly introducing participants to urban environmental elements.

Future steps

The findings of this study could be highly useful for taken into account for further research efforts with emphasis at harmonisation and compilation of certain sets of vehicle parameters by the application of statistics and/or algorithms for developing scenarios with increased physiological and context validity These harmonised variables' sets could be tailored made for diverse interaction type scenarios such as the ones studied in these experiments (i.e. focus on the interaction between drivers and riders in specific accident prone road types). There is no

validation study specifically oriented on interaction. Thus, its findings could be used to improve existing simulators and considerations to be made for conspicuity research. In other words, the general pattern of changes on road was present in the simulator as it is evident by the overall small mean differences in some of the vehicle parameters. The controllability of those factors -that apparently are more context dependent -could allow for alterations and improvements that could facilitate refinement of driving simulator. A next step would be to directly address conspicuity with these results at hand and focus on gaze control with identification of "failed- to-see" instances in simulated tasks with taking in to account previous PTW related training to drivers. Training could include both PTW and rider specific information. In addition, subjective evaluations included in this study could provide the grounds for furthering comparability aspects with other conducted simulator studies, at least on qualitative and discussion aspects. This could be an opportunity to see if simulation of certain road conditions has higher inherent aspects for success.
Finally, controlling for variations and deviations in future in-depth studies focusing on complex environments that target interactions could be a next step. The driving simulator could yield similar findings (e.g. speed adjustment issues and lower lateral control) for relevant studies (e.g. interaction studies) but its sensitivity with regards to potential variations should be further researched.

5. REFERENCES

[1] National Highway Traffic Safety Administration. 2008 Traffic Safety Annual Assessment Highlights DOT 811 172. Washington DC: National Highway Traffic Safety Administration, 2009. http://www-nrd.nhtsa.dot.gov/Pubs/811172.pdf

[2] Clarke, D.D., Ward, P., Bartle, C. and Truman, W. (2004). In-depth study of motorcycle accidents. Road Safety Report No. 54. London: Department for Transport.

[3] ACEM (2004). MAIDS: In-depth investigation of accidents involving power-two wheelers. Assocation de Constructeurs Européens de Motorcycles (ACEM). Brussels. BG. http://www.maids-study.eu.

[4] Gershon, D.D., Ward, P., Bartie, C., Battle, C., Truman, W. (2004). In-depth study of Motorcycle Accident. Road Safety Research. Report No. 54. Department for Transport. London.

[5] Shinar, D. (2007). Traffic safety and human behavior. Motorcyclists and Riders of Other Powered Two-Wheelers. 16. 657–694.

[6] Crundall, D., Clarke, D., ward. P.. Bartle. C. (2008). Car drivers' skills and attitudes to motorcycle safety: A review. Road Safety Research Report No. 85. Department for Transport. London.

[7] Cannon, W.B. (1929). Bodily changes in pain, hunger, fear, and rage. New York: Appleton-Century-Crofts.

[8] Horswill, M.S., Helman, S., Ardiles, P.and Wann, J.P. (2005). Motorcycle accident risk could be inflated by a time to arrival illusion. Optometry and Vision Science. 82. 8. 197-230.

[9] Hurt, H.H. Jr., Ouellet, J.V., Thom, D.R. (1981). Motorcycle Accident Cause Factors and Identification of Countermeasures. 1. Washington DC: US Department of Transportation. National Highway Traffic Safety Administration.

[10] Brenac, T., Clabaux, N., Perrin, C., Van Elslande, P. (2006). Motorcyclist conspicuity-related accidents in urban areas: A speed problem? Advances in Transportation Studies. 8. 23-29.

[11] Brooks, A., Chiang,. D., Smith, T., Zellner, J., Peters, J., Campagne. J. (2005). A driving simulator methodology for evaluating enhanced motorcycle conspicuity. 19th International Technical Confrence on teh Enhanced Safety of Vehicles. National Highway Safety Administration.

[12] de Lapparent. M. (2006). Empirical Bayesian analysis of accident severity for motorcyclists in large French urban areas. Accident Analysis and Prevention. 38. 260-268.

Time Series Analysis of Heartbeat-Interval at the Subjects Ranging from Crustacean Animal to Human

Toru Yazawa
Biophysical Cardiology Research Group, Neurobiology Laboratory,
Tokyo Metropolitan University, Hachioji, Tokyo 192-0397, Japan

and

Tomoo Katsuyama
Department of Physics, Numazu National College of Technology, Numazu, 4108501, Japan

ABSTRACT

The aim of our study was to quantify the condition of the heart: sick or not in numerical order. In the present study, we focused attention on the period-2 heartbeat. "Alternans" is an arrhythmia exhibiting alternating amplitude/interval from beat to beat on the electrocardiogram and was first described in 1872 by Traube. Recently alternans was finally recognized as the harbinger of a cardiac disease, when an ischemic heart exhibited alternans. The pattern, alternans, arises spontaneously. As-yet-unidentified mechanisms must contribute. In animal models we detected alternans at various experimental conditions, including the heart under emotional stress and the heart of a dying specimen. We have tested the detrended fluctuation analysis (DFA) on alternans and revealed that in both, animal models and humans, alternans rhythm lowers the scaling exponent that was computed by the DFA. We concluded that the scaling exponent can reflect a risk for the "failing" heart, especially when the low scaling exponent and alternans are concurrently present.

Keywords: heartbeat, fuluctuation, DFA, animal model, alternans rhythm.

1. INTRODUCTION

Japanese persimmon trees bear rich fruits every other year. Atmospheric oxygen on the earth has bistability [1]. Period-2 is an intriguing rhythm in nature. The cardiac “alternans” is another period-2 phenomena. In cardiac period-2, the heartbeat is alternating the amplitude/interval from beat to beat. It can be seen on the electrocardiogram (EKG). Alternans has remained an electrocardiographic curiosity for more than three quarters of a century [2, 3]. To date, alternans is recognized as a marker for patients at an increased risk of sudden cardiac death [2, 3, 4, 5, 6,7].

We have been studying neurobiology of crustacean cardio-vascular system. In our physiological experiments on the hearts in the 1980’s, we have noticed that alternans is frequently observable with the “isolated” hearts (Note; the heart sooner or later dies in the experimental dish). We soon realized that it is a sign of future cardiac cessation. Nowadays, some authors believe that it is the harbinger for sudden death [2, 6]. So, we came back to the crustacean physiology. As-yet-unidentified mechanisms must contribute to this abnormal heartbeat. Details of alternans have not been studied in crustaceans. But, we considered that we may study this intriguing rhythm by the detrended fluctuation analysis (DFA), since we have demonstrated that the DFA can distinguish a normal heart (intact heart) from an unhealthy heart (isolated heart) in animal models [8]. In the present investigation, we describe that period-2, alternans, lowered scaling exponent.

2. MATERIALS AND METHODS

DFA Methods: Background

The DFA is based on the concept of “scaling” and “universality” [9]. It is a method to understand a “critical” phenomena [9, 10, 11]. Systems near critical points exhibit self-similar properties, and therefore, in physics, they are invariant under a transformation of scale.

Stanley and colleagues have considered that the heartbeat fluctuation is a phenomenon, which has the property of scaling. They first applied the concept to a biological data, the DNA and the EKG in the late 80’s to early 90’s [10, 11]. They emphasized on its potential utility in life science [11]. Technologically, it seems not matured, but practical use of nonlinear technology is widely accepted and increasingly

advancing.

DFA Methods
We made our own programs for measuring beat-to-beat intervals, and for calculating the approximate scaling exponent of the interval time series (K. Tanaka [13]). Those DFA-computation methods have already been explained elsewhere [12, 13]

Heartbeat recording
From human subjects we used the finger pulse recording with a Piezo-crystal mechanic-electric sensor, connected to a Power Lab System (AD Instruments, Australia). From model animals we used electrophysiological recording with two metal electrodes implanted to the dorsal carapace (pictures in Figs. 1 and 3). The attached electrodes often stayed on animal carapace until ecdysis.

Volunteers and ethics
Human subjects were selected from colleagues in our university, volunteers who were voluntarily visited us desiring their heart to be checked. All subjects were treated as per the ethical control regulations of our university, Tokyo Metropolitan University.

Model animals
It is very important that animal models are healthy before an investigation. We confirmed that all animals used were naturally healthy before starting any experiments. We observed with our own eyes when we captured all specimens from a natural habitat by ourselves. The EKG recordings from crustacean model animals were done by implanted permanent metal electrodes, which are connected to the Power Lab System. By this recording, animals were usually walked around in the container.

3. RESULTS AND DISCUSSION

It is known that the human heart rate goes up to over 200 beats per min (BPM), when life comes to an end. During the period through the brain death, a figure data is shown in a reference (see Fig. 1 of [14]).

In an animal model, an increase of heart rate during the dying period was observable, for example, crustaceans (Fig. 1) and insects (data not shown). This demonstrates a strong resemblance of a cardiac control mechanism between lower animals and humans. The heart of those animals is innervated from the autonomic nerve system. Evolutionally, this similarity was noticed (see, "Natural selection 150 years on", Nature insight Feb 12, 2009 pp808-811). There is another reason to use lower animals in medical investigations. Crab/shrimp are our foods. Ethics is of course a big requisition. Furthermore, we nowadays know Gehring's discovery of a gene, named homeobox: To our surprise at that time of his discovery, an identical gene named "pax-6" was found to work both, in fly's eyes and mouse's eyes at an embryonic stage for developing an optical sensory organ [15]. Fundamentally creatures are all the same in basic physiology. Both the crustaceans and humans have "the autonomic nerves," i.e., acceleratory and inhibitory cardio-regulatory nerves.

EKG from a dying crab/shrimp exhibited alternans (Fig. 1B). This alternans was followed by a period of high-rate heartbeats (Fig. 1C), which is an indication of terminal condition. Alternans is thereby the harbinger of death in crustacean animals. Interestingly no alternans is seen at the terminal condition (Fig. 1C).

The DFA of alternans revealed that the alternans exhibits a low approximate scaling exponent (see below). Alternans and low exponents would be a sign of illness. We have long noticed that the isolated heart, which can repeat contractions for hours in a dish, often exhibits alternans (Fig. 2B1). The DFA revealed that the scaling exponent of alternans is low, i.e., slope is far lower than one (Fig. 2B2). We tested another three-isolated hearts of this lobster species, all of which exhibited alternans (data not shown), and we found that the scaling exponent of the alternans' heart was low. A healthy lobster, however, exhibits a normal scaling exponent (see Figs. 2A1 and 2A2).

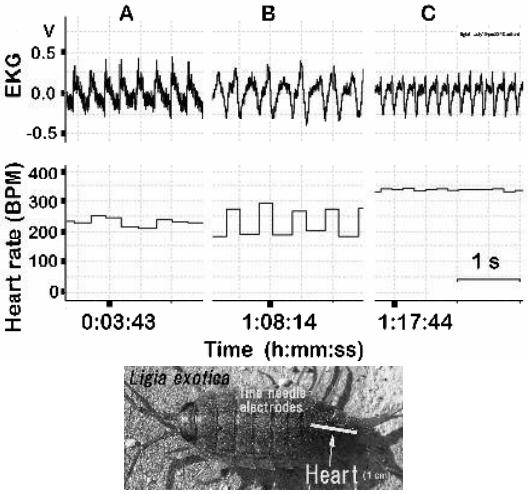


Figure 1. EKG from a dying isopod crustacean, *Ligia exotica*. (A) A recording started at time zero. The base line heart rate is about 250 beat per minute. (B) One hour after (A), an irregular rate and alternans can be seen. (C) About 1.2 hours after (A), no alternans is seen. The heart rate increased to about 350 beats per minute. This crab died 1.3 hours after the recording (A).

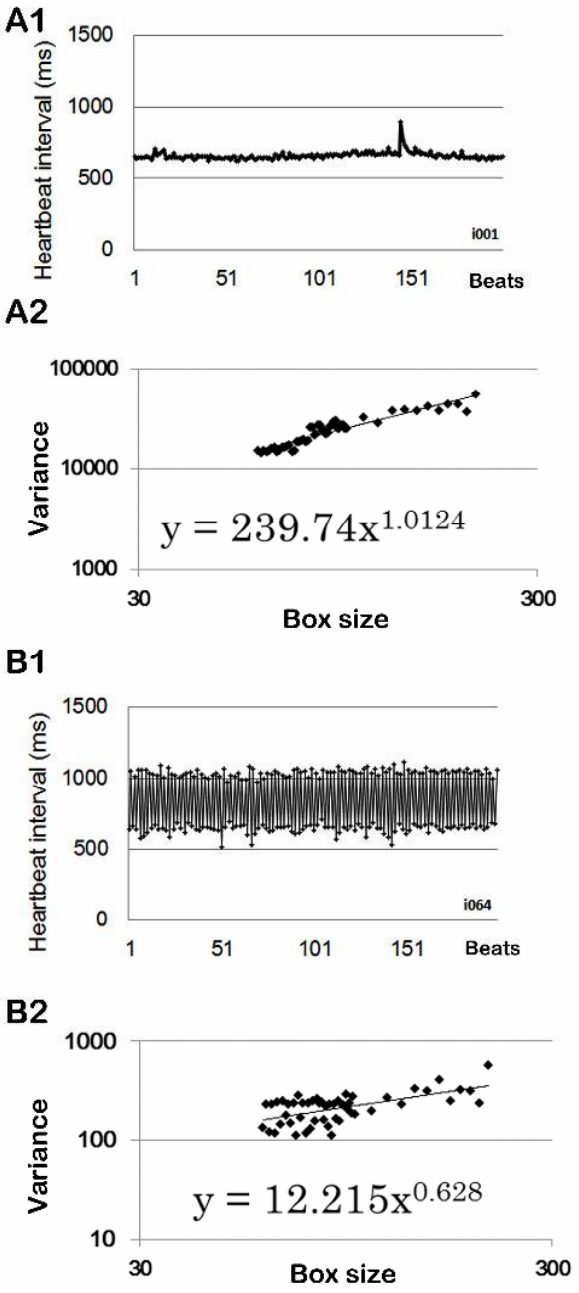
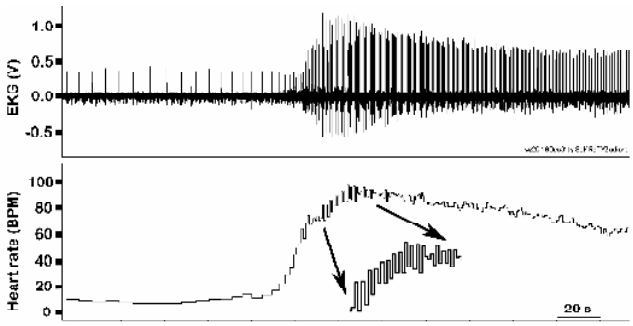


Figure 2. Intact (A) and isolated (B) heart of a spiny lobster, *Panulirus japonicus*. A1 and B1, 100 heartbeats. A2 and B2, DFA computation. A2, Intact heart exhibited a scaling exponent 1.0124. No alternans appeared (see A1). B2, Alternans appeared here all the way down from the first beat to 4000th beat (not shown all). The scaling exponent was found to be low, 0.628, which was computed at box size 60-210.

We once drew a conclusion that freely moving animals without stress do not exhibit alternams and they exhibit a normal scaling exponent. However, further studies uncovered that a crayfish can spontaneously exhibit alternans when it had an emotional arousal (Fig. 3). This alternans is apparently not dying. We finally confirmed alternans is recorded especially at the top speed of its racing heartbeat (Fig. 3). So far we do not know the mechanism of generation of alternals. It seems that alternans comes out from complex interactions between the heart and brain, such as psychological excitement. This type of alternans does not last for long, 30 second at the most (Fig. 3). We cannot apply the DFA to such an interesting psychological or emotional phenomena, because the DFA needs about 2000 heartbeats. If it could last for 5 minutes, we could do calculate approximate scaling exponent of alternans, and we did it (Fig. 2).



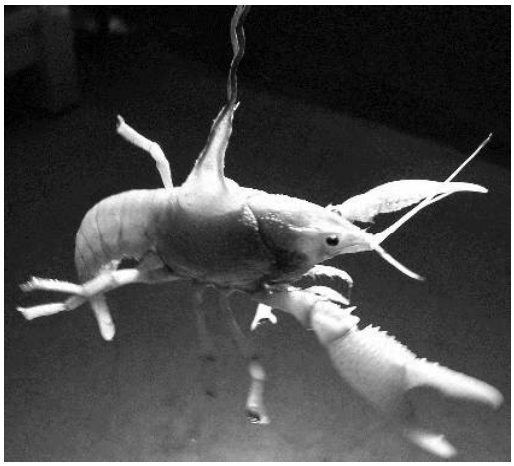


Figure 3. Crayfish *Procambarus clarkii* exhibits alternans. It occurs at a top speed of a cardiac acceleration. This occurs spontaneously when the animal was in the shelter. Upper, EKG. Lower, Heart rate in beat per min. Inset, Enlargement of Alternals. Both amplitudes alternans and intervals alternans can be seen (see A). Picture: Crayfish hanging in the air with EKGs recording wires.

Finally, we studied the human alternans. The finger pulse of a volunteer was tested (Figs. 4 and 5). Similar to the models, human alternans exhibited a low exponent (Fig. 6). This subject, a 65 years old female is physically weak and she cannot walk a long distance. However, she talked with an energetic attitude. She was at first nervous because of us, but finally she got accustomed to our finger pulse testing task, and then she became relaxed. Hours later, we were surprised to note that her alternans decreased in numbers. The heart reflects the mind. We observed that alternans is coupled with the psychological condition, probably with an impulse discharge frequency of the autonomic nervous system. This is fundamentally similar to the animal models.

Alternans is believed to be the harbinger for a sudden death [2, 6]. A low scaling exponent accompanied with alternans seems to be a serious case. In contrast, it is believed that healthy human hearts exhibit a scaling exponent of 1.0 [11]. We obtained the same results (data not shown).

Figure 7 shows an example of alternans in which the subject's heartbeats show a period-2 rhythm (Fig. 7A). We met him in 2007 at an exhibition called Innovation Japan 2007. At that time he at first told us that he knew that his heart was not normal and he regularly went to see a doctor. He visited us because we presented our DFA method at the exhibition. We recorded his heartbeat (Fig. 7A). His scaling exponent in September 2007 was 0.6709 (see Figure 7B and 7C). We explained to him that he had alternans, and what is alternans as far as we knew.

In September 2008 and 2009, we returned to the exhibition. We did not expect he visited us again but he came. We recorded his heartbeat and calculated his scaling exponent. Especially in September 2009, to our surprise, the alternans was almost gone, and his scaling exponent increased than in 2007 (see Fig. 8, the scaling exponent was 0.733). He said that since learning the results shown in Fig. 7 in 2007, he had been walking to work instead of driving, every day. In September 2010, we returned to the exhibition and he visited us. He was mandatory retired from work in 2009, and he said he is staying at home. Figure 9 shows that an increase of the scaling exponent from 2007 to 2009 stopped in 2010 because of retirement.

His scaling exponents were compared in Fig. 9. Three different DFA computations at different box size are shown. Years 2007 - 2010 are compared. Figure 9 summarizes that his scaling exponents were improved although retirement stopped it. When he was working at Toshiba Co. Ltd, the exponents shifted to the direction toward to the good health state which is ultimately 1.0. We would say: "Efforts bear fruit." "DFA is useful."

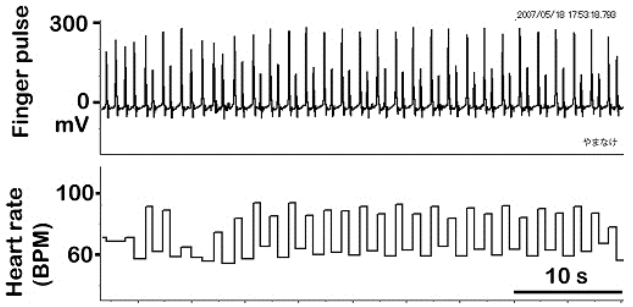


Figure 4. Human alternans. A volunteer woman age 65. Upper trace, recording of finger pulses. Lower trace, heart rate.

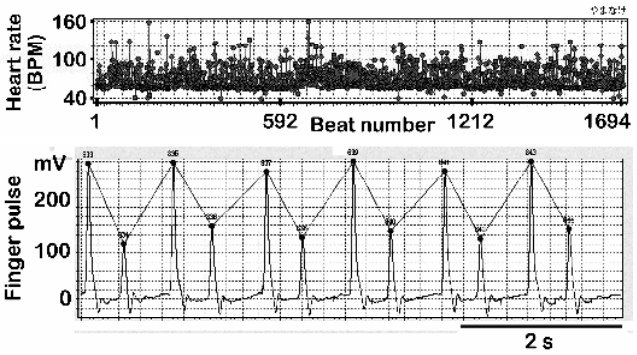


Figure 5. Human alternans. The same volunteer woman shown in Fig. 4. Upper trace, Time series, 1701 beats. Lower trace, Example recording, beat number 833 – 844. Both amplitudes alternans and intervals alternans can be seen.

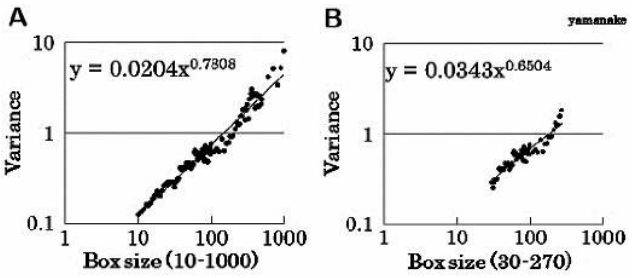


Figure 6. Result of the DFA of human alternans shown in Figs. 4 and 5. A, Box size, 10-1000. B, Box size, 30-270. Straight lines represent a slope obtained from different box-size-length. Theoretically, the slop determines the approximate scaling exponent. The 45-degree slope (not shown here) gives the scaling exponent 1.0 which represents that the heart is totally healthy. The alternans significantly lowers the scaling exponent in human heart as well as in crabs.

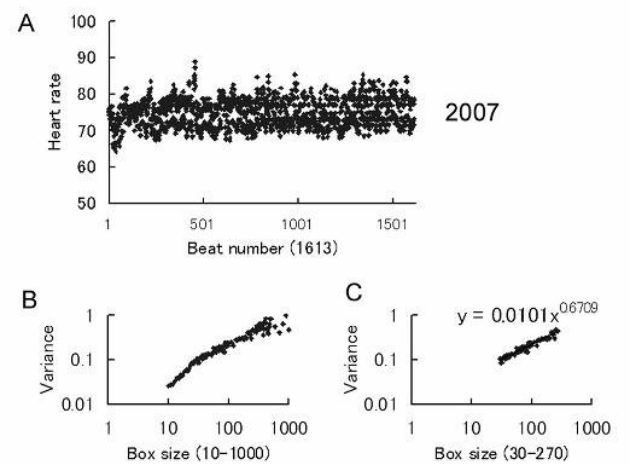


Figure 7. DFA of alternans. Recorded in September 2007 in Tokyo in his age 60s.

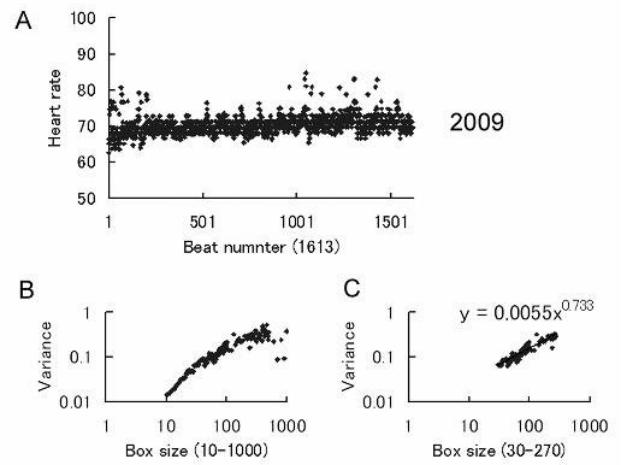


Figure 8. DFA of alternans. Recorded again in September 2009 in Tokyo.

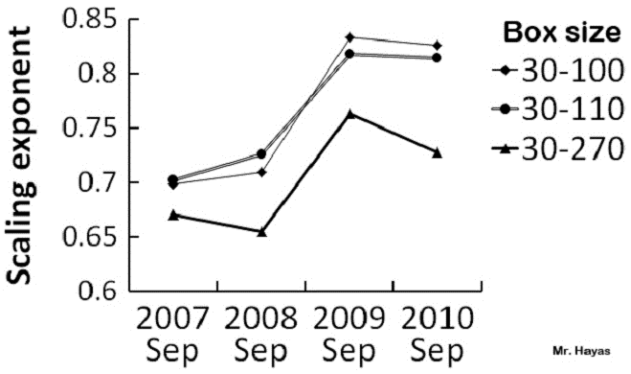


Figure 9. Comparison of the scaling exponents at different box size. Computed from the data from the subject shown in Figures 7 and 8.

A real time observation of both, EKG-signal and biological condition of a specimen, is important to interpret the physiological meaning of the scaling exponent. We did all investigations in front of us together with the on-line EKG observation. We thus were able to interpret/speculate the direct relationship between EKG-changes to behavioral changes. It seems that, until recently, clinicians took recordings and physicists analyzed the recordings. We performed both, recordings and analysis by the same scientists. This is one important point of advantage in this study, compared with the previous DFA report.

Other important points of our present study are that we made our own PC program (program by K. Tanaka [13]), which assisted the accuracy of the peak-identification of heartbeats, and then the calculation of the scaling exponent. Furthermore, supported by the real time observation, we were able to distinguish numerically and quantitatively normal hearts from abnormal hearts. It is said that alternans is the harbinger of a sudden death of humans. That was true in dying models. However, alternans was detectable everywhere in models; for example, during emotional changes (Fig. 3). Therefore stressful psychological circumstances may invoke autonomic acceleratory commands in the brain and then the commands finally trigger alternans in the heart.

4. CONCLUSION

We have previously made an automated DFA computation program which enabled us to perform variety of analysis on the heartbeat obtained both from model animals and human subjects. Using the original program, we were able to determine fine or poor health for each subjects: An appropriate box size is about 30-270 beats. Finally, we reached the conclusion that alternans lowers the approximate scaling exponent. Furthermore, alternans appears not only in dying conditions but also in emotionally stressful conditions.

5. ACKNOWLEDGEMENT

This research was supported by a Grant-in-Aid for Scientific Research No. 23500524 to TY. We thank DVX Co. Ltd. for financial support for the research.

6. REFERENCES

[1] C. Goldblatt, T. M. Lenton, and A. J. Watson. “Bistability of atmospheric oxygen and the great oxidation”, Nature, Vol. 443, 2006, pp. 683-686.

[2] D. S. Rosenbaum, L. E. Jackson, J. M. Smith, H.

Garan, J. N. Ruskin, and R. J. Cohen, “Electrical alternans and vulnerability to ventricular arrhythmias”, The New England J. of Medicine, Vol. 330, 1994, pp. 235-241.

[3] B. Surawicz, and C. Fish. “Cardiac alternans: diverse mechanisms and clinical manifestations”, Journal of American College of Cardioloy, Vol. 20, 1992, pp. 483-99.

[4] M. R. Gold, et al. A comparison of T-wave alternans, signal averaged electrocardiography and programmed ventricular stimulation for arrhythmia risk stratification. Journal of the American College of Cardiology. Vol. 36, No. 7, 2000, pp. 2247-2253.

[5] A. A. Armoundas, D. S. Rosenbaum, J. N. Ruskin, H. Garan, R. J. Cohen. Prognostic significance of electrical alternans versus signal averaged electrocardiography in predicting the outcome of electrophysiological testing and arrhythmia-free survival. Heart. Vol. 80, 1998, pp. 251-256.

[6] B. Pieske, and K. Kockskamper. Alternans goes subcellular. A “disease” of the ryanodine receptor? Circulation Research, Vol. 91, 2002, pp. 553-555.

[7] K. Hall, D. J. Christini, M. Tremblay, J. J. Collins, L. Glass, and J. Billette. Dynamic control of cardiac alternans. Physical Review Letter. Vol. 78, 1997, pp. 4518-4521.

[8] T. Yazawa, K. Kiyono, K. Tanaka, and T. Katsuyama. Neurodynamical control systems of the heart of Japanese spiny lobster, Panulirus japonicus. Izvestiya VUZ. Applied Nonlinear Dynamics. Vol.12, No. 1-2, 2004, pp. 114-121.

[9] H. E. Stanley. Phase transitions. Power laws and universality. Nature, Vol. 378, 1995, p. 554.

[10] C. -K. Peng, S. Havlin, H. E. Stanley, and A. L. Goldberger, "Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series", Chaos, Vol. 5, 1995, pp. 82-87.

[11] A. L. Goldberger, L. A. N. Amaral, J. M. Hausdorff, P. C. Ivanov, and C. –K. Peng, Fractal dynamics in physiology: Alterations with disease and aging. PNAS, Vol. 99, 2002, suppl. 1, pp. 2466-2472.

[12] T. Katsuyama, T. Yazawa, K. Kiyono, K. Tanaka, and M. Otokawa. Scaling analysis of heart-interval fluctuation in the in-situ and in-vivo heart of spiny lobster, Panulirus japonicus. Bull. Univ. Housei Tama Vol. 18, 2003, pp. 97-108 (Japanese).

[13] T. Yazawa and K. Tanaka, “Scaling exponent for the healthy and diseased heartbeat. Quantification of the heartbeat interval fluctuations”, Advances in Computational Algorithms and Data Analysis. Chap. 1. pp1-14. Springer-Verlag New York Inc (C) (2008/11) ISBN-13: 978-140208918

[14] F. Conci, M. Di Rienzo, and P. Castiglioni, “Blood pressure and heart rate variability and baroreflex sensitivity before and after brain death”, Journal of Neurol. Neurosurg. Psychiatry, Vol. 71, 2001, pp. 621-631.

[15] W. J. Gehring. Master Control Genes in Development and Evolution: The Homeobox Story, New Haven: Yale University Press, 1998.

Decomposition of Radar Data Through Intelligent Successive Regression

Michael Anthony Johnson
ECE Department, The University of Alabama in Huntsville
Huntsville, Alabama 35899, USA

and

Reza Adhami
ECE Department, The University of Alabama in Huntsville
Huntsville, Alabama 35899, USA

Abstract

This paper details the decomposition of the amplitude track of a low-angle target to allow independent analysis of target radar cross section (RCS), multipath, and target roll dynamics. This analysis is applicable to ballistic test range radar that tracks surface-to-surface targets. Through the method developed in this paper, surface parameters may be obtained by analysis of the amplitude track. A method of successive fitting and subtraction called A Priori Mode Decomposition (AMD) is developed. This method is similar to Empirical Mode Decomposition (EMD) but with a priori knowledge of the system. AMD was applied to a simulated radar test environment in systematic fashion over a portion of parameter space showing robustness to changes in the modeled system parameters and noise.

Keywords: Successive Regression, Empirical Mode Decomposition, EMD, Multipath, Radar, A Priori Mode Decomposition, AMD

1 Introduction

This paper details the decomposition of the amplitude track of a low-angle target to allow independent analysis of target radar cross section (RCS), multipath, and target roll dynamics. This analysis would be applicable to a ballistic test range radar that tracks surface-to-surface targets. Typically amplitude data is not used beyond detection, but through the method developed in this paper additional surface parameters may be obtained. In Section II, an amplitude model of a target over a rough surface is developed for use in simulation of the target return at low-angle. Radar tracks are then analyzed using Fourier techniques and Empirical Mode Decomposition (EMD) is used in an attempt to isolate the basic target return from the multipath and target fluctuation effects. A method of successive fitting and subtraction is developed similar to EMD but with a priori knowledge of the system. This method is then applied to tracks generated over a range of system parameters to determine the robustness of the method.

2 Simplified Signal Model

Overview

We will consider the case shown in Figure 1 where a monostatic radar is located a distance R_z above the ground on the z-axis. The target to be tracked will fly a path parallel to the x-axis offset in the y dimension by T_y at an altitude of T_z above the ground. The terrain is approximated by the xy plane. The radar will be used for range r, and received power estimation P_{rx}.

Table 1: Radar Parameters

Parameter	Value
Radar Transmit Power	$P_t = 40.0\,dBm$
Antenna Gain	$G_r = 20.0\,dB$
Wavelength	$\lambda = 7.5mm$
Detection Bandwidth	$B = 3.0\,kHz$
System Noise Figure	$F = 10.0\,dB$
Radar Altitude	$R_z = 3.0\,m$

Table 2: Target Parameters

Parameter	Value
Target Cross Section	$\sigma_{rcs} = -15.0\,dBsm$
Target Altitude	$T_z = 2.0\,m$
Target Cross Range	$T_y = 1.0\,m$
Target Roll Rate	$F_s = 0.10\,{}^1/_m$
RCS Flutter	$A_r = 0.2\,dB$

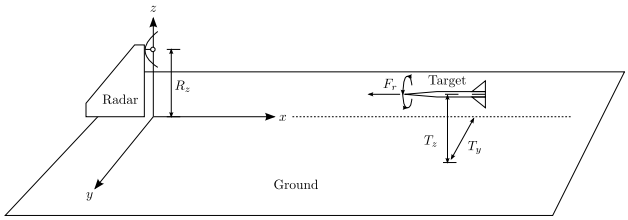


Figure 1: Coordinate System

The parameters considered for the baseline simulation are summarized in Table 1, 2, and 3.

Radar Range Equation

If we consider the target to be a non-fluctuating target, then return is not a random variable and will be considered simply to be the

Table 3: Terrain Parameters

Parameter	Value
Conductivity	$\mho = 1.0 \times 10^{-3}\, \mho/m$
Relative Dielectric Constant	$\epsilon = 10$
RMS Surface Roughness	$\sigma_h = 40.00\, mm$
Foliage Attenuation	$\alpha = 0.24\, Np/m$

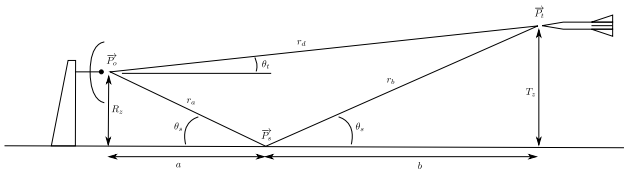


Figure 2: Multipath Paths

received power P_s as calculated by the radar range equation.[1]

$$P_s = \frac{P_t 2 G_r \lambda^2 \sigma_{rcs}}{(4\pi)^3 r^4} \tag{1}$$

Using this equation and the values from Tables 1 and 2, we can generate the simulated received power vs range plot shown in Figure 3a.

Multipath

When the target is either low or distant, the amount of multipath effect may be significant. If we consider specular multipath only, the four primary signal paths we will consider are shown in Figure 2 and are: the direct path $l_d = 2r_d$, the indirect path through the specular point outgoing $l_a = r_a + r_b + r_d$, the indirect path from the target back to the radar $l_b = r_d + r_b + r_a = l_a$, and the indirect path both directions $l_c = r_a + r_b + r_b + r_a$.

For a trajectory over a perfect electrical conductor (PEC), we would expect the received signal to receive interference from each path equally resulting in the received signal modulated by a propagation factor:

$$F_{pec} = e^{jl_d 2\pi/\lambda} + 2e^{jl_a 2\pi/\lambda} + e^{jl_c 2\pi/\lambda} \tag{2}$$

This has a "chirp" characteristic as a function of range, the signal can destructively cancel and also constructively add, since the signal can be increased by up to four times the propagation factor is as much as $12dB$.

We are unlikely to encounter data taken over a perfect conductor. Initially we will consider a perfectly smooth, perfectly flat dielectric slab and consider its reflection coefficient similar to the development in [2]. The terrain will be assumed to have a relative dielectric constant

$$\epsilon_{rc} = \frac{\epsilon}{\epsilon_o} - 60i\lambda\sigma \tag{3}$$

where we will initially assume the nominal value of $\epsilon = 10\,\epsilon_o$ and $\sigma = 1.0 \times 10^{-3}\, \mho/m$. Additionally, the permeability shall be assumed to be $\mu_{rc} = 1$ and the normalized admittance:

$$Y = \sqrt{\frac{\epsilon_{rc}}{\mu_{rc}}} \tag{4}$$

A vertically polarized system was modeled using reflection coefficients R_o derived in [3][2] and given below:

$$R_o(\theta_g) = \frac{Y^2 \sin\theta_g - \sqrt{Y^2 - \cos^2(\theta_g)}}{Y^2 \sin\theta_g + \sqrt{Y^2 - \cos^2(\theta_g)}} \tag{5}$$

A target flying over this smooth terrain is not affected much (less than 3dB) by the terrain at the small grazing angles ($\theta_g < 3^o$) that we see over most of the flight path. No terrain is perfectly flat, so we must modify the reflection coefficient due the the effect of uneven rough terrain. For terrain with a Gaussian distribution of terrain height where the rms roughness is $\sigma_h = 40.00\, mm$ we can use the following relation derived in [2] and given below:

$$R_s = \rho_s R_o \tag{6}$$

where

$$\left\langle |\rho_s|^2 \right\rangle = e^{-\left(\frac{4\pi\sigma_h \sin\theta_g}{\lambda}\right)^2} \tag{7}$$

The reflection coefficient rapidly approaches zero for even small angles at our wavelength.

Finally at extremely low grazing angles on rough foliage-covered surfaces such as grass, the incident and reflected rays must pass through a significant amount of biomass. In this case, the path length traversed by the reflected ray is determined by the height of the ground cover which will be taken to be the same as the RMS surface roughness σ_h resulting in a path length:

$$d = \frac{2\sigma_h}{\sin(\theta_g)} \tag{8}$$

which results in a path attenuation of:

$$\langle A \rangle = e^{-2\alpha d} = e^{-4\alpha\sigma_h/\sin(\theta_g)} \tag{9}$$

where α is the path attenuation constant. We will use $\alpha = 0.24\, Np/m$ for the frequency of interest found in[4] and a mean foliage cover height of$\sigma_h = 40.00\, mm$.

If we use these reflection coefficients and apply them to Equation 2 we get

$$\begin{aligned} F_p =& e^{jl_d 2\pi/\lambda} \\ &+ 2R_o(\theta_g)\,\rho_s(\theta_g)\,A(\theta_g)\,e^{jl_a 2\pi/\lambda} \\ &+ R_o^2(\theta_g)\,\rho_s^2(\theta_g)\,A^2(\theta_g)\,e^{jl_c 2\pi/\lambda} \end{aligned} \tag{10}$$

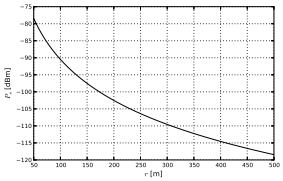
which results in a ripple seen in Figure 3b. We can see in this figure that since the effect of surface roughness drives the reflection coefficient to zero at steep angles and the attenuation drives it to zero at extremely shallow angles. The surface roughness effect dominates at steep angles presented at closer ranges and the foliage absorption dominates at the extremely shallow angles presented at distant ranges.

RCS Modulation

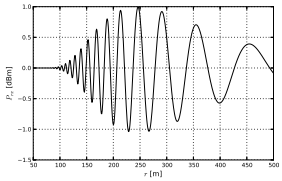
Projectiles are commonly spin-stabilized to enhance stability during flight. The presence of fins can make the RCS of the target vary as a function of roll angle.[5] If we assume the target is traveling parallel to the x-axis and it rolls once every R_d, the spatial frequency is $F_s = {}^1/_{R_d} = 0.10\,{}^1/_m$ and raises or lowers the RCS by $A_r = 0.2\, dB$. We can model this fluctuation with:

$$RCS_{roll}(x) = RCS_o + A_r \sin(2\pi x F_s + \phi) \tag{11}$$

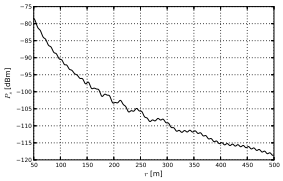
If the target is spin-stabilized by fins once the roll rate achieves steady state, the spatial frequency of the RCS variation is primarily dependent on velocity. If we observe the target over a short distance the roll rate will be constant. It should be noted that R_d is the distance required to roll to a similar symmetric orientation, e.g., if the target has 4 fins it would roll through a cycle 4 times per actual revolution of the target. If we add the fluctuating cross section according to the parameters in Table 2, we get the result shown in Figure 3d.



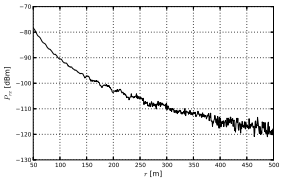
(a) Target Return Power vs Range



(b) Propagation Factor Over Rough Ground with Foliage

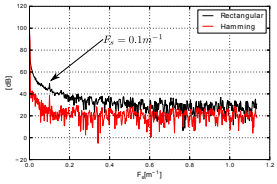


(c) Rolling Target Over Rough Terrain

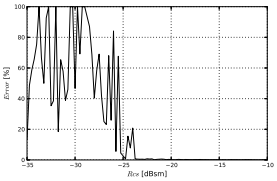


(d) Fully Modeled Power Return

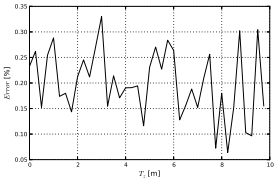
Figure 3: Target model



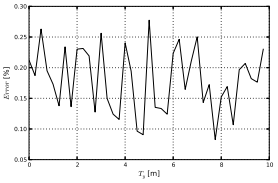
(a) DFT of target track



(b) Estimate over varied RCS



(c) Estimate over variedT_z



(d) Estimate over variedT_y

Figure 4: Median estimate of F_s using Fourier methods

Noise

We will additionally assume that the I and Q channels have identical independently distributed [i.i.d.] white noise with power $\sigma_g^2/2$ in each for a total noise power of σ_g^2 in the signal z. The total noise power in ν is σ_g^2 which is given by

$$\sigma_g^2 = kTBF \tag{12}$$

where thermal noise density is $N_o = 10\log(kT) = -174.0\,dBm/Hz$. This allows us to model the received power at the detection of

$$P_{rx} = P_s F_p + \nu \tag{13}$$

The received power of the signal corrupted by all sources including noise in Figure 3d.

3 Signal Decomposition

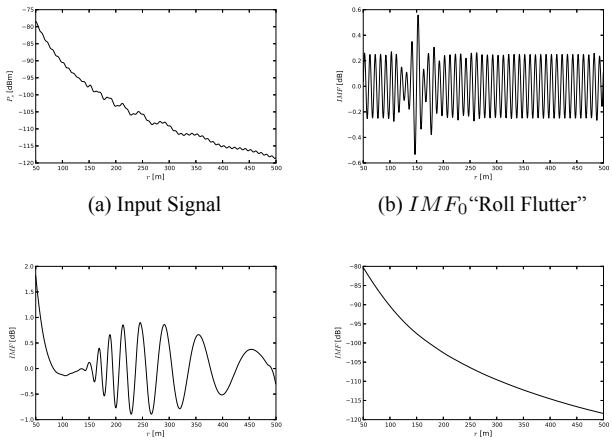
The signal represented in Figure 3d is a linear combination of well defined modes when considered in logarithmic (dB) scale. The roll fluctuation is a constant amplitude sinusoidal oscillation so should yield good results with Fourier techniques. Fourier analysis is severely limited when applied to an intensely non-stationary signals such as the radar amplitude track and multipath propagation factor. With the limitations of Fourier methods it looked promising to explore application of Empirical Mode Decomposition (EMD) to the problem. Introduced by Norden E Huang in 1998 [6], EMD is singular in its ability to derive basis functions (Intrinsic Mode Functions) from the data set itself rather than pulling from a fixed family of basis functions like the Fourier or Wavelet transform. It is well-suited to decomposition of strongly non-stationary and non-linear data sets owing to its method of defining frequency content based explicitly on local zero crossings in the data rather than fitting to an ensemble of oscillatory functions. This ability to deal with data sets with non-stationary frequency and large trends made it look like a good choice to use with this type of data set.

FFT

Using the target model described by Equation 13, we will attempt to analyze the target track using Fourier methods. Initially we will perform the DFT on the track shown in Figure 3d with and without Hamming weights applied, the result is shown in Figure 4a. Since the track has a large $A_r = 0.2\,dB$ fluctuation at a spatial frequency of $F_s = 0.10\,{}^1/_m$, we can easily see a peak in the windowed FFT; it is however much less obvious in the rectangular FFT. Having seen the ability of the windowed FFT to reveal the roll flutter peak, a simple peak interpolator was used to estimate the frequency and amplitude of the roll flutter using the FFT. The simulation was swept through parameter space by varying the RCS of the target from -35 to -10 dBsm and then adjusting the cross range of the target T_y and target altitude T_z from 0 to 10 meters at the nominal RCS of $-15.0\,dBsm$. At each point in parameter space estimates of the roll parameters were made in Monte Carlo fashion by recording the median result of 10 noise realizations. The results of this are shown in Figure 4. It can be seen that the method estimates the roll frequency well when the RCS is bigger than -27dBsm. The estimate quality does not appear to be impacted by the altitude or cross-range position of the target as long at the RCS is large enough.

EMD

The noise-free version of the simulated received signal shown in Figure 3c was decomposed using EMD into a set of IMFs and the result is shown in Figure 5. The results shown in Figure 5 exhibit very clear separation of the modes allowing clear association of each IMF with the basic ${}^1/_{r^4}$ power roll-off, multipath propagation factor, or the roll-induced flutter. Clear as they are, there was some coupling between the ${}^1/_{r^4}$ roll-off and the multipath modes. Once noise is added and the signal shown in Figure 3d was decomposed, the results changed dramatically. It appears that decomposition in the presence of noise results in over sifting, which was exacerbating boundary effects which resulted in highly non-physical IMFs. These IMFs are very dependent on the particular realization of noise, little of the characteristics of the signal were discernible. This inconsistency of the transform result was be quantified by looking at the orthogonality index θ, one thousand noise realizations of the baseline were decomposed and only 68% had $\theta > 1.0$, compared to $\theta = 0.02$ for the noise free case. In addition to poor mode orthogonality the number of IMFs found varied between 6 and 8 suggesting non physical mode choices.



(a) Input Signal (b) IMF_0"Roll Flutter"

(c) IMF_1 "Multipath" (d) Residual "$1/r^4$"

Figure 5: EMD Derived IMFs of Noise Free Signal

4 AMD

The strength of EMD is that no knowledge of the underlying signal is required to decompose the signal. This is a significant asset when attempting to decompose signals where there is an incomplete understanding of the underlying process, such as with the sunspot cycle or turbulent pressure data.[7] In the case of the radar signal considered in this paper, the governing equations of the components of the signal are know a priori. It seems reasonable that we should be able to use that information to form a better decomposition of the signal.

In essence, the EMD method sifts high-frequency information from the signal, assigns it as an IMF, and then subtracts it from the signal. This process is then repeated until no oscillatory data remains. Inspired by the EMD method, an algorithm was developed to attempt to fit a model of each underlying process, form a signal for this "a priori mode function" (AMF) and then subtract it from the signal. This process is repeated until all the known modes have been estimated and removed. The residual should then consist of unmodeled dynamics and noise.

Justification for the method

The most straightforward method is to look at the residual between the signal and a mode function

$$g\left(x,\vec{p}\right)=R_{i-1}\left(x\right)-m_i\left(x,\vec{p}\right) \tag{14}$$

where $R_i\left(x\right)$ is the current signal, $m_i\left(x,\vec{p}\right)$ is the mode constructed with parameters $\vec{p}$ over x. A cost function such as the ℓ_1 norm then assigned and a minimization program is run

$$\min\sum_{x=x_{min}}^{x_{max}}\left|g\left(x,\vec{p}\right)\right|=\min\ \ell_{1g} \tag{15}$$

the $\vec{p}$ that forms the minimum is then assigned as a mode

$$a_i\left(x\right)=m_i\left(x,\vec{p}_{min}\right) \tag{16}$$

and subtracted from the signal.

$$s_i\left(x\right)=R_{i-1}\left(x\right)-a_i\left(x\right) \tag{17}$$

At this point we already diverge from EMD; EMD estimates the IMFs using the envelope of the signal. This has the effect of estimating the mode with the highest frequency content first

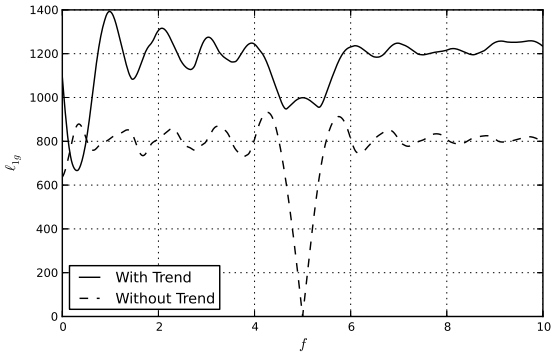


Figure 6: ℓ_1 Norm vs Frequency with and without Trend

and removing it in roughly decreasing order of frequency until a residual is left. In order for above algorithm to succeed, $g\left(x\right)$ must have a minimum. This is made more likely by assigning $m_i\left(x,\vec{p}\right)$ that estimates the local mean or trend of the signal. To make clear why this important step is necessary, let us consider the simple tone with a linear trend

$$s\left(t\right)=2t+\sin\left(2\pi ft\right) \tag{18}$$

where $f=5$ and $0.0\leq t<1.0$.

If we go on to form the cost function $g\left(t,f\right)=s\left(t\right)-\sin\left(2\pi ft\right)$ and generate the value of the ℓ_1 norm over frequency we would get the solid line in Figure 6 which would have directed an unconstrained minimization program to identify $f=0.3$ as the global minimum. Even if we had constrained the search near the value of $f=5.0$, the program would still have chosen either $f=5.3$ or $f=4.7$. If the trend-line is removed first, then the behavior shown by the dashed line in Figure 6 is achieved which would clearly choose the proper answer of $f=5.0$. This simple example illustrates the importance of removing the low-frequency trend information before attempting to fit to higher frequency oscillatory modes.

Description of the Algorithm

The algorithm can be described by the following 6 steps:

1. The known modes should be arranged by inspection from lowest frequency content to highest $m_0\left(x,\vec{p}\right)\ldots\ m_n\left(x,\vec{p}\right)$
2. Set $k=0$ and designate the input signal $R_0\left(x\right)$
3. Fit $m_k\left(x,\vec{p}\right)$ to $R_k\left(x\right)$ to form an estimate $\hat{\vec{p}_k}$ of the mode parameters
4. Form the first IMF, $IMF_k=m_k\left(x,\hat{\vec{p}_k}\right)$
5. Form the working residual $R_{k+1}\left(x\right)=R_k\left(x\right)-IMF_k$
6. Set $k=k+1$ and repeat at step 3 until all known modes are complete

Application to Model

To explain the algorithm more fully, an example decomposition of the radar track signal generated with radar parameters from Table 1 as well as the target parameters shown in Table 2 and terrain parameters in 3 will be performed. It is assumed that the radar data collected is in a test range environment, so the radar operating parameters and flight path is known via the normal radar range and bearing operation. Thus R_z, T_y, T_z are known but the rest of

the parameters must be estimated from the amplitude return data shown in Figure 3d.

Arrangement of Modes: Step one is to arrange the a priori modes into ascending order of frequency content. The overall amplitude return signal is composed of the r^{-4} roll off in the radar range equation, the chirped frequency pulse of the multipath components, the steady tone of the roll flutter, and noise.

Since the radar range equation has no oscillating components it is assigned first. The equation is given in Equation 1 and used as m_0 where the only unknown parameter to estimate is σ_{rcs}:

$$m_0\left(r, \langle \sigma_{rcs} \rangle\right) = \frac{P_{tx} G_r G_t \lambda^2 \sigma_{rcs}}{(4\pi)^3 r^4} \tag{19}$$

Next, the multipath signal described in Equation 10 is used as m_1 where the unknown parameters are σ_h, and α. Since the target track and is known $\theta_g\left(r, T_x, T_z, R_z\right)$ and the rest of the radar parameter such as P_{tx} and λ are constant.

$$\begin{aligned} m_1\left(r, \langle \sigma_h, \alpha \rangle\right) =& e^{jl_d 2\pi/\lambda} \\ &+ 2R_{\rho o}\left(\theta_g\right) \rho_s\left(\theta_g, \sigma_h\right) A\left(\theta_g, \sigma_h, \alpha\right) e^{jl_a 2\pi/\lambda} \\ &+ R_o^2\left(\theta_g\right) \rho_s^2\left(\theta_g\right) A^2\left(\theta_g\right) e^{jl_c 2\pi/\lambda} \end{aligned} \tag{20}$$

The roll-induced amplitude flutter described in Equation 11 is used as m_2, where the unknown parameters are the amplitude A_r, frequency F_s, and phase ϕ.

$$m_2\left(r, \langle A_r, F_s, \phi \rangle\right) = A_r \sin\left(2\pi r F_s + \phi\right) \tag{21}$$

Range Equation: Once the modes have been ordered, steps 2-5 can be performed. The signal shown in Figure 3d is assigned to R_0 and then fit to $m_0\left(r, \langle \sigma_{rcs} \rangle\right)$ using the downhill simplex algorithm. This resulted in an estimate of RCS $\hat{\sigma}_{rcs} = -15.0\, dBsm$ which is then used to generate a mode function IMF_0 shown overlaid onto R_0 in Figure 7a.

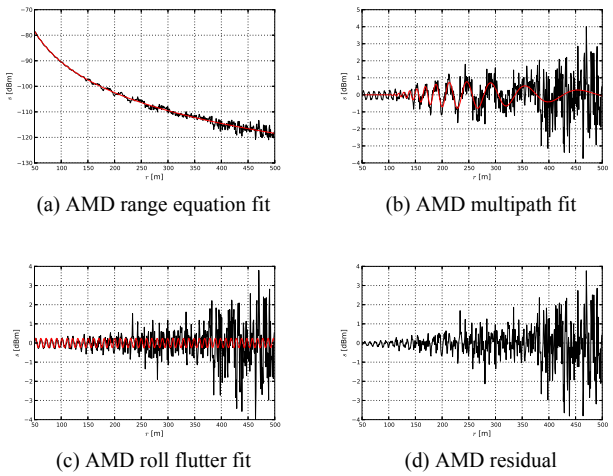
Multipath: After the first mode estimate is made, steps 2-5 can be performed again for the second mode. IMF_0 is subtracted from R_0 to form R_1 which is then fit to $m_1\left(r, \langle \sigma_h, \alpha \rangle\right)$ using the same downhill simplex algorithm used by the previous step. This resulted in an estimate of the surface roughness $\hat{\sigma}_h = 39.86\, mm$, and attenuation rate $\hat{\alpha} = 0.26\, Np/m$ which is then used to generate the mode function IMF_1, shown overlaid onto R_1 in Figure 7b.

Roll Induced Flutter: After the second mode estimate is made, steps 2-5 can be performed again for the third and final mode. IMF_1 is subtracted from R_1 to form R_2, which is then fit to $m_2\left(r, \langle A_r, F_s, \phi \rangle\right)$. Since this is a simple sinusoidal signal, Fourier methods are employed. The result is an estimate $\hat{A}_r = 0.10\, dB$, and $\hat{F}_s = 0.25\, 1/m$ which is shown overlaid onto R_2 in Figure 7c.

Residual: In this case, the residual shown in Figure 7d consists primarily of the system noise. Since the log power signal was decomposed to allow the other signal features to be a linear combination of modes, the noise appears to be non-stationary.

5 Performance

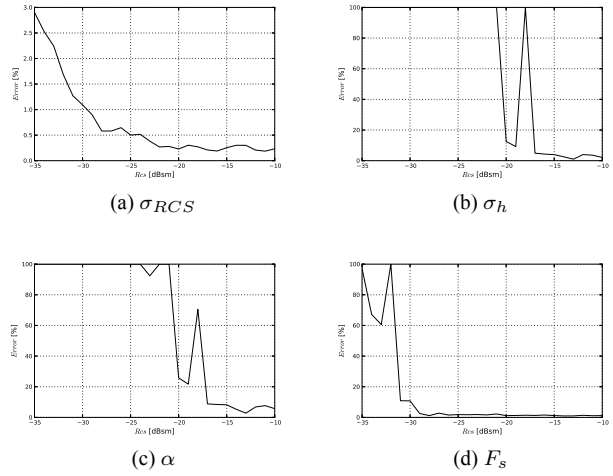
To determine the robustness of the method against different system parameters, first the signal was generated using the values in the previous section with one parameter swept in over a range. At each parameter value the method was applied to 10 different noise realizations and the median percentage error was recorded. First the target RCS (σ_{RCS}) was varied from -35 to -10 dBsm. The median error in the estimates of the system parameters are shown in Figure 8. In this figure, it can be seen that the estimation of σ_{RCS}



(a) AMD range equation fit (b) AMD multipath fit

(c) AMD roll flutter fit (d) AMD residual

Figure 7: AMD Example



(a) σ_{RCS} (b) σ_h

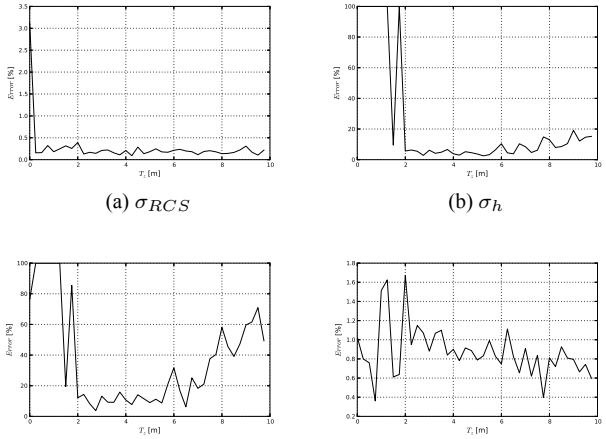
(c) α (d) F_s

Figure 8: Parameter Estimation Error vs Target RCS

is better than 3% throughout the range. Estimation of surface parameters σ_h and α require cross sections greater than $-20 dBsm$, and estimation of the roll rate requires a cross section greater than $-30 dBsm$. Next, the target altitude and cross range was swept from 0 to 10m, and the results shown in Figures 9 and 10. In these plots we see low to moderate estimation errors spread uniformly over the range. Combined with the previous RCS sweep, this suggests the algorithm is not "tuned" for a particular realization or point in parameter space. Parameter estimation success seemed to be more dependent on target altitude than cross range.

6 Conclusions

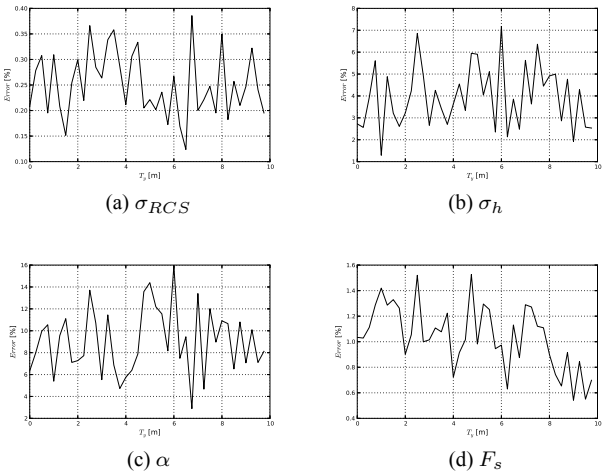
In this paper, a detailed amplitude model of a target over rough flat terrain was developed. This amplitude track data was then analyzed with Fourier methods which had some success identifying the roll rate of the target but gave no insight into the multipath behavior. EMD was then investigated which had some success decomposing a pristine track but did not work in the presence of noise. Finally, EMD was used as inspiration to develop a method of successive regression and fitting to the data dubbed AMD. AMD was applied not only to a single simulated environment but was applied in Monte Carlo fashion to a large portion



(a) σ_{RCS} (b) σ_h

(c) α (d) F_s

Figure 9: Parameter Estimation Error vs Target Altitude



(a) σ_{RCS} (b) σ_h

(c) α (d) F_s

Figure 10: Parameter Estimation Error vs Target Cross Range

of parameter space which showed it to be effective and robust to changes in the modeled system parameters.

At a very high level, EMD and AMD are similar and part of the greater family of transforms which attempt to decompose a signal into a set of functions of which the linear combination is the original signal. Both EMD and AMD do not require the component functions to be related and attempt to have the modes have physical meaning rather than mathematical purity. EMD decomposes a signal by extracting the highest frequency signal into the first IMF. This occurs because the upper and lower envelopes ride on top of the high frequency signal and hug the underlying low frequency components allowing the high frequency signal to be stripped off. This behavior is inherent in EMD. In contrast, AMD requires that the current mode be the lowest frequency of the remaining modes so that the least square fit can ignore the remaining modes by allowing them to average out over cycles. Though EMD is more flexible because it can attack problems where the underlying physics may be completely unknown or unknowable, AMD was much more useful for the signal explored in this paper.

References

[1] M. Richards, **Fundamentals of radar signal processing**. McGraw-Hill, 2005.

[2] P. Beckmann and A. Spizzichino, **The scattering of electromagnetic waves from rough surfaces**. Artech House, Inc., 1987.

[3] J. Stratton, **Electromagnetic Theory**. McGraw Hill, 1941.

[4] A. Nashashibi, K. Sarabandi, S. Oveisgharan, M. Dobson, W. Walker, and E. Burke, "Millimeter-wave measurements of foliage attenuation and ground reflectivity of tree stands at nadir incidence," **Antennas and Propagation, IEEE Transactions on**, vol. 52, no. 5, pp. 1211–1222, 2004.

[5] Y. Wang, R. Zhang, and J. Sun, "On separating the information of slow-mode angular motion of projectile," in **Signal Processing (ICSP), 2010 IEEE 10th International Conference on**. IEEE, pp. 2125–2128.

[6] N. e. a. Huang, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," **Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences**, vol. 454, no. 1971, pp. 903–995, 1998.

[7] N. Huang, Z. Shen, and S. Long, "A NEW VIEW OF NONLINEAR WATER WAVES: The Hilbert Spectrum 1," **Annual Review of Fluid Mechanics**, vol. 31, no. 1, pp. 417–457, 1999.

CREID: Development of reliable and scalable DHCP system for carrier IP networks

Katsuhiro Naito
Department of Electrical and Electronic Engineering, Mie University,
1577 Kurimamachiya, Tsu, 514-8507, Japan
Email: naito@elec.mie-u.ac.jp
Makoto Nishide
Net Step Inc.
213 Obatachomiyamae, Ise, 519-0504, Japan
Email: mnishide@netstep.co.jp
Eiji Miyazoe
OSS BroadNet Inc.
3-5-7 Hisamoto, Takatsuku, Kawasaki, 213-0011, Japan
Email: emiyazoe@ossbn.co.jp

***Abstract*— Dynamic host configuration protocol (DHCP) is essential service to configure information about networks at user terminals in Internet service providers (ISPs). Therefore, many DHCP programs are released in Internet. However, few free software DHCP programs can achieve required reliability and scalability in ISP's usage. In this paper, we develop the reliable and scalable DHCP system called CNR Emulator on ISC-DHCP (CREID) based on free software such as ISC-DHCP, Duplicated Replicated Block Device (DRBD), and Pacemaker. ISC-DHCP supports Internet Protocol (IP) v4 and IPv6 that are required in commercial ISP services. DRBD and Pacemaker can construct clustering systems over some physical computers. From the numerical results, we can find that the developed DHCP system can achieve enough DHCP transaction performance which is required in commercial ISPs, and high service availability over 99.999%.**

***Keywords*— DHCP, Carrier IP networks, ISC-DHCP, DRBD, Pacemaker**

I. Introduction

Dynamic host configuration protocol (DHCP) [1], [2] becomes more important functions to setup network configuration automatically according to the development of computer networks. Especially, Internet Service Providers (ISPs) require the DHCP function to manage user terminals and assign IP addresses [3], [4], [5], [6]. Therefore, they require more reliable and scalable DHCP systems to provide stable network services. As the results, high-end DHCP systems for commercial ISPs have achieved high reliability and scalability [7]. In these systems, they can achieve high transaction performance of DHCP requests and support fail-over mechanisms against server troubles. On the contrary, the price of these systems becomes expensive due to the lack of competitors. Therefore, it is difficult to employ them in small ISPs due to lack of capital-investment spending even if they have a better and stable performance.

ISC-DHCP [8] is the well known free DHCP software provided by Internet Systems Consortium. The implementation of ISC-DHCP is distributed over various kinds of UNIX OS. Therefore, many users use the software in local networks. On the contrary, the transaction performance of ISC-DHCP is not high comparing to the requirement in commercial large ISPs. Additionally, the fail-over mechanisms of ISC-DHCP are not enough to achieve a stable DHCP service in commercial ISPs [9], [10], [11].

In this paper, we develop the reliable and scalable DHCP system called CNR Emulator on ISC-DHCP (CREID) based on free software such as ISC-DHCP, Duplicated Replicated Block Device (DRBD) [12], and Pacemaker [13]. ISC-DHCP supports Internet Protocol (IP) v4 and IPv6 that are required in commercial ISP services. DRBD and Pacemaker can construct clustering systems over some physical computers. Therefore, the CREID can provide required functions in commercial ISPs, and can achieve reliability by using cluster fail-over mechanisms. From the evaluation experiments, we can find that the developed DHCP system can achieve enough DHCP transaction performance which is required in commercial ISPs, and high service availability over 99.999%.

II. Fundamental performance of ISC-DHCP

The developed DHCP system employs ISC-DHCP as the fundamental DHCP software. Therefore, it is important to evaluate the fundamental performance of ISC-DHCP to design the developed system. Table I shows the hardware specifications in fundamental evaluation of ISC-DHCP.

Generally, IP address ranges are registered in a DHCP server beforehand. Transaction performance of assigning an IP address deteriorates due to increase of IP address range. In this evaluation, we use dhcperf [14], which is an evaluation software for DHCP services. In the measurements, we measure the DHCP discover transaction performance including Discover, Offer Request, Ack, and Release.

Fig. 1 shows the DHCP transaction performance versus the IP address range. From the results, we can find that the

TABLE I
HARDWARE SPECIFICATIONS IN FUNDAMENTAL EVALUATION

	DHCP Server	DHCP Client
CPU	Intel Xeon(R) E5620 2.40GHz	Intel Celeron G1101 2.26GHz
Memory	8GB	1GB
HDD	300GB 6G SAS 15000rpm	250GB SATA2 7200rpm
Network	1Gbps	
OS	Scientific Linux 6.1	
DHCP	ICS-DHCP 4.1-ESV-R2	

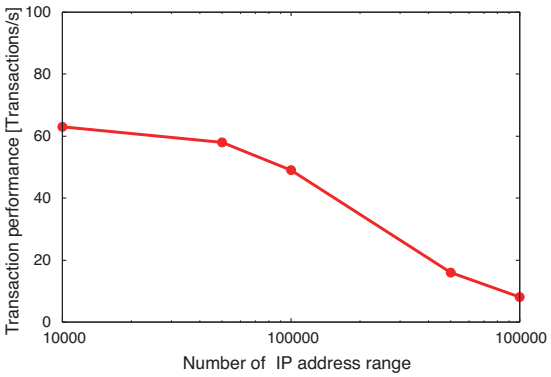


Fig. 1. DHCP transaction performance of ISC-DHCP.

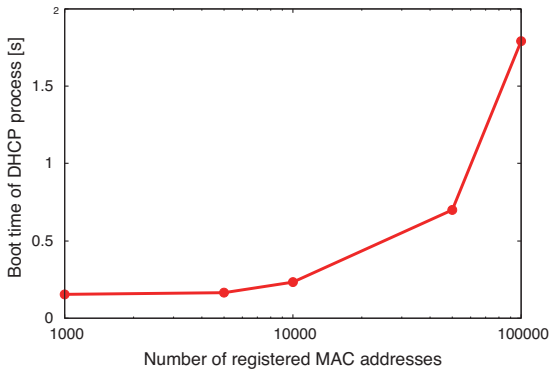


Fig. 2. Boot time of DHCPD process with registered MAC address.

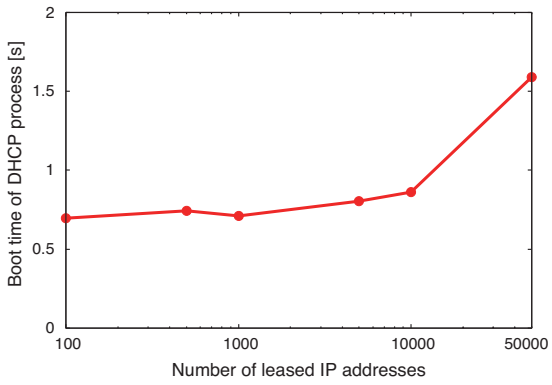


Fig. 3. Boot time of DHCPD process with leased records.

transaction performance deteriorates drastically according to the increase of the IP address range. For example, Common Antenna TeleVision (CATV) provides 50,000 Cable Modems (CMs) with one Cable Modem Termination System (CMTS). Therefore, the IP address range should be set as twice number of CMs. As the results, the transaction performance of ISC-DHCP in large networks is not enough in commercial ISPs.

Some ISPs register subscriber's MAC address information into DHCP systems to provide adequate services. ISC-DHCP always read the MAC address information when DHCPD processes boot up. Therefore, boot time depends on the number of registered MAC addresses. ISC-DHCP cannot provide DHCP service during booting of a DHCPD process. Therefore, boot time is an important factor to achieve high service availability.

Fig. 2 shows the boot time of DHCPD process with registered MAC address. The results shows that the boot time increases up to a few second when the number of MAC address increases. DHCP client software generally implements retry mechanisms when DHCP service is temporally unavailable. Therefore, the short boot time does not effect the availability of DHCP service.

ISC-DHCP servers add lease information of IP addresses into lease files when DHCP clients request a new IP address or renew an assigned IP address. Therefore, the size of the lease files increases according to the service period. In ISC-DHCP, DHCPD processes optimizes the lease files when they boot up.

Fig. 3 shows the boot time of DHCPD process with leased records. From the results, we can find that the boot time increases when the number of leased IP address increases. However, the maximum boot time is short, and does not effect the availability of DHCP service when the retrial mechanisms are implemented at DHCP clients.

Fig. 4 shows the boot time of DHCPD process with timeout records. From the results, the boot time increases drastically according to the number of timeout records. Fig. 5 shows the boot time of DHCPD process with renewed records. From the results, we can find that the boot time increases when the number of renewed records increase. During the boot time of the DHCPD process, the DHCP service is not available. Therefore, the reduction of boot time is important in commercial usages.

III. SERVICE AVAILABILITY

Generally, system managers of DHCP services edit configuration files according to user contracts and maintenance of networks. DHCPD processes should be restarted to reload the latest configuration files when the configuration files are edited. During the restating process, DHCP services are down. Additionally, DHCP services are also down when the DHCP system switches a main server to a backup server due to main server troubles. Therefore, the service availability of DHCP services is an important factor in commercial usages. In this section, we consider the service availability ratio for DHCP

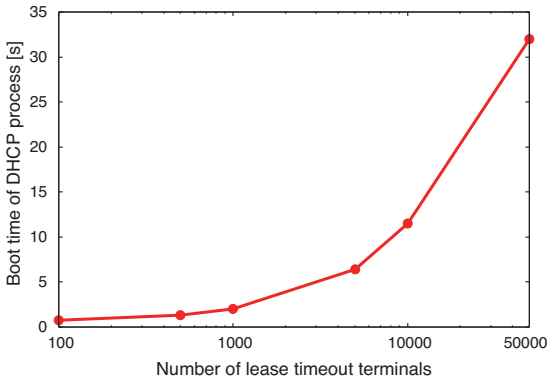


Fig. 4. Boot time of DHCPD process with timeout records of ISC-DHCP.

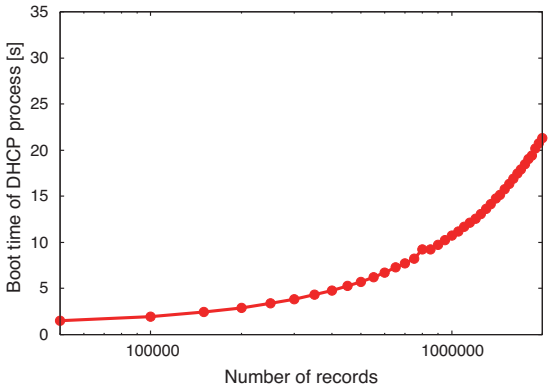


Fig. 5. Boot time of DHCPD process with renewed records of ISC-DHCP.

services.

At first, we classify reasons of restart process into four categories: short-term maintenance, long-term maintenance, system troubles, optimization of lease files.

- Short-term maintenance
  System managers edit configuration files according to user contracts day by day. We define the short-term maintenance frequency as P_{short}, and required boot time of processes as T_{short}.
- Long-term maintenance
  System managers edit configuration files according to network maintenance, policy changes, and redefinition of subnet, IP address ranges, etc. We define the long-term maintenance frequency as P_{long}, and required boot time of processes as T_{long}.
- Server troubles
  DHCP systems should change a main server to a backup server to continue DHCP services to user terminals when some troubles happen on the main server. We define the occurrence frequency of main server troubles as P_{fail}, and required switching period from the main server to the backup server as T_{fail}.
- Optimization of lease files
  ISC-DHCP requires long period to optimize the lease files when the lease files include timeout or renewed records. Therefore, system managers should restart DHCP processes to reduce the size of the lease files periodically. We define the periodical restart frequency as P_{opt}, and required boot time of DHCPD processes as T_{opt}.

The optimization process is performed when the DHCPD process is restarted. Therefore, we redefine the required time including optimization process as T'_{short}, T'_{long}, and T'_{fail}.

DHCP clients renew an assigned IP address when the lease time is end, and try to renew several times when it cannot receive DHCP messages from DHCP servers. During retrial period, it continues to use the assigned IP address. Therefore, we should consider the timeout period for the DHCP renew process. We define the timeout period at DHCP client as $T_{timeout}$. The actual service down period depends on each period and the timeout period. For example, the actual service down period due to the short maintenance T''_{short} is

$$T''_{short} = \begin{cases} 0 & (T'_{short} \leq T_{timeout}) \\ T'_{short} - T_{timeout} & (T'_{short} > T_{timeout}). \end{cases} \quad (1)$$

Then, we can calculate the total service down period per day as

$$T_{down} = P_{short}T''_{short} + P_{long}T''_{long} + P_{fail}T''_{fail} + P_{opt}T''_{opt}. \quad (2)$$

When we assume that each user terminal boots up independently, the number of DHCP clients that request to renew an assigned IP address is

$$N_{lease} = N_{cm}/T_{lease}, \quad (3)$$

where N_{cm} is the number of user terminals and T_{lease} is the lease period of DHCP services.

Then we can obtain the service available ratio per user terminals as

$$P_{available} = 1 - N_{lease}T_{down}/N_{cm}. \quad (4)$$

IV. CREID SYSTEM

A. System model

Fig. 6 shows the system model of CREID. This figure assumes that CATV networks where CREID provides DHCP services to cable modems (CMs). In order to improve the transaction performance of ISC-DHCP, CREID employs multiple processes of DHCPD. Therefore, each DHCPD process can provide independent DHCP services for CMs. Additionally, we construct a cluster system to achieve fail-over mechanisms by using DRBD and Pacemaker. As the results, CREID can switch from the main server to the backup server when some troubles happen in the main server.

B. Experimental results

We perform experimental measurements of CREID to evaluate the performance. In the measurements, we use the same hardware shown in Table II.

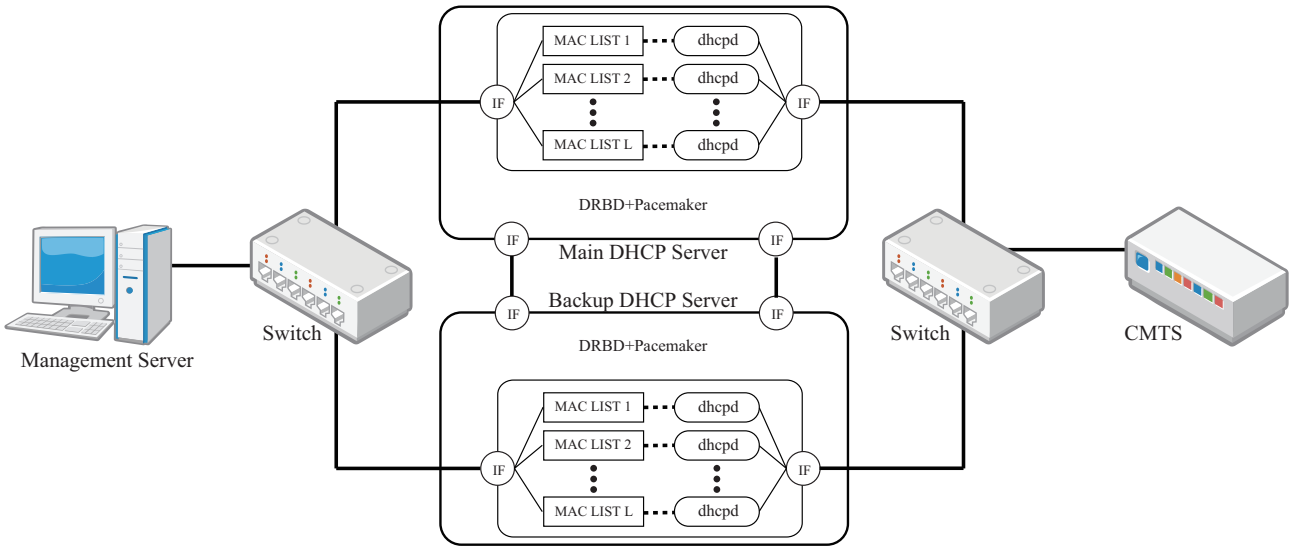


Fig. 6. System model of CREID.

TABLE II
HARDWARE SPECIFICATIONS IN CREID

	DHCP Server	DHCP Client
CPU	Intel Xeon(R) E5620 2.40GHz	Intel Celeron G1101 2.26GHz
Memory	8GB	1GB
HDD	300GB 6G SAS 15000rpm	250GB SATA2 7200rpm
Network	1Gbps	
OS	Scientific Linux 6.1	
DHCP	ICS-DHCP 4.1-ESV-R2	
Pacemaker	1.0.11	
Heartbeat	3.0.5	
DRBD	8.4.1	

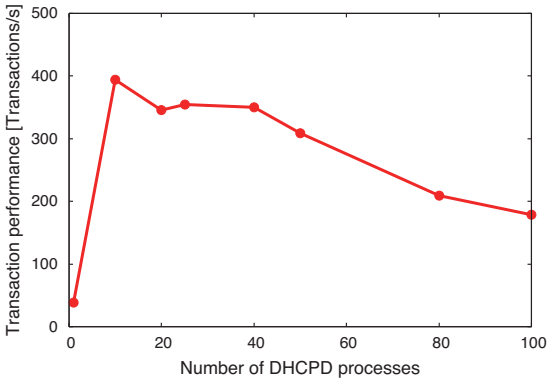


Fig. 7. DHCP Discover transaction performance of CREID.

1) Clustering performance: CREID employs DRBD and Pacemaker to achieve fail-over mechanisms of DHCP service. Clustering mechanisms generally cause performance overhead due to additional processing of clustering layer. CREID requires accessing to hard disks when an IP address is leased. Therefore, I/O performance is an important factor in CREID. In the measurements, CREID mounts a disk with synchronous option. Then, we measure the period for copying sixty files from the primary server to the backup server. The size of each file is 10 [MBytes].

The results show that the throughput performance without DRBD is 56.34 [Mbps], and that with DRBS is 3.81 [Mbps]. Since lease files for DHCP processes include only text information, the throughput performance with DRBS is enough to exchange the text information.

2) Transaction performance with multiple DHCPD processes: Fig. 7 shows the DHCP discover transaction performance of CREID. Fig. 8 shows the DHCP renew transaction performance of CREID. From the results, we can find that we can improve the performance by selecting the adequate number of DHCPD processes. In CATV networks, required transaction performance is between 100 and 300 transactions per second [15]. Therefore, CREID can achieve the required performance by using free software.

3) Boot time performance with multiple DHCPD processes: Fig. 9 shows the boot time of DHCPD process with registered MAC address of CREID. From the results, we can find that the boot time is less than 200 [ms] even if the number of registered MAC addresses increases. CREID allocates subset of MAC addresses to each DHCPD process. Therefore, the number of registered MAC address for each DHCPD process can be decreased.

Fig. 10 shows the boot time of DHCPD process with leased records. From the results, CREID can keep short boot time when the number of lease IP addresses increases.

Fig. 11 shows the boot time of DHCPD process with timeout records. From the results, we can find that the boot time is quite short even if the number of timeout records increases. Fig. 12 shows the boot time of DHCPD process with renewed

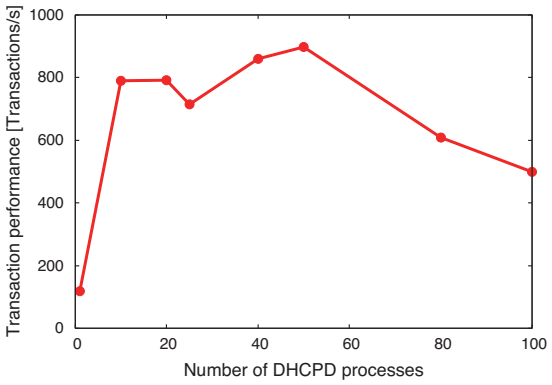


Fig. 8. DHCP Renew transaction performance of CREID.

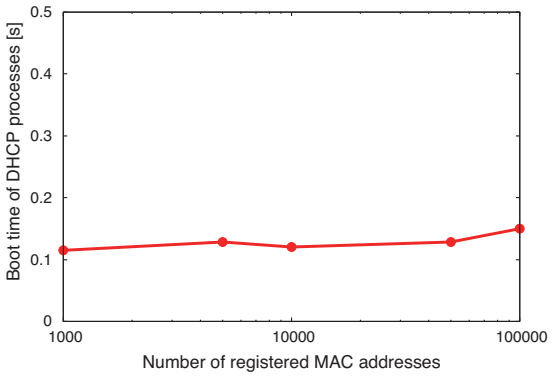


Fig. 9. Boot time of DHCPD process with registered MAC address of CREID.

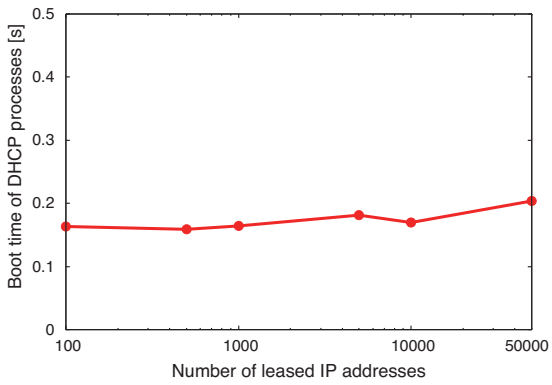


Fig. 10. Boot time of DHCPD process with leased records of CREID.

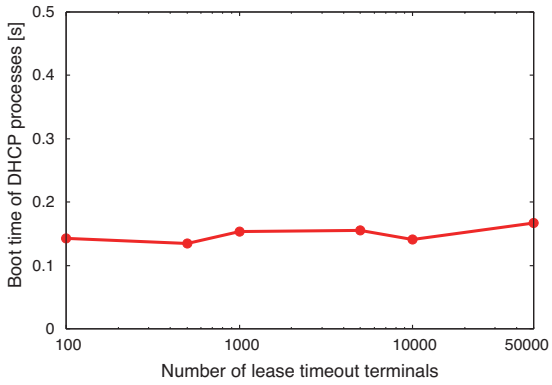


Fig. 11. Boot time of DHCPD process with timeout records of CREID.

records. From the results, CREID can reduce the boot time when the number of renewed records increase. From the measurement results, we can find that CREID can achieve scalable DHCP system when the large number of user terminal is required in commercial ISPs.

4) Processing performance: Tab. III shows the memory usage and swap memory usage with multiple DHCPD processes. The results show that the memory usage increases when the number of DHCPD processes increases. However, the increased memory amount is not large. Therefore, the system will work stably when the enough physical memories are implemented.

Tab. IV shows the CPU load with multiple DHCPD processes. From the results, the maximum value of r increases when the number of DHCPD processes increases. However, the average value of r is smaller than the maximum value. Therefore, CPU load is not constantly high. Additionally, the maximum value of b increases. However, the value of b is smaller than the value of r. A server system generally works well when the average CPU load is small. Therefore, CREID also works well on the real hardwares.

C. Service available ratio

We evaluate the service available ratio of CREID according to the experimental results. We assume that CREID provides DHCP services to 50,000 user terminals. The lease time is set as 24 hours, P_{short} is set as 10, P_{long} is set as once per three months, P_{opt} is 1. From the numerical results, the boot time is about 0.2 [s]. Therefore, we set T_{short}, T_{long}, and T_{opt} as 0.25 [s].

Additionally, the measured switching period of fail-over mechanisms is about 23.2 [s]. Therefore, we set T_{fail} as 23.5 [s] and P_{fail} as once per a year . In the implementation of ISC-DHCP client, timeout period is 10 [s]. But, we evaluate more critical situation. So, we set the $T_{timeout}$ as 0 [s]. The summary of the parameters is shown in Tab. V.

From the equations in the section III, we can obtain that T_{down} is 5.3 [s], and $P_{available}$ is 99.9991%. Therefore, we can find that CREID can achieve high service availability according to the experimental measures.

V. Conclusions

This paper proposed a reliable and scalable DHCP systems called CREID. The developed system consists of free software such as ISC-DHCP, DRBD and Pacemaker. From the experimental measurements, we can find CREID can achieve high transaction performance that is required in commercial ISPs and high service available ratio more than 99.999%.

TABLE III
MEMORY AND SWAP MEMORY USAGE.

Number of processes	1	10	20	25	40	50	80	100
Memory usage (KB)	830,144	1,003,468	1,141,088	1,214,228	1,424,456	1,577,352	1,999,016	2,289,708
swap memory usage (KB)	0	0	0	0	0	0	0	0

TABLE IV
CPU LOAD.

Number of process	1	10	20	25	40	50	80	100
Maximum value of r	1	7	19	22	25	34	58	66
Average value of r	0	1	1	2	1	7	12	7
Maximum usage ratio of users (%)	7	4	5	6	7	9	13	15
Average usage ratio of users (%)	6	3	4	4	6	7	11	14
Maximum value of b	3	2	3	3	2	3	3	5
Average value of b	1	1	1	2	1	1	2	2
Maximum CPU wait ratio (%)	9	19	18	17	17	19	17	17
Average CPU wait ratio (%)	6	16	16	16	15	17	15	14
Increased value of bi	1,202	746	220	452	208	630	179	347
Increased value of bo	28,902	96,571	105,216	111,363	124,097	137,204	161,086	184,383

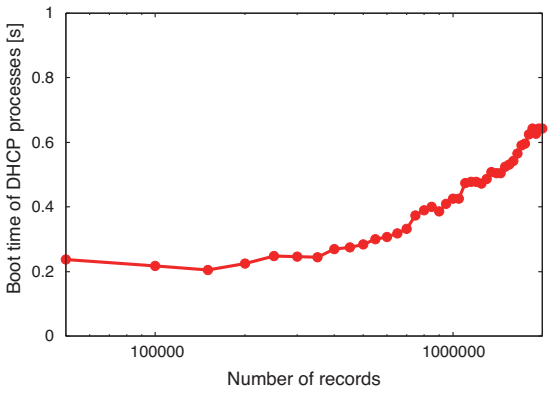


Fig. 12. Boot time of DHCPD process with renewed records of CREID.

TABLE V
PARAMETERS FOR EVALUATING AVAILABILITY

Number of CMs	50,000
Lease period of DHCP	24 hours
P_{short}	10 times / day
T_{short}	0.25 s
P_{long}	1 time / 3 months
T_{long}	0.25 s
P_{fail}	1 time / year
T_{fail}	23.5 s
$P_{optimization}$	1 time / day
$T_{optimization}$	0.25 s
$T_{timeout}$	0 s

ACKNOWLEDGMENT

Authors thank Casa Systems, Inc. and OSS BroadNet Inc. for experimental measurements.

REFERENCES

[1] R. Droms, "Dynamic Host Configuration Protocol," IETF RFC2131, March 1997.
[2] R. Droms, J. Bound, B. Volz, T. Lemon, C. Perkins, and M. Carney, "Dynamic Host Configuration Protocol for IPv6 (DHCPv6)," IETF RFC3315, July 2003.
[3] S. Alexander and R. Droms, "DHCP Options and BOOTP Vendor Extensions," IETF RFC2132, March 1997.
[4] J. Littlefield, "Vendor-Identifying Vendor Options for Dynamic Host Configuration Protocol version 4 (DHCPv4)," IETF RFC3925, October 2004.
[5] Cable Television Laboratories Inc., "Operations Support System Interface Specification," Cable Television Laboratories, Inc., CM-SP-OSSIv3.0-I07-080522, May 2008.
[6] Cable Television Laboratories Inc., "CableLabs' DHCP Options Registry," Cable Television Laboratories, Inc., CL-SP-CANN-DHCP-Reg-I02-080306, March 2008.
[7] Cisco "Cisco Network Registrar," http://www.cisco. com, retrieved: January 2012.
[8] Internet Systems Consortium, ISC-DHCP, http://www. isc.org, retrieved: January 2012.
[9] R. Droms, K. Kinnear, M. Stapp, B. Volz, S. Gonczi, G. Rabil, M. Dooley, and A. Kapur, "Draft, DHCP Failover Protocol," IETF INTERNET DRAFT, March 2003.
[10] http://paulroberts69.wordpress.com/2011/10/27/isc-dhcp-failover-is-just-too-complex/, retrieved: January 2012.
[11] http://www.accumuli.com/using-infoblox-dhcp-failover-part-1-i-3232.php, retrieved: January 2012.
[12] B. Hellman, F. Haas, P. Reisner, and L. Ellenberg, "The DRBD Userfs Guide," , http://www.drbd.org/users-guide/, retrieved: March 2012.
[13] A. Beekhof, "Pacemaker 1.0 Configuration Explained," http://www.clusterlabs.org, retrieved: January 2012.
[14] dhcperf, http://www.nominum.com, retrieved: January 2012.
[15] http://cn.teldevice.co.jp/product/infoblox/ib_spec.html, retrieved: January 2012.

Multi-rate routing protocol based on bottleneck link throughput for mesh networks

Katsuhiro Naito, Yusuke Hiei, Kazuo Mori, and Hideo Kobayashi
Department of Electrical and Electronic Engineering, Mie University,
1577 Kurimamachiya, Tsu, 514-8507, Japan
Email: {naito, kmori, koba}@elec.mie-u.ac.jp, {hiei}@com.elec.mie-u.ac.jp

Abstract— **IEEE 802. 11 is a major candidate device that supports a multiple transmission rate for constructing wireless mesh networks. Various transmission rate control methods have been proposed to select an appropriate transmission rate according to link status. Cross layer mechanisms between routing and transmission rate control have good benefit to improve throughput performance because data packets through some links to achieve communication between end-nodes in wireless mesh networks. Therefore, various cross layer mechanisms have been proposed. However, simple selection mechanisms for transmission rate are not enough to be considered. As the results, almost all conventional protocols seem to be difficult to be implemented in real wireless mesh networks. In this paper, we employ AODV(Ad hoc On-Demand Distance Vector) as the base routing protocol because AODV is simple and good mechanisms for dynamic fluctuation of network condition. Then, we extend it to take into account end-to-end average throughput and bottleneck link throughput to select better wireless link and avoid the worst bottleneck link. The features of the proposed protocol are to estimate the achievable transmission rate according to link status and to select an appropriate link as a route. From the simulation results using QualNet, we show the proposed mechanisms can improve the whole throughput performance in mesh networks.**

Keywords— **Multi-rate transmission, bottleneck link throughput, Multi-hop networks**

I. Introduction

Mesh networks have been attracted attention recently according to the development of wireless communication devices. IEEE 802. 11 is a candidate standard device to construct mesh networks, and supports multi-rate transmission such as from 6 to 54 [Mbps]. The well known link-rate-control mechanisms are Auto Rate Fallback (ARF) [1], Receiver Based Auto Rate (RBAR) [2], and can improve throughput performance. These mechanisms can control transmission rate according to each link status.

Various multi-hop routing protocols have been proposed [3], [4]. In Ad hoc On-Demand Distance Vector (AODV), a source node starts to construct a route to a destination node when an application tries to start communication. Almost all routing protocols construct a route based on a hop count information to minimize route distance. Therefore, constructed routes sometimes go through erroneous wireless links. Generally, erroneous wireless links tend to select lower transmission rate to improve bit error rate performance on links. As the results, the end-to-end throughput deteriorates due to an erroneous part with low transmission rate over the route [5], [6].

Recently, cross layer mechanisms between transmission rate control and routing mechanisms have been proposed to improve throughput performance. In the conventional research, nodes estimate the channel condition by measuring Signal to Noise Ratio (SNR), and select a high quality link to construct a route [7] to achieve high transmission rate. HT-AODV extends the routing mechanisms that take into account the quality of wireless links [8]. However, main target of these research is to employ high transmission rate links to construct a route. Therefore, the transmission rate on bottleneck link is not enough to be considered. However, the end-to-end throughput performance depends on transmission rate of bottleneck links because the bottleneck links cannot transfer the data with enough transmission rate. Additionally, the signal with the lower transmission rate on the bottleneck links intersects neighbor communication frequently because low transmission rate requires a long transmission period.

In this paper, we propose a multi-rate routing protocol for mesh-networks. Our protocol is based on AODV, which is on-demand type routing protocols. Therefore, our protocol can handle the variation of channel status because routes are constructed when applications try to start communication. The feature of our protocol is to focus on the bottleneck link throughput, which dominates the end-to-end throughput. Additionally, nodes receive some route request message to find

more appropriate route in the proposed protocol. From the numerical results by using the network simulator QualNet [9], we can find that the proposed protocol can achieve higher end-to-end throughput performance and improve the bottleneck link throughput.

II. Proposed routing protocol

A channel condition between each node generally changes frequently. Therefore, reactive routing protocols tend to handle the variation of channel condition. We employ AODV as the base protocol to develop the multi-rate transmission routing protocol because a route is constructed when applications try to start communication. In AODV, a source node broadcasts a route request (RREQ) message to a whole network to find a destination node. RREQ message is forwarded at each node, which is called flooding, in the whole network. The destination node replies a route reply (RREP) message to the source node when it receives the first RREQ message. In the proposed protocol, intermediate nodes and a destination node should estimate transmission rate by measuring signal-to-interference and noise power ratio (SINR) Then, the destination node selects the maximum transmission rate on bottleneck links to avoid deterioration of end-to-end throughput.

A. Estimation of transmission rate

In the proposed protocol, intermediate nodes and a destination node measure SINR of RREQ messages. Then, they estimate an available transmission rate by evaluating with the analytical model of modulation schemes. The estimated transmission rate is recorded in the special field in RREQ messages. Therefore, we extend the message format of RREQ such as Fig. 1.

Intermediate nodes estimate transmission rate between a sender node and own node according to the following equation.

$$R_{ave,j} = \frac{R_{sum,j}}{HopCount} \tag{1}$$

$R_{sum,j}$ is the summation of transmission rate at each link between a source node and node j.

Additionally, intermediate nodes compare the estimated own transmission rate to the transmission rate at bottleneck link R_{bottle}, and replace the bottleneck transmission rate when the own link is the bottleneck link. Fig. 2 shows the example operation of RREQ forwarding.

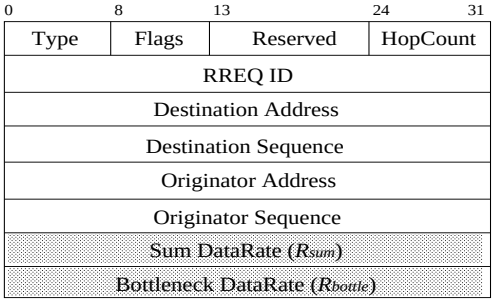


Fig. 1. Message format of RREQ.

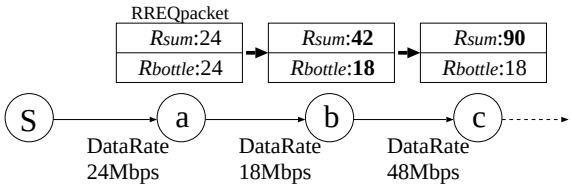


Fig. 2. Example operation of RREQ forwarding.

B. Selection of RREQ

In conventional AODV, intermediate nodes and a destination node receives a first RREQ message. On the contrary, intermediate nodes and a destination node receive some RREQ messages from neighbor nodes to find a link with higher transmission rate in the proposed protocol. Finally, they select an appropriate RREQ through links with higher transmission rate to construct a route.

Selection at intermediate nodes: An intermediate node j starts a timer when it receives a first RREQ from neighbor nodes. Until the timeout of the timer, it continues to receive RREQs from its neighbor nodes. The timeout period is set according to the following equation.

$$T_{wait,j} = D_{R,ij} \times T_{rand,max} + T_{rand} \tag{2}$$

$D_{R,ij}$ is a multiplication factor according to the estimated transmission rate between node i and node j. T_{rand} is a random value to avoid contention, and $T_{rand,max}$ is the maximum value of T_{rand}.

Fig. 3 shows the example operation of RREQ selection at intermediate nodes. We employ the average transmission rate and the transmission rate at a bottleneck link as the criteria for RREQ selection.

1) Intermediate node j estimates the transmission rate according to the equation 2 when it receives a first RREQ.

TABLE I

MULTIPLICATION FACTOR ACCORDING TO TRANSMISSION RATE R.

Transmission rate R [Mbps]@	54	48	36	24	18	12	9	6
Multiplication factor D_R	1	2	3	4	5	6	7	8

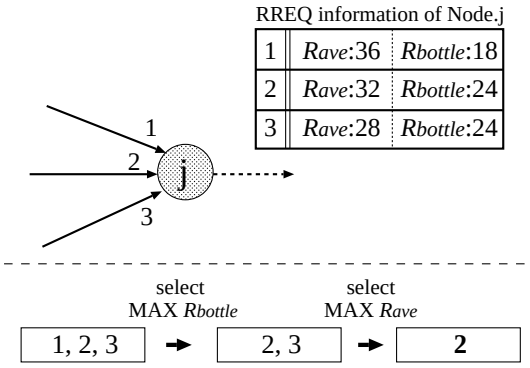


Fig. 3. Example operation of RREQ selection at intermediate nodes.

2) Intermediate node j selects the RREQs with the maximum $R_{bottle,j}$.
3) Intermediate node j selects the RREQ with the maximum $R_{ave,j}$.
4) Intermediate node j rebroadcasts the selected RREQ.

Selection at a destination node: Destination node D starts the timer when it receives a first RREQ. Then, it continues to receive the other RREQs for $T_{wait,D}$. In order to select an appropriate RREQ, we employ the average transmission rate, transmission rate at a bottleneck link and hop count as the criteria. Fig. 4 shows the example of RREQ selection at the destination node.

1) Destination node D continues to receive RREQs for $T_{wait,D}$.
2) Destination node D selects RREQs with the maximum $R_{bottle,D}$ after reception of RREQs.
3) Destination node D selects the RREQs with the minimum hop count to avoid over-lapped routes.
4) Destination node D selects the RREQ with the maximum $R_{ave,D}$.
5) Destination node D creates the RREP for the selected RREQ.

C. Example operations

Fig. 5 show the example operations of the proposed routing protocol. In the examples, we set the $T_{rand,max}$ as 20[msec],

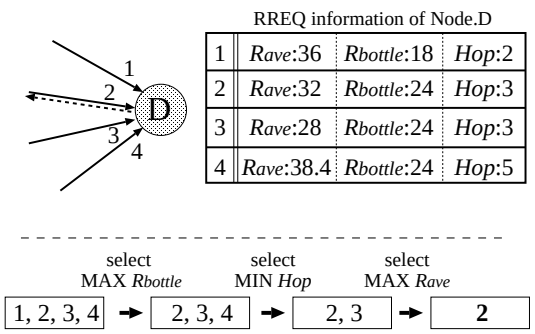


Fig. 4. Example operation of RREQ selection at a destination node.

and T_{rand} as 0[msec] in the equation 2.

1) Node S starts to find the route to node D when the traffic occurs in node S.
2) Node S broadcasts RREQ to its neighbor nodes. The neighbor intermediate nodes W, X, Y estimate the maximum transmission rate according to SINR. Then, they set the timer period T_{wait} for RREQ forwarding according to the equation 2. They calculate the summation of throughput R_{sum} and the bottleneck throughput R_{bottle} after the estimation of transmission rate. In the example figures, the intermediate node W estimates the available transmission rate as 48 [Mbps], and sets the timer period $T_{wait,W}$ as 40[msec]. In the similar fashion, the node X and Y set the timer period as 120 [msec] and 60 [msec] respectively.
3) Fig . 5 (a) shows the transmission status at 40 [msec]. The timer of the intermediate node W timeouts at 40 [msec]. Therefore, it selects the appropriate RREQ from the received candidate RREQs. In the example, it broadcasts the RREQ because it received only one RREQ. Then destination node D receives the RREQ which is transmitted by node W. It sets the timer period $T_{wait,D}$.
4) Fig . 5 (b) shows the transmission status at 120 [msec]. The timer of intermediate node X timeouts at 120 [msec]. Then, node X selects the appropriate RREQ according to the criteria in the proposed protocol. In this example, RREQ from node W is selected as the appropriate RREQ. Then, node W rebroadcasts the RREQ, and discards RREQs from nodes S and Y.
5) Fig . 5 (c) shows the transmission status at 40+$T_{wait,D}$ [msec] where the timer of destination node D timeouts. In the example, destination node D receives three

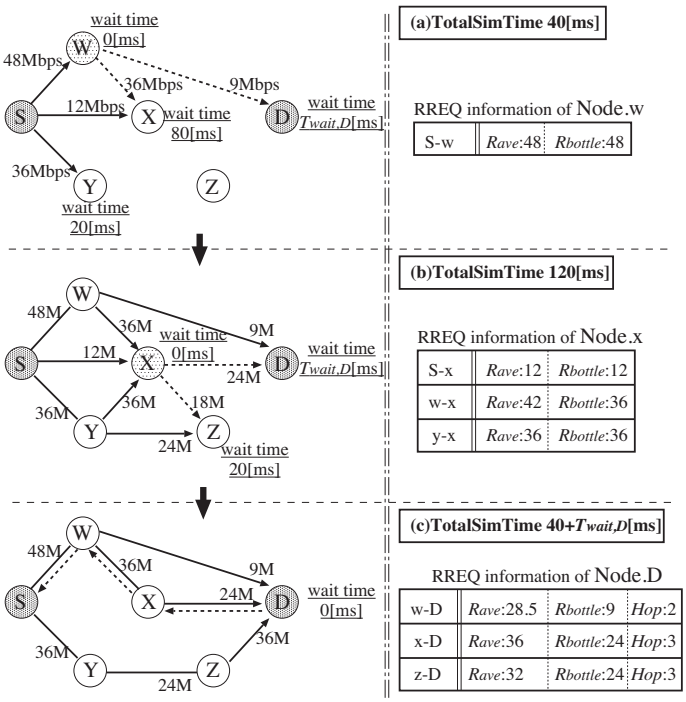


Fig. 5. Example operations

RREQs from different routes. It selects RREQ from node X as the appropriate route, and discards the other RREQs.

6) Finally, destination node D replies RREP via the route where the selected RREQ was forwarded. As the results, the route via nodes W and X is constructed.

III. Numerical results

We perform computer simulation by QualNet to evaluate the proposed protocol. As the conventional protocol, we evaluate the performance of AODV with ARF, which is the auto rate transmission control mechanisms according to channel conditions. We implemented the proposed protocol by extending AODV module in QualNet. In the simulations, we assume that 200 nodes are randomly distributed within 1500 squares meters area, and nodes do not move in the simulation period. As the wireless device, we employ IEEE 802.11g device which supports 6, 9, 12, 18, 24, 36, 48 and 54 [Mbps]. Therefore, each node selects a transmission rate from these rate according to the SINR. FTP communication with 1024 bytes packets are assumed between source nodes and destination nodes. The numerical results are the average of 100 simulations. Detail simulation parameters are shown in Tab. II.

A destination node waits for $T_{wait,D}$ to receive multiple

TABLE II
Simulation parameters

Simulator	QualNet 5. 0
Simulation area	1500[m]×1500[m]
Number of nodes	200
Node location	Random
Wireless device	IEEE 802. 11g
Wireless channel	AWGN
Application	FTP
Packet size	1024[byte]
Number of connections	5, 10, 15, 20, 25
Simulation period	60[s]
Number of trials	100

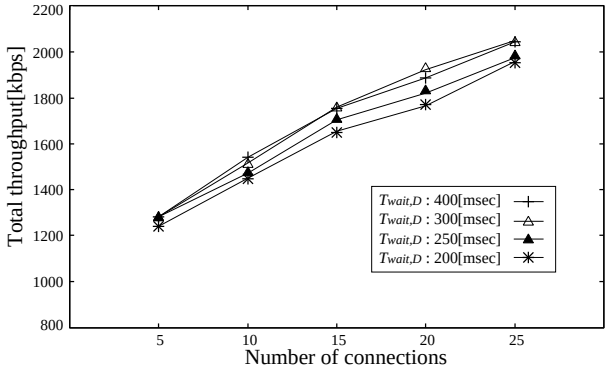


Fig. 6. Total throughput performance.

RREQs. The longer $T_{wait,D}$ increases probability to find more appropriate route, but makes long delay to construct a route. Fig. 6 shows the throughput performance when $T_{wait,D}$ is changed for 200, 250, 300 and 400 [ms]. From the results, we can find the throughput performance degraded in case of less than 300 [ms]. Therefore, the proposed protocol requires 300 [ms] to find an appropriate route at a destination node in this situation. In the following simulations, we employ 300 [ms] as $T_{wait,D}$, which is the waiting period to receive multiple RREQs.

The feature of the proposed protocol focuses on transmission rate of bottleneck link because the bottleneck link is a dominant factor for end-to-end throughput. Route selection mechanisms are also important to select an appropriate route with higher transmission rate of bottleneck link. Fig. 7 shows the transmission rate of bottleneck link when the proposed protocol without route selections at a destination node is employed. Therefore, intermediate nodes select a more appropriate route, but a destination node selects a first

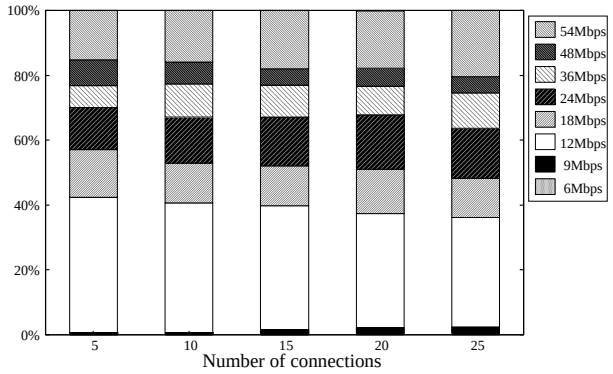


Fig. 7. Bottleneck transmission rate in proposed protocol without destination selection

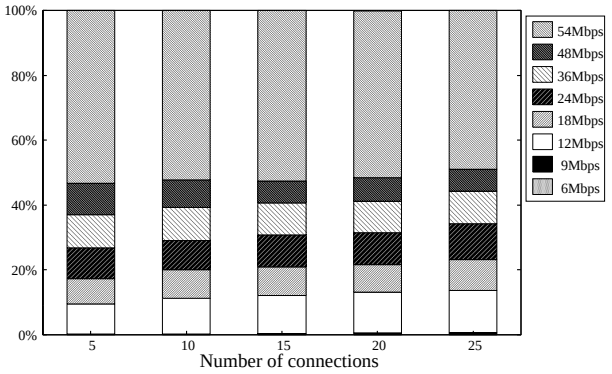


Fig. 9. Bottleneck transmission rate in proposed protocol

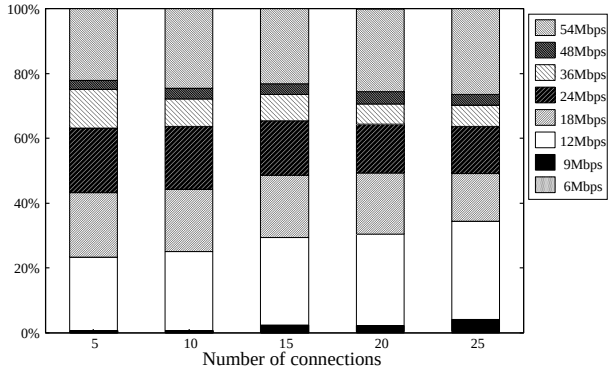


Fig. 8. Bottleneck transmission rate in proposed protocol without intermediate selections

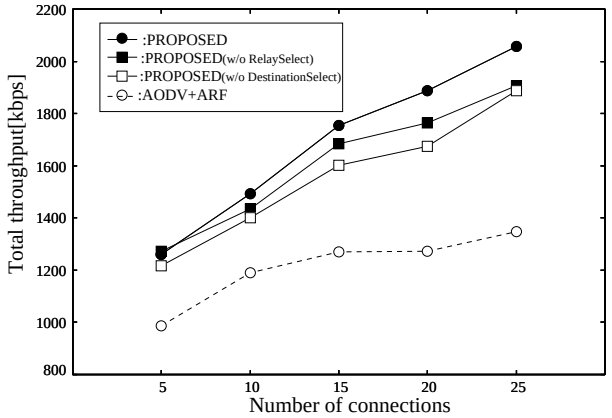


Fig. 10. End-to-end throughput.

RREQ as the route. The results show that the transmission rate of about 40% links is less than 12 [Mbps]. Additionally, the percentage of lower transmission rate increases when the number of connections increases. Therefore, the route selection at destination nodes is important to improve the transmission rate of bottleneck-link.

Fig. 8 shows the transmission rate of bottleneck link when the proposed protocol without route selections at intermediate nodes is employed. In this protocol, intermediate nodes select a first RREQ as the route and a destination node selects an appropriate route by receiving some RREQs. From the results, we can find that the transmission rate of more than 20% link is less than 12 [Mbps].

Fig. 9 shows the transmission rate of bottleneck-link when the proposed protocol is employed. In the proposed protocol, intermediate nodes and a destination node select an appropriate route by receiving some RREQs for determined period. The results show that the transmission rate of bottleneck link in the proposed protocol is more than 12 [Mbps]. Additionally, the higher transmission rates such as 54 [Mbps] increases comparing to Figs. 7 and 8. Therefore, the route selection at intermediate nodes and a destination node is effective to improve the throughput.

Fig. 10 shows the end-to-end throughput performance of the proposed protocol, the proposed protocol without RREQ selection at intermediate nodes, the proposed protocol without RREQ selection at a destination node, and AODV with ARF. From the results, we can find that the proposed protocol can achieve higher throughput performance. Especially, employing the RREQ selection at a destination node is the most important to improve throughput performance because the destination node can select an adequate route from different candidate routes. On the contrary, AODV with ARF achieves lower throughput performance because some bottleneck links cause deterioration of end-to-end throughput.

IV. Conclusions

In this paper, we proposed a multi-rate routing protocol for mesh networks. Our protocol is based on AODV, and can handle the variation of channel status because routes are constructed when applications try to start communication. The feature of our protocol is to focus on the bottleneck link throughput, which dominates the end-to-end throughput. Then, intermediate nodes and a destination node receive some RREQ messages to find an appropriate link with higher transmission rate. From the numerical results, we can find that the proposed protocol can achieve higher end-to-end throughput performance by using the network simulator QualNet.

Acknowledgment

This work was supported by Grant-in-Aid for Young Scientists (B)(23700075), Japan Society for the Promotion of Science (JSPS).

References

[1] A. Kamerman and L. Monteban, "WaveLAN-IemI: A High-Performance Wireless LAN for the Unlicensed Band", Bell Lab Technical Journal, Vol. 2, No. 3, pp. 118-133, August 1997.

[2] G. Holland, N. Vaqidya, P. Bahl, "A Rate-Adaptive MAC Protocol for Multi-Hop Wireless Networks", ACM International Conference on Mobile Computing and Networking (Mobicom), pp. 236-251, July 2001.

[3] C. E. Perkins, E. M. B-Royer and S. R. Das, "Ad Hoc On-demand Distance Vector(AODV) Routing", IETF, RCF3561, July 2003.

[4] P. Jacquet, P. Muhlethaler, A. Qayyum, A. Laouiti, L. Viennot, and T. Clausen, "Optimized link state routing protocol", Draft-ietf-manet-olsr-02. txt, IETF Internet Draft, January 2000.

[5] D. S. J. De Couto, D. Aguayo, B. A. Chambers, and R. Morris, "Performance of multihop wireless networks: Shortest path is not enough", Computer Communication Review, 33(1):83-88, 2003.

[6] D. H. Kown, W. J. Kim, Y. J. Suh, "Performance comparisons of two on-demand Ad-hoc routing protocols in dynamic rate shifting WLANs", Proceedings of IEEE International Conference on Communications(ICC'03), Anchorage, AK, USA. Piscataway, NJ, USA: IEEE2003 512-516, 2003.

[7] M. M. Rahman, D. S. Kim, C. S. Hong, S. Han, J. P. Hong, "A routing protocol for multi-rate wireless ad-hoc networks: Cross layer approach", ICOIN'09 Proceedings of the 23rd international conference on Information Networking, 2009.

[8] Y. Wei, X. Guo, M. Song, and J. Song, "High throughput route selection in multi-rate wireless mesh networks", J. China Universities of Posts and Telecom. , vol. 15, no. 3, pp. 13-18, 2008.

[9] Network Simulator QualNet,
http//www. scalable-networks. com/

Public civil action as a instrument for protection of environment and brazilian energy resources

Tércio de Sousa Mota
Universidade Federal de Campina Grande - UFCG
Pós Graduação em Recursos Naturais – CTRN
Campina Grande – Paraíba, Brazil

Erivaldo Moreira Barbosa
Universidade Federal de Campina Grande - UFCG
Pós Graduação em Recursos Naturais – CTRN
Campina Grande – Paraíba, Brazil

Gabriela Brasileiro Campos Mota
Universidade Federal de Campina Grande - UFCG
Pós Graduação em Engenharia de Processos – CCT
Campina Grande – Paraíba, Brazil

Maria de Fátima Nóbrega Barbosa
Universidade Federal de Campina Grande - UFCG
Centro de Ciências Jurídicas e Sociais – CCJS
Campina Grande – Paraíba, Brazil

Cristiele de Sousa Mota
União de Ensino Superior de Campina Grande – FAC-CG
Curso de Graduação em Direito
Campina Grande – Paraíba, Brazil

ABSTRACT

Environmental degradation and exploitation of energy resources, often dominated the neglect of real society, which occurred just at a time where the economic and social development were not compatible with the preservation of nature, meant that there was a need to be created legal means to protect the environment and energy resources. Thus, laws aimed at environmental protection have emerged, such as the Law n. 6938/81, Law no. 9.605/98, Law no. 7.347/85, which regulates the implementation of the Public Civil Action for consumer protection, and cultural environment, thus the shape framework for the protection of collective and diffuse interests. The Public Civil Action is means typical and most important procedural protection of the environment and resources Brazilian natural energy. Given this premise, we need to evaluate effectiveness of the institutional functions of the prosecutor, after the holder civil public action to protect public health and social, environmental and other diffuse and collective interests, including the exploitation of resources energy, without prejudice to the legitimacy of others, the right to be assigned to public and private organizations to protect the interests of non-individual as well as point the Public Civil Action as a means of protecting the environment.

Key words: protection, environment, exploitation, energy resources.

1. INTRODUCTION

Environmental degradation and exploitation of natural energy resources, combined to the real neglect of society, which occurred precisely at a time when the economic and social development were not consistent with the preservation of nature, it made with that there was the need to create legal ways for the protection of the environment and energy resources. It appeared, this way, laws aimed at environmental protection. Among these are the Law no. 6.938/81 - National Environmental Policy, Law no. 9.605/98 - Environmental Crimes Law and the Law no. 7.347/85, which regulates the implementation of the Public Civil Action for the protection of the environment, consumer and cultural heritage, so a legal means for the protection of collective and diffuse interests. The Public Civil Action is the typical and most important procedural way of protection of the environment and natural energy resources in Brazil. The Public Civil Action is headquartered in the U.S. class action, which also derive from the French post intérêt d'action, the representative action odhasionprozess English and German. The U.S. class action it is a process initiated in a state or federal court by a group of people with the same legal interest, making it more practical to resolve the dispute, especially in matters of consumer law, in which the

interest is purely individual too small to excite an individual action. To protect, however, the individual interests of those represented by the association that promotes the action, the U.S. Supreme Court established rules restricting the use of class action, the judge two leading cases in 1973 and 1974, concerning the jurisdiction based on an estimate of the injury individual and scientification of potential victims. In Brazil, the first legal text to have on the Public Civil Action was the Federal Complementary Law 40, December 14, 1981, former National Organic Law of the prosecution. However, the Public Civil Action was not the first instrument of defense of diffuse interests in Brazil, with a view to establishing the popular action, disciplined by Law No. 4717 of June 29, 1965. Procedurally, the Consolidation of Labor Laws, approved by Decree-Law 5452 of May 1, 1943, has brought in its wake, the possibility of introduction of collective bargaining before the Labor Court, the unions, the Presidency of the Court of Working for Justice and Attorney General Labour (Ministry of Labor). However, he sought an instrument to safeguard more effectively the general interest. The Federal Constitution of Brazil in 1988, in his art. 129, III, came to admit the defense of "other diffuse and collective interests" is not explicitly listed, which did raise the device previously vetoed. Thus, the Law 8.078/90, by introducing the Code of Consumer Protection, gave the item IV of article. 1 of Law of Public Civil Action your essay originally approved by Congress, which enables you to include diffuse and collective interests to labor among those liable to be brandished by this procedural tool.

2 OBJECTIVE

Given this premise, we have as objective to evaluate the effectiveness and frequency of the institutional functions of the Public Ministry, then the titular of the civil public action to protect the public and social heritage, environmental and other diffuse and collective interests, including the exploration of energy resources, without prejudice to the legitimacy of others, being a justice garanted to government and private agencies for the guardianship of non-individual interests, as well as pointing the Public Civil Action as a instrument of protecting the environment. It was still intended to verify if the environment and the rational exploitation of energy resources in Brazil has been effectively defended, if the legal instruments of protection are able to reverse the increasing environmental degradation; also focus if the administrative measures are sufficient to contain the environmental liabilities; evaluate whether the public civil action can change the reality of environmental protective system, and also demonstrate the importance and effectiveness of public civil action, focusing on the legitimate and the procedure.

3 METHODOLOGY

Then within a scientific line, we adopted the deductive method, where the production of knowledge is based on the criticism of environmentalists and civil counselor. Data collection involved literature survey, as well as environmental laws, Constitution, Code of Protection and Consumer Protection, and the Civil Code of 2002 and the law n° 7.347/85 and Laws no 9.985/00, to thus ensure that scientific research is lead by a system of referrals of legal and scientific knowledge.

4 RESULTS

Society as a whole over time has evolved, and many things have emerged to facilitate their survival in society. The man has been promoting the development of not only technical, but technology, are discovering new ways to grow, be it economic or social.

So it brought an unbridled struggle for breaking barriers inherent in any process in society and has made the man "when it comes to the discovery of machinery and equipment, sure to enhance the space where it is so you can draw your creations.

The environment in these long years of "development" has been forgotten, or rather neglected by those who are embedded in it. What is not seen is that, as part of a context, the man will suffer the consequences arising out of that environment degradation.

Given this reality, the environmental issue has assumed overall size and unlimited, with respect to physical space, since there is anywhere in the subjective right to a healthy environment and balanced.

The Public Civil Action was created for the benefit of all rights for the protection of trans, diffuse and homogeneous. Currently, despite the obstacles that your filing provides, this action presents itself as a major element in the defense of the environment. Therefore, it is important to ensure effectiveness.

For this, we suggest an improvement, compared to the Judiciary, the performance of members and helpers of Justice in the administrative procedures involved in the process through a greater commitment from both the prosecutor, as stakeholders, as well as inform the population of the possibilities that this action provides, so that thus there is a greater public engagement, as well as the contribution to the construction of a stance on surveillance of everyone in the conduct of proceedings, since the interests defended in this action, covering the whole society .In this study, we observed that the Public Civil Action when used in environmental protection and energy resources in Brazil is beneficial because at the same time that repress the practice of harmful acts to the environment, it also aims repair the damage generated by the causative agent.

5 REFERENCES

ATHIAS, Jorge Alex Nunes. Responsabilidade Civil e Meio Ambiente - Breve Panorama do Direito Brasileiro. In: Dano Ambiental, Prevenção, Reparação e Repressão. São Paulo: Revista Editora dos Tribunais, 2011.

ANTUNES, Paulo de Bessa. Direito Ambiental. São Paulo: Lúmen Juris, 2010.

BRASIL. Constituição da República Federativa do Brasil. Brasília: Senado Federal, 2010.

BRASIL. Decreto de Lei n° 6.938. De 31 de Agosto de 1981. Dispõe sobre a Política Nacional do Meio Ambiente, seus fins e mecanismos de formação e aplicação de outras providências.

BRASIL. Decreto de Lei n° 7.347 de 24 de Julho de 1985. Disciplina a ação civil pública de responsabilidade por danos causados ao meio-ambiente, ao consumidor, a bens e direitos de valor artístico, estético, histórico, turístico e paisagístico (VETADO) e dá outras providências.

BRASIL. Decreto de Lei n° 9.605 de 12 de Fevereiro de 1998. Dispõe sobre as sanções penais e administrativas derivadas de condutas e atividades lesivas ao meio ambiente, e dá outras providências.

BRASIL. Decreto de Lei n° 9.985 de 18 de julho de 2000. Regulamenta o art. 225, § 1o, incisos I, II, III e VII da Constituição Federal, institui o Sistema Nacional de Unidades de Conservação da Natureza e dá outras providências.

CARVALHO, Carlos Gomes Introdução ao Direito Ambiental. São Paulo: Conceito Editorial, 2008.

CORRÊA, Jackson. Proteção Ambiental & Atividade Minerária. Curitiba: Juruá. 2011.

DERANI, Cristiane. Direito Ambiental Econômico. São Paulo: Max Limonad, 2010.

FERRAZ, Sérgio. Responsabilidade Civil por Dano Ecológico. Revista de Direito Público. Vol: 49-50. São Paulo, 2011.

FARIAS, Talden. Introdução ao Direito Ambiental. Belo Horizonte: Ed. Del Rey, 2009.

FIORELLO, Celso Antonio Pacheco. Curso de Direito Ambiental Brasileiro. 10 ed. São Paulo: Saraiva, 2011.
GIANNINI, Massimo Severo. El Poder Publico: Estados Y Administraciones Publicas. Milano: Civitas, 2010.
MACHADO, Paulo Afonso Leme. Direito Ambiental Brasileiro, 6ª ed. São Paulo: Malheiros Editores Ltda, 2009.
MILARÉ, Edis. A ação civil pública na nova ordem constitucional. São Paulo: Saraiva, 2010.
_____ "Tutela jurídico-civil do ambiente". Revista de Direito Ambiental. São Paulo, 2010.
MONTIBELLER Filho, Gilberto. O Mito do Desenvolvimento Sustentável. Florianópolis: UFSC, 2011.
SILVA, José Afonso. Direito Ambiental Constitucional. 2. ed. São Paulo: Malheiros, 2009.

OFDM cooperative flooding mechanisms for Multi-hop networks

Katsuhiro Naito, Tomoyuki Hagino, Kazuo Mori, and Hideo Kobayashi
Department of Electrical and Electronic Engineering, Mie University,
1577 Kurimamachiya, Tsu, 514-8507, Japan
Email: {naito, kmori, koba}@elec.mie-u.ac.jp, {hagino}@com.elec.mie-u.ac.jp

***Abstract*— Flooding mechanisms, where all nodes retransmit same information to a whole network, are used in various routing protocols and dissemination schemes in multi-hop networks. It is well known that flooding mechanisms consume a lot of wireless resources. Additionally, broadcast storm problems occur when a number of nodes or traffic increase in a network. In this paper, we focus on features of orthogonal frequency-division multiplexing (OFDM), which is employed in IEEE 802.11a standard. In OFDM, receivers can demodulate some same OFDM signals within guard interval (GI) period. Therefore, some senders transmit a same OFDM signal to a receiver synchronously in the proposed method. In addition, the paper proposes autonomous synchronous method for the proposed transmission method. Simulation results show that the proposed method can reduce consumed a wireless resource and improve a packet delivery ration and delay performance.**

***Keywords*— OFDM, Cooperative communication, Flooding, Multi-hop communication**

I. Introduction

Various studies about wireless multi-hop networks have been proposed according to the development of wireless communication technologies. In multi-hop networks, broadcast mechanisms called flooding have been used in various purposes [1], [2].

In unicast routing protocols for multi-hop networks, flooding mechanisms have been used to search a destination node. In AODV [3] and DYMO [4], a source node broadcasts a route request message to its neighbor nodes. Then, other nodes rebroadcast the route request message again. Finally, the route request message is broadcasted in the whole network.

In multicast routing protocols for multi-hop networks, flooding mechanisms have been used to construct a delivery tree for data messages [5]. Additionally, data messages are rebroadcasted on the delivery tree. Therefore, multicast data delivery is almost semi-flooding mechanisms.

In applications, flooding mechanisms have been used for ITS communication with reliable and short delay communication [6]. In ITS communication, it is difficult to construct a route due to frequent change of network topologies. Additionally, emergency safety messages for active safety applications require high reliability and short delivery delay. Since flooding mechanisms do not require any routing protocols, they are used for delivery of emergency safety messages.

As mention above, flooding mechanisms are important function in multi-hop networks. But, it is well known that flooding mechanisms suffer from packet corruption when the number of nodes in a network increases [7]. The problem is called the broadcast storm problem. It occurs due to rebroadcasting process by all nodes. This process consumes a lot of wireless resources. Therefore, it is known that flooding mechanisms are simple and easy way to deliver data in a whole network, but is not scalable.

In this paper, we propose a new cooperative communication scheme for OFDM. OFDM is a well known modulation technique, which is employed in IEEE 802.11a, g, WiMax, and 3G systems [8], [9]. Therefore, the proposed cooperative scheme can be applied in various types of wireless communication systems. Additionally, we propose a new flooding mechanism based on the OFDM cooperative communication. In the proposed flooding mechanism, neighbor nodes can transmit a same message simultaneously. Therefore, we can reduce the consumed wireless resource and delivery delay by simultaneous transmission. In the numerical results, we show that the proposed cooperative scheme can improve transmission performance and the proposed flooding mechanisms can improve message delivery ratio and delivery delay performance.

II. OFDM cooperative communication

OFDM is employed in IEEE 802.11a, g which is the candidate device for multi-hop communication. It is well known that OFDM has high tolerance for multi-path communication

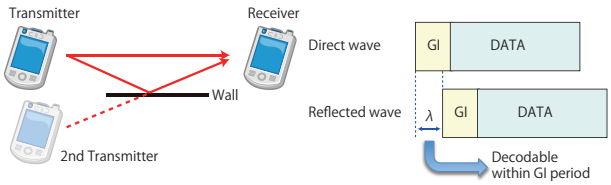


Fig. 1. Concept of cooperative communication in OFDM.

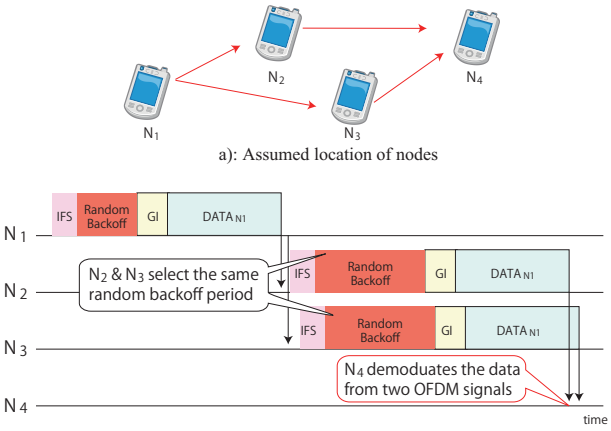


Fig. 2. Concept of proposed access control mechanism.

environment.

Fig. 1 shows the high tolerance of OFDM communication for multi-path waves. In this example, the receiver receives the direct wave and the reflected wave from the transmitter. In the situation, the reflected wave arrives with λ delay comparing to the direct wave. Generally, the transmitter adds guard interval (GI), which is a posterior part of an OFDM signal, to the head of OFDM signal. As the results, the receiver can demodulate OFDM signals including the direct wave and some reflected waves within GI period.

In the proposed communication, the other transmitter transmits the same OFDM signal, which seems to be same as the reflected wave at the receiver. In Fig. 1, the second transmitter transmits the same OFDM signal in synchronism with the first transmitter. The receiver receives two OFDM signals from the first and the second transmitters. Then, it can demodulate the OFDM signals correctly when the arrival difference is within GI period. In this situation, we can reduce the consumed wireless resource by transmitting the same OFDM signal simultaneously, and improve communication performance by using different wireless path.

III. Access Control for OFDM Cooperative Communication

The proposed cooperative communication requires autonomous synchronization mechanisms for transmission of same OFDM signals. Fig. 2 shows the concept of the proposed access control mechanism. The proposed mechanism is based on CSMA/CA (Carrier Sense Multiple Access/Collision Avoidance), and is extended to synchronize with neighbor nodes. In the example, node 1 transmits the OFDM signal, and nodes 2 and 3 retransmit the OFDM signals simultaneously. At the retransmission process at nodes 2 and 3, they initialize the random back-off value for CSMA with information such as a source address, a hop count, a packet sequence in the packet. Therefore, they can obtain the same back-off value according to the information in the packet. Finally, they can synchronize the transmission timing of the OFDM signals, and node 4 can demodulate the OFDM signals from nodes 2 and 3 correctly because the arrival difference of the OFDM signals is within GI period.

Fig. 3 shows the flow chart of the proposed access control mechanism. We extend CSMA mechanisms to achieve group access control because the proposed cooperative communication requires the synchronous transmission by some neighbor nodes. In general CSMA mechanisms, nodes select a random back-off value independently to disperse transmission timing of each node and avoid contention of signals. In the proposed cooperative communication, different packets are broadcasted simultaneously in networks. Therefore, nodes are required two conditions to transmit packets. First one is that nodes should select a random back-off value to disperse transmission timing and avoid contention between different packets because neighbor nodes may transmit a different packet. Second one is that nodes that transmit a same packet should select a same random back-off value to synchronize the transmission timing to achieve the reception of packet within GI period. In the proposed access control mechanisms, the following procedures are performed.

- The procedures start when nodes receive a packet from neighbor nodes.
- The nodes check the transmission record to avoid duplicate retransmission of packets. They discard the packet when the packet has been transmitted yet.
- The nodes initialize the Network Allocation Vector (NAV) to set Back-Off Index (BOI) value where CW_{min} is the minimum contention windows size.
- The nodes initialize the random sequence by using the

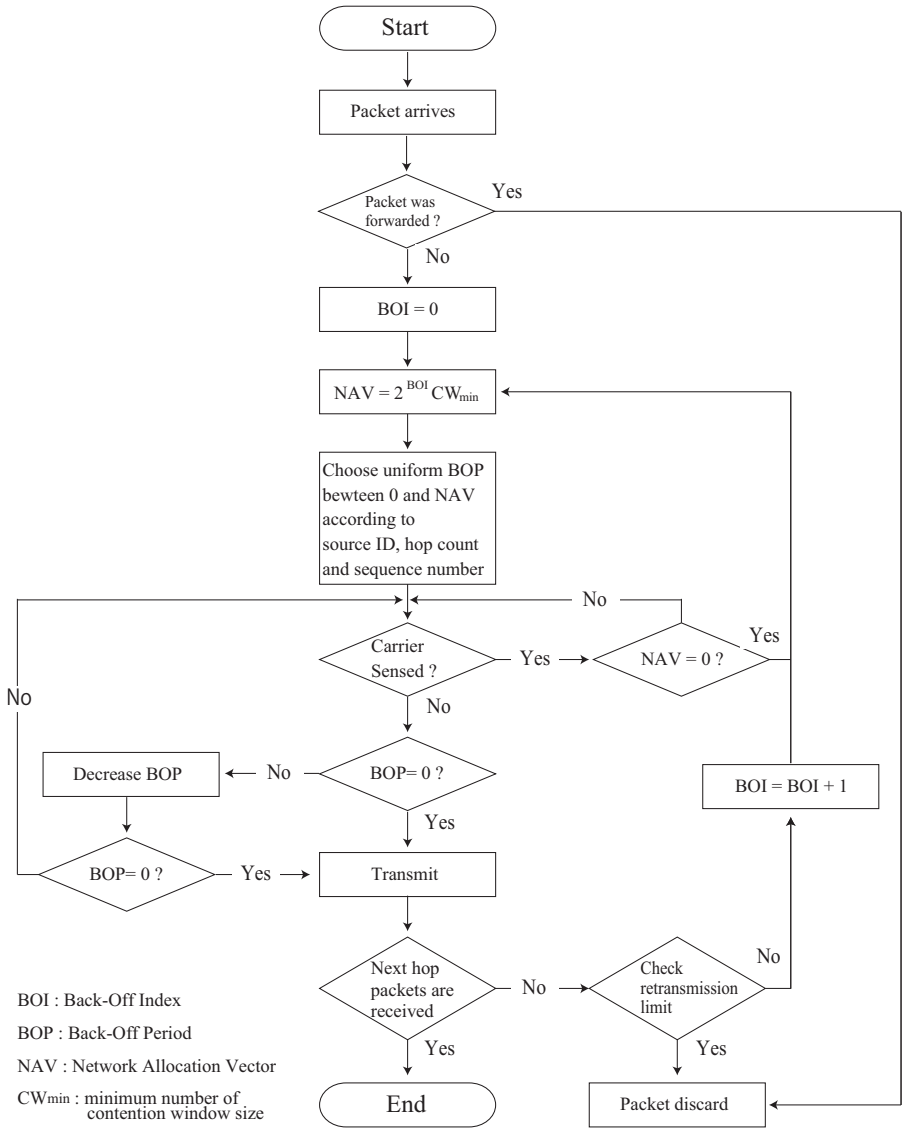


Fig. 3. Flowchart of media access control.

packet information such as a source address, number of hop counts, packet sequence number. Therefore, they can obtain the random value according to the initial value of random sequence. This means that the nodes with the same packet information can obtain the same random value according to the packet information. Then, they generate the random value, which is the same value between the nodes with same packet information, and set the generated random value as the back-off period for CSMA mechanisms.

- The nodes try to sense the channel to check the signal transmission from neighbor nodes. They continue to sense the channel when the channel is used by neighbor nodes. Finally, they check the BOP value when the channel is not busy.
- The nodes decrease the BOP value when the value does not equal to 0. Finally, they transmit an ODFM signal when the BOP value equals to 0.
- The nodes sense the channel after transmission of the OFDM signal. They finalize the procedures when they receive the same packet from neighbor nodes because this means that the neighbor nodes can receive the transmitted OFDM signal.

IV. Numerical results

A. BER performance in OFDM cooperative communication

We perform the simulation of the proposed OFDM cooperative communication by Matlab. In the simulation, we assume

TABLE I

SIMULATION PARAMETERS IN PHYSICAL LAYER SIMULATION

Simulator	Matlab 6.5
Number of FFT points	64
Number of Subcarriers	52
Number of pilot subcarriers	14
Bandwidth	20 [Mhz]
Modulation scheme	16QAM
Symbol period	2.6 [μs]
GI period	0.52 [μs]
Channel model	Rayleigh fading
Number of multi-path	5

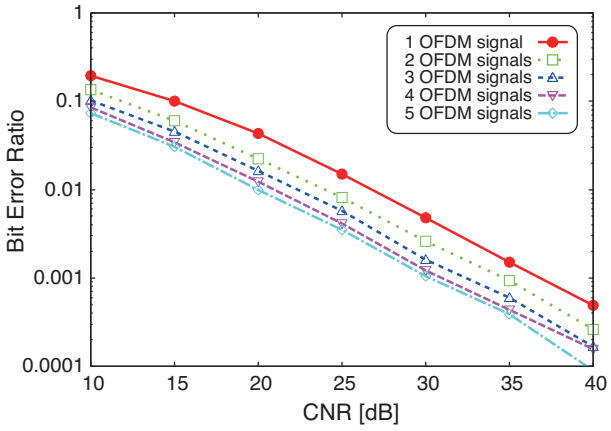


Fig. 4. BER performance in OFDM cooperative communication.

that 52 subcarriers including 14 pilot subcarriers, the number of FFT points is 64, the bandwidth is 20 [Mhz]. We employ 16 QAM as the modulation scheme, and set the symbol period as 2.6 [μs] and the GI period as 0.52 [μs]. In the simulation, we assume rayleigh fading environments and the distance between transmitters and a receiver is same. Detail simulation parameters are shown in Tab. I.

Fig. 4 shows the bit error ratio performance in the proposed cooperative scheme. From the results, we can find that the proposed scheme can improve the BER performance when the number of transmitters increases because a receiver can receive larger signal power according to the increasing of transmitters. Additionally, the channel condition between transmitters and a receiver is different. Therefore, the proposed scheme can obtain the performance improvement by path diversity effect.

B. Flooding performance based on cooperative communication

We perform the simulation of flooding mechanisms based on the proposed OFDM cooperative communication by QualNet. We implemented the proposed cooperative communication by extending IEEE 802.11a module in QualNet. In the simulations, we assume that 100 nodes are randomly distributed within 1000 squares meters area, and nodes moves based on the random way point model. Source nodes transmit 512 bytes packets with one second interval. The results are the average of 100 simulations. Detail simulation parameters are shown in Tab. II.

TABLE II

SIMULATION PARAMETERS IN NETWORK LAYER SIMULATION

Simulator	QualNet 5.01[10]
Simulation time	150 [s]
Simulation trial	100 [times]
Number of nodes	100 [nodes]
Node movement	Random waypoint Pause time : 5 [s] Speed : 0.1 - 2 [m/s]
Node position	Random
Simulation area	1000 x 1000 [m]
Transmission interval	1 [s]
Packet size	512 [Bytes]
Communication device	IEEE 802.11a
Transmission rates	6 [Mbps]
Transmission power	19 [dBm]
Channel frequency	5 [GHz]
Antenna gain	0 [dB]
Antenna type	Omni directional
Antenna height	1.5 [m]
Propagation path loss model	Free Space
Wireless environment	Rayleigh fading

Fig. 5 shows the packet reception ratio, which means that the number of correctly received packets is divided by the number of detected signals. From the results, the packet reception ratio of full flooding is the lowest. In the full flooding, nodes can receive only 50 % of packets because nodes transmit a same packet in rotation after confirmation of channel free by channel sensing mechanisms. In IEEE 802.11 system, the flooding mechanism can be achieved by broadcast transmission mode, which employs only channel sensing mechanisms, does not employ request to send / clear to send (RTS/CTS) mechanisms. Therefore, the broadcast transmission mode may suffer from hidden node problems. In the flooding, many nodes should transmit a same packet. As the results, signals for many packets tend to conflict.

In the probabilistic flooding, nodes can reduce the number of transmitted packets with predefined probability. Therefore, the probabilistic flooding alleviates the traffic congestion. However, the total traffic also increases when the number

of nodes or the number of source packets increases even if the nodes employ probabilistic reduction of transmission packets. In the simulation condition, we can find that the packet reception ratio of the probabilistic flooding can be improved a little. Therefore, the probabilistic flooding is not enough to reduce the packet contention and improve the packet reception ratio.

In the proposed cooperative flooding mechanisms, the packet reception ratio is improved up to 80 %. In the proposed mechanisms, nodes can demodulate an OFDM signal correctly when packets are received within GI period. Additionally, nodes can transmit a same OFDM signal simultaneously. Therefore, the proposed mechanisms can reduce the number of transmitted packet virtually because these packets are transmitted at same time and required wireless resource is also decreased. As the results, we can improve the packet reception ratio by demodulating some OFDM signals and reducing network traffic.

Fig. 6 shows the delivery delay between a source node and each receiver node. Therefore, accurate delivery delay depends on the relative position between these nodes. The results are the average of delivery delay between a source node and all receiver nodes. From the results, the delivery delay in full flooding increases dynamically according to the increase in the number of source nodes because total traffic in the network also increases when the number of source nodes increases. In the heavy traffic condition, each node cannot obtain the transmission chance in a short time sensing period. Therefore, the full flooding mechanisms do not suit for short delivery delay service.

In the probabilistic flooding, we can find that the delivery delay can be reduced according to the predefined forwarding probability such as 75 % because the available wireless resource is one of the dominant factor for delivery delay. The probabilistic flooding mechanisms can reduce the number of transmitted packets and increase the available wireless resource. Hence, these mechanisms are effective to reduce the delivery delay even if these mechanisms cannot improve the packet reception ratio.

In the proposed mechanisms, the delivery delay can be reduced to 1/3 comparing to the full flooding mechanisms because nodes can transmit a same OFDM symbol simultaneously. Therefore, nodes tend to obtain the transmission chance according to increasing of available wireless resource. As the results, the proposed mechanisms are effective to reduce the delivery delay and improve the packet reception ratio.

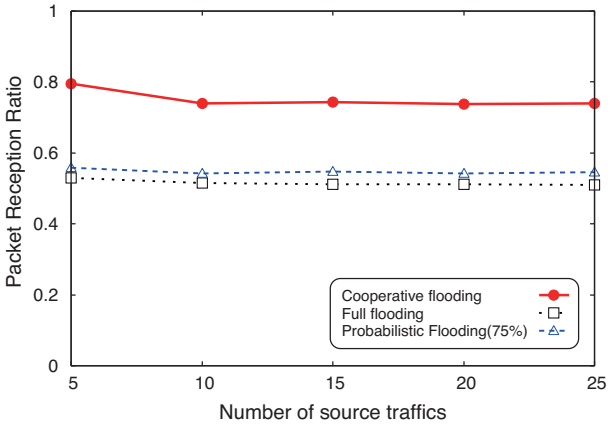


Fig. 5. Packet reception ratio.

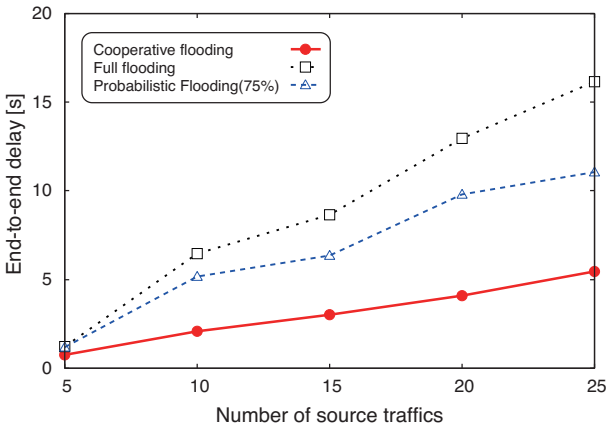


Fig. 6. End-to-end delay.

V. Conclusions

This paper focuses on the characteristic of OFDM, which is the well known modulation scheme for recent wireless devices. The original point is that nodes transmit a same OFDM symbol simultaneously to achieve cooperative packet reception within a guard interval period in OFDM at receiver node. Additionally, we have proposed the channel access control mechanisms to achieve autonomous group synchronization of packet transmission timing by extending CSMA mechanisms. Therefore, the proposed mechanisms can achieve synchronous transmission in multi-hop networks that do not have any infrastructure for transmission timing control. From the physical layer simulations, we can find that the BER performance can be improved by cooperative transmission. Additionally, the proposed mechanisms can improve the packet reception ratio and reduce the delivery delay from the network simulations.

ACKNOWLEDGMENT

This work was supported by Grant-in-Aid for Young Scientists (B)(23700075), Japan Society for the Promotion of Science (JSPS).

REFERENCES

[1] K. Zeng; Z. Yang; W. Lou, "Location-Aided Opportunistic Forwarding in Multirate and Multihop Wireless Networks," IEEE Transactions on Vehicular Technology, Vol. 58, Issue 6, Jul. 2009.

[2] M. Naghshvar, T. Javidi, "Opportunistic Routing with Congestion Diversity in Wireless Multi-hop Networks," in Proc. of the IEEE Conference INFOCOM 2010, November 2010.

[3] C. Perkins, E. Belding-Royer, and S. Das, "Ad hoc On-Demand Distance Vector (AODV) Routing," IETF Request for Comments 3561, Jul. 2003.

[4] I. Chakeres, C. Perkins, "Dynamic MANET On-demand (DYMO) Routing," IETF internet draft (work in progress), Jul. 2010.

[5] Y. Yi, S. Lee, W. Su, M. Gerla, "Optimized Link State Routing Protocol (OLSR)," IETF draft-ietf-manet-odmrp-04, Nov. 2002.

[6] J. Mittag, F. Thomas, J. Härri, and H. Hartenstein, "A comparison of single- and multi-hop beaconing in VANETs," ACM VANET '09, pp. 69–78, Sep. 2009.

[7] Y.-C. Tseng, S.-Y. Ni, Y.-S. Chen, and J.-P. Sheu, "The broadcast storm problem in a mobile ad hoc network, " Wireless Networks, Vol. 8, pp. 153–167, 2002.

[8] D. Lee and K. Cheun, "A new symbol timing recovery algorithm for OFDM systems," IEEE Transactions on Consumer Electronics, Vol. 43 No. 3, pp. 767–775, Aug. 1997.

[9] L. Thibault and M. T. Le, "Performance evaluation of COFDM for digital audio broadcasting. I. Parametric study," IEEE Transactions on Broadcasting, Vol. 43, No. 1, pp. 64–75, Mar. 1997.

[10] QualNet, URL:http://www.scalable-networks.com

Virtual and augmented environments and realistic user interactions to achieve embedded accessibility designs, the VERITAS project Use Cases methodological framework and outcomes

Eleni Chalkia
CERTH/HIT
Athens, 15231, Attica, Greece
and
Evangelos Bekiaris,
CERTH/HIT
Athens, 15231, Attica, Greece

ABSTRACT

VERITAS (Virtual and augmented environments and realistic user interactions to achieve embedded accessibility designs), is a 7th Framework project, which core concept is the research and development of an open framework for providing inbuilt accessibility support at all the stages of realisation of mainstream ICT and non-ICT technologies. The project aims at delivering to product and software developers 'generic' instructions - embedded in an empowering virtual reality platform, for exploring new concepts, designing new interfaces and testing interactive prototypes that will inherit universal accessibility features, including compatibility with established assistive technologies.

When developing interactive systems, within a specific oriented project like VERITAS, the correlation between the system and the users is a challenging task, especially when the main purpose is providing solution to a problem considering users' environment. The Use Cases come to provide a representation of the contract between the stakeholders and the system's behavior. The Use Cases are used as an assistive tool to the designers, to help them understand and create computer systems and applications as artifacts of human activity. In this paper we will present the VERITAS project, the methodology followed for the extraction of its Use Cases providing selected examples of the procedure followed.

Keywords: Use Cases, Virtual environment, user interaction, simulation and accessible design.

1. INTRODUCTION

One of the highest priorities of the European Commission, as well as of the European Society, during the last years, is the equal and yet cost-efficient support of the growing disabled and senior citizens. People with disabilities are not just a tiny minority of the population of the European Union. The lowest estimate, based on the currently defined disablement categories, estimates their total number at around 74 Million persons. However, when including in this group also the people with cognitive impairments, as well as people who are in the hinterland between fully able bodied and classically impaired, the percentage of people with disabilities across Europe reaches 15% [1].

In VERITAS, our target is to create tools that assist the developers in creating accessible ICT and non-ICT products for these disabled users that are embedded in the application they are using for designing these products, so as to have an inbuilt accessibility environment. The main VERITAS innovation lies in the fact that, even if there have been some limited and isolated attempts to support accessibility testing of novel products and applications, there is a clear lack of a holistic framework that supports comprehensively virtual user modelling, simulation and testing at all development stages and realistic/immersive experience of the simulation.

To this end, VERITAS aims to develop tools for inbuilt accessibility support at all stages of ICT and non-ICT product development. The goal is to introduce simulation-based and virtual reality based all-inclusive models at all stages of product design and development into Automotive, Smart living spaces, Workplace, Infotainment and Healthcare application domains.

Thus, the goal of VERITAS is to ensure that future products and services are being systematically designed for all people, including those with disabilities and functional limitations, as well as older people. In the current paper we will present the way the end users interact with the outcomes of VERITAS as initially introduced, in the form of narratives which constitute the project's Use Cases and Application Scenarios.

2. USE CASES CONCEPT

When developing interactive systems, within a specific oriented project like VERITAS, we create possibilities for learning, work, and leisure, for interaction and information. In the design process of these systems, the correlation between the system and the users is a challenging task. This task is very demanding, especially when the main purpose is to provide a solution to a problem considering the users' environment. The Use Cases come to provide an indicative solution and representation of the contract between the stakeholders and the system's behaviour. The Use Cases are used as an assistive tool to the designers, so as to help them understand and create computer systems and applications as artefacts of human activity, as things to learn from, as tools to use in one's work, as media for interacting with other people [2].

Use Cases were initially presented by Ivar Jacobson in 1967 as usage scenarios and became immediately attractive because the term implies "the ways in which a user uses a system" [3]. In the mid-1980s, Jacobson coined the Swedish term "anvendningsfall", which roughly means "situation of usage" or "usage case", but when publishing into English translated it in use case [4].

Since Jacobson, the Use Cases have been used extensively in different points of most of the system development. They can be used to stimulate discussion within a team about a system-to-be. They might be used in order to report the actual system requirements, which will emerge from the user needs. Additionally, the systems final design can be documented using the same use case form. The Use Cases can be used for a system as large as an entire company, or as small as a piece of a software application program.

The main purpose of the Use Cases is to present, in a detailed and also clear and easy-to-understand way, the functional requirements of a system in a non-technical manner, but in a way that will describe the user's environment and yet will be easy from the developer to translate into technical characteristics of the system. The Use Cases can also be considered as a description of a system's behaviour, written from the point of view of a user who has told the system to do something specific. In this way, the Use Cases have the unique ability to help teams to understand the value that the system provides to its stakeholders [5]. In a more simple approach, Use Cases describe who is doing what and when, and also what is expected from the system for each request. To this end, Use Cases comprise a powerful tool to capture functional requirements for software systems, in order to evaluate them [6].

There are many ways in which a researcher can write the Use Cases. They may be presented as simple scenarios, like narrative stories as it has been proposed from Carroll & Rosson in 2002 at the "Scenario-based design" [2], or it can consist of various different parts that decompose the scenarios, as proposed from Ivar Jacobson in 1986 with the detailed templates (fully dressed or casual) [4]. In all cases, the Use Cases should be well-written and easy to be read from the designer. They should be consisted of sentences written in only one grammatical form, a simple action step, in which an actor achieves a result or passes information to another actor.

In VERITAS case we have used the scenario based design in order to identify our Use Cases and during the design process of the project, we have enriched their descriptions using the Ivar Jacobson fully dressed template. The fully dressed template is characterized by:

- One column of text (not a table).
- Numbered steps.
- A numbering convention in the extensions sections that involves combinations of digits and letters (e.g. 2a, 2a1, 2a2, etc.).

The fully dressed template includes the most basic elements for a Use Case representation like the Use Case description, which is a small narrative that gives the goal of the Use Case, the scenarios that decompose the Use Case in smaller parts, the actors that participate in the Use Case, the priority level, the system input and output as well as the interaction steps between the user and the system in order to achieve the goal. Thus, a Use Case is a collection of related success and failure scenarios that describe actors using the system to support a goal. So, a goal holds together all the scenarios (success and failure). Scenarios and Use Cases go until goal success or abandonment.

But what is the deference between a scenario and a Use Case? The scenario is a sequence of interactions happening under certain conditions, to achieve the primary actor's goal, and having a particular result with respect to that goal. The interactions start from the triggering action and continue until the goal is delivered or abandoned, and the system completes whatever responsibilities it has with respect to the interaction. A Use Case is a collection of possible scenarios between the system under discussion and external actors, characterized by the goal the primary actor has towards the system's declared responsibilities, showing how the primary actor's goal might be delivered or might fail.

Another very important element of the Use Case is the actors involved, namely the stakeholders. A stakeholder is usually an actor of the system, primary or secondary. A primary actor is a stakeholder that calls upon the system to deliver one of its services. The primary actor has a goal with respect to the system, one that can be satisfied by its functions. The primary actor is usually the actor, who triggers the use case, so the use case starts because the primary actor sends a message, pushes a button, enters a

keystroke, or in some other way initiates the story. There are two common situations in which the initiator of the use case is not the primary actor. The first is when a Use Case is triggered from another Use Cases, like a sequence of events, and the second is when the Use Case is triggered by time. The primary actors are very important for the system under development because they are the ones that the system is actually designed for. They are the users of the system, which should cover their needs and requirements. The list of the primary actors of the system should be done in the very beginning of the project and this list should be used for the whole duration of the project.

Despite the fact that the primary actor is very important during the requirements gathering process, it could appear to be not so important for the Use Cases itself. What happens is that, over time, the use case writers discover that a use case can be used by multiple sorts of actors, whose needs are more or less the same. So, in the Use Cases when we refer to a primary actor, we mainly refer to the user who will do the specific task and might have, or not, the profile of the actor that has been set during the User Needs phase. To this end, sometimes, in a Use Case, by an actor we mean an individual, but we also mean the general category of individuals who can play that role. Of course, the primary actor becomes important again just before the finalization of the system, when the different actors are called to test and evaluate the system prototypes.

3. VERITAS USE CASES METHODOLOGICAL FRAMEWORK

Use Cases generally are designed to be used by the designer/developer. Despite that, in VERITAS, when developing the Use Cases, we need to bear in mind both the end users (developers and designers), but also the beneficiaries (elderly and disabled). Even if the Use Cases, since are technical driven, will only address the end users (developers, designers), the beneficiaries (elderly and disabled) and their requirements should be also included, since the final product, which will be developed using the VERITAS tools, will be used by them. Under this framework, each technical detail of the system must be clearly defined and correlated with the actors that are involved in it and, furthermore, there must be a clear and understandable explanation of the interaction between the user and each technological component-function.

The basis of the Use Cases of a system under development lies upon the user requirements. According to Brooks, requirements always change, especially when designs incorporate rapidly-evolving technologies, requirements change rapidly every 2-3 years [7]. The more successful a design is, the more widely-adopted is, and the quicker its users expect more from it in terms of functionalities and design requirements. So, the Use Cases need to be updated constantly during the project, in order to fit the updated user requirements.

Additionally, the Use Cases can be used in various steps of the project duration. The iterative evaluation and updated procedure for the evolvement of the Use Cases during the project are presented and explained in this section, showing in which state of the process we are now, what has been done in the previous years and what has still to be done.

As it has initially identified by Rubin, Use Cases are applied in many steps of the projects lifecycle [8]. This stands also in VERITAS, where the Use Cases have and will be applied in various steps of the development process. In the Analysis domain, the research started with the definition of the problem scenarios, which defined in order to assist on the methodology for the capture of the user and the industrial needs. The problem scenarios have been reformed according to the outcomes of the user and industrial needs, as well as the benchmarking of the existing models, into the initial list of the Use Cases which has been prioritized, discussed and assessed by end-users and beneficiaries, as well as by internal and external to the project experts, iteratively until its finalization. Finally, the Use Cases have been reformed into pilot scenarios that will be used from the external developers during the pilot testing phase in order to assess the functionality, usability, usefulness, satisfaction and interaction of the tools developed. A more detailed representation of the participation of the Use Cases in the different project steps, that has been adopted and followed in VERITAS, is presented in the scheme that follows.

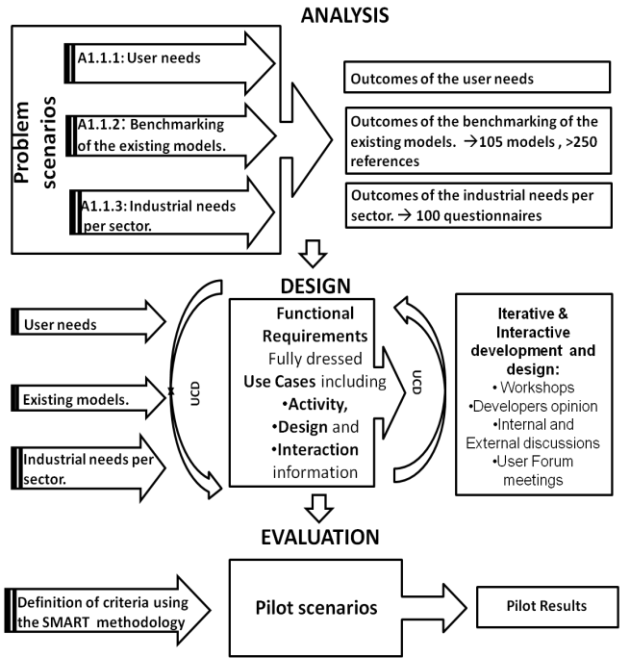


Figure 1: VERITAS Use Cases design framework

According to the figure above, the Use Cases are initially used in the Analysis of the problem of the project, in a more abstract format. Then they take their integrated format for the design process and finally are reformed into pilot scenarios in order to be used in the assessment and the evaluation of the developed systems. The outcome of this methodology at the moment consist of a set of problem scenarios, a set of 20 Use Cases and a set of preliminary pilot scenarios.

In these following sections, we will refer to all the development steps, by giving also concrete examples of the form of the Use Cases where possible.

4. ANALYSIS DOMAIN

The analysis domain has been mainly designed, discussed and developed in the beginning of the project when the main stakeholders' groups have been identified, the involved designers have been highlighted, as well as the beneficiaries groups and clusters have been reported; the industrial needs of designers and developers within the VERITAS application sectors has been identified, by pinpointing also in detail the similarities and the differences between them.

In the beginning of the analysis domain, and in order to have a clear vision of what is the problem that we are dealing with, some problem scenarios for the "before" situation have been drafted. Using the Use Cases, in the form of problem scenarios, at the analysis stage, can prevent the occurrence of costly error correction at later stages of the development cycle. In the following sections we will present one example of each application domain.

Below we present an indicative problem scenario for the automotive domain.

Problem scenario:
Paul has a company specialized in adjusting vehicles for use by people with disabilities. His team has extensive expertise in this field; however they still do depend heavily on personalization case by case since every user has a different set of needs. He therefore works together with driver rehabilitation specialists who perform comprehensive evaluations to identify the adaptive equipment most suited to the needs of the driver with disabilities. A complete evaluation includes vision screening and, in general, assesses:

- Muscle strength, flexibility, and range of motion
- Coordination and reaction time
- Judgment and decision making abilities
- Ability to drive with adaptive equipment

Upon completion of such an evaluation, the driver receives a report containing specific recommendations on driving requirements or restrictions, and a complete list of recommended vehicle modifications.

Equally, Paul's evaluators also consult on compatibility and transportation safety issues for passengers with disabilities. They assess the type of seating needed and the person's ability to exit and enter the vehicle. They provide advice on the purchase of modified vehicles and recommend appropriate wheelchair lifts or other equipment.

With the outcomes of the evaluation, Paul's team would start adjusting a car that seems to fit the needs of the user. While the vehicle is being modified, the user/driver must be available for fittings. This avoids additional waiting time for adjustments once the equipment is fully installed. Without proper fittings, problems might arise with the safe operation of the vehicle, and might results in the vehicle to be brought back for adjustments.

While Paul's team never ended in a situation that the car is not useful for the driver, it has happened many times that modifications still had to be carried out after the first presentation to the user of the completed vehicle. Paul's "ethical company bible" clearly states that no car leaves the premises without 100% satisfaction from the driver/user, thus iterations often happened, till the user/driver was fully satisfied.

The problem scenarios developed in this step are design-neutral and present the situation as it is trying to describe the problem of the users. Even if the problem scenarios have been developed during the early analysis of the project the team had a clear yet flexible idea of how technology might enhance current practice. The high level of abstraction of the problem scenarios has allowed the team to reform its visions without influencing the extraction of the user requirements or the extraction of the Use Cases list.

The outcomes of the Analysis domain have formed the basis for the update of the evolvement of the problem scenarios into the initial version of the Use Cases that is used in the Design domain that follows.

5. DESIGN DOMAIN

In the design domain, the protagonist is the functional requirements of the system-to-be. The functional requirements capture the intended behavior of the system. This behavior may be expressed as services, tasks or functions the system is required to perform. The systems functional requirements, either basic or exclusive are the features that characterize the system, making it useful and usable for the target groups and allow it to penetrate into the market [6].

The Use Cases are an effective tool for gathering the user requirements and presenting the functional requirements

of a system, when they are developed in a disciplined (systematic) and coherent manner, as part of a methodology that first creates a well defined domain-model.

For extraction of the Use Cases in the design domain the following realization steps too place.

1. Define a Use Cases Model (capable of supporting the variety of situations that VERITAS will support). The model is comprised of:
   - **Template** for the Use Cases descriptions (based on the fully-dressed format), in which the elements that are necessary according to VERITAS objectives should be inserted gradually. Important is to keep Use Cases very simple, in order to produce a Use Cases Index.
   - **Diagrams** accompanying each of the Use Cases textual descriptions (on the template). Diagrams are designed with the same capabilities of the template, i.e., providing the possibility of attaching details later. The diagrams are realized in a close to final step of the Use Cases, but are part of the Use Case model.
2. Define the Use Cases Index (comprised of a set of Use Cases titles). This step is achieved before the previous one, even if it comes second at the methodology. It corresponds to the Use Cases Index (comprised of a set of UC names) and it is the first level of detail to start writing Use Cases. The reason why this step comes second is the need to have a model before (from 1st step).
3. Discuss the initial list of Use Cases among the partners in order to come to a first common version, having in mind the following success criteria:
   - Containing all elements: such as UCs aim/scope, the trigger for the UC, the primary actor and possibly other stakeholders, all the interests of the stakeholders, preconditions, success and failure conditions.
   - The UC should contain a template with the necessary elements and diagrams, as well as the UC textual descriptions.
   - The UC should clearly show under what conditions the VERITAS functionalities are successful in relation to the problem/goal of the primary user.
   - The UC should clearly show what the minimal functionalities should be in relation to the problem/goal of the primary user for successful results.
4. Present the Use Cases in the 1st VERITAS Pan-European Workshop and User Forum with users (designers-developers) and with beneficiaries (elderly and disabled). Create a template for the Use Cases prioritization and a methodology to be followed. Prioritization of the Use Cases from both the users and the beneficiaries and notes of their comments.
5. Update of the Use Cases according to the 1st VERITAS Pan-European Workshop and Use Forum, correlate them to the different disability types and connect them to specific personas.
6. Discuss the Use Cases with the project partners and organize a review from external developers.
7. After gathering the commends from the internal and external developers the Use Cases have been updated and presented again to the users and the beneficiaries during the 2nd Pan-European Workshop and User Forum, where they have been discussed and commended.
8. After taking under consideration the commends from the 2nd Pan-European Workshop and User Forum of VERITAS, the Use Cases have been updated and discussed once more with the project partners. Their final format was structured and the same stands for their content.
9. Design of the UML diagrams accomplishing each of the UCs textual descriptions (on the template).

On the basis of the aforementioned framework the Use Cases of VERITAS were extracted and the extended list is presented below.

Category 1: Use Framework	
UC 1.1: User model generator. UC 1.2: Model platform. UC 1.3: Intelligent avatar editor. UC 1.4: Interaction adaptor. UC 1.5: Core simulation. UC 1.6: Multimodal interfaces. UC 1.7: Interaction manager and immersive simulation.	
Category 2.a: Automotive desktop design	**Category 2.b: Automotive immersive design**
UC 2.1.a: Car interior desktop design	UC 2.1.b: Car interior immersive design
UC 2.2.a: Motorcycle handling desktop design	
UC 2.3.a: ADAS/IVIS desktop design	UC 2.3.b: ADAS/IVIS immersive design
UC 2.4.a: ARAS/OBIS design	
Category 3.a: Smart living Spaces desktop design	**Category 3.b: Smart living Spaces immersive design**
UC 3.1.a: Home interior desktop design.	UC 3.1.b: Home interior immersive design.
UC 3.2.a: Domotics desktop design.	UC 3.2.b: Domotics immersive design.
Category 4.a: Workplaces desktop design	**Category 4.b: Workplaces immersive design**
UC 4.1.a: Office desktop design.	UC 4.1.b: Office immersive design.
UC 4.2.a: Collaborative tools desktop design.	
Category 5: Infotainment	
UC 5.1: Accessible metaverses design.	
UC 5.2: Collaborative games design.	
Category 6: Health Care	
UC 6.1: Remote Patient solutions design.	
UC 6.2: Mobile application design.	
UC 6.3: Health coach application design.	

Table 1: VERITAS project Use Cases list.

6. EVALUATION DOMAIN

The evaluation domain, as depicted in Figure 1 is the one that follows the design domain in which we are in the current stage of the project. In the evaluation domain the final pilot application scenarios that will be extracted from the Use Cases will derive.

In the current early phase of the project, user feedback may be obtained in rather informal settings, as defined also from the UCD methodology for example a participatory design session where designers-developers can be included in discussion and envisionment of activity scenarios, as it has already happened in VERITAS 1st and 2nd Pan-European Workshop and User Forum.

In the final test and evaluation of the VERITAS tools, a carefully designed summative evaluation will be carried out, with the goal of assessing how well the system performs with respect to its usability specifications. In that stage the pilot application scenarios will provide a realistic task context that will be analyzed in subtask so as to provide expected or desired usability outcomes. In that time the prototype will be robust enough to measure subtask times more detailed. Usability specifications developed in this way have two important roles in evaluation. First, they provide concrete usability objectives that can be serve as a management tool in system development and then the team's usability engineers are able to insist that redesign and improvement continue until they are met ([9], [10]). Second, the specifications tie the results of empirical evaluation directly to the usability issues raised during design. The pilot applications scenarios will be

7. CONCLUSIONS

On the basis of the aforementioned methodological framework the Use Cases of VERITAS were extracted to various forms, either descriptive in problem and application scenarios or specific for the development process using the fully dressed template. All the aforementioned have been realized keeping the user of the system in the centre of the development having as a top priority his/her needs and expectations.

To this end, the problem scenarios have been extracted in the beginning of the project so as to identify the core description of the situation as is for the under development system. The problem scenarios are concrete yet flexible, since their target is to assist developers envisage their system to be and not entrap them in non achievable solutions.

The development phase that followed required a more detailed description of the problem scenarios that also includes the correlation between the system and the users. Since the finalization of the Use Cases descriptions, the developers had identified the basic functionalities of the system, so it was possible to format the final Use Cases of the project following the fully dressed template format. The iterative procedure also promoted the iterative communication among developers, users and stakeholders, helping to make design activities more accessible to many sources of expertise.

Finally, the Use Cases will be used for the extraction of the final pilot application scenarios, when the prototype is robust enough to identify specific tasks to be objectively and subjectively measured, so as to evaluate the actual systems from the developers' perspective, as well as the developed application from the beneficiaries' perspective.

References

[1] Report of the Inclusive Communications (INCOM) subgroup of the Communications Committee (COCOM) COCOM04-08, 2004.

[2] Mary Beth Rosson and John M. Carroll, "Scenario-Based Design", Department of Computer Science and Center for Human-Computer Interaction Virginia Tech, Blacksburg VA, 2002

[3] Ivan Jacobson, "Object Oriented Software Engineering: A Use Case Driven Approach", Book, 1992

[4] Alistair Cockburn, "Use cases, ten years later", Paper, Originally published in STQE magazine, Mar/Apr 2002.

[5] Bittner, K., Spence, L, (2006). Managing Iterative Software Development Projects. Addison-Wesley Professional, June 27.

[6] A. Fantechi, S. Gnesi, G. Lami and A. Maccari Applications of linguistic techniques for use case analysis. In Requirements Engineering, Springer London Volume 8, Number 3 / 2003.

[7] Brooks F, (1995) The Mythical Man-Month: Essays on Software Engineering, Addison-Wesley, Reading, Massachusetts (Anniversary Edition, originally 1975).

[8] D. M. Rubin, "Uses of Use Cases", Methodologies and Practices – White Paper, Softstar Research, Inc., 1998

[9] Carroll JM, and Rosson MB, (1985) "Usability specifications as a tool in iterative development," Advances in Human-Computer Interaction, HR Hartson, ed., Ablex, Norwood, New Jersey., (New Orleans, USA), 225-231. ACM Press: NY.

[10] Good, M., Spine, T.M., Whiteside, J., & George, P. 1986. User-derived impact analysis as a tool for usability engineering. In M. Mantei & P. Oberton (Eds.), Proceedings of Human Factors in Computing Systems: CHI '86 (pp. 241-246). New York: ACM.

Applications of Discrete-Event Simulation for Mining Process Plants in Chile.

Andrea GONZÁLEZ
Technological Development Leader, JRI Engineering,
Luis Uribe 2343, Ñuñoa, Santiago, Chile.

Juan RAYO
Technical Manager, JRI Engineering,
Luis Uribe 2343, Ñuñoa, Santiago, Chile

and

Juan CAMPOS
Project Engineer, JRI Engineering,
Luis Uribe 2343, Ñuñoa, Santiago, Chile.

ABSTRACT

The article describes the development and application of a new discrete time-event simulation modeling methodology used in the copper mining industry in Chile. The proposed methodology creates a simulation model that represents a mine-mill operation, and it is used to evaluate different improvement options to maximize the system performance. Two applications of this methodology to existing copper plants in Chile are presented in this work.

Keywords: Simulation model, mining operations, mine-mill process, copper industry, transport model.

1. INTRODUCTION

The copper mining in Chile is the largest in the world. The economic importance of this activity has originated important developments and research, resulting in a better knowledge of the operation of the metallurgical process units. However, when an industrial facility is complex, or has too many types of equipment, originated by a number of expansions, or the case when surge bins between stages are of limited capacity, then the interrelationship between the different process units is not well known, thus inefficiencies may occur, producing a remarkable lack of process capacity. To overcome this issue, simulation has gained increased attention as a technique for planning and analyzing new mining operations, or modifying and improving existing ones, around the world. The simulation of a system corresponds to the numerical evaluation of a model by using a computer. This process enables the observation and measurement of the performance of a system under existing or proposed configurations of interest, over long periods of time. Thus it makes possible to carry out experiments without altering the real-life system, measuring the outcome to help the decision-making process with valuable information. Furthermore, for a complex system simulation is often the only feasible method to investigate the problem and evaluate alternative operational conditions, hence becoming a tool for justifying investment costs. On the other hand, simulation models can also study the behavior of a system before it is implemented, improving the design stage of the project.

In [5] Sturgul and Li (1997), the authors emphasize that the advantages of simulation in the mining industry are not only to provide management with a detailed look into the future but also to allow the company to make critical decisions and understand a variety of issues about the current system. Several experiences for simulation models in mining operations in Canada ([7] Vagenas, 1999), Europe ([4] Panagiotou, 1999), South Africa ([6] Turner, 1999) and Australia ([1] Basu & Baafi, 1999) have been reported in the technical literature. Most of these applications have been made in the research of surface and underground mining, with special interest in truck/ship transport. However, there are no reported experiences in the use of this tool in the analysis and development of the mining copper industry in Chile or South America, in spite of it increasing importance.

Lynch and Morrison in [3] mention that economic factors have become more important in recent simulation modeling, since consideration of the cost of items such as equipment, power, labor, suppliers, and of the contract conditions for sales of products are commonly considered in simulation models to improve economic performance.

This increasing importance and economic advantages of simulation motivated this research work, which focuses on the development of a methodology based on the use of a *dynamic stochastic discrete event model*, namely models evolving with time. These models include random input components and for which the states variables change instantaneously in response to certain events taking place at separate points in time (discrete events models). This simulation model will allow a decision maker to define the dynamics of the transportation/crushing/comminution operation of a mining facility, to make better operational decisions, resulting in an increase of mineral processing flow, increasing also the operation revenues.

The proposed methodology is used to improve the mineral throughput in two existing Chilean mining facilities, a large concentration plant and the dry-area of a large hydrometallurgical plant. In this article, simulations of the

current operation conditions were developed, and several improved scenarios were run (with marginal investment requirements) in order to increase the mineral flow in both cases.

The simulation models were built and run by using ProModel^{TM}, discrete event software widely used in the manufacturing industry.

2. BASIS OF THE SIMULATION

The information required to build a simulation model based on discrete event software is listed as follows. An ideal case model is needed to ensure that the mass balance of the studied system is correct.

Data Gathering

In order to build a robust simulation model, a thorough understanding of the system is needed, and therefore detailed information must be collected. The required information to build this model are the flow diagram, a detailed description of the plant's operation, the equipment characteristics (capacity, length, velocity, residence time, etc.). Also the operational statistics play a crucial role to obtain a realistic model, with information related to the maintenance and failure duration of different events, and the daily/hourly flow measurements. Benchmarking information of several mining facilities shows that it is required operational information of at least a thousand shifts to obtain a reliable data to build a simulation model.

Modeling the Ideal Case

In order to create an affordable simulation model, it is required to simplify the system by creating discrete blocks representing sets of mining/transport/process units. The constraints imposed by the simulation software set the precision, units, and time of simulation. Once this criterion is fixed, the design data (equipment characteristics, ore feed rate, etc.) requires to be escalated in order to create a realistic simulation model. The fundamental components of the simulation are:

Entities: the elements which travel through the system (tons of ore).

Locations: the objects that build the transport system (storing places, processing units, conveyor belts, etc).

Arrivals: entry points of entities to the system.

Process: definition of how the entities interact with each location, and their routing rules.

First, all the process' blocks are defined in the graphic interface provided by the software, defining the locations of the model. Then, by using the design information of the system, residence time and maximum process capacity is given for every location.

The relationships between the locations of the model are defined by considering the information from the flow diagram and further plant operational information provided by the plant operators.

The entity arrivals are defined by the mining facility mineral throughput. The initial condition for the simulation is defined with a warm-up routine, which allows each location of the model to be completely loaded with entities before running the simulation.

For this ideal case, no maintenance and/or failure time of the locations is considered, in order to check the consistency of the amount of entities that enter and exit the system.

This step of the proposed methodology is completed when the simulation runs in a reasonable time frame and the entities that come into the system are equal to the entities that exit it (mass balance).

Statistical Analysis of the Operational Data

The second step of the proposed method considers the construction of a more realistic model, where locations present delay periods due to maintenance or failure events. These periods are created as random functions in the model, based on the operational information available.

The operational information must be analyzed and refined, dismissing "outliers" and defining stable periods to assure a 95% confidence interval for the data. Maintenance and failure events of each location are classified, defining only the events that are directly related to the defined location. The interferences caused by other locations in the system are dismissed.

After processing the operational information, frequency tables of the "Time Between Failures" (TBF) and "Time To Repair" (TTR) of each location are created. By using statistical software, empirical functions are fitted to define TBF and TTR functions for the model. For both cases studied in this paper, the results show that most TTR functions follow the Pearson distribution, whereas TBF functions fits better the Log-Logistic distribution, as it is shown in Figure 1.

3. CONSTRUCTION OF THE CURRENT SITUATION CASE

In order to create a realistic simulation model, it is required to use a stochastic methodology that represents accurately the current situation of the studied facility.

Monte Carlo Method

The operational statistics, previously analyzed and processed, is incorporated to the Ideal Case in the form of TBF's and TTR's functions, creating a Stochastic Model. The probabilistic nature of the simulation hence results in a variability range for down-time events (namely events that stop a location, make it unavailable for the transport of entities).

According to the Monte Carlo Theory (see [2] for example), the average of an estimated quantity obtained by a stochastic simulation should get closer to the real average of this quantity as the number of experiments increase. On the other hand, the Law of Large Numbers and the Central Limit Theorem state that the distance between the real and the simulated quantity (error) should decay to zero proportionally to $1/\sqrt{n}$ where n is the number of replications. This theoretical bound is not very realistic, and in practice the error decay rate is much faster if a "good" random numbers generator is provided, which turns out to be the case of ProModel^{TM} (by using the prime modulus multiplicative linear congruential random number generator [8]).

In order to ensure that the difference between the average simulated quantity and the real average quantity is small, a stability approach is considered in the proposed methodology. With this approach, the minimum number n of replications of the experiment to ensure a robust approximation can be determined. Basically, a fairly large number of replications are run to observe the evolution of the averaged quantities simulated of interest as n increases. After reaching a difference between the real and estimated quantity is smaller that 2% it is considered that the simulation is stable, and the minimum number of replications for the case studied is set.

Calibration of the Current Situation Case

To validate the results obtained by the simulation model, a comparison between simulated and real control variables is performed. If the difference between these two values is bigger than 2% then the model needs to be calibrated. Since the input information of the model usually involves truncation of some

quantities, there is a certain degree of freedom to adjust the model parameters. These parameters are chosen to achieve an error of the approximation below 2% in every control variable. When the desired range of precision is reached, then the model is considered calibrated.

Sensitivity Analysis

After the calibration process is finished, a sensitivity analysis is carried out to measure the relative importance of the down time effect of each location on the global system. To determine the relative importance of some location, simulations are run considering all the downtimes of the system turned off but the one associated to the location of interest. The system throughput for each run is compared to the ideal case throughput.

This methodology allows identifying bottle-neck situations in a specific location and detecting maintenance engineering improvement opportunities. The results of this analysis are summarized and displayed usually as a Tornado Chart.

4. CASE SCENARIO EVALUATION

Using the calibrated simulation model as starting point, improvement options can be simulated and evaluated, in order to maximize the plant throughput. Some potential scenarios to evaluate are listed as follows:

1) Improvement of availability and use of critical equipment, implemented through variations of the TBFs and/or TTFs.
2) Adding new accumulation and/or process units.
3) Variation of other model parameters such as mine production rate or maximum treatment capacity of certain equipment.

The study of improvement alternatives is the most interesting and valuable part of the whole process, since the throughput increase produced by proposed alternative is quantified, with no modification of the real system. Given the complexity of the plant, the outcome of these simulations is not easy to predict because of the non-linear stochastic nature of the system, making this methodology remarkably useful for management decisions. Furthermore the investment/profit of each scenario is estimated by using available benchmarking information. The most cost-effective improvement option is therefore determined. In order to evaluate the proposed methodology, two case studies are presented, where the main concern is to increase the current ore throughput of the system with the minimum investment.

5. CASE STUDY Nº1: AN EXISTING COPPER CONCENTRATION PLANT.

The first case study is applied to an existing large copper concentrator plant in Chile. The industrial complex receives ore from an underground mine and produces copper concentrate. Figure 2 shows a diagram of the principal process units considered for this case studied.

The material flow starts from the mine, where ore is crushed and transported by a series of conveyors belts and bins to the mill site. There are several kilometers of distance between the crushing and milling sites, therefore the importance of the conveyor system is emphasized. The processed mineral has two possible grinding paths. The preferential path considers a grinding system associated to a semi-autogenic (SAG) mill. If this line capacity is full, a secondary path is utilized (secondary/tertiary crushing circuit). The secondary crushing circuit consists of a secondary crusher, in which the fine ore is ground to a finer size. If it does not meet the size requirements, it is sent to a tertiary crushing system and is screened again. A single large ball mill helps to decrease the particle size of this circuit. The ore produced from both grinding paths feed the semi autogenic mill where water is included to form slurry, to later proceed to the flotation process. The floatation process separates minerals by taking advantage of differences in their hydrophobicity prior to refinery process.

The model was built following the design information for each unit process. The statistical operational information was provided by the plant's management, which accounts for about 500 labor shifts. The failure and maintenance data was analyzed and filtered. The duration and frequency of each equipment´s down time was analyzed according to the different types of failures. The statistical analysis was carried out consistently to the process blocks considered for this study, and condensed into TBF and TTR functions for each process unit. The daily flow through Bin A was chosen as a control variable, as well as the ore feed to both mills. Other quantities of interest were also computed, such as the average total days a month for which the system's total production is over a certain threshold fixed value, and the utilization factor of critical equipment. These last variables allow a better understanding of the system and serve as a comparison tool between alternatives, rather than validation or calibration purposes.

After the stability analysis and calibration process, the current situation model was run. Several improvement alternatives to the current situation model were evaluated:

Case 1: Using the current situation model, a slight increase in Bin A capacity was considered.
Case 2: Built up from Case 1, adding an improvement of programmed maintenance of critical equipment.
Case 3: Built up from Case 2, adding a major improvement of the availability of the largest conveyor in the system.
Case 4: Built up from Case 2, adding a slight improvement of availability of the largest conveyor in the system and new Bin parallel to Bin A.
Case 5: Built up from Case 4, adding a major improvement of capacity to the SAG Mill´s Bin.

The results of these simulations are summarized in Table 1.

	Flow through Bin A	Comparison against Current Situation Model
Operational Value	100%	-
Current Situation Model	101%	-
Case 1	104%	3%
Case 2	109%	8%
Case 3	112%	12%
Case 4	113%	13%
Case 5	116%	15%

Table 1. Case Study Nº1 Results.

After an economic evaluation, Case 4 was selected due to its large increment of ore throughput (13%) and its reduced cost estimation (in comparison with Case 5, which presents a larger flow increment).

6. CASE STUDY Nº2: AN EXISTING HYDROMETALLURGICAL PLANT´S DRY CRUSHING AREA.

The proposed methodology was also applied to study an existing hydrometallurgical plant´s dry crushing area, which is depicted in Figure 3. In this case the ore is fed directly into the feeders at the primary crusher, which are located at the entry points of the plant. The ore processed by the primary crusher is carried by conveyor to the secondary crushing process where it is previously screened. Particles fine enough to bypass the

secondary crushing process go directly onto the grinding process, while large pieces of ore are crushed by the secondary crusher. If the ore is still oversize, it is sent to one of two tertiary crushers and continues in this cycle until the particles are fine enough to pass through the screen and enter the agglomeration process. Once the ore is sufficiently fine, it is conveyed to the fine ore bins and it is stored. Later on it is send to the agglomeration tank, where water and sulfuric acid are added to be piled up in lixiviation heap. The produced solution goes later to a solvent extraction-electro winning system to obtain copper cathodes.

The information used in the operational statistical analysis accounts for one year of operation. The definition of the process' block arrangement is motivated by the format of the statistical information of maintenance and failure. The plant's throughput is usually measured in one of the conveyors, hence a control variable was created to measure the flow through this block, and it was also used as the calibration parameter.

The stability analysis and calibration process were followed by the sensitivity analysis, where the importance of the relative influence of each block of equipment down time effect on the system was measured. The associated tornado chart is shown in Figure 4.

After defining the current situation model, it was run and its results defined the actual ore throughput of the system. Several alternatives were proposed to increase the system's throughput:

Case 1: This case considers: increase of power/velocity of several conveyor belts, increase of ore feed to the system and construction of a second agglomeration tank.
Case 2: Built up from Case 1, adding a new conveyor identical and parallel to the existing one, for receiving the undersize of the secondary sieve.
Case 3: Built up from Case 2, adding the use of more explosives in the mine, the so-called "Mine-to-Leach".
Case 4: Built up from Case 3, adding a new sixth tertiary screen.
Case 5: Built up from Case 1, adding a large stock pile right after the primary crusher.
Case 6: Built up from Case 1, adding a new complete third line of secondary crushing.
Case 7: Built up from Case 1, adding a second fine ore silo parallel and similar to the existing one.
Case 8: Defined as a combination of Cases 4 and 5.
Case 9: This case summarizes all the proposed improvements options with the exception of the large stock pile.
Case 10: It contains all the modifications described in all the cases.

The results of the simulations of all the cases are summarized in Table 2.
An economical evaluation was performed, considering the increment of ore throughput of the plant and the investment cost associated to each evaluated case. Case 4 presented the better profit-cost ratio, and it was proposed as the best alternative to increase the plant throughput.

	System's Throughput	Comparison against Current Situation Model
Operational Value	100%	-
Current Situation Model	98%	-
Case 1	105%	7%
Case 2	106%	8%
Case 3	106%	8%
Case 4	113%	15%
Case 5	110%	12%
Case 6	108%	10%
Case 7	108%	10%
Case 8	118%	20%
Case 9	116%	18%
Case 10	119%	21%

Table 2. Case Study Nº2 Results.

7. CONCLUSIONS AND FINAL REMARKS.

The described methodology can be used not only by the Chilean mining companies but also for others. The mining processing plant optimization model can be used as a general production planning tool for any mine-mill operation. Optimization of the existing system can be evaluated, and investment/operational decisions can be supported. The model is flexible enough to have different arrival rates from mine production, and inventory capacity limits. It also can model different kinds of processing/transporting units.
The proposed improvement options evaluated for the two studied cases represents an increase up to 20% of the production throughput, which implies close to 20% of production increase (cathodes or copper concentrate), but normally more than 25% of revenues. To the best of the authors' knowledge, the application of this methodology is unprecedented in the copper mining industry in Chile and South America and represents a powerful tool for designers and decision makers in the mining industry.

REFERENCES

[1] Basu, A, J. & Baafi, E,Y. (1999). **Discrete Event Simulation of Mining Systems: Current Practice in Australia**. International Journal of Mining, Reclamation and Environment, 13(2), 79.
[2] James, F. (1980). **Monte Carlo Theory and Practice**. Reports on Progress in Physics, Volume 43, Number 9.
[3] Lynch A.J. & Morrison R.D. (1999). **Simulation in mineral processing: history, present status and possibilities**. The Journal of The South African Institute of Mining and Metallurgy, October-December, 283.
[4] Panagiotou, G. N. (1999). **Discrete Mine System Simulation in Europe**. International Journal of Mining, Reclamation and Environment, 13(2), 43.
[5] Sturgul, J. R., & Li, Z. (1997). **New Developments in Simulation Technology and Applications in the Minerals Industry**. International Journal of Mining, Reclamation and Environment. 11(4), 159.
[6] Turner, R. J. (1999). **Simulation in the Mining Industry of South Africa. International Journal of Mining, Reclamation and Environment**. 13 (2) , 475
[7] Vagenas, N. (1999). **Applications of Discrete-event Simulation in Canadian Mining Operations in the Nineties**. International Journal of Mining, Reclamation and Environment 13(2), 77.
[8] Law. A & Kelton W. **Simulation, Modeling and Analysis**. New York. Mc Graw Hill. Publ., 2000.

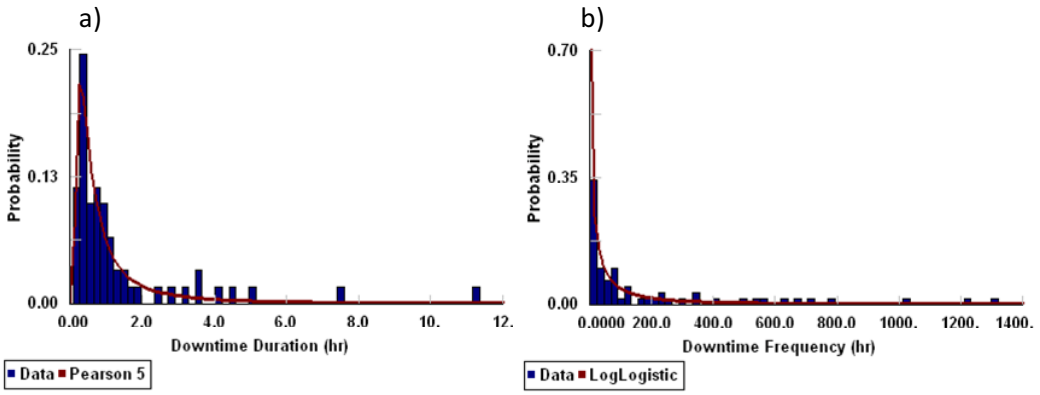


Figure 1: TTR or Down-time duration (a) and TBF or Frequency (b) Example.

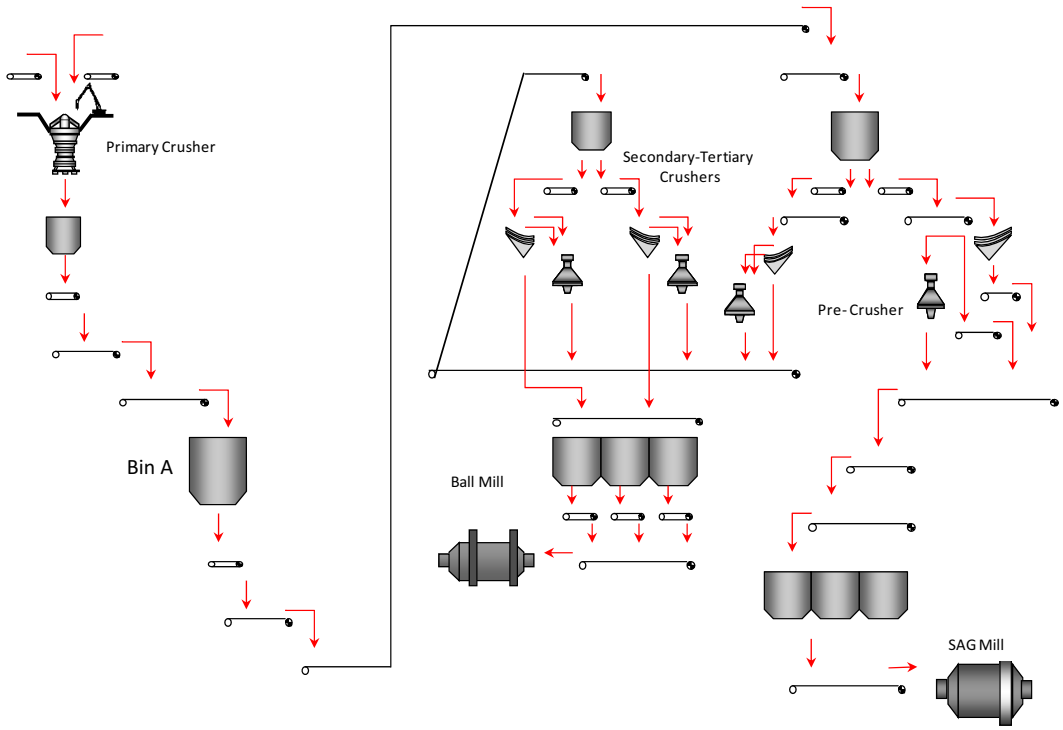


Figure 2: First Case Study. Mine-Mill Process. Copper Concentration Plant.

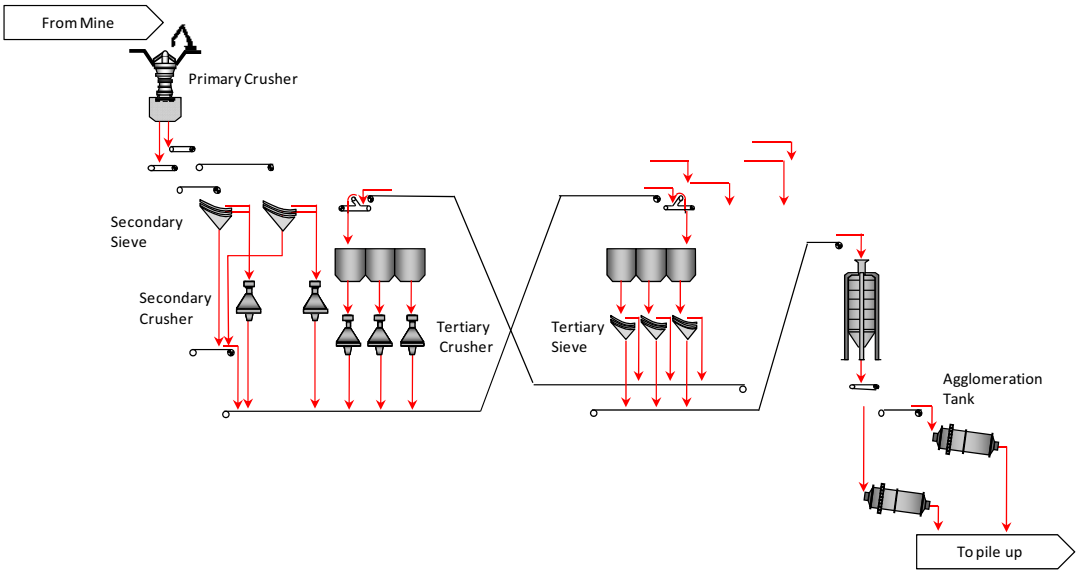


Figure 3: Second Case Study. Dry Grind Area. Hydrometallurgical Copper Plant.

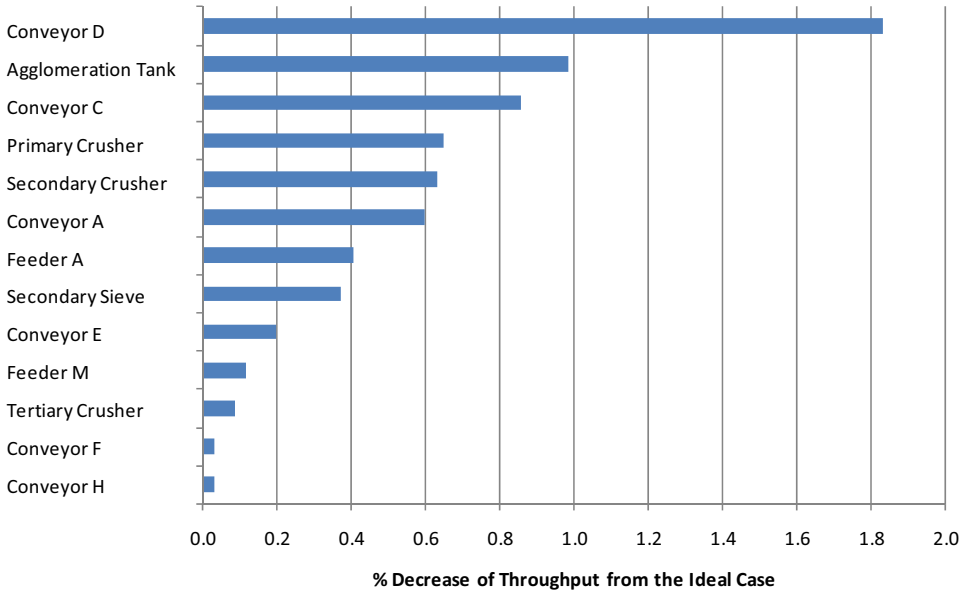


Figure 4: Sensitivity Analysis for Different Equipment Down-Time.

Fuzzy AHP approach for the Selection of Groundwater Recharge Alternative: Sensitivity Analysis

Masengo Ilunga,
Department of Civil Engineering and Chemical Engineering,
University of South Africa,
Pretoria, South Africa.

ABSTRACT

Fuzzy Analytical Hierarchical Process (FAHP) has been introduced in the recent decade to accommodate vagueness in the AHP. In a previous study by the above-mentioned author, a fuzzy set theory through AHP was used to select the most suitable groundwater artificial recharge (AR) alternative or technique for Kharkams Village, in Namaqualand of South Africa. ASR and ASTR were the two techniques used considering a normal judgment (i.e. decision maker or water resource manager's preference of 0.5). ASR and ASTR are Aquifer Storage and Recovery and Aquifer Storage Transfer and Recovery respectively. Five criteria were used in the selection, among those the legal and regulatory issues were considered as the fifth criterion associated with vagueness or uncertainty. The current study explores the impact of the different values for the water manager's preference-ranging between a pessimistic situation to an optimistic situation; on the consistency of the judgment matrix. The results showed that the weights of the five criteria remained almost unchanged as compared with the previous study. Throughout FAHP process the judgment was shown to be consistent. From a pessimistic situation to an optimistic situation, the overall preferences (recharge suitability) for ASR and ASTR remained closer to 55 % and 45 % respectively for FAHP likewise in the previous study. This supported that the decision maker or water resource manager's preference does not influence the legal and regulatory framework for selecting artificial groundwater recharge techniques.

Keywords: Fuzzy analytic hierarchical process, artificial recharge, sensitivity analysis

1. INTRODUCTION

For a semi-arid country like South Africa, artificial groundwater recharge alternative can be beneficial for sustainable development of groundwater. The two techniques or alternatives ASR and ASTR were evaluated in the selection of groundwater recharge for Kharkams in Namaqualand of South Africa [1], [2]. ASR and ASTR are Aquifer Storage and Recovery and Aquifer Storage Transfer and Recovery respectively. Analytic Hierarchy Process (AHP) and FAHP were used respectively by [1] and [2]. In the FAHP, the water resource manager's preferences were presented in terms interval judgment in a similar way done by [3]. A fuzzy set defuzzification technique to address vague data (i.e. fish activity data) was then proposed [3]. The application of AHP and FAHP specifically to groundwater recharge methods remains very sparse, except for example articles few studies [4] and [2]. In this study, alternatives and criteria for Kharkams Village are evaluated through a sensitivity analysis by considering cases where the decision maker or water resource manager's preferences range from pessimistic to optimistic using fuzzy triangular numbers (TFN). The fuzzy set defuzzification is carried out in a similar way done by [3].

2. FAHP FORMULATION

AHP method was developed by Prof. Thomas L. Saaty in the 1980s and is known as one of the Multi Criteria Decision Making methods (MCDM). AHP uses pairwise comparisons and derives ratio scales either from actual measurements or from subjective considerations (e.g. preference). Due to human judgment, AHP may allow some small inconsistency [5]. The technique has been applied to several fields including engineering and recently to the selection of artificial groundwater recharge techniques [1].

AHP technique can be traced generally in the following steps as replicated by [1]:

1. Formulate the problem as a hierarchy: A goal (objective), alternatives and criteria are contained in the hierarchy.
2. The pairwise comparison of the criteria based on a nine-point scale is done for the elements of the hierarchy. Table 1b shows the scale. Intensities are allocated based on the human judgments-experts' experiences or individual experience who have knowledge of relevant topics.
3. Establish the judgment matrix for the hierarchy: The information derived from the previous step is summarized in a comparison matrix.
4. Consistency test: The validity of hierarchy structure is tested by computing the consistency ratio from the judgment matrix. The consistency ratio CR is the constant index CI divided by random index RI. CR is less than 10 % for a consistent judgment. Mathematically, CR is expressed as follows:

$$CR = \frac{CI}{RI} \quad (1)$$

Where:

$$CI = \frac{\lambda_{MAX} - n}{n-1} \quad (2)$$

λ_{MAX} is the maximum Eigen value and n is the dimension of the judgment matrix

Normally, RI are fixed values that exist in the literature and are associated with the size of judgment matrix, for example, see Table 1a.

5. The selection of alternative(s) is based on the computation of normalized principal priority vector (Eigen vector) obtained from a comparison as built matrix.

In general, FAHP uses the same steps as listed above, however variables are expressed in terms of fuzzy data. Hence Table 2 is a similar table used by [3]. Two parameters λ and α were introduced and are the decision maker's preference and the risk tolerance respectively [3].

In this study, λ values vary from 0.1 to 0.9 to evaluate the impact on the goal of the hierarchy structure. When λ tend to 0.1, the decision maker or water resource manager is considered to be pessimistic otherwise, he is considered to be optimistic about his preference

Table 1a. Random index (R I)as function size of matrix (n)

n	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49

***Table 1b. Scale for pairewise comparison in AHP analysis* [1]**

Intensity of Importance	Definition
1	Equal importance
3	Moderate importance of one element over another
5	Strong importance of one element over another
7	Very strong importance
9	Extreme importance
2,4, 6, 8 are intensities to express intermediates values	

Table 2. Triangulation fuzzy numbers and linguistic variables

Intensity of Importance	Definition	TFN	Linguistic variables
1	Equal importance	(1,1,1)	Least importance
3	Moderate importance of one element over another	(2,3,4)	Moderate importance
5	Strong importance of one element over another	(4,5,6)	Essential importance
7	Very strong importance	(6,7,8)	Demonstrate importance
9	Extreme importance	(9,9,9)	Extreme importance
2,4, 6, 8 are intensities to express intermediates values		(1,2,3), (3,4,5), (5,6,7) and (7,8,9)	intermediates values

In general, FAHP uses the same steps as listed above, however variables are expressed in terms of fuzzy data. Hence Table 2 is a similar table used by [3]. Two parameters λ and α were introduced and are the decision maker's preference and the risk tolerance respectively [3].

3. IMPLEMENTATION OF FAHP FOR KHARKAMS

Kharkams is a small village in the semi-arid Namaqualand region of South Africa and depends solely on groundwater. Namaqualand is in the northwestern part of the Northern Cape. The lowest yielding of the village's three production boreholes is artificially recharged with ASR technique whenever surface runoff is available. For more details, the reader can be referred to [6].

The implementation of FAHP for Kharmans was done through the following hierarchy structure [2]:

*Goal: Defining the most suitable artificial groundwater recharge

*Criteria: need for an artificial recharge scheme, water source, aquifer permeability and water quality. In addition to these four top criteria, "legal and regulatory issues" are considered as the fifth criterion. The legal and regulatory issues received a strong importance as compared to other criteria since in South Africa all artificial recharge schemes need to be licensed and obtaining the necessary permits is thus crucial to the success of new projects. Fuzziness is only associated with the last criterion since it is ambiguous to quantify with precision this last criterion. Fuzzy numbers and linguistic variables are presented to address the inherent vagueness, imprecision or uncertainty associated with this last criterion.

*Alternatives: Aquifer Storage and Recovery (ASR) and Aquifer Storage Transfer and Recovery (ASTR).

In addition to the above, the following values for the decision marker's preference λ were considered (Table 2): 0.1; 0.2; 0.3; 0.4; 0.5; 0.6; 0.7; 0.8; 0.9

Each value of the decision maker's preference leads finally to the determination of the judgement matrix. Through numerical computation, the consistency ratio can be approximated using the principal Eigen value for the judgement matrix by summing the products between each element of Eigen vector and the sums of columns of the reciprocal matrix. Using MathLab, the maximum Eigen value can also be computed for each value of the water resource manager's preference.

4. PAIREWISE COMPARISONS

Table 3 shows the criteria pairwise comparison with corresponding preferences. The following criteria for the selection of artificial groundwater recharge techniques were considered: need for an artificial recharge scheme (A), water source (B), aquifer permeability (C), water quality (D) and legal and regulatory issues (E).

Based on the author's subjective consideration or preference, the legal and regulatory issues can be considered to have stronger importance than the rest of criteria [2]. For a subjective consideration to hold, the consistency ratio should be is less than 10%. Table 4 shows the weights of alternatives with respect to criteria.

Table 3. Pairwise comparison of criteria

X	Y	Importance	Intensity
A	B	Y	3
A	C	X	2
A	D	X	3
B	C	X	5
B	D	X	3
C	D	Y	2
E	A	X	4
E	B	X	4
E	C	X	4
E	D	X	4

X, Y: represent criteria in first column and second column of table 3 respectively, any particular raw.

The pairwise comparison between artificial recharge techniques, i.e. ASR and ASTR with respect to criteria was carried out in a similar way done by [2].

5. RESULTS AND DISCUSSION

Table 5 shows the weights of the five criteria for selecting the artificial groundwater recharge techniques for $\lambda = 0.1$ .Calculations of criteria weight and overall preference of alternatives (which are not shown here) were conducted in a similar way as done previously (see Appendices 1, 2 and 3 of article published in [2]). The results revealed that the goal weights on the five criteria for different values of λ were very close (see Table 6). In all cases the legal and regulatory issues have a strong preference (varying between 41 % and 46 %).

This could be justified by the fact that the legal and regulatory issues received highest priority since in South Africa all artificial recharge schemes need to be licensed and obtaining the necessary permits is thus crucial to the success of new projects. This criterion in particular will draw the attention of the water manager as a decision maker. Then he will prioritize other criteria.

Table 7 shows the calculations of values of the consistency ratio for different values of λ. The maximum Eigen values were estimated using MathLab. For a 5 x 5 matrix, (n = 5 and CI = 1.12). In all cases, CR values were less than 1 (or 10 %). These results support that the judgment was logical and consistent during FAHP process.

Table 4. weights of alternatives

	A	B	C	D	E
ASR	0.5	0.67	0.5	0.5	0.5
ASTR	0.5	0.33	0.5	0.5	0.5

Table 5. Criterion weights, $\lambda = 0.1$

	A	B	C	D	E	Average
A	0.112	0.056	0.142	0.259	0.125	0.14
B	0.337	0.168	0.355	0.259	0.125	0.25
C	0.056	0.034	0.071	0.043	0.125	0.07
D	0.037	0.056	0.142	0.086	0.125	0.09
E	0.458	0.687	0.290	0.352	0.125	0.45
sum						1

Table 6. Criteria weights matrix for different values of λ

λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
A	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.15
B	0.25	0.25	0.25	0.25	0.25	0.25	0.26	0.25	0.27
C	0.07	0.06	0.06	0.06	0.07	0.07	0.07	0.07	0.07
D	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.1
E	0.45	0.46	0.46	0.46	0.45	0.45	0.44	0.45	0.41

Table 7. Computed consistency ratio (CR) values for different values λ

λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
CR	0.089	0.085	0.085	0.0.088	0.0.089	0.088	0.063	0.088	0.0.85

Table 8. Overall preference of artificial recharge techniques, i.e. ASR and ASTR

λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
ASR	0.5425	0.5425	0.5425	0.5425	0.5425	0.5425	0.5442	0.5425	0.5459
ASTR	0.4575	0.4575	0.25	0.25	0.25	0.25	0.4558	0.4575	0.4541

Table 8 shows the overall preferences for ASR and ASTR, for different values of λ. The overall preferences were closer to 55 % and 45 % for ASR and ASRT respectively during FAHP. These results confirmed the suitability of ASR for the AR groundwater for the Kharmans Village. These results translate that the choice of the groundwater artificial recharge technique is less influenced by the fact that the decision maker (water manager) is more pessimistic or optimistic with regards to the legal and regulatory issues. In other words, these results are justified by the higher priority during FAHP. In any circumstances, the water resource manager as a decision maker has to comply with the South African legal framework before any artificial recharge schemes need to be licensed and obtaining the necessary permits. This is crucial to the success of new projects.

The overall preferences obtained could be explained by the fact that the legal and regulatory issues were ranked higher than the rest of criteria.

6. CONCLUSIONS

A sensitivity analysis on FAHP was carried out in the selection of artificial groundwater recharge alternatives/techniques i.e. ASR and ASTR to Kharkams Village of Namaland region in South Africa. It was shown that the decision maker or water resource manager's preference does not influence the legal and regulatory framework for artificial groundwater projects. In general, ASR outperformed ASTR. This confirmed conclusions from previous studies. The application of FAHP in this study is a relatively simple case (i.e. Kharmans Village). For problems of multi-hierarchy and multi-variable system, the power of AHP is very much acknowledged in decision-making process. Further work may include a sensitivity analysis of FAHP with more fuzzy variables for groundwater artificial recharge problems.

REFERENCES

[1] M. Ilunga **Application of Analytic Hierarchy Process (AHP) for the Selection of Groundwater Recharge Alternative.** Proceedings of the scientific conference of the NASAC-KNAW collaboration initiative on "W*ater management issues in Africa*", 28 – 31 march 2012, Boname Hall, Msiri, Reduit, Mauritius. Accepted.

[2] M. Ilunga Fuzzy AHP approach for the Selection of Groundwater Recharge Alternative, Proceedings of the *Fifth International Groundwater Conference (IGWC-2012)* on Assessment and management of groundwater resources in Hard Rock systems with special reference to Basltic Terrain at Aurangabad, Maharashtra, India, December 18-21, 2012, submitted.

[4] L. Dong, M. Fanxiang, F. Qiang, **Application of analytic hierarchical process in optimisation selection of groundwater artificial recharge methods in Sanjiang Plain**, Proceedings of the International Conference on Management and Service Science, MASS 2009 Wuhan, 20th September-22 September 2009. Policy and Management, Chinese Academy of Science, IEEE Wuhan Section.

[3] M. A. Ashari, **Using Fuzzy Analytic Hierarchy Process for Southern Johor River Ranking**. Int. J. Advance, Soft Comput. Appl., Vol. 1, No. 1, July 2009

[5] M. Ilunga and E.K. Onyari **Application of Hierarchical Process (AHP) for ANN model selection in streamflow prediction.** *The 15th World Multiconference on Systemics, Cybernetics and Informatics (WMSCI 2011)*, Orlando, USA, July 19 th-22 nd, 2011, Proceedings Volume III, pp. 215-219

[6] Department of Water Affairs (DWA), Strategy and Guideline Development for National Groundwater Planning Requirements. Water Banking: A practical guide to using Artificial Groundwater recharge, dated November 2010, pp. 24.

Machining Characteristics of Dry Grinding with cBN Structured Wheels

Taghi Tawakoli
Institute of Grinding and Precision Technology, KSF, Hochschule Furtwangen University
Jakob-Kienzle-Str.17, Villingen-Schwenningen, 78054, Germany

Dal Ho Lee (corresponding author)
Institute of Grinding and Precision Technology, KSF, Hochschule Furtwangen University
Jakob-Kienzle-Str.17, Villingen-Schwenningen, 78054, Germany

and

Amir Daneshi
Institute of Grinding and Precision Technology, KSF, Hochschule Furtwangen University
Jakob-Kienzle-Str.17, Villingen-Schwenningen, 78054, Germany

Abstract

Due to the demand for environmental friendly machining process, the dry grinding has lately become a subject of special interest. Despite its ecological and economical advantages, dry grinding process has not been widely introduced into industrial environment. The main reason is the problem that challenges the reduction of the heat generated in the grinding contact zone or transferred to the workpiece without using coolant-lubricants. In order to solve this problem, a new concept using the structured grinding wheel has been developed in the last few years. This paper describes the effects of process parameters on dry grinding using a structured grinding wheel. The results prove that dry grinding with the structured wheel leads to lower grinding forces, area-specific grinding energy, residual stresses and thermal damages to the workpiece as compared to the case of grinding with the normal wheel.

Keywords: grinding, dry grinding, dressing, wheel, structuring, cBN wheel

1. Introduction

Many attempts and efforts have been made to reduce the coolant consumption in grinding process. Some methods such as “dry grinding” and “minimum quantity lubricants”, (MQL) have been suggested and investigated by researchers to approach reducing the coolant-lubricants in grinding process. MQL is the application of a mist of the lubricant into a high pressure air stream. The air is generally not cooled and fluid flow rate is extremely low [1-3].

Despite the great interests in investigations aiming at developing dry grinding [4-6], it has not yet been employed in industrial environments to a great extent. This is due to its economic inefficiency caused by additional required equipment and high rate of the deteriorated parts owing to the thermal damages. In order to perform an efficient dry grinding operation which entails a product with no thermal damages, the chip formation needs to be optimized so that the friction between the grinding wheel and the workpiece is minimized. Optimal chip formation depends on the workpiece material and grinding parameters as well as the grain size. Besides, the optimal chip size largely depends on the number of active grains. It is generally stated that less number of active grains, higher chip thickness and lower specific grinding energy. In order to minimize the number of active grains, a new concept using structured grinding wheels has been developed in the last few years [7].

Grinding with special structuring is a method which has recently presented a very good potential to reduce significantly the use of coolant-lubricant in grinding operations through several experimental investigations [8-11]. However, a comprehensive and systematic study toward industrial application of dry grinding is lacking. Aiming for lessening this gap, this paper presents an investigation on characteristics of dry grinding with cBN structured wheels.

2. Creation of the structure

Figure 1 shows the schematic of the external cylindrical plunge grinding operation and flat surface grinding operation by a normal grinding wheel and a structured grinding wheel. The form of the structure, showed in this figure, is an example amongst various structures which may be created on the wheel surface through an individual conditioning.

The desired structures on the wheel surface are created by individual setting of the dressing kinematic parameters consisting of the shape of dresser, dressing depth of cut, a_{ed}, and the dressing feed, f_{ad}. As illustrated in figure 2, the wheel is flattened under ordinary dressing conditions (step 1).

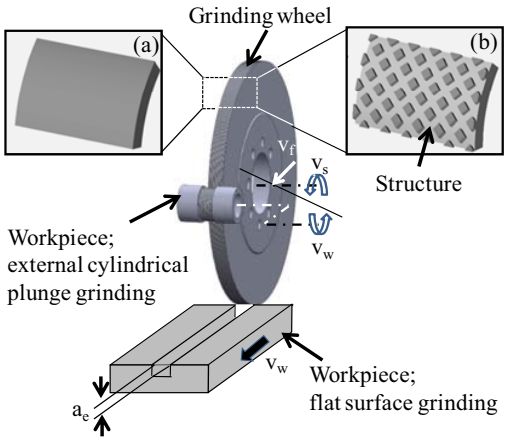


Fig. 1. Schematic views: (a) normal grinding wheel with full contact layer, (b) structured grinding wheel with reduced contact layer

On the flattened wheel surface, the desired structure is created through a special conditioning, in which the dressing feed, f_{ad}, is set bigger than the active width, b_d (step 2). The most important point in this step is that the operational parameters ($f_{ad\text{-}stru.}$ and $a_{ed\text{-}stru.}$) in special conditioning must be bigger than those in ordinary dressing (f_{ad} and a_{ed}).

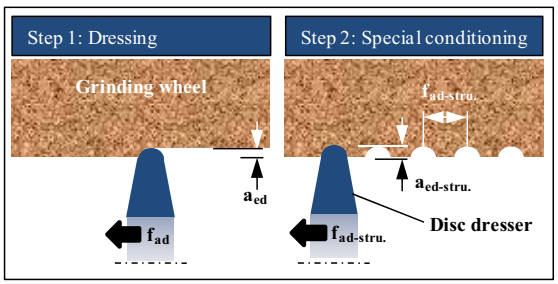


Fig. 2. Schematic overview of contact layer before and after structuring

3. Experimental setup

The experiments to evaluate the effects of employing the structured wheels on the efficiency of dry grinding were performed on a cylindrical grinding machine and surface grinding. In this investigation, the cBN grinding wheels (vitrified and resin bond) with an outer diameter of 400 mm and a width of 15 mm were used to grind a bearing steel (100cr6) with a hardness of about 60 HRC. The wheels on which surfaces structured with 25%, 50% and 75% contact layer were applied for comparison with those with full contact layer. A diamond dressing disc was used for dry grinding wheel dressing and special conditioning.

The workpiece for the cylindrical plunge grinding test was fixed on a rotational dynamometer (Kistler Instruments company) assembled on the workpeice spindle. The dynamometer is able to measure three orthogonal forces (F_x, F_y, F_z) and the torque around the Z axis. The grinding forces for the surface grinding test were measured by a piezoelectric dynamometer positioned under the workpiece clamping system.

The surface roughness of the workpiece was measured with the surface roughness measurement instrument from the Hommel-Etamic GmbH company, model "Wave system™ Hommel Tester T8000".

The residual stress measurement of the ground workpieces was carried out based on X-ray diffraction using the $\sin^2\Psi$-method on the {211}-lattice plane of the ferrite phase in the steel samples. Table 1 lists the test conditions during grinding operations in detail.

4. Results and discussion

4.1. Effects of the reduced contact layer

Figure 3 illustrates the effects of the reduced contact layers with 25%, 50%, 75% and full contact layer on the specific tangential grinding force at the specific material removal rate of Q'_w = 0.33 mm³/mm·s in flat surface dry grinding with the vitrified cBN wheel (B126 c125). It can be seen that reduction of the contact layer leads to reduction in specific tangential grinding force. The reduced contact layer leads directly to the increase of uncut chip thickness and proportion of the cutting regime. This figure also shows that the specific tangential grinding force in grinding with a structured wheel can be influenced by the percentage of the contact layer. The specific tangential grinding force, F'_t, is reduced from 1.05 N/mm in 75% contact layer to 0.63 N/mm in 25% contact layer. This is caused by the bigger number of active grains when the contact layer is larger.

Table 1. Test condition during the experiments

Cylindrical plunge grinding	
Grinding parameters	Q'_w = 0.5 – 12 mm³/mm·s q_s = -180 (up grinding) v_s = 20-80 m/s a_p = 10 mm z = 0.04 mm Coolant-lubricant : Grinding oil
Dressing parameters	a_{ed} = 3 µm b_d = 0.426 mm q_d = +0.8 (down dressing) U_d = 4
Flat surface grinding	
Grinding parameters	v_s = 50, 60 m/s v_w = 1, 1.5 m/min a_p = 15 mm a_e = 15-25 µm
Dressing parameters	a_{ed} = 5 µm q_d = +0.9 U_d = 2

Nomenclature

a_e	depth of cut
a_{ed}	depth of dressing cut
$a_{ed\text{-}stru.}$	depth of dressing cut for special conditioning
a_p	width of cut
b_d	active width of dressing tool
d_s, d_w	grinding wheel and workpiece diameter
E''_c	area related grinding energy
f_{ad}	axial dressing feed (per wheel revolution)
$f_{ad\text{-}stru.}$	axial dressing feed for special conditioning
F'_n, F'_t	specific normal and tangential grinding force
q_d	dressing speed ratio
q_s	grinding speed ratio
Q'_w	specific material removal rate
R_a, R_z	average roughness and mean peak-to-valley height
t_a	spark-out time
U_d	dressing overlap ratio
v_s	grinding wheel circumferential speed
v_w	workpiece speed
z	stock removal
σ_{ll}	Residual stress (Parallel to grinding direction)
Ψ	Glancing angle

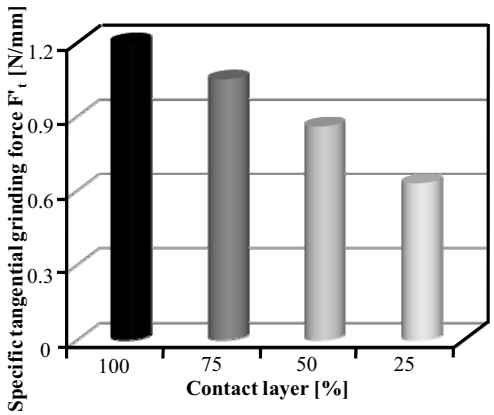


Fig. 3. The effect of the reduced contact layer on the specific tangential grinding forces (Q'_w = 0.33 mm³/mm·s, v_s = 60 m/s, v_w = 1m/min, flat surface dry grinding, workpiece 100cr6 hardened)

In figure 4, the influence of the reduced contact layers on the surface roughness parameters, R_a and R_z, at the specific material removal rate of Q'_w = 0.33 mm³/mm·s in flat surface dry grinding with vitrified cBN wheel (B126 c125) has been represented. It is clear that the finer surface roughness will be achieved using normal wheel with full contact layer comparing with the structured wheel with reduced contact layer. This can be due to the reduction of rubbing action, which improve the surface roughness when grinding with the structured wheels. In other words, the chip thickness becomes relatively higher when the contact layer is decreased and the portion of cutting action is increased in grinding operation. As a result, higher surface roughness values appear on the workpiece.

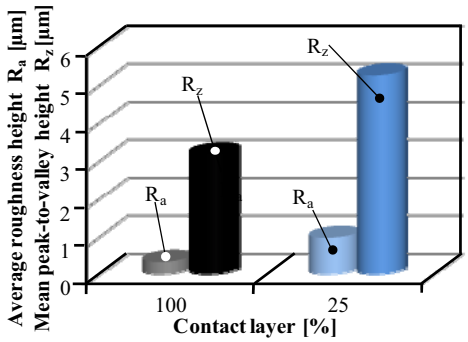


Fig. 4. Influence of the reduced contact layer on the surface roughness (Q'_w = 0.33 mm³/mm·s, v_s = 60m/s, v_w = 1 m/min, flat surface dry grinding, workpiece 100cr6 hardened)

4.2. Grinding forces and energy

In figure 5, the influence of the structuring of the vitrified cBN grinding wheel (B126 c125) on the grinding forces for various material removal rates in cylindrical plunge grinding has been represented. The workpiece diameter is 60mm in this experiment. The value of -180 has been set for the grinding speed ratio, q_s, which represents the ratio of the wheel speed to the workpiece speed, for all the cylindrical plunge dry grinding experiments in the present study. In dry grinding with 25% contact layer the forces are lower as compared with grinding with 100% contact layer in both dry and wet grinding cases. As already mentioned, the reduction in grinding forces was caused by the new structure in which the number of kinematic cutting edges is reduced. This provides an optimized chip formation process.

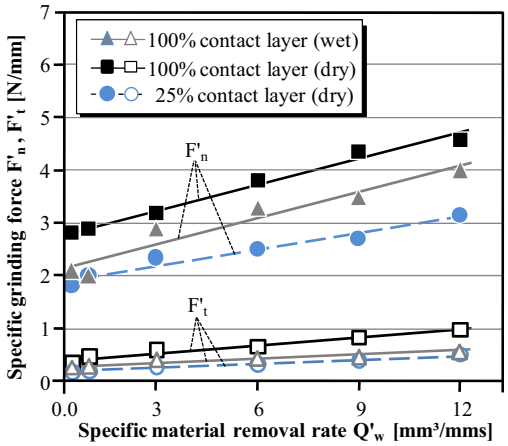


Fig. 5. Influence of the specific material removal rate, Q'_w, on the specific grinding forces (v_s = 60 m/s, q_s = -180, cylindrical plunge grinding with vitrified cBN wheel, workpiece 100cr6 hardened)

The influence of the structuring on heat generation in the grinding contact zone can be determined by the area-specific grinding energy E''_c. The area-specific grinding energy is an energetic process parameter which describes

the flow of energy into the component based on one unit area element and therefore can be used as a reference of thermal damages of the workpiece. This parameter is defined as follows:

$$E''_c = F'_t \cdot v_s / v_w \quad (1)$$

Where F'_t indicates the grinding specific tangential force, and v_s and v_w represent the wheel and workpiece speed, respectively.

As shown in figure 6, the area-specific grinding energies in cylindrical plunge dry grinding with 25% contact layer are almost in all the specific material removal rates much smaller than those in both cylindrical plunge dry and wet grinding with 100% contact layer.

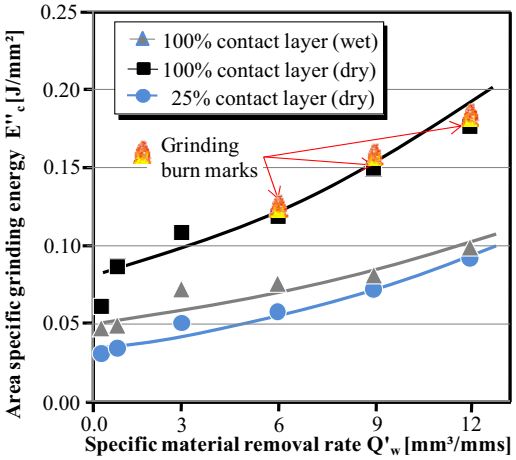


Fig. 6. Reduction of area-specific grinding energy caused by the wheel surface structuring (v_s = 60 m/s, q_s= -180, cylindrical plunge grinding with vitrified cBN wheel, workpiece 100cr6 hardened)

4.3. Surface roughness quality

Figure 7 illustrates the effect of reducing the contact layer on the average surface roughness height in cylindrical plunge grinding with vitrified cBN wheel (B126 c125). The workpiece diameter is 60 mm. The results indicate higher average surface roughness values for 25% contact layer compared with those for full contact layer. These are caused, in turn, by a smaller number of kinematic cutting edges and a larger resultant chip thickness when using the structured grinding wheel.

The workpeice roughness value, however, can be reduced by a spark-out process. A spark-out process is a special finishing process in which the feed rate is zero. The influence of the spark-out time on the surface roughness of the workpiece is represented in Figure 8. It can be seen that the spark-out leads to a reduction in workpiece roughness values to a certain level, regardless of the workpiece roughness values before spark-out.

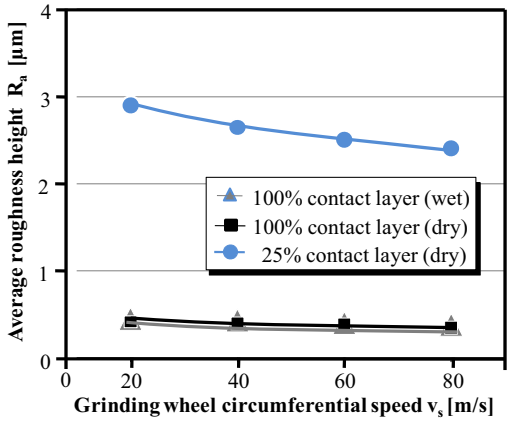


Fig. 7. Reduction of area-specific grinding energy caused by the wheel surface structuring (Q'_w = 3mm³/mm·s, q_s = -180, cylindrical plunge grinding with vitrified cBN wheel, workpiece 100cr6 hardened)

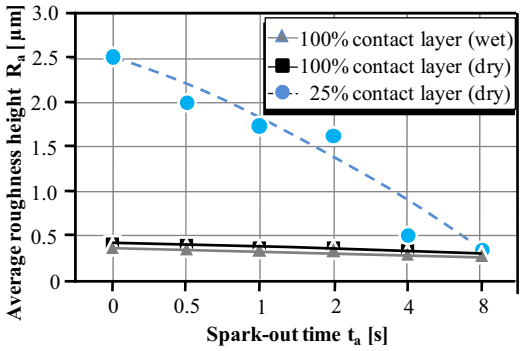


Fig. 8. Reduction of the surface roughness through the spark-out time (Q'_w = 3 mm³/mm·s, v_s = 60 m/s, q_s = -180, cylindrical plunge grinding with vitrified cBN wheel, workpiece 100cr6 hardened)

4.4. Structuring effect on the residual stress

In figure 9, the influence of the reduced contact layers on the residual stress in flat surface dry grinding with the resin bond cBN wheel (B126 c125) has been represented. The results indicate that the residual stresses are compressive in the case of dry grinding with the structured wheel with 25% contact layer, compared to conventional wheel where the residual stresses are tensile. This can be due to the lower grinding specific energy resulted from employing the structured wheels instead of the normal wheels. The lower heat generation, workpiece surface temperature, thermal stresses and eventually tensile residual stresses are the consequences of the lower grinding specific energy in dry grinding with the structured wheel.

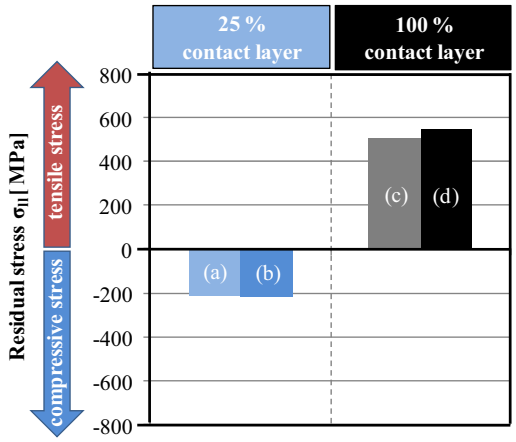


Fig. 9. Influence of the reduced contact layer on the residual stress in the flat surface dry grinding with the resin bond cBN wheel, workpiece 100cr6 hardened: (a) $v_s = 50$ m/s, $v_w = 1$ m/min, $a_e = 15$ µm, (b) $v_s = 50$ m/s, $v_w = 1.5$ m/min, $a_e = 15$µm, (c) $v_s = 60$ m/s, $v_w = 1$ m/min, $a_e = 15$ µm, (d) $v_s = 60$ m/s, $v_w = 1.5$ m/min, $a_e = 15$ µm

5. Summary

In order to determine the characteristic of the dry grinding with the structured cBN grinding wheels, a systematic investigation has been carried out in this study. The results can be summarized as follows:

- The grinding forces and the area-specific grinding energy decrease with the reduction in the contact layer.
- The structured wheel produces less thermal loading and tensile residual stresses.
- The roughness values of the workpieces ground with the structured wheel are higher than those ground with the normal wheel. The workpiece roughness, however, can be greatly reduced by the spark out.

References

[1] Brinksmeier, E. ; Brockhoff, T. ; Walter, A., 1997, Minimalmengenkühlschmierung und Trockenbearbeitung beim Schleifen, Härterei-Technische Mitteilungen, 52/3: 166-170.

[2] Tönshoff, H.K. ; Wobker, H.G. ; Brunner, G. ; Kroos, F., 1995, Möglichkeiten und Grenzen des Trockenschleifens gehärteter Stähle, Härterei-Technische Mitteilungen, 50.

[3] Weinert, K., 1998, Trockenbearbeitung und Minimalmengenkühlschmierung, VDI Buch.

[4] Brinksmeier, E. ; Brockhoff, T. ; Walter, A., 1997, Minimalmengenkühlschmierung und Trockenbearbeitung beim Schleifen, Härterei-Technische Mitteilungen, 52/3: 166-170.

[5] Tönshoff, H.K. ; Wobker, H.G. ; Brunner, G. ; Kroos, F., 1995, Möglichkeiten und Grenzen des Trockenschleifens gehärteter Stähle, Härterei-Technische Mitteilungen, 50.

[6] Weinert, K.; Buschka, M. ; Johlen, G. ; Willisch, C., 1998, Schleifen mit Graphit als Schmierstoff, VDI-Z 140 1/2:.46-49.

[7] Tawakoli, T. ; Rabiey, M., 2006, Trockenschleifen, Möglichkeiten und Grenzen, 6. Seminar „Moderne Schleiftechnologie und Feinstbearbeitung“, Stuttgart.

[8] Tawakoli, T.; Westkämper, E.; Rabiey, M, 2007, Dry grinding by special conditioning, International Journal of Advanced Manufacturing Technology, 33: 419-424.

[9] Tawakoli, T. ; Rabiey, M.,2008, Trockenschleifen mit CBN Schleifscheiben, der Effekt der Makrotopographie, 7. Seminar „Moderne Schleiftechnologie und Feinstbearbeitung“,Stuttgart.

[10] Tawakoli, T. ; Rabiey, M., 2008, An innovative concept and its effects on wheel surface topography in dry grinding by resin and vitrified bond CBN wheel, Machining Science and Technology, 12/4: 514 – 528.

[11] Tawakoli, T. ; Lee, D.H.; Rasifard, A.: Trockenbearbeitung beim Rundschleifen, Vortragsband zum 8. Seminar Moderne Schleiftechnologie und Feinstbearbeitung, Stuttgart, Vulkan-Verlag, 11.05.2010, pp. 4c1/13

RPD: Reusable Pseudo-id Distribution for a Secure and Privacy Preserving VANET

Sulaiman Ashraph
asulaiman@nur.ac.rw
National University of Rwanda
Butare, Rwanda

Dawoud S Dawoud
shenouda.dawoud@vu.ac.ug
Victoria University
Kampala, Uganda

and

Auxeeliya Jesudoss
ajesudoss@nur.ac.rw
National University of Rwanda
Butare, Rwanda

ABSTRACT

In any VANET, security and privacy are the two fundamental issues. Obtaining efficient security in vehicular communication is essential without compromising privacy preserving mechanisms. Designing a suitable protocol for VANET by having these two issues in mind is challenging because efficiency, unlinkablity and traceability are the three qualities having contradictions between them. In this paper, we introduce an efficient Reusable Pseudo-id Distribution (RPD) scheme. The proposed protocol is characterized by the Trusted Authority (TA) designating the Road Side Units (RSUs) to generate *n* reusable pseudo ids and distribute them to the On Board Units (OBUs) on request. RSUs issue the aggregated hashes of all its valid pseudo-ids along with a symmetric shared key and a particular pseudo-id to each vehicle that enters into its coverage range. Through this the signatures and certificates attached to the messages can be eliminated and thus resulting in a significantly reduced packet size. The same anonymous keys can then be re-distributed by the RSUs episodically to other vehicles. We analyze the proposed protocol extensively to demonstrate its merits and efficiency.

Keywords: VANET, privacy, security, unlinkability, traceability, pseudonyms.

1. INTRODUCTION

Vehicular communication (VC) systems are developed as a means to enhance road safety, traffic management and infotainment facilities for drivers and passengers. In vehicular ad hoc networks each vehicle is equipped with a communication device known as On Board Units (OBUs) that facilitate them to communicate with other vehicles, RSUs located on the road at different points and the TA (trusted authority) as well. In general, OBUs frequently broadcasts routine traffic related messages [1] with information about its position, current time, direction, speed, acceleration/deceleration, traffic events, etc. This helps the vehicle to be warned with critical situations such as accidents, traffic jams and so on, in addition with predicting the movements of the nearby vehicles.

Though this communication helps the driver community, it has a critical side effect of privacy. An attacker can easily track the physical location of a vehicle using these messages just by eavesdropping the communication. Tracking the movements of a vehicle such as "Big brother syndrome" is another case. One approach to solve this problem is that the vehicles broadcast their messages under pseudonyms that they change with some frequency [2]. The pseudonym based approach that has been proposed by [3, 4] is an idea to help the vehicles exchange their communications without revealing their real identity. Many studies have contributed for this approach. One of them is Baseline Pseudonyms (BP) approach that stores a huge number of pseudonyms in the OBU [5, 4]. Other is the Hybrid approach (HP) which is the combination of BP and Group Signature (GS) approaches [6] that generates pseudonyms on board and uses it for sending messages by attaching a group certificate.

In all the pseudonym based approaches that are previously discussed, pseudonyms generated are discarded soon after their life time. This cause the pseudonym providers to generate pseudonyms every now and then upon the request from the vehicles. Though generation of pseudonyms by the TA or RSUs is not an issue with their high computation and storage capacity, the computation cost of OBUs on signature and certificate generation and verification grows linearly with the traffic density, since every message comprises of a public key, a signature using its private key and a certificate on the public key essentially. In order to address this problem we propose a reusable anonymous key distribution scheme in which the pseudo-ids are generated in bulk by the RSUs and issued to the OBUs in its coverage zone by attaching a token with it. This token contains a hashed value of the given pseudo-id sealed with the long term public key of the vehicle that receives the pseudo-id. RSU also disclose the aggregated hashes of all its valid pseudo-ids generated by it to the vehicles in its range in order to facilitate them knowing the authentication of the pseudo-ids the messages are sent from. Session keys are generated by the RSUs to communicate with vehicles to share this information. Therefore, the proposed scheme avoids the attachment of signature and certificate with every message and by this way cuts down the cost of message generation and verification.

On the other hand, the token provided along with the pseudo-id will be attached by the vehicles while sending safety messages. This token is embedded with messages merely for future traceability.

2. OUR APPROACH

2.1 System Model

Prior to the network deployment, the TA generates a set of basic cryptographic materials for each OBUs and RSUs such as q, G, G_T, ê, P_1, P_2. TA randomly selects a master secret key $s \in Z_q$ and computes $U_1 = sP_1$ and $U_2 = sP_2$ as its public keys. TA also chooses a cryptographic hash functions H: $(0, 1)^* \rightarrow G$. Each RSU and vehicle are preloaded with the public parameters q, G, G_T, ê, P_1, P_2, U_2, H.

Notation	Description
s	TA's master secret key
U_1, U_2	TA's public keys
V_i	the i-th vehicle
R	the RSU
RID_v	real ID of the vehicle
RID_R	real ID of the RSU
PK_{vi}	long term public key of V_i
sk_{vi}	corresponding private key of PK_{vi}
T_{exp}	time expiry
WR_{PKvi}	warrant on the public key of V_i
$Cert_{TA}[PK_{vi}]$	TA's certificate on the public key of V_i
LOC_R	Location of RSU
LID_R	Location ID of RSU
SID_i	i^{th} short term multipliable pseudo-id
PID	pseudo-id
WR_{SIDi}	warrant on SID_i
PK_R	delegated public key of RSU
sk_R	delegated private key of RSU,
K_s	shared session key between V and RSU
E_x	encryption using the key x
D_x	decryption using the key x
T_{issue}	issue time
T_{return}	return time
τ	token
$ack_{termination}$	acknowledgement message for termination request
$h(.)$	a one way hash function such that SHA-1[24]

Table 1: Notations

The proposed protocol could be explained in five stages: registration and anonymous key generation, distribution of aggregated hashes, message generation, message validation and id traceability and revocation list. The key generation and mutual authentication between RSUs and vehicles of this protocol is based on [21]. For easy understanding the notations used throughout this paper are listed in table1.

2.2. Registration and Anonymous key generation

a) Key generation by TA:

All the vehicles and RSUs must register themselves with the TA before they join in the VANET. Each vehicle is assigned with a real identity $RID_{vi} \in G$. We assume that the TA is in-charge of checking the vehicle's identity, generating a long term public/private key pair for each vehicle and loading it into its OBU. TA chooses a random private key $sk_{vi} \in Z_q$ for the vehicle and computes $PK_{vi} = sk_{vi}P_1$ as its long term public key. TA also sets a warrant WR_{PKvi} for the issued public key to denote the expiry of the key. It stores RID_v, PK_{vi}, sk_{vi} and WR_{PKvi} in its database for future traceability and returns PK_{vi} , sk_{vi} ,WR_{PKvi}, $Cert_{TA}[PK_{vi}]$ to the vehicle.

The registration of RSUs with TA is very similar to that of vehicle registration. Firstly, the RSU sends its real-id RID_R and its location information LOC_R to the TA. TA selects $SID_i = LID_R \oplus H(a.U_1)$ for the RSU as its short term multipliable pseudo-id by choosing 'a' as a random nonce and sets a warrant WR_{SIDi} for SID_i by including its time expiry. It also computes a delegated key pair (PK_R, sk_R) where the public key $PK_R = H(RID_R)P_2 + rP_2 \in G$ and the private key $sk_R = -sH(PK_R) - r \in Z_q$ in order to prove itself as a trusted RSU to the vehicles. TA stores LID_R, SID_i, WR_{SIDi}, PK_R, sk_R in its database and returns SID, LID_R, WR_{SID}, PK_R, sk_R to the RSU. This SID can be periodically altered by the RSU (may be once in a day) by making a request to the TA.

b) Key generation by RSUs:

In this phase, the RSU is responsible for generating 'n' number of pseudo-ids PID_{i1}, PID_{i2}….. PID_{in} based on the short term pseudo-id SID_i of RSU acquired from the TA. The PID is encrypted here using ElGamal encryption algorithm [22] over the Elliptic curve Cryptography [23]. Each pseudo id $PID_{ik} = SID_i \oplus H(b.U_2)$, where k={1…n}, '$b$' is a random nonce and is changed each time to guarantee a distinct PID.

3.3. Distribution of token and aggregated hashes of pseudo-ids

The proposed RPD protocol comprises of four phases: pre-authentication phase, mutual authentication phase, key distribution phase and token return phase as illustrated in figure 3. The detailed explanation of the proposed protocol is as follows:

Pre-Authentication Phase:

In this phase, the RSU generates 'n' pseudo-ids and stores them in its pseudo-id table as shown in table2. It also computes the hash values of all the pseudo-ids generated by it and aggregates all the hashes, i.e., $h_{aggr} = h(PID_1) \| h(PID_2) ….. \| h(PID_n)$.

pseudo_id	T_{gen}	SID	Status
PID_{i1}	t_1	SID_i	1
PID_{i2}	t_1	SID_i	0
⋮	⋮	⋮	⋮
PID_{in}	t_1	SID_i	

Table 2: pseudo-id table

Authentication Phase:

At regular intervals, RSU broadcasts a *hello* message M, its real id RID_R and its delegated public key PK_R by signing them using its delegated private key sk_R. R computes its signature σ_R using (ω, Q) on the *hello* message M as follows.

$Q = nP_2$, $n \in Z_q$

$\omega = sk_R - nH(Q) \in Z_q$

$msg_1 = (RID_R, PK_R, M, \sigma_R(M), Q, \omega, T_s)$

When the vehicle V_i enters into the communication range of the RSU R, it detects the public key PK_R of R through this message. Note that V_i uses this message only at the first time to obtain the symmetric key with R, other vehicles that are already inside the RSU range ignores the message. V_i verifies the location information attached by default with the RSU message by matching the location information of R through GPS. If both are matching then,V_i checks the public key of R for its trustworthiness.

Once the RSU is authenticated by V_i , V_i generates a random number $r_1 \in Z_q$ and computes $r_1P_1 \in G$ as its share for the session key K_s , $H(msg_1)sk_{vi}$ as its signature σ_{vi} and forms a request for the session key and pseudo-id, signs them using its private key. V_i then submits its credentials that includes its long term public key obtained from the TA and its request to R after encrypting them using the public key PK_R of R.

$$msg_2 = E_{PK_R}(V_{id} \| N \| T_s)$$

where, $N = (r_1P_1, req_1, \sigma_{vi}(r_1P_1 \| req_1))$, $r_1 \in Z_q$ and V_{id} is the vehicle's credential (see phase II of RPD protocol) and req_1 is the pseudo-id request.

The authentication on the other hand is as follows: RSU R scans the revocation list each time a new vehicle tries to associate with it. Thus, on the reception of msg_2 from V_i, R decrypts msg_2 using its delegate private key sk_R and checks V_i's public key in the revocation list and the freshness of the timestamp attached with the message. If the public key is not revoked and its warrant is valid, R checks whether the signature σ_{vi} of V_i is legitimate.

Key Distribution Phase:
After authenticating V_i , R randomly picks a pseudo-id (whose status is 0) from the pseudo-id table, and chooses r_2 for the selection of session key K_s. R then computes a token $\tau = H(PID_i \| PK_{vi} \| T_{issue})$ to bind the long term public key of V_i with the pseudo-id PID_i temporarily. R stores the token, pseudo-id, V_i's public key along with the token issue time as shown in the first four columns of table3. Note that, the records of the token table are wiped out after a certain period of time (may be once in a week or two) in order to avoid the table growing linearly. Then R encrypts the pseudo-id, token and the aggregated hashes of all its pseudo-ids by using the shared session key and sends to V_i. Once V_i receives the message from R it calculates the session key K_s and decrypts the message using it. Vehicle V_i now holds the pseudo-id and uses it for sending messages to other vehicles.

Token	**pseudo id**	**V's public key**	T_{issue}	T_{return}
τ_1	PID_{i1}	PK_{vi}	t_1	t_1+t
τ_2	PID_{i2}	PK_{vj}	t_2	t_2+t
⋮	⋮	⋮	⋮	⋮
τ_n	PID_{in}	PK_{vn}	t_n	t_n+t

Table 3: token table

Token return Phase:

Since the long term public key of the vehicle is bound with the pseudo-id it is provided, vehicle V_i must return the token to the RSU after its usage. In this phase a vehicle may pass two types of request to the RSU. *req1* is a request for new pseudo-id, which must be sent to the RSU when V_i wants to change its pseudo-id. In such case the RSU extracts the vehicles's public key from the old pseudo-id and rebinds the public key with another pseudo-id to construct a new token and issues the new id along with its token to the vehicle by encrypting it using the shared session key. *req2* is another type of request the vehicle sends to RSU when it goes out of the range of the RSU or when it receives a *hello* message from another RSU. This request can be called as a handover request to make itself free from bonds with that RSU. In either case, the RSU will respond by giving a new pseudo-id or a handover acknowledgement message based on the type of the request it received.

Key Reusability is the main advantage of the proposed scheme when compared to other studies as the RSU's burden on continuous pseudo-id generation is considerably reduced because of reusing the same key for many vehicles. Upon receiving *req1,* the RSU uses the token to extract the public key of the vehicle, resets the status of the corresponding pseudo-id of token (i.e status= 0 for the corresponding pseudo-id in table 2). This pseudo-id can then be reused by being bound with another vehicle's public key upon request.

I. Pre-Authentication Phase:

R : computes PID_{ij} where j={1…n}
R : computes $h_{aggr} = h(PID_1) \| h(PID_2) \ldots \| h(PID_n)$

II. Mutual Authentication Phase:

R : computes $Q = nP_2$, $n \in Z_q$
R : computes $\omega = sk_R - nH(Q) \in Z_q$
R : computes σ_R as (Q, ω)
R : broadcasts $msg_1 = (M, PK_R, \sigma_R(M \| PK_R), Q, \omega, T_s)$
V_i : checks σ_R to authenticate R

V_i : computes $V_{id} = (PK_{vi} \| WR_{PKvi} \| Cert_{TA}[PK_{vi}])$
V_i : computes σ_{vi} as $H(msg_1)sk_{vi}$
V_i : computes $N = (r_1P_1, req_1, \sigma_{vi}(r_1P_1 \| req_1))$, $r_1 \in Z_q$
V_i : computes $msg_2 = E_{PK_R}(V_{id} \| N \| T_s)$

$V_i \rightarrow$ R : msg_2
R : $D_{skR}(msg_2)$
R : verifies PK_{vi} and authenticates V_i

III. Key Distribution Phase:

R : computes session key $K_S = r_1r_2P_1$
R : picks PID_{ik} from pseudo-id table
(where, $1 \le k \le n$ and status(PID_{ik})=0)
R : sets status(PID_{ik})=1
R : computes $O = (PID_{ik}, T_{issue}, r_2)$
R : sets $\tau = h(PID_{ik} \| PK_{vi} \| T_{issue})$ where $k \in j$
R : computes $msg_3 = E_{Ks}(O, \tau, h_{aggr})$
R$\rightarrow V_i$: msg_3 , r_2
V_i : computes $K_s = r_1r_2P_1$
V_i : computes $D_{Ks}(msg_3)$
V_i : holds PID_{ik} and h_{aggr}

IV. Token return Phase:

V_i : computes $msg_4 = E_{Ks}(req_i, \tau)$
$V_i \rightarrow$ R : (PID_{ik}, msg_4)
R : if $req_i = req_1$ then
map τ in table 3 and set status(PID_{ik})=in table 2 for the corresponding τ
sets a new $\tau = h(PID_{il} \| PK_{vi} \| T_{issue})$ for V_i , where $l \in j$
$msg_5 = E_{Ks}(PID_{il}, \tau)$
else if $req_i = req_2$ then
map τ in table 3 and set status(PID_{ik})=0 in table 2 for the corresponding τ
$msg_5 = ack_{termination}$
R$\rightarrow V_i$: msg_5

(note: req_1 = new pseudo-id request ; req_2 = termination request)

Figure 1: Reusable Pseudo-id Distribution (RPD) Protocol

3. RELATED WORK

The IEEE 802.11p task group is working on the Dedicated Short Range Communications (DSRC) standards, which aims to enhance the 802.11 protocol to support wireless data communications for vehicles and road side infrastructure [7]. Many studies have been reported on the security and privacy-preservation issues for VANETs [3, 4, 8-12]. The privacy and security issues for VANET can mainly be classified into three categories.

First is based on a huge number of pseudo-anonymous key based (HAB) protocols [3, 4, 8]. Though this is a simple and straight forward solution, there found three main disadvantages [12] in HAB: (a) each OBU has to take large storage space to store a number of anonymous key pairs; (b) very time consuming for the authority to track for any problematic certificate due to the long revocation list; (c) once some OBUs' anonymous keys are revoked, it takes a long time for each OBU to update the certificate revocation list.

The second one is based on group signature (GSB) which was first introduced in [14] which allows a group member to sign messages anonymously on behalf of the group. The identity of a signer can still be revealed by the group manager in case of a dispute. Although the group signature can achieve anonymity on conditional privacy preservation, the time for message verification grows linearly with the number of revoked vehicles [15]. Worse, the unrevoked have to update their private keys and group public keys with the group manager when the number of revoked vehicles surpasses some predefined threshold. In [16], application of the short group signatures is suggested. Authors in [8] propose an efficient security protocol called GSIS which is based on the group signature scheme. With this protocol only a private key and group public key are stored in the vehicle, and the messages are signed according to the group signature scheme without revealing any identity information to the public. However the verification of each group signature requires at least two pairing operations which might not be scalable when the density of traffic is increased. Finally, a hybrid pseudonym based approach [5] has been proposed by combining the baseline pseudonym scheme [3] and the group signature scheme [8] together. However this approach is also categorized as GSB, since it suffers with the same drawbacks.

The third one employs the RSUs to assist with message authentication [12, 13]. Authors in [17] propose an authentication algorithm called Group-ID Tree. In this protocol the vehicle is able to connect the RSU after proving its membership in a group. However, this leads to additional overhead in managing group membership. The protocol proposed by the authors of [18] elects a group leader who then communicates with the RSU on behalf of the group. This protocol also suffers with the disadvantage of the overhead associated with the Revocation List (RL) management required to authenticate group membership. ECPP (Efficient Conditional Privacy Preservation) [12] protocol was proposed to solve the storage requirements by using the RSU to manage the vehicle's certificate. In this protocol the RSU issues only an ephemeral certificate for valid vehicles at the time of authentication to eliminate the need for the vehicles to manage the certificates and the RL. In [13] the authors introduced a RSU aided message authentication scheme called RAISE. RAISE is responsible for verifying the authenticity of the messages sent from the vehicles and for notifying the results back to the vehicles. They also adopted *k-anonymity* [19] to protect user identity privacy where the RSUs assign a common pseudo id to *k*-vehicles. Our work complements the RAISE and ECPP works by providing another protocol to furnish a conditional privacy preserving and a secure VANET environment.

4. EVALUATION

In this section, we use the ns-2 simulator 2.34 to evaluate the performance of our RPD protocol. Since the proposed protocol focuses on the signing and verification overhead, we are more concerned in the system performance of RPD in terms of throughput, message loss ratio and average end to end 1message delay. We simulate a traffic scenario with high vehicle density of 30-180 vehicles. The ECDSA and the group signature verification delays are 3.87ms and 11 ms respectively [25]. The simulation script is written in TCL using DSDV protocol. The traces are recorded and analyzed using *awk* utility in Linux (Fedora 14).

Simulation Setup	
Physical and MAC model	IEEE 802.11a standard
Nominal bit rate	2Mbps
Transmission Range	300m
Number of nodes	30-180nodes
Simulation duration	1000 seconds
Simulation area	1500m x 300m
Traffic Type	CBR
Routing Protocol	DSDV
Packet Size for OBU message	166

Table 4: NS-2 Simulation Parameters

a) Throughput

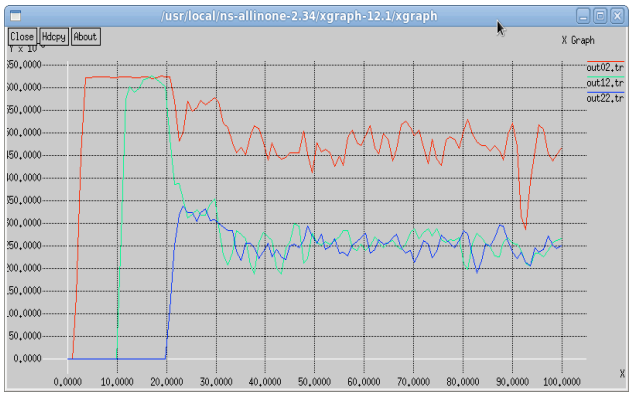
Throughput is the average rate of successful message delivery over a communication channel. Throughput is usually measured in bits/sec or data packets/sec. The throughput of a protocol varies based on the cryptographic operations involved in securing the message and the transmission overhead. The additional overload caused by security is mainly by the length of the authority certificate on the public key and the digital signature attached to every signed message. In ECDSA [20] which is accepted as the most appropriate candidate for VANET in terms of packet overload and verification delay, the total length of a signed packet is around 281 bytes, in which the additional overhead for each message is 181 bytes due to the cryptographic operations. With the group signature based scheme GSIS [8] the additional communication overhead is 184 bytes [26]. According to IBV scheme [9] a short length pseudo-id $PID_{ij} = PID_{ij}^1 + PID_{ij}^2$ posses a total length of 42 bytes. With RPD the total message length is calculated as follows:

$$L_{msg} = L_M + L_{PID} + L_{\tau} + L_{TS}$$
$$= 100 + 42 + 20 + 4$$
$$= 166 \text{ bytes}$$

According to [1], the message M occupies 100 bytes. Therefore, the additional overhead is only 42+20 bytes, which is very low for the OBU. In addition, RPD does not require the revocation list stored in OBUs, which makes the protocol free from increase in its storage overhead with the increase in number of revoked public keys. On the other hand, the additional transmission overhead on RSU is $(20*n)/m$ bytes along with the parameters for mutual authentication, where 20 B is the length of a $h(PID_{ij})$ sent by the RSU which is multiplied by n for n aggregated PIDs. This $20*n$ are shared by m messages, because in RPD the n pseudo-ids are hashed and sent as an aggregated hash (h_{aggr}) only once for an RSU range during the symmetric key establishment and thus it is considered as negligibly small.

Figure 2 shows the throughputs of Group based, PKI and RPD schemes over a period of 100 sec with a traffic density of 50 vehicles. We can see that when compared to the traditional PKI based ECDSA scheme and the group signature based scheme, RPD has very high throughput. This is because, the signature and the certificate attached dominates the length of the overhead and thus

reduces the throughput. The advantage gained by the proposed scheme is obvious, since no signature or certificate is attached with the message. (*x-axis: Time in seconds, y-axis: Data Packets in bits)*



out02.tr=RPD Scheme
out12.tr=PKI Based Scheme
out22.tr=Group Based Scheme

Figure 2: Protocol throughputs of Group Based, PKI and RPD schemes(100secs)

b) Message Loss ratio

One among the main performance metrics considered is the average message loss ratio which is to be denoted as MSG_{L_ratio}. A message is lost only if the queue of messages is full when the message verification rate is much lower than the message arrival rate. As defined in [25] the MSG_{L_ratio} can be expressed as,

$$MSG_{Lratio} = \frac{1}{N}\sum_{n=1}^{N}\frac{M_{cons}^{i}}{M_{recv}^{i}}$$

Where N_s represents the total number of vehicles in the simulation and N_c represents the number of vehicles in one hop communication range of the vehicle *i*. M_{recv}^{i} represents the total number of messages received by the vehicle *i* in the medium access control layer, M_{cons}^{i} represents the total number of messages received by the vehicle *i* in the application layer. Here, we only consider the message loss incurred by the security protocol rather than the loss caused by the wireless communication between the RSU and vehicles.

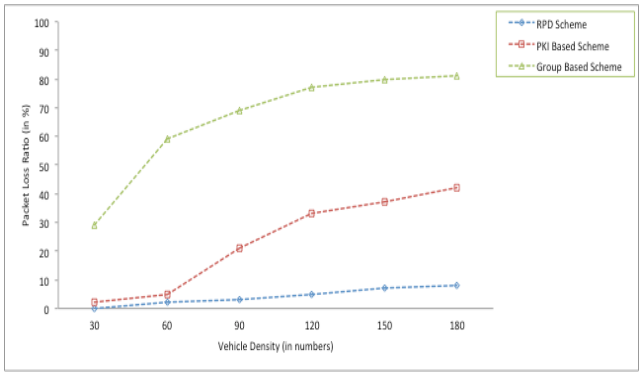


Figure 3: Message Loss Ratio vs Vehicle Density

Figure 3 shows the relationship between the PLratio and the number of vehicles, which is represented for the traffic load. We can observe that the message loss ratio of the three schemes increases as the traffic load increases. The group signature based scheme has the highest PLratio, the PKI based scheme grades the second place, whereas RPD has the lowest PLratio. This is because, the message verification rate is absolutely based on the h_{aggr} comparison computation cost is neglected when compared to the PKI based signature scheme [3].

c) Average end to end message delay

The average end to end message delay which we denote MSG_{delay} as can be defined as the difference between the time V_i sends the m^{th} message and the time V_j receives it. Considering N as the total number of vehicles in the simulation, M as the number of messages sent by the vehicle, and J as the number of adjacent vehicles within the communication range of vehicle V_i. If $T_{send}^{i\_j\_m}$ represents the time instant V_i in the application layer sends the m^{th} message to V_j and $T_{recv}^{i\_j\_m}$ represents the instant V_j in the application layer receives the m^{th} message then, according to [8], the average message delay is expressed as follows:

$$MSG_{delay} = \frac{1}{(N.M_n.J_n)}\sum_{j=0}^{J_n}\sum_{m=1}^{M_n}\times\left(T_{sign}^{i\_j\_m} + T_{transmission}^{i\_j\_m} + T_{verify}^{i\_j\_m}\right)\times\left(L_{i\_j\_m}+1\right)$$

Where $T_{sign}^{i\_j\_m}$, $T_{transmission}^{i\_j\_m}$ and $T_{verify}^{i\_j\_m}$ denotes, the time taken by the i^{th} vehicle to sign the m^{th} message, the time taken for the m^{th} message to get transmitted from i^{th} vehicle to j^{th} vehicle and the time taken by j^{th} vehicle to verify m^{th} message respectively. M_n is the number of messages sent by V_i and J_n is the number of vehicles within the one hop communication range of V_i. Since RPD does not require the message to be signed and the verification can be neglected as the h_{aggr} comparison computation is very fast, the message end to end delay is exclusively depends on the transmission delay which does not vary a lot with the increase of traffic load such like for a city scenario of 20 to 150 vehicles the message end to end delay is around 22ms[8] which is smaller than the maximum allowable message end to end transmission latency of 100ms[7]. $L_{i\_j\_m}$ (denotes the queue length in V_j when *m* from V_i is received) is neglected in RPD as well, for the above said reason. Therefore, the message delay for RPD can be reformulated as follows:

$$MSG_{delay} = \frac{1}{(N.M_n.J_n)}\sum_{j=0}^{J_n}\sum_{m=1}^{M_n}\times\left(T_{transmission}^{i\_j\_m}\right)$$

$$= \frac{1}{(N.M_n.J_n)}\sum_{j=0}^{J_n}\sum_{m=1}^{M_n}\times\left(T_{receive}^{i\_j\_m} - T_{send}^{i\_j\_m}\right)$$

Figure 4 shows the relationship between the MSG_{delay} and the traffic load. We can see that group signature scheme has the

highest MSG_{Lratio} due to the high verification delay whereas RPD yields the minimum MSG_{delay}. This demonstrates the effectiveness of the proposed protocol.

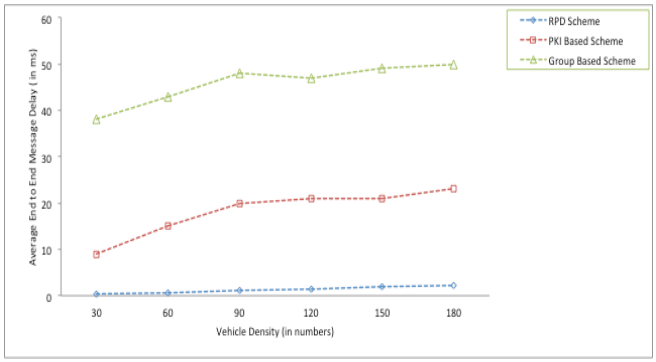


Figure 4: Average End-End Message Delay vs Vehicle Density

5. CONCLUSION

In this paper, a novel reusable pseudo-id distribution (RPD) scheme has been proposed. With RPD, RSUs are responsible to generate the anonymous ids in bulk and issue them one at a time to the requesting vehicles. The token which binds the long term public key of the vehicle with the given pseudo-id facilitates traceability. Also this makes the vehicle accountable for messages from the pseudo-id and insists the token return to get a new pseudo-id. The RPD protocol has many advantages because of the cost cut down of signing and verifying messages. Extensive simulation has been conducted to demonstrate the quite low transmission delay, message loss ratio and the message end-to-end delay. For future research, we will contribute to reduce the signature verification cost for vehicle to vehicle communication when the fixed infrastructures such as RSUs are absent in the network.

5. REFERENCES

[1] U.S. Dept. Transp., Nat. Highway Traffic Safety Admin., *Vehicle Safety Communications Project,* 2006. Final Rep.

[2] Raya, M. Hubaux, J.P.: *In: Proc. of Third ACM Workshop on Security of Ad Hoc and Sensor Networks (SASN 2005)*, Alexandria, November 2005.

[3] M. Raya and J,P, Hubaux, Securing Vehicular Ad Hoc Networks, *Journal of Computer Security, Special Issue on Security of Ad Hoc and Sensor Networks,* Vol. 15, pp. 36-68 , January 2007.

[4] X. Lin, X. Sun, X.Wang, C. Zhang, P.–H. Ho and X. Shen, TSVC: Timed Efficient and Secure Vehicular Communications with Privacy Preserving, *IEEE Transactions on Wireless Communications*, vol. 7, no, 12, pp. 4987-4998, December 2008.

[5] G.Calendrillo, P. Papadiitratos, J.-P. Hubaux, and A. Lioy. Efficient and Robust Pseudonymous Authentication in VANET. *In proceedings of the fourth ACM international workshop on Vehicular Ad hoc Networks (VENET)*, pp. 19-28, New York, NY, USA, 2007, ACM press.

[6] G.Calendrillo, P. Papadiitratos, J.-P. Hubaux, and A. Lioy. On the Performance of Secure Vehicular Communication Systems, *IEEE Transactions on Dependable and Secure Computing,* 2010.

[7] National highway traffic safety administration, U.S. Department of Transportation, *Vehicle Safety Communications Project – Final Rep.,* April 2006.

[8] X. Lin, X. Sun, P.–H. Ho and X. Shen, GSIS: A secure and privacy preserving protocol for vehicular communications, *In proc. IEEE INFOCOM*, Phoenix, AZ, 2008, pp: 1229-1237.

[9] C. Zhang, R. Lu, X. Lin, P.–H. Ho and X. Shen, An efficient identity based batch verification scheme for vehicular sensor networks, *In proc. IEEE INFOCOM*, Phoenix, AZ, 2008, pp: 246-250.

[10] K. Ren, W. Lou, R.H. Deng, and K. Kim, A novel privacy preserving authentication and access control scheme in pervasive computing environments, *IEEE Trans. Veh. Techonol.,* vol. 55, no. 4, pp. 1373-1384, July 2006.

[11] Y. Xi, K. Sha, W. Shi, L. Schewiebert, and T. Zhang, Enforcing privacy using symmetric random key-set in vehicular networks, *In proc. ISADS,* 2007, pp. 344-351.

[12] R. Lu, X. Lin, H. Zhu, P.–H. Ho and X. Shen, ECPP: Efficient conditional privacy preservation protocol for secure vehicular communications, *In proc. IEEE INFOCOM*, Phoenix, AZ, 2008, pp: 1229-1237.

[13] C. Zhang, X. Lin, R. Lu, and P.–H. Ho, RAISE: An efficient RSU aided message authentication scheme in vehicular communication networks. *IEEE International Conference on Communications (ICC '08),* Beijing, China, May 2008.

[14] D. Chaum and E. Ven Heyst, Group Signatures, *In Advances in Cryptology – EUROCRYPT 1991*, LNCS 547, pp. 257-265, Springer-Verlag, 1991.

[15] Hu. Xiong, Konstantin Beznosov, Zhiguang Qin, Matei Ripeanu, Efficient and Spontaneous privacy preserving protocol for secure vehicular communications, *In proc. IEEE ICC,* Cape Town, May 2010, pp. 1-6.

[16] D. Boneh and X. Boyen, and H. Shacham, Short Group Signatures, *In Advances in Cryptology – EUROCRYPT 2004,* LNCS 3152, pp. 41-55, Springer-Verlag, 2004.

[17] K. Sha, Y. Xi, W. Shi, L. Schewiebert, and T. Zhang, Adaptive privacy-preserving authentication in vehicular networks, *In proc. of the International Workshop on Vehicular Communication and Application,* pp. 1-8, October 2006.

[18] K. Sampigethaya, M. Li. L. Huang, and R. Poovendran, AMOEBA: robust location privacy scheme for VANET, *IEEE Journal on Selected Areas in Communications,* vol. 25, no. 8, pp. 1569-1589, 2007.

[19] L. Sweeney, K-ANONYMITY: a model for protecting privacy, *International Journal on uncertainty, Fuzziness and Knowledge Based Systems,* vol. 10, no. 5, pp. 557-570, 2002

[20] IEEE Trial-use Standard for Wireless Access in Vehicular Environments – Security Services for Applications and Management Messages. *IEEE Standard 1609.2-2006,* July 2006.

[21] Hyoung-Kee Choi, In-Hwan Kim and Jae-Chern Yoo, Secure and Efficient Protocol for Vehicular Ad Hoc Network with Privacy Preservation, *EURASIP Journal on Wireless Communications and Networking,* vol. 2011.

[22] T. Elgamal, A public-key cryptosystem and a signature scheme based on discrete logarithms. *IEEE Transaction on InformationTheory,* Vol. 31, Bo. 4, pp. 469-472, 1985.

[23] V. S. Miller, Use of elliptic curves in cryptography, *in proc. of Advance in Cryptology, pp. 417-426, Aug. 1985.*

[24] D. Eastlake and P.Jones, US Secure hash algorithm 1 (SHA 1), *IETFRFC* 3174, 2001.

[25] Chenxi Zhang, Xiagong Lin, RongXing Lu, Pin-Han Ho, Xuemin Shen, An efficient message authentication scheme for vehicular communications, *in IEEE Transaction on Vehicular Technology,* vol, 57, No. 6, Nov 2008.

[26] X. Sun, Anonymous secure and efficient vehicular communications, M.S. thesis, University of Waterloo, ON, Canada, 2007.

Electricity production during wastewater treatment in a mediator-less MFC inoculated with aged anaerobic sludge

Ali J. Jaeel
Department of Environmental Engineering, University of Baghdad,
Baghdad, Iraq

and

Zainab Ziad Ismail, PhD
Department of Environmental Engineering, University of Baghdad,
Baghdad, Iraq
Email address: zismail3@gatech.edu; zismail9@gmail.com

ABSTRACT

In this study, electricity generation integrated with organic substrate biodegradation was investigated in a continuous mediator-less microbial fuel cell with graphite electrodes and a selective type of membrane separating the anodic and cathodic compartments of MFC from each other. The performance of MFC was evaluated to treat actual domestic wastewater using aged anaerobic sludge. Results revealed that COD and BOD removal efficiencies were up to 88% and 75%, respectively. At an external resistance value of 100 Ω, a maximum power and current values of 318 mV and 3.11 mA, respectively were obtained.

Keywords: MFC, anaerobic sludge, electricity generation, and wastewater treatment.

1. INTRODUCTION

The high energy requirement of conventional sewage treatment plants are demanding for an alternative treatment technology which is cost effective and requires less energy for efficient performance. Among the new sources of clean energy, microbial fuel cells (MFCs) represent a novel technique which has gained lots of attractions in recent years. Microbial fuel cell (MFC) is a device that directly converts the metabolic power of microorganisms into electricity using electrochemical technology. It is a promising green method to treat organic effluents and produce electricity at the same time. In MFC, Microorganisms oxidize organic matter in the anode chamber producing electrons and protons. Electrons transfer via an external circuit to the cathode chamber where electrons, oxygen and protons combine to produce water [1, 2].

In the anaerobic chamber, the substrate is oxidized by bacteria and the electrons transferred to the anode either by an exogenous electron carrier, or mediator, such as potassium ferric cyanide, or neutral red [3, 4], or directly from the bacterial respiratory enzyme to the electrode. In the latter case, the MFC is known as a mediator-less MFC [5, 6].

MFCs have operational and functional advantages over the technologies currently used for generating energy from organic matter. First, the direct conversion of substrate energy to electricity enables high conversion effeciency. Second, MFCs operate efficiently at ambient, and even at low, temperatures distinguishing them from all current bio-energy processes. Third, an MFC does not require gas treatment because the off-gases of MFCs are enriched in carbon dioxide and normally have no useful energy content. Fourth, MFCs do not need energy input for aeration provided the cathode is passively aerated. Fifth, MFCs have potential for widespread application in locations lacking electrical infrastructure and also to expand the diversity of fuels we use to satisfy our energy requirements [7].

The main objective of this work was to evaluate the performance of an up flow microbial fuel cell (UMFC) inoculated with aged anaerobic sludge and fed with actual domestic wastewater. The performance of the UMFC was considered with respect to COD removal and power generation.

2. MATERIALS AND METHODS

The up flow MFC consisted of a dual rectangular chambers made of transparent acrylic parallelepiped having dimensions of 52 x 9.4 x 9.4 cm. The cathode chamber (26 cm height) was located on the top of the anode chamber (26 cm height) (Figs. 1 & 2). The anode and cathode chambers contained graphite plain electrodes; each had a surface area of 60 cm^2. The graphite electrodes were abraded by sand paper to enhance bacterial attachment. The two chambers were separated by a cation

exchange membrane (CEM) type CMI-7000, supplied by membrane international INC., NJ. The CEM sheet of dimensions 10X10 cm was placed between two perforated glass sheets containing 25 pores, each of 6.77 mm diameters (Fig. 2).

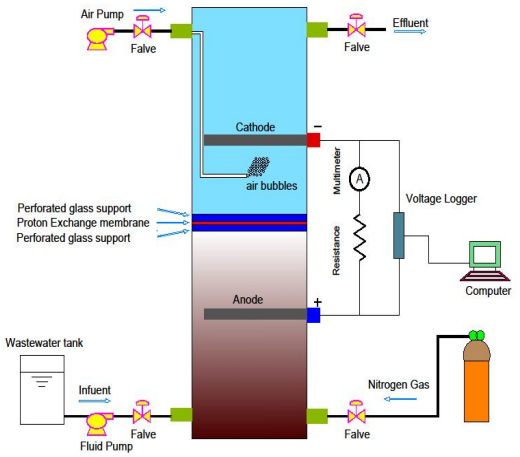


Fig. 1 Schematic diagram of the MFC.

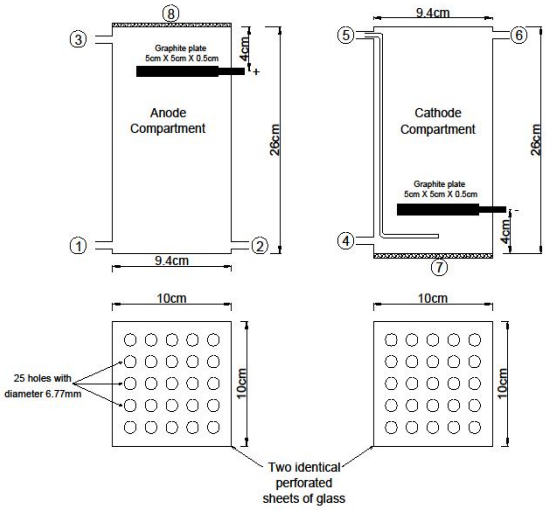


Fig. 2 Dimensions of anode and cathode chambers.

Operating conditions

The bioreactor were operated at room temperature, approximately 30 ^{o}C and continuously fed with actual domestic wastewater at a rate 0.1 mL/min until stable power output. The total hydraulic retention time (HRT) was 216 h. Wastewater fed to the bio-electro reactor had a pH ranging from 7.1 to 7.4 and average initial concentrations of COD and BOD of 380 and 200 mg/L, respectively. Table 1 presents the quality of the actual raw wastewater, freshly obtained from the main sewer pipe of Al-Kut city, Iraq. Granular aged anaerobic sludge which was collected from the bottom of a local septic tank was used to inoculate the UMFC.

Analysis

The concentrations of chemical oxygen demand (COD) and biological oxygen demand (BOD) were determined according to the procedures described in the *Standard Methods* [8]. Voltage was continuously measured by a multimeter with a data acquisition system and converted to power according to P=IV, where P = power, I = current, and V= voltage. The power was normalized by the surface area of the anodes. Columbic effeciency was calculated as the total coulombs measured divided by the moles of COD removed assuming 4 mol of electrons/mol of COD.

Table 1. Quality of the raw actual wastewater.

Constituent	Unit	Value
COD	mg/L	380
BOD	mg/L	200
TSS	mg/L	370
TDS	mg/L	1731
PO_4^{-3}	mg/L	12.8
Cl^-	mg/L	29.2
SO_4^{-2}	mg/L	206
NO_3^-	mg/L	9.3
pH	-	7.1-7.4

3. RESULTS AND DISCUSSION

After the anode chamber of the UMFC was inoculated with anaerobic mixed consortia, the fuel cell was operated with actual wastewater as feed to support the formation of biomass and subsequent adaptation to the new micro-environment. Constant substrate (COD) removal effeciency and voltage output were considered as indicators to assess the stable performance of the MFC. Microbial Fuel cell was operated continuously for 45 days. Results proved the feasibility of bio-electricity generation from wastewater treatment in MFC. However, the performance of MFC was mainly evaluated with respect to COD and BOD removal (Figs. 3 and 4).

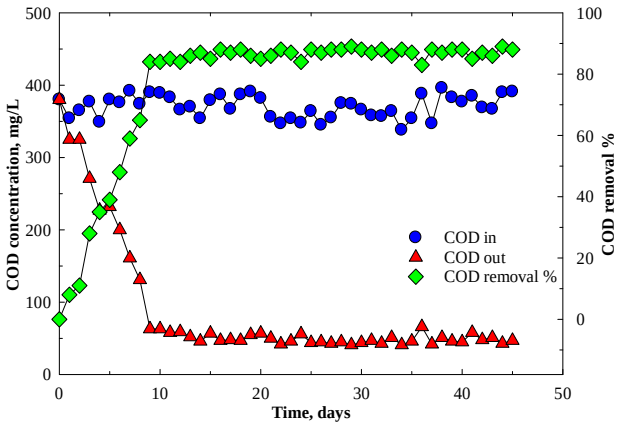


Fig. 3 COD removal profile.

The decrease in COD concentration indicated effective functioning of the enriched mixed culture in metabolizing the carbon source present in wastewater as electron donors. It is well observed that the anodic chamber acted as anaerobic suspended growth reactor normally used for wastewater treatment with respect to COD removal effeciency along with power generation.

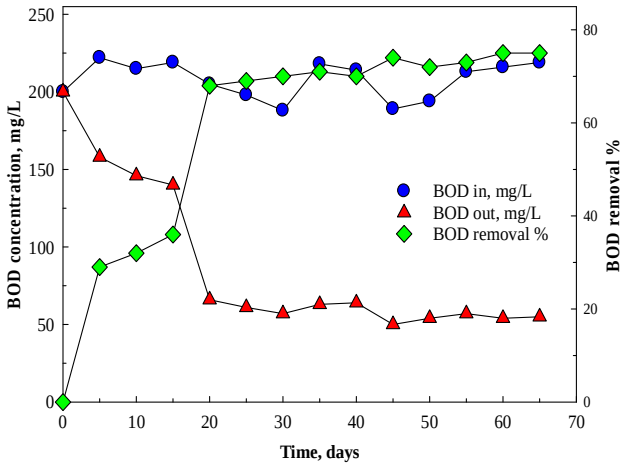


Fig. 4 BOD removal profile.

The performance of UMFC with respect to the current and voltage output during the operation are given in Fig. 5. During the initial step after inoculation, a gradual rise in the voltage up to 316 mV was noticed on the 10th day (Fig. 5). Maximum current output of 3.11 mA was observed after 10 days operation, measured at 100Ω external resistant. Kim et al. [9] reported the generation of a maximum current up to 3.66 mA at 50Ω in a dual anode-chambered MFC inoculated with a pure culture of *Shewanella oneidesis* MR-1and fed with lactate-based artificial wastewater. Huang et al. [10] investigated the electricity generation integrated with xylose degradation in a two-chamber mediator-less MFC inoculated with mixed culture from a primary clarifier using carbon paper electrodes. Maximum power density achieved by Huang et al. was 2.6 mW/ m^2. He et al. [11] reported the electricity generation while simultaneously treating an artificial sucrose-based wastewater in an up-flow microbial fuel cell (UMFC). Maximum power density was 170 mW/ m^2 at 84 Ω internal resistances. The dissimilarity between the results obtained in this study and the previously reported data could be attributed to the difference in several parameters including but not limited to the quality of wastewater with respect to the organic content concentration as well as the origin of the organics, wastewater being real or synthetic, type and source of inoculum electrodes material, and the resistance.

Polarization curves

Once MFC stabilized at the maximum steady voltage, the polarization curve was obtained by recording the voltage via varying the external resistance from 5 to 10000 Ω. Fig. 6 presents the polarization curve as a function of current density, potential and power density measured at variable resistances (5-10000 Ω).

Bio-film growth on the anode

Small piece of the bio-film thick layer was scratched from the anode after 45 days, operation showing black color indicating that the anode was covered with bacteria.

Selective enrichment procedure adopted in this study might have caused the growth of specific bacteria capable of producing bioelectricity, which had electrochemical activity. When MFC initially inoculated with the sludge, slight current was generating. Along with time, the current gradually increased indicating the electro-activity of bacteria. However, biofilms formation on the anode was the critical factor for electricity generation.

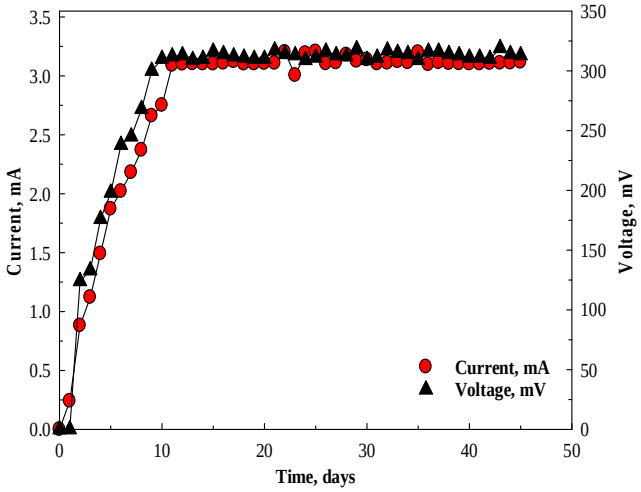


Fig. 5 Current generation and voltage profiles.

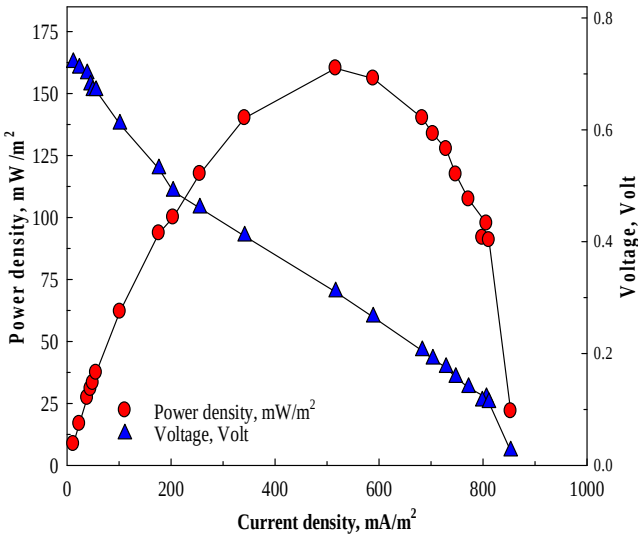


Fig. 6 Polarization curve during stable performance.

5. CONCLUSION

This study reported the evaluation of an up-flow microbial fuel cell for simultaneous bio-electricity generation from anaerobic wastewater treatment. The experimental work results revealed the feasibility of using MFC fabricated with low-cost non-coated plain graphite electrodes, without using any toxic mediators in the anode. Anaerobic conditions were maintained during the operation of the UMFC using anaerobic aged sludge for bio-electricity generation. COD removal effeciency in the UMFC proved the performance of microbial fuel cell as alternative wastewater treatment method in addition to the renewable energy generation. Provided the biological understanding increases, the electrochemical technology advances and the overall electrode prices decrease, this technology might qualify as a new technology for organics conversion to electricity in the future.

6. REFERENCES

[1] Y. Fan, H. Hu, H. Liu, "Enhanced Columbic effeciency and power density of air-cathode microbial fuel cells with an improved cell configuration", **Journal of Power Sources**, Vol. 171, 2007, pp. 348-354.

[2] Z. Li, X. Zhang, Y. Zeng, L. Lei, "Electricity production by an overflow-type wetted-wall microbial fuel cell", **Biosource Technology**, Vol. 100, 2009, pp. 2551-2555.

[3] K. Rabaey, N. Boon, V. Dendf, M. Verhaege, M. Höfte, W. Verstraete, "Bacteria produce and use redox mediators for electron transfer in microbial fuel cells", **Extended abstract, Proceedings of 228th ACS Meeting, American Chemical Society**, Philadelphia, PA, 2004a.

[4] K. Rabaey, N. Boon, S.D. Siciliano, M. Verhaege, W. Verstraete, "Biofuel cells select for microbial consortia that self-mediate electron transfer", **Applied Environmental Microbiology**, Vol. 70, 2004b, pp. 1-10.

[5] J.K. Jang, T.H. Pham, I.S. Chang, K.H. Kang, H. Moon, K.S. Cho, B.H. Kim, " Construction and operation of a novel mediator-and membrane-less microbial fuel cell", **Process of Biochemistry**, Vol. 39, 2004, pp. 1007-1012.

[6] B. Min, S. Cheng, B.E. Logan, "Electricity generation using membrane and salt bridge microbial fuel cells", **Water Research**, Vol. 39, 2005, pp. 1675-1686.

[7] K. Rabaey, W. Verstraete, "Microbial fuel cells: novel biotechnology for energy generation", **TRENDS in Biotechnology**, Vol. 23, 2005, pp. 291-298.

[8] APHA, AWA, WPCF. In **Standard Methods for the Examination of Water and Wastewater,** 20th ed**,** American Public Health Association, Washington, DC, 1998.

[9] M.H. Kim, I.J. Iwuchukwu, Y. Wang, D. Shin, J. Sanseverino, P. Frymier "An analysis of the performance of an anaerobic dual anode-chambered microbial fuel cell", **Journal of Power Sources**, Vol. 196, 2011, pp. 1909-1914.

[10] L. Huang, R.J. Zeng, I. Angelidaki, " Electricity production from xylose using a mediator-less microbial fuel cell", **Biosource Technology**, Vol. 99, 2008, 4178-4184.

[11] Z. He, S.D. Minteer, L.T. Angenent, "Electricity generation from artificial wastewater using an up flow microbial fuel cell" **Environmental Science and Technology**, Vol. 39, 2005, pp. 5262-5267.

GRAS: A Group Reliant Authentication Scheme for V2V communication in VANET

Auxeeliya Jesudoss
jauxeeliya@nur.ac.rw
National University of Rwanda
Butare, Rwanda

Sulaiman Ashraph
asulaiman@nur.ac.rw
National University of Rwanda
Butare, Rwanda

and

Ratnakar Kotnana
ratnakar@nur.ac.rw
National University of Rwanda
Butare, Rwanda

ABSTRACT

Unlike fixed or wired networks, mobile ad-hoc networks pose a number of challenges for peer-to-peer communication due to their dynamic nature. This paper presents a novel framework for vehicle-to-vehicle communication controlled and facilitated by a group leader within a group of vehicles. A communication model for a pure ad-hoc network is developed with much concern about the privacy and security of the system, for the ease of effective communication between vehicles with a reduced communication and computational overhead when no fixed infrastructure is present in the roadsides. In the proposed protocol, vehicles within a radio frequency form a group. They elect their leader based on some criteria who is then responsible for generating a group public and private key pair. Each vehicle is equipped with a tamper resistant OBU which is capable of generating public/private keys pairs and also self-certifies the generated keys based on one way hash chaining technique. Any vehicle joins the group communicates the group leader, authenticates itself to obtain the group key. Later, the vehicle uses the group key to send traffic related messages to the group leader who is responsible for batch verifying the authenticity of the message from different sources and one hop broadcast them to reduce the computation overhead on message verification in each vehicle. In addition, our scheme adopts the k-anonymity approach to protect user identity privacy, where an attacker cannot associate a message with the sending vehicle. Extensive analysis and simulations show that the proposed architecture provides an efficient and fully self organized system management for car-to-car communication without the need of any external infrastructure.

Keywords: VANET, V2V, group communication, hash chaining, security, privacy.

1. INTRODUCTION

VANET technology enables the communication between vehicle to vehicle (V2V) and vehicle to roadside infrastructure (V2I) units [5][6]. All vehicles use a communication device called onboard units (OBU) to communicate with the RSUs. In most cases of V2V, the system topology is divided into smaller partitions of disjoint groups of hosts, who can act with independent control [1].

In general, groups can be categorized into predefined groups and dynamic groups. In predefined groups, vehicles that belong to specific category are part of specific groups. This type of approach is not very scalable and too inflexible and thus not suitable for VANETs. In case of dynamic groups, the groups are formed dynamically, presumably based on how close they are or what their driving pattern is [2].

In this paper we propose a group based authentication scheme in which vehicles use self generated anonymous public keys and certificates to authenticate themselves to a group leader when they want to join in a group or to any vehicle in normal situation. The group leader verifying the trustworthiness of a vehicle, issues a group key to the joining vehicle, which will be used later for sending safety messages within the group. At the time of sending messages, every vehicle sends its message to the group leader who will aggregately verify all the signatures of all the messages and broadcast a single aggregated message to all its members at a particular time interval. Here the OBUs (On Board Units) of each vehicle generates the anonymous keys and they themselves certify those keys with the help of a check value computed by the TA (Trusted authority) based on the one-way hash chaining mechanism. The same check values are issued to multiple vehicles by the TA in order to avoid the vehicle being tracked by any adversary during the time of message sending in a network. But at the same time, each OBU automatically attaches a tracking hint to all the traffic related messages it generates in VANET. This tracking hint is again computed by the TA using its master secret key and therefore useful for later tracking in case of any dispute.

2.OUR APPROACH

2.1 Network Model

For the proposed scheme, we address only vehicle-to-vehicle communication (V2V) for safety related applications. In this, vehicles are arranged into non-overlapping groups which are dynamic in nature. More precisely, vehicles form groups with their closest neighbors in their mobility. In addition to that, each vehicle is equipped with a GPS and tamper proof hardware devices such as OBUs which are capable of generating short lived anonymous public/private key pairs for message authentication. Moreover, our scheme does not depend on any fixed infrastructures such as RSU

for key establishment or certification services, as the OBUs themselves are able certify their anonymous public keys which can be authenticated by the receiver (*in this case group leader is the receiver*) and could be revoked by the trusted authority as well based on the check values issued during system initialization.

2.2 Notations

In this paper, we use the notations that are described in Table 1 and also adopt the system parameters that are used by [20]:

G= (*G*, ∗). Finite cyclic group of order *q* (for some large *q*), $g \in G$ is a generator of G, and we assume that computing discrete logarithms in G with respect to *g* is computationally infeasible. For example, G might be a large multiplicative subgroup of Z_p^* for some large prime *p*, where *q* is a large prime dividing *p*− 1; alternatively, G could be the group of points on an elliptic curve (usually written additively).

h Cryptographic (one way) hash function mapping arbitrary length binary strings to strings of a fixed length *l* (where a typical value for *l* might be 224).

f Cryptographic (one way) hash function mapping the set {0, 1,... ,*q* − *1*} onto itself; in practice, *f* might, forexample, be derived from *h*.

m ≥ 1, Positive integer that determines the maximum number of key pairs that can be generated by a vehicle.

Notation	Description
sk_{TA}	TA's master secret key
PK_{TA}	TA's public keys
V	the vehicle
OBU_v	the OBU equipped by the vehicle V
$Cert_{TA}$	TA's certificate
Ћ	Tracking hint
cv	Check value
PK_x	Public key *x* at time T_x
sk_x	Corresponding private key *x* at time T_x
H_x	Helper value for the public key PK_x
jreq	Group joining request
Pu	Short lived public key of vehicle V
Pr	Corresponding private key
$Cert_{OBU}$	OBU certificate on Pu and the signed check value *cv*
$Encr_x$	Encryption on key x
Dcr_x	Decryption on key x

Table 1: Notations

2.3 Assumptions

Our proposal is based on some assumptions:

- Each vehicle contains a tamper proof hardware also known as on-board units (OBU) and a global positioning system (GPS).
- The OBU equipped in each vehicle is able to generate anonymous public and private keys and self certifies the generated public keys.
- We also assume that, the OBUs are configured in such a way that they automatically attach a tracking hint provided by the TA with every message it sends in the network.
- The revocation is controlled by the TA using RTC protocol by mapping the tracking hint to the vehicle ID issued during the registration phase.
- We also assume a straight road scenario for representing a highway. Vehicles in their mobility, forms dynamic groups and elect their leaders who can later batch verify and aggregate the messages sent by different vehicles.

2.4 Role of OBUs

In this paper, we espouse the one way hash chaining scheme proposed by [20]. During the vehicle registration TA generates a check value for each OBU and certifies them. This is done through the following steps.

1) TA randomly chooses $sk_{TA} \in Z_p^*$ as its master secret key and computes its public key $PK_{TA} = g^{sk_{TA}}$

2) During registration, Vehicle V equipped with OBU_V submits its unique vehicle-id V_{ID} to the TA. The TA chooses a secret key *s* for the OBU_V and generates a positive integer *P* such that

$$P = f^m(s) \leftarrow f^{m-1}(s) \ldots\ldots f^{m-i}(s) \leftarrow f^{m-i-1}(s) \ldots\ldots f^2(s) \leftarrow f(s) \leftarrow s$$
$$= \prod_{j=0}^{m} f^j(s)$$

3) TA also chooses two positive integers *a, b* and computes a large positive integer *Q* such that *Q*=(*P* **a*)+ *b* and stores *Q* as secret. TA then calculates the check value *cv* and a unique tracking hint Ћ for OBU_V as follows:

$cv = h(g^{R*Q})$, where *R* is a large positive integer.

$$Ћ = Encr_{PK_{TA}}(V_{ID} * g^{sk_{TA}})$$

4) Then the following parameters are transferred from TA to OBU_V

$$Cert_{TA}\{cv\}, Ћ, P, R, a, g^b, s, TS$$

where $Cert_{TA}\{cv\}$ is the CA's certificate for the check value *cv* and TS is the validity time of each public key that are going to be generated by the OBU.

5) In each time interval T_x ,OBU_V chooses a random integer r_x in such a way that, $r_x \neq r_{x+n}$ for $x \in \{0 \ldots m\text{-}1\}$ and $n \in \{1 \ldots m\text{-}x\text{-}1\}$, and $(R * r_x * b) \bmod P = 0$ and generates a private/public key pair sk_x, PK_x, and a helper value H_x for the generated public key as follows.

$$sk_x = R * r_x * a * \prod_{(j=0)}^{(m-x-1)} f^j(s) and$$

$$PK_x = g^{(R*r_x*a*\prod_{(j=0)}^{(m-x-1)} f^j(s))}$$

$$H_x = \prod_{(j=0)}^{(m-x-1)} * r_x^{(-1)}$$

6) When Vehicle V wants to join in a nearby group, OBU_V generates a public/private key pair *Pu/Pr* and the

corresponding certificate $Cert_{OBU}$ that contains the signed check value *cv*, where sk_x is used by the OBU to sign the certificate $Cert_{OBU}$ it issues for the generated public key *Pu.*

7) The vehicle V then sends the following parameters together with the group joining request *jreq* to the leader of the group to which it wants to join, intending to authenticate itself to the leader in order to obtain the group key.

jreq, Pu,$Cert_{OBU}$*,* PK_x*,* H_x and g^y,
where $y = R * (\frac{r_x * b_i}{P_i}) * \prod_{(j=0)}^{(m-i-1)} f^j(s)$

2.4 Overview of GRAS

Using PKI for vehicle-to-vehicle communication is costly and inefficient. Whereas, using symmetric keys between vehicles is an efficient procedure bur rather, establishing them is more complicated [22]. Hence, for our protocol we choose vehicle groups such that, vehicles that are moving in close proximity of each other forms a group. Soon after group formation, the members of the group elect a leader for the group who is then responsible for generating a group key and shares it with the other group members.

a. Group construction and Leader selection

We restrict the group formation in such a way that, vehicles can form a group if every member in a group hears the broadcasts of all other members in that group. Based on the assumptions made in 4.1, the vehicles in a group will have almost the same velocity on average and will move relative to each other in the same direction. Such a type of group can then be represented by a head called group leader. As we mainly focus on the message authentication and aggregation, any existing optimal algorithms can be adopted by the proposed GRAS scheme for group construction and leader election.

b. Member join and authentication

In general, every vehicle V broadcasts periodic safety messages once they join in the network. At the same time, V looks for any nearby group and attempts to join if it exists. If V hears any hello message from a group leader GL, it will immediately communicate the GL with a joining request *jreq* along with its self generated public key *Pu,* its OBU certificate $Cert_{OBU}$ signed with the private key sk_x, the corresponding public key PK_x and helper value H_x permitting the verification of $Cert_{OBU}$, in order to obtain a membership in the group. The group leader then verifies the credentials submitted by the requesting vehicle for its authenticity and discloses the group public key of its group. The algorithm works as follows:

Algorithm: Member join and authentication

if (V hears "*hello*" message from a group leader GL)
{

1. V_i computes

$$sk_x = R * r_x * a_i * \prod_{(j=0)}^{(m-i-1)} f^j(s)$$

$$PK_x = g^{(R*r_x*a_i*\prod_{(j=0)}^{(m-i-1)} f^j(s))}$$

$$H_x = \prod_{(j=0)}^{(m-i-1)} * r_x^{(-1)}$$

2. $V_i \rightarrow$ GL : $Encr_{PK_{gl}}$ *(jreq, Pu,*$Cert_{OBU}$*,* PK_x*,* H_x*,* g^y)

3. GL computes

$$cv' = (PK_x * g^y)^{H_x}$$

4. GL checks

if ($cv = h(cv')$)
{
GL stores *Pu,cv*
GL → V_i : $Encr_{PK_x}\left(PK_{gp} \| \{PK_{gp}\}_{sk_{gp}}\right)$
}
else
{
GL : no reply
}
}

c. Batch verification

In GRAS, the safety message *Msg* will hold the following format:

$$Msg = Pu \| M \| \text{Ꮒ} \| TS \| \{Pu \| M \| TS \| \text{Ꮒ}\}_{\sigma_{pr}}$$

A vehicle V with the group public key *Pu* can generate a valid signature σ_{pr} which is composed of (μ, γ) for a given message *M* as follows.

1) OBU selects a random number $x \in Z_p^*$
2) It then computes α, λ and γ such that
$\alpha = g^x, \lambda = f(M \| PK_x \| \alpha \| TS) \in Z_p^*, \gamma = Pr + x\lambda$
(α, γ) is a valid signature on message *M*

Any verifier (group leader in this case) can verify the above signature and accept the message if the following equation holds:

$$\hat{e}(g,\alpha)^{\gamma} = \hat{e}(g,\alpha)^{(Pr+x\lambda)}$$
$$= \hat{e}(g,\alpha)^{Pr}.\hat{e}(g,\alpha)^{x\lambda}$$
$$= \hat{e}(g^{Pr},\alpha).\hat{e}(g^x,\alpha^{\lambda})$$
$$= \hat{e}(g^{(R*r_x*x_i*\prod_{(j=0)}^{(m-i-1)} f^j(s))},\alpha).\hat{e}(\alpha,\alpha^{\lambda})$$
$$= \hat{e}(Pu,\alpha).\hat{e}(\alpha,\alpha^{\lambda})$$
$$= \hat{e}(,Pu.\alpha^{\lambda})$$

Consider the group leader receives $(\alpha_1,\gamma_1)\,(\alpha_2,\gamma_2)\ldots..(\alpha_n,\gamma_n)$ which are the signatures on the messages M_1, M_2,.. M_n, respectively. Then, those signatures can be aggregately verified as follows.

1) GL calculates $\lambda_1, \lambda_2 \ldots..\lambda_n$ and α', γ' such that

$$\alpha' = \prod_{i=1}^{n} \lambda_n \alpha_n, \gamma' = \sum_{i=1}^{n} \gamma_n$$

2) GL accepts the message if the following equation holds:

$$\hat{e}(g,\alpha')^{\gamma'} = \hat{e}(\alpha', \prod_{i=1}^{n} Pu_i \alpha_i^{\lambda_i})$$

d. Message Aggregation

Once the group leader GL batch verifies all the message sources that sent messages to it at time *t*, it aggregates all the messages as follows:

$$Aggr_{msg} = (Pu_1 \| M_1 \| \hbar) \| (Pu_2 \| M_2 \| \hbar) \| \ldots \| (Pu_n \| M_n \| \hbar)$$

Then it signs the above with the group private key sk_{gp} and one hop broadcasts $Aggt_{msg} \| \{Aggt_{msg}\}_{sk_{gp}}$ to vehicles in its group who are within its communication range.

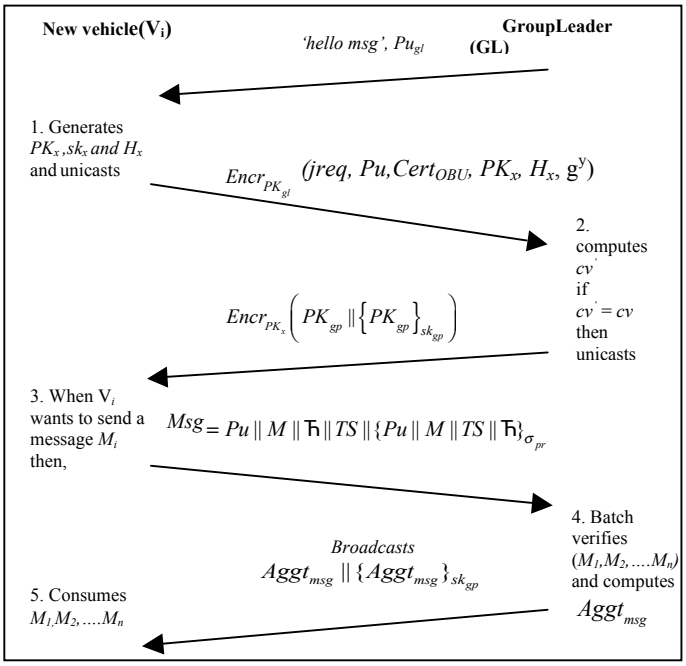


Figure 1: Protocol: GRAS

2.5 *k*-anonymity property

For identity preservation, we employ the concept of k-anonymity [24] in the proposed GRAS scheme to mix *k* vehicles. With GRAS, the TA assigns a common check value *cv* to *k* vehicles, where the *k* number of vehicles will take the same check value during their registration. When an adversary (including the group leader) intends to trace a specific vehicle through its check value, it cannot succeed as more than one vehicle may have the same check value in a group. The highest value of *k* can be determined by the TA which could be less than or equal to the total number of vehicles registered to the TA.

2.6 Member joining and leaving

When a vehicle V leaves a group, the GL simply removes V's information from its member list but rather do not make any changes in the group key. While at the same time, when a GL wants to leave the group, it sends a *leave* message to other members of the group before leaving, which will allow the group members to elect the node with next highest weight as the new GL.

However, if there are very few vehicles in an area, a group may not form, due to the lack of the leader [21]. But these shortcomings do not affect the functionality of the VANET, since a vehicle falls back automatically into the digital signature mode when it cannot join a group. In such case, any vehicle can send its messages directly to the network using traditional PKI model by attaching the corresponding helper value, check value and its certificate together with the self generated public key and it's tracking hint.

3. RELATED WORK

3.1 RSU based authentication

As the message authentication with digital signature put forth a big question on privacy, some solutions were proposed by researchers relying on an infrastructure to provide a certification service for vehicles. Authors in [15] proposed a conditional privacy preservation scheme, which talks about three layers of privacy. First, an RSU is responsible for issuing temporarily anonymous certificate to vehicles, and thus, it can map the vehicle with the issued anonymous certificate. Second, however, for an IVC, a vehicle's identity is absolutely anonymous to other vehicles. Third, the trust authority (TA) can track the vehicle's real identity. Furthermore, this scheme addresses certificate revocation issues, since the expiration date in the certificate indicates the validity period of the certificate. However, it does not take scalability issues into consideration. RAISE [17] explores an important feature of VANETs by employing RSUs to assist vehicles in authenticating messages. In this protocol, the messages sent by neighboring vehicles are stored in a temporary buffer of the receiving vehicles. When the RSU broadcasts the aggregated hashes of all received messages, the receiving vehicles compare them with the stored messages and consume them if they find a match. Though this significantly reduces the verification overhead, this approach still leaves a lot of room for improvement.

3.2 Pseudonym based authentication

Message authentication and anonymity are the two major security challenges identified in [9]-[13]. To achieve message authentication and anonymity, [14] proposed that each vehicle should be preloaded with a large number of anonymous public and private key pairs together with the corresponding public key certificates. Traffic messages are signed with a public-key-based scheme. To achieve privacy, each public and private key pair has a short lifetime, and a pseudo ID is used in each public key certificate. However, this scheme requires a large storage capacity to store this security information. The pseudonym based approach that has been proposed by [14, 19] is an idea to help the vehicles exchange their communications without revealing their real identity.

3.3 Group based authentication

Group based protocols have been proposed as a complementary approach to privacy preservation. The key idea is to hide in a group a vehicle's explicit identity and location. This is a tradeoff between privacy preservation and information accuracy. The protocol proposed by the authors of [23] elects a group leader who then communicates with the RSU on behalf of the group. This protocol also suffers the disadvantage of the overhead associated with the RL (Revocation List) management required to authenticate group membership. Group signature (GSB) which was

first introduced in [18] which allows a group member to sign messages anonymously on behalf of the group. The identity of a signer can still be revealed by the group manager in case of a dispute. Lin *et al.* [3] proposed a group signature- based scheme to sign each message. In this protocol, the recipients can verify a message's signature with the group's public key. If the signature is authentic, the recipient can confirm that the sender is a group member but cannot identify a specific person. Since there is no identity information included in messages, this approach can also achieve identity privacy preservation, reduced storage cost and low bandwidth consumption. They considered the short group signature scheme that was introduced by Boneh et al. [4], which is secure and considered to be best suited to the V2V application. Authors in [3] propose a geographic based group formation that is a hybrid approach of both predefined and dynamic group formation. In this approach, the map is divided into overlapping cells using GPS. So every vehicle is aware of its current group at any moment by knowing its current location. Also the vehicle which is closest to the cell center is will be assigned as the group leader. A symmetric cryptographic key is shared between group members for mutual communication. However, this approach meets a serious communication overhead as the members have to communicate each other to agree upon the symmetric key.

To reduce the overhead of the group signature-based scheme, a similar scheme [16] was developed in which a vehicle can generate public and private key pairs by itself by using a group key. This scheme can achieve a tradeoff between the group-signature-based scheme and the traditional PKI-based scheme. The scheme proposed in [8] proposes the privacy preserving group communication scheme for VANETs to satisfy forward and backward secrecy, authentication, protection against collusion, and privacy under symmetric-key cryptography. Additionally, a node can calculate the new group key as well as update its compromised key list (i.e., keys which are utilized by misbehaving nodes) even if the node misses the group rekeying process. This assumes that every node has knowledge of the revoked node's key set in advance and retains the information. This increases the storage requirement of each node as well as the bandwidth consumed by communications between each node and the key server. Though, group signature is a stronger property than pseudonymous authentication, as any two group signatures generated by a node cannot be linked [7], group formation in VANET is yet not clear as it has not been clearly discussed in many papers.

4. PERFORMANCE EVALUATION

A. Communication Overhead

In this section, we calculate the communication overheads of the proposed GRAS protocol. We consider the Tate pairing implementation on an MNT curve with embedding degree 6. Accordingly, each point on this MNT curve is represented by 21 bytes. With GRAS, the communication overhead of GRAS is 21+98+21+21+21+77 bytes, where public keys *Pu, PK*$_x$, helper value H_x and the parameter g^y takes 21 bytes each, along with the 98 bytes OBU certificate and TA's certificate on the check value as 77 bytes are shared for *n* message generated by one public key *Pu*, because vehicles submit those information to the GL once per the self generated public key.

Also it is indicated in section VI that, an OBU_i with $Cert_{OBU}$ can generate a valid signature (α,γ) while sending an arbitrary message *M* to the GL. Since α and γ are points on the elliptic curve, the size of a signature in GRAS takes 42 bytes. Consequently, the communication overhead incurred in a signed message transmitted by an OBU is 62 bytes, which is the signature size plus the size of the tracking hint encrypted using SHA-1 algorithm which is also shared by *n* vehicles as the GL verifies every message on behalf of all the other members in the group. Note that the other group members need to verify only the group signature generated by the GL for *n* aggregated messages where *n* is determined by the number of vehicles with the packet release interval of 300 ms to the GL to broadcast a batched packet. This shows that the GRAS scheme is more feasible with respect to the incurred communication overhead.

B. Verification Delay

We compare the verification delay of the GRAS batch signature verification with the signature verification of ECDSA and CAS. CAS is a certificateless aggregate signature scheme [26]. The time needed to verify one ECDSA signature is $2T_{mul}$, and that for CAS is $3T_{pair}+2T_{mtp}$, where T_{pair} represents the time required to perform a pairing operation, T_{mtp} denote the time required for a hash function to perform one map to point operation, and T_{mul} corresponds the time required to perform one point multiplication. In [25], T_{pair}, T_{mtp}, and T_{mul} are found for a super singular curve with embedding degree $k = 6$ to be equal to 4.5 *msec*, 3.9 *msec*, and 0.6 *msec*, respectively. For CAS, there is no certificate; however CAS takes $2T_{pair}$ to perform check process in order to verify the sender. For the GRAS scheme, the verification delay of a message signature requires $(3T_{pair} + T_{mul})/n$ which include an additional $2T_{pair}$ to verify the sender once per public key and *n* is the number of vehicles in the group (every message is verified by the GL alone on behalf of the group with *n* members).

Scheme	One signature verification	*K* signature verification
ECDSA	$2T_{mul}$	$2KT_{mul}$
CAS	$5T_{pair}+2T_{mtp}$	$(4K+1)T_{pair}+2KT_{mtp}$
GRAS	$(3T_{pair} + T_{mul})/n$	$(3T_{pair} + KT_{mul})/n$

Table 2: Signature verification delay

Table 2 shows a summary of the signature verification delays for ECDSA, CAS and the GARS schemes.

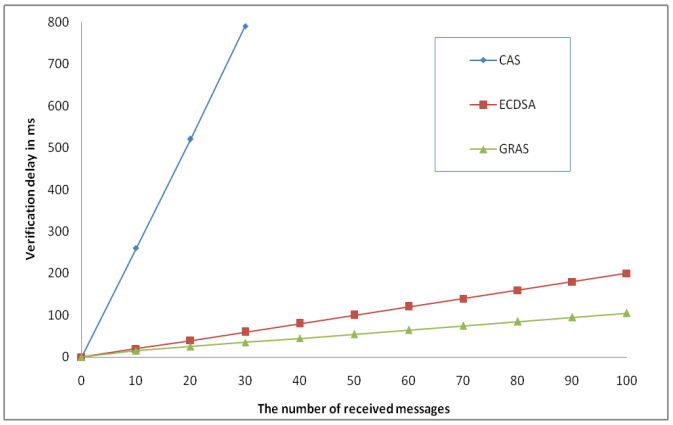


Figure 2: Verification delay for CAS, ECDSA and GRAS schemes

Figure 2 shows the verification delay for ECDSA, CAS and the GRAS schemes. It can be seen that the GRAS scheme has the lowest verification delay. In GRAS, the number of the pairing operations required for signatures verification is independent on the number of the signatures to be verified.

5. CONCLUSION

In this paper, we proposed a group based pseudonymous authentication scheme GRAS which can work effectively during the absence of any fixed infrastructure in the road sides. The proposed protocol vehicles with relative velocity and direction form groups and shares traffic related messages in the network. For that, they rely on the support of a group leader who can verify the messages collectively and broadcasts them by aggregating all the messages after performing enough authentication of the message sender. While preserving both privacy and authenticity of the message originator, GRAS drastically reduces the signature verification cost by freeing the other group members to do so. Simulations have been conducted to demonstrate the effective performance of the proposed protocol.

6. REFERENCES

[1] Chunhung Richard Lin and Mario Gerla, Adaptive Clustering for Mobile Wireless Networks, *IEEE Journal on Selected Areas in Communications,* 15(7): 1265-1275

[2] A. Stampoulis and Z. Chai. Survey of Security in Vehicular Networks. Technical report, 2007. Project CPSC 534

[3] X. Lin, X. Sun, P.-H. Ho, and X. Shen, GSIS: A secure and privacy preserving protocol for vehicular communications, *IEEE Transaction on Vehicular Technology*, vol. 56, no. 6, pp. 3442–3456, Nov. 2007.

[4] Dan Boneh and Hovav Shacham. Group Signatures with Verifier-Local Revocation. *In Proceedings of CCS 2004,* pages 168–177. ACM Press, 2004.

[5] Routing in Vehicular Ad hoc Networks: *A survey Vehicular Technology Magazine, IEEE Issue Date: June 2007,* Vol. 2 Issue 2, pp. 12-22.

[6] Sardari, M.Hendessi F. Fekri F, A New Paradigm for Collaborative Content Distribution from Roadside Units to Vehicular Networks, *IEEE conference on Senor, Mesh and Ad hoc communication and Networks 2009,* pp. 1-9

[7] G. Calandriello, P. Papadimitratos, J.-P. Hubaux, and A. Lioy. Efficient and Robust Pseudonymous Authentication in VANET. *In Proceedings of the ACM International Workshop on Vehicular Ad hoc Networks (VANET)*, pages 19–28, September 2007.

[8] A. Wasef and X. Shen, PPGCV: Privacy Preserving Group Communications Protocol for Vehicular Ad Hoc Networks, *in Communications, 2008. ICC'08. IEEE International Conference* on, pp. 1458{1463, May 2008.

[9] M. E. Zarki, S. Mehrotra, G.Tsudik and N. Venkatasubramanian, Security issues in a future vehicular network, *in Proc. Euro Wireless Conference,* 2002, pp. 270-274

[10] B. Parno, A. Perrig, Challenges in securing vehicular networks, *in Proc. 4th Workshop HotNets,* 2005.

[11] P. Papadimitratos, V. Gligor and J-P. Hubaux, Securing vehicular communications- Assumptions, requirements and principles, *in Proc. Workshop ESCAR,* 2006, pp. 5-14.

[12] A. Aijaz, B. Bochow, F. Dotzer A. Festag, M. Gerlach, R.Kroh and T. Leinmuller, Attacks on inter vehicle communication systems-An analysis, *in Proc. 3rd Int. Workshop Intell. Transp.,* 2006, pp.189-194.

[13] P. Papadimitratos, L.Buttan, J-P. Hubaux, F. Kargl, A. Kung and M.Raya, Architecture for secure and private vehicular communications, *in Proc. 7th Int. Conf. ITS Telecommunication,* 2007, pp.1-6.

[14] M. Raya and J.-P. Hubaux, Securing vehicular ad hoc networks, *Journal on Computer Security, vol. 15, no. 1,* Jan. 2007, pp. 39–68.

[15] R. Lu, X. Lin, H. Zhu, P.-H. Ho, and X. Shen, ECPP: Efficient conditional privacy preservation protocol for secure vehicular communications, in *Proc. IEEE INFOCOM*, Phoenix, AZ, 2008, pp. 1229–1237.

[16] G. Calandriello, P. Papadimitratos, A. Lioy, and J.-P. Hubaux, Efficient and robust pseudonymous authentication in VANET, in Proc. Int. Workshop VANET, 2007, pp. 19–28.

[17] Chenxi Zhang, Xiagong Lin, RongXing Lu, Pin-Han Ho, Xuemin Shen, An efficient message authentication scheme for vehicular communications, *in IEEE Transaction on Vehicular Technology,* vol, 57, No. 6, Nov 2008.

[18] D. Chaum and E. Ven Heyst, Group Signatures, *In Advances in Cryptology – EUROCRYPT 1991*, LNCS 547, pp. 257-265, Springer-Verlag, 1991.

[19] X. Lin, X. Sun, X.Wang, C. Zhang, P.–H. Ho and X. Shen, TSVC: Timed Efficient and Secure Vehicular Communications with Privacy Preserving, *IEEE Transactions on Wireless Communications*, vol. 7, no, 12, pp. 4987-4998, December 2008.

[20] G. Kounga, T. Walter, and S. Lachmund, Proving reliability of anonymous information in VANETs, *IEEE Trans. Veh. Technol.*, vol. 58, no. 6, pp. 2977–2989, Jul. 2009.

[21] M. Raya and J,P, Hubaux, Securing Vehicular Ad Hoc Networks, *Journal of Computer Security, Special Issue on Security of Ad Hoc and Sensor Networks,* Vol. 15, pp. 36-68 , January 2007.

[22] Mayank Verma, Dijiang Huang, SeGCom : Secure Group Communication in VANETs,

[23] K. Sampigethaya, M. Li, L. Huang, and R. Poovendran, AMOEBA: robust location privacy scheme for VANET, *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 8, pp. 1569–1589, 2007.

[24] M.Raya, P. Papadimitratos, I. Aad, D. Jungels and J-P. Hubaux, Eviction of misbehaving and faulty nodes in vehicular networks, *IEEE journal on selected areas of communication,* Vol 25, No. 8, Oct 2007.

[25] C. Zhang, R. Lu, X. Lin, P.-H. Ho, and X. Shen, An efficient identity based batch verification scheme for vehicular sensor networks, *Proc. IEEE INFOCOM 2008*, pp. 6–250, 2008.

[26] Z. Gong, Y. Long, X. Hong, and K. Chen, Two certificateless aggregate signatures from bilinear maps, *Proc. 8th ACIS Inter. Conf. on SNPD 2007*, vol. 3, pp. 188–193, 2007.

Towards a Common Research Strategic Roadmap for the Transportation Sector in Europe and beyond

Dr. Evangelos Bekiaris,
Matina Loukea,
CERTH / HIT
Athens, 15231, Attica, GREECE

Jean-Pierre Medevielle,
IFSTTAR
Lyon-Bron, F-69675, Bron Cedex, FRANCE

Steve Phillips,
FEHRL
Brussels, 1200, BELGIUM

Wolfgang Steinicke
EURNEX
Berlin, 14193, GERMANY

ABSTRACT

DETRA (Developing a European Transport Research Alliance) is a 7th Framework project, whose concept derives from the so-called Lyon Declaration and concerns the deepening of the European Research Area objectives in transport in order to address the Grand Challenges. Key priorities of this Alliance is to examine the strengths, weaknesses, opportunities and threats (SWOT) in the domain and develop common understanding and approaches to reducing fragmentation and overcoming barriers. The DETRA project aims to meet and exceed the requirements and objectives of the call for an Analysis of the state of ERA development within the transport domain and to develop recommendations for the EC, member states and other stakeholders as well as for the DETRA partner organisations themselves. In this study, particular emphasis is given to the part of DETRA concerning the development of a single trans-European research program, which can be used as a compass for the future research activities of the whole transportation area.

Keywords: Transportation Area, European Research Area, Research Priorities, Transport Research Database, Pan-European Research Program

1. INTRODUCTION

Transport is a key area, affecting most of the Great Challenges of Europe, according to the Lisbon Treaty. Although substantial progress in improving the transport sector has been made across Europe during the last decade, it is clear that different countries are improving at different rates, gains in some countries are becoming increasingly difficult to obtain, some sectors of road users are seeing a decrease in the development and the demographic and socio-economic influences on road user behaviour will drastically influence transport in the coming years. The continuing objective of further evolution in the EU transportation system should be properly supported by ambitious, cost-effective and targeted research activities. Increasing effectiveness and efficiency of research in the transport area is a mission having a direct impact on the quality of life in Europe.
At a time when Europe faces economic constraints, resources devoted to research must be spent wisely. In order to help this process it is necessary to describe and determine priorities. If these priorities are widely known and disseminated it helps both those who provide the research and those who fund it. Research organisations can consider how to address the topics and define where alliances of complementary skills need to be made, what assets best need to be developed and what techniques to be applied. Research funding organisations should consider budgets, determine where contingencies arise and set the appropriate frameworks for implementation, including regulation and legislation that might be needed to maximise the benefits.

In this context, an EC co-funded project called DETRA, has been initiated, in order to contribute to the government of the research in the transportation area. The aim of the DETRA project is to analyze and then guide ERA development within the transport domain. DETRA's concept derives from the Lyon Declaration, which was signed in 2008. Lyon Declaration's signatories, i.e. ECTRI, FERSI, FEHRL, EURNEX, HUMANIST, ISN and NEARCTIS organisations committed themselves to work together on the deepening of the European Research Area in the transport domain. From this commitment grew the objective to create a European Transport Research Alliance (ETRA) that would strengthen research in the transport domain. Key priorities of this Alliance are to examine the strengths, weaknesses, opportunities and threats (SWOT) in the domain and develop common understanding and approaches for reducing fragmentation of research and overcoming barriers of implementation.

The most important key priority of DETRA and of the organizations that collaborate in this project is to address the major challenges, as they were defined by the European Union. These challenges are related to all sectors of our society (e.g. public health, climate change, globalisation etc.) and according to the perspective of DETRA and its members, in each of these challenges the transportation sector can contribute to its solution. Some examples of such a transportation effect in the major challenges of Europe are the following:

- to make transport infrastructure and transport systems more resilient to a changing climate as well as directly reducing the climate change impacts of transport;
- to reduce energy consumption in the transport system, increasing the security of supply to reduce transport system impacts on ground water supplies and consider transport and water in land-use planning;
- to raise the standards of public health by increasing access to health facilities (including developing countries), increasing the resilience of the transport system in pandemics and reducing traffic accidents fatalities and injuries;
- to shape and maintain a transport system that reflects the needs of the developing globalization and to help European transport system stakeholders to adapt accordingly;
- to increase the effectiveness of transport necessary for food production (including developing countries) and to improve/innovate in transport logistics to reduce food waste;
- to adapt transport and mobility systems to an aging population.

2. OPTIMISING RESEARCH PROGRAMMES AND IDENTIFYING PRIORITIES

Within the context of the DETRA project a significant emphasis is placed on well-coordinated research programmes and priorities, including a jointly-programmed public research investment at European level involving common priorities, coordinated implementation and joint evaluation. This is the objective of the project's WP titled "Optimising research programmes and identifying priorities".

This goal has been initially pursued by a survey on the existing research programs and roadmaps developed in single European states, at the EC level, as well as internationally (e.g. in Japan, USA), related to all areas and all modes of transport. The purpose of this process is to collect all the research priorities for each transport subarea that concern any existing gaps or emerging for any transport mode.

Currently, more than 40 such research priority recommendations and/or roadmaps have been analyzed, stemming from several types of organizations, such as technology platforms (e.g. ERTRAC, ERRAC, etc.), research institutes associations (e.g. FERSI, ECTRI, FEHRL, etc.), industrial associations (e.g. EUCAR, CLEPA, ACEM) and citizen representive organizations (e.g. POLIS, FIA, FEMA, etc.). Of course, special emphasis has been given to the roadmaps and documents of the European Commission, as for example the "White Paper - Roadmap to a Single European Transport Area – Towards a competitive and resource efficient transport system" and the "Europe 2020 Flagship Initiative Innovation Union".

The collection of the material, which is a continuous process throughout the duration of this project, has been followed by the classification and categorization of the information, that has been based upon an electronic template for the classification of each research document's information, according to the four (4) major categories of transport designated by the beginning of the DETRA project (Safety & Security, Congestions & Capacity, Environment & Energy, Globalization). In addition, this template has also helped to analyze the key elements of the documents.

Research Priorities

Starting from these 4 key areas that constitute the backbone of the DETRA project, and according to the information collected by all the analysed documents, a list of research priorities has been emerged, containing so far 20 research areas and nearly 92 themes. This categorization of the transport research priorities concerns all the transport modes (road, rail, maritime and aeronautics) and reflects their research needs. Below, the research priorities of Safety & Security and Congestion & Capacity domains are being indicatively presented, while this kind of categorization has been also done for the other major areas.

I. SAFETY & SECURITY

- **Road Transport**
  - Passive Vehicle Safety
  - Active Vehicle Safety
  - Training
  - Simulation
  - Infrastructure Safety Measures
  - Cooperative Systems
  - Testing Platforms & FOTs
  - Vulnerable Road Users

- Advanced road surface and bridge materials

➢ **Railway Safety**
- Threat Analysis
- Active & Passive Safety Systems
- Training
- Design of Infrastructure
- Crisis Management Procedures
- Terrorism

➢ **Maritime Safety**
- Development of Prevention Strategies (Risk – Based Design)
- Human Machine Interface
- Development of Robust Ships and Reliable Equipment
- Strategies, Methods & Procedures for Safeguarding Security
- Development of Decision Support Systems

➢ **Aeronautics Safety**
- Human Machine Interface
- Threat Analysis / Risk Prevention
- Active and Passive Vehicle Systems
- Simulation
- Mitigation Strategies
- Training

➢ **Security**
- Risk Management
- Protective Measures
- Mitigation Strategies
- Resilient Infrastructures
- Network Components Critically Computation
- Supernetworks

II. CONGESTION & CAPACITY

➢ **Road Infrastructure**
- Monitoring & Maintenance
- V2I Communications
- Dedicated Infrastructure
- Electronic Tolling Systems

➢ **Traffic Management**
- ITS
- Advanced land and waterborne transport management systems
- Active Traffic Management

➢ **Urban Planning**
- Car free Zones/Areas and urban road pricing
- Transit-oriented Development
- Non-motorized Mobility (e.g. Pedestrians & Bicycles)
- Traffic Calming Measures and Shared Space

➢ **Supply and Demand**
- Dynamic Traffic Assignment Models
- Fusion of Micro & Macro Traffic Simulation and Planning Models
- Activity based Models

➢ **Multimodality**
- Freight
- Infrastructure
- Interoperability
- Single European Transport Area

➢ **Aeronautics**
- Interoperability Principles
- Development of New Management Systems
- Single European Sky (SESAR)
- Improvement of the Efficiency and capacity of airports
- Simulation & Modelling
- Novel Concepts of Automated Aircrafts

➢ **Rail**
- Tracking including Sub-grade and Rail
- Interoperability
- Wheel/Rail Interface
- Signalling and Control Systems - Line-side Equipment
- Intelligent Mobility, e.g. Telematics (Galileo), Intermodality, Customer Information Systems, Web-based Information Systems

➢ **Maritime**
- e-Maritime Initiative
- Inland waterway transport

Transport Research Work Programmes and Roadmaps Database

The determination of Research Priorities has also provided the structure of the "Transport Research Work Programmes and Roadmaps Database" which has been developed for the collection, sorting and analysis of the research programs and roadmaps.

The objective of this database is to provide a tool which will facilitate the detailed and clear presentation of all the relevant research documents. In this web-based tool (http://160.40.63.90/detra/), anyone has the opportunity to search for the documents that he/she is interested in, as well as to suggest further documents regarding research priorities and roadmaps in transport. This database is considered as being a living structure, which grows and develops at least until the end of the project. The content of this database, which is continuously updated, was collected through a detailed literature survey made by all DETRA partners, as well as by collecting research

programmes that have been developed by the partners themselves.

The DETRA tool is a fully web-based application. A web-based solution was selected, as this one guarantees easy access, without the need of special software for the end-user. A combination of technologies has been used for the successful development of this tool. The technologies that have been used are the following:

- **HTML, PHP 5** and **CSS** are the technologies used for the design of a dynamic web application.
- **MySQL 5**, which is a powerful open source database tool for the design and management of databases.
- **NetBeans IDE early access for PHP**, which is an open design application by which a fully functional and complete website can be created.
- **APACHE HTTP Server 2.2**, which is an open –source HTTP server for modern operating systems. It is a secure, efficient and extensible server that provides HTTP services in sync with the current HTTP standards.

This tool has been developed in order to support the DETRA EU project and, as mentioned above, the main idea was the design and development of a tool which will be able to collect research programmes. It consists of three areas. These are i) Home ii) Insert and iii) Search areas.

Illustration 1: Home page of "Transport Research Work Programmes and Roadmaps Database"

In more detail, the "Home" area contains the login form together with all the appropriate functionality of a login page, e.g. "Register", "Forgot your username", etc. The "Search" area is available to all visitors (registered or not). It provides a search engine which is able to present the stored information in many different ways according to the criteria selected by the user. On the other hand, only the registered users are able to insert information to the database of the tool. This can be done via the "Insert" area of the tool.

3. TOWARDS A MODULAR, PAN-EUROPEAN TRANSPORT RESEARCH PROGRAMME

The final outcome of this whole process, as mentioned above, will be the development of a single trans-European Transport Research Roadmap, containing all the relevant research priorities for each separate field of the transportation research area.

A first draft of this document has already been prepared, presenting the main research priorities of the road transport area as far as the sectors of Safety & Security, Environment & Energy and Congestion & Capacity are concerned. For the presentation of this first draft, a workshop has been held in the beginning of September 2011, in which nearly 20 stakeholders attended, apart from the project's partners. During this workshop, a fruitful and interesting brainstorming took place, which leaded to interesting and valuable feedback. This Roadmap will be updated several times, until it reaches to its final form in May of 2012.

Analysis of the Procedure

The first draft of the "Optimised pan-European Transport Research Work programme, Roadmap and priorities" has been developed by the analysis of the various s transport roadmaps and work programmes derived mainly from EU research and industry organisations.

Based upon the benchmarking database results, initially a set of commonalities and differences between the different transport research work programmes has been structured. These commonalities and differences have been obtained relying on the research priorities that have been presented above.

More specifically, during the analysis of each document, emphasis has been given to the research needs and priorities described in it, which were related to the different domains and sub-domains of the transport area. All these research needs and priorities have been categorised according to the research priorities' list defined in the DETRA project and by this procedure a set of commonalities has come up concerning all the same or relevant needs and requirements that have been designated by the several different organizations as far as the fields of the whole transport domain is concerned.

			ERTRAC - "Strategic Research Agenda"	EUROPEAN COMMISSION- ""Towards a European Road Safety Area: Policy Orientations on Road Safety 2011-2020"	FERSI - ECTRI - "The Sustainable Safety Approach to Road Transport and Mobility"	EUCAR - "Challenges and Priorities for Automotive R&D"
SAFETY & SECURITY	Road Safety	*Active Safety*	Research on **Advanced Driver Assistance Systems** such as lane and distance keeping and warning for inappropriate speed.	Promotion of the use **of modern technology in order to increase road safety**.	Driver/ rider **monitoring technologies and tools, to prevent accidents** due to driver/rider inattention, excessive workload, fatigue etc.	Development of **driver assistance functions** to optimise overall potential, maximise comfort and safety for the driver.

Illustration 2: Example of commonalities regarding the Active Safety theme

The Example of the Fields of Active and Passive Safety

Regarding the section of Road Safety, 35 documents from 20 key stakeholders have been taken under consideration and have been analysed in respect to their proposals for the future research requirements.

An indicative example of the results of the procedure described above, as well as a model of the basic structure of the "Optimised pan-European Transport Research Work programme" is the research priorities defined for the theme of Active Safety of the Road Safety section, which is presented below:

- **ROAD SAFETY**
  - **Research Priorities in Active Safety**

After the categorization and the extraction of commonalities from all the relevant documents regarding the Active Safety theme, the following research priorities have been emerged.

At first, the prioritization from the ERTRAC strategic agenda is being presented, which accepted and agreed by almost all the other organizations, either they are ERTRAC members or not.

- Safe integration of nomadic devices into the driving task and interfacing them to the CAN [ERTRAC];
- Further research on ADAS, such as lane and distance keeping, warning for inappropriate speed, longitudinal and lateral distance sensing, collision avoidance and improved vision [ERTRAC];
- Active safety measures for accident consequences mitigation [ERTRAC];
- Intelligent systems for occupants' detection and protection and interior adjustments [ERTRAC];
- Integrated and adaptive HMI for driver workload minimization [ERTRAC];
- Emergency services research, with emphasis on info on accident severity, location, passenger and eventually the presence of any dangerous goods [ERTRAC];

In addition, some more specific ideas have been included from various stakeholder representative organizations.

- Emphasis on further development of ADAS for driver's/rider's inattention, excessive workload, fatigue, drowsiness, alcohol abuse, illegal drugs use, medical drugs misuse (i.e. follow up of SENSATION and DRUID IPs [FERSI – ECTRI, FEHRL AND EPoSS];
- Individualized and personalized active safety systems, based upon driver's/rider's behaviour and driving/riding style [FERSI – ECTRI]. This is very close to the Adaptive HMI Concept [EPoSS]. Emphasis on ADAS adaptation for the elderly [HUMANIST];
- Active safety systems tailoring for clean vehicles [eSafetyForum];
- Adaptive light projection (using e.g. turning lights, projection, automatic light bean) for better road illumination for cars [EPoSS] and PTWs [FERSI].

The work is still in progress.

4. CONCLUSIONS

The aim of this paper is to present the DETRA project and in particular the part of it related to the development of a pan-European Transport Research Work Programme. In order for this report to be finalized, more than 40 well-coordinated roadmaps and research work programmes for research and industry organizations have been collected and analysed. This procedure, which is an ongoing one, has already led to the first draft of this report containing research priorities regarding the fields of Safety & Security, Congestion & Capacity and Environment & Energy of the Road Transport mode. The completion of this "Optimised pan-European Transport Research Work Programme" is expected to be finalised up until May of 2012.

References

1. The Lisbon Treaty, http://www.lisbon-treaty.org/wcm/index.php
2. European Commission: Europe 2020 Flagship Initiative Innovation Union. Communication from the commission to the European Parliament, the Council, the European Economic and Social Committee and the committee of the Regions. Brussels, SEC(2010) 1161, 06/10/2010
3. European Commission: Roadmap to a Single European Transport Area – Towards a competitive and resource efficient transport system". White Paper, 28/03/2011
4. Lyon Declaration, 2008 - http://eufp-event.inrets.fr/docs/Lyon%20Declaration_ETA.pdf
5. "Strategic Research Agenda", ERTRAC.
6. "Towards a 50% more efficient road transport system by 2030 ", ERTRAC.
7. "The electrification approach to Urban Mobility and Transport", ERTRAC – EPoSS
8. "The Sustainable Safety Approach to Road Transport and Mobility", FERSI – ECTRI.
9. ECTRI Position on FP8", ECTRI.
10. "White Spots in European Transport Research Traced by ECTRI's Thematic Working Groups", ECTRI.
11. "Vision: Road Transport in Europe 2025", FEHRL.
12. "Strategic Research Agenda of the European Technology Platform on Smart Systems Integration", EPoSS.
13. "European Roadmap - Electrification of Road Transport", EPoSS.
14. "Strategic Research Agenda- ICT for Intelligent Mobility", EsafetyForum.
15. eSafetyForum - "New Recommendations"
16. "Humanist research Topics - Work Program 2011-2013", Humanist.
17. "Secondary Safety Research Action Plan", APSN (Advanced Passive Safety Network).
18. "Global Challenges - Local Solutions", JPI "Urban Europe".
19. "Challenges and Priorities for Automotive R&D", EUCAR.
20. "Transportation Roadmap 2010", EICOSE.
21. “A Vision for Integrated Road Transport Research”, EARPA.
22. "Further Advances in Automotive Safety Importance for European Road Transport Research", EARPA.
23. VAGVERKET (Swedish transport Administration) - "Road Safety"

EVALUATION OF KNOWLEDGE SHARING IN A COLLABORATIVE SOFTWARE

Douglas Juliani[12], Jordan Juliani [3], Aline França de Abreu[1], João Artur de Souza[1]

[1] Postgraduate Programme in Engineering and Knowledge Management, Federal University of Santa Catarina - UFSC.
[2] Department of Computer Science, Federal Institute of Santa Catarina – IFSC.
[3] Department of Library and Information Management, State University of Santa Catarina, Florianopolis, Brazil - UDESC.

douglas@webpack.com.br, {jordanjuliani, afdeabreu, jartur}@gmail.com

ABSTRACT

The widespread use of collaborative applications as a means of communication, research and collation of information, among others, has significantly accelerated the acquisition, dissemination, and especially the sharing of information and knowledge among individuals. Thus, understanding the factors influencing this process is crucial to the success of a software based on the contribution of people. In this research, among the various factors analyzed, it was found that the usability and the characteristics of individuals are crucial to assess knowledge sharing in a knowledge-based system, in this case, the social software PreçoPúblico. We chose to use questionnaire with 23 items as an evaluation tool that, based on two existing, refurbished and was validated by pre-testing and statistical technique. Being 19 closed questions (categorical measures) and 4 open, descriptive. Results showed that users are predisposed to share their knowledge and information about product prices in the software. Still, it was shown that the system has a high degree of usability, despite some difficulties encountered by most respondents coming from the fact it is the first system to access them. The comments collected from the open questions suggest system improvements.

Key words: *knowledge sharing, usability, individual factors, evaluation, PreçoPúblico.*

1. INTRODUCTION

Collaborative tools are part of the everyday life of hundreds of millions of users all over the world, pointing to a significant change in the way people interact with applications in the internet, like the social networks facebook, MySpace, LinkedIn, Orkut and many others that collectively make up what is known as *WEB 2.0*. In this way, dozens of organizations in all the economic sectors are adopting said technologies to generate value to its stakeholders – be they actionists, clients, employees, citizens or societies generally speaking. It's clear then, that we reached the limits between the simple and a big revolution: Simple due to the extreme ease of using, implementing and low cost of the tools; potencially revolutionary by the ease of reaching and engaging thousands or even millions of users to collaborate and interact (TERRA, 2009).

Han & Anantatmula (2006) affirm that the ease of access to the technological tools, specifically information and communication tools (*TICS*), are important to stimulate the collaborators to share knowledge, concluding that the ease of use and necessary training can persuade and incite the use of available technologies to share knowledge.

In this direction, afforded mainly by the ease of use and informal traits, Tseng, S.-M., & Huang, J. -S (2010) relate that Wikipedia has a relevant efect in the process of the organizations' knowledge sharing.

Many dimensions are influential to knowledge sharing, among the individual factor (attitudes and behaviors) and the organization's rules, the TICs (technological factors) can help the employees to receive knowledge, but not necessarily to garantee their commitment, that's why investing in technology can facilitate the flux of explicit knowledge, but will hardly contribute to the exchange of tacit knowledge, which has a bigger value and allows the generation of new ideas. (ORDAZ et al, 2009).

Systems of knowledge management, used to capture and distribute knowledge, normally require the users' contribution instead of keeping the knowledge to themselves or transmiting it directly to the closest ones through conversations or annotations. These systems should, thus, stimulate the participation of individuals through rewards, be they tangible or intangible (KING & MARKSJR, 2008).

As it's been introduced up to this point, various studies investigate the factors that influence knowledge sharing, however, this research aims to separately measure those atributes to determine the intensity of knowledge sharing in a system based on knowledge. In this incumbency, the comportamental traits of the users and of the system itself, specifically in terms of usability, are examined.

This work destinates itself, thus, to investigate under the aspects of individual traits and usability, a system based on knowledge, aiming to potentialize the sharing and use of knowledge by the users that utilize it.

Based on the above exposed, the question of this research presents itself:

What's the intensity of knowledge sharing in a system based on knowledge with the users' individual traits and usability as a basis?

2. KNOWLEDGE BASED SYSTEMS AND COLLABORATIVE TOOLS

To study knowledge sharing in a system, it's necessary to learn what a technological artefact that transfers knowledge or informations between users understands, its characteristics and distinctions compared to other kinds of systems.

Firstly, one must highlight, as said in a literature revision about applications and knowledge managing technologies elaborated by Liao (2003), that there is a diversification of concepts according to the line of research, expertise and specialty of each author, being that some terminologies have concepts in common, as for example, specialist systems/artificial inteligence and systems based on knowledge.

Another take on this defines any information technology application which helps in any way managing actives of knowledge. Specialist systems, collaborative tools like *groupware*, *data warehouses* or even *intranets* are included in this class of software (Hendriks, 1999). On the other hand, Currie & Kerrin (2004), Hayes & Walsham (2000) and Mackinlay (2002) *apud* Tseng (2008), put forth that knowledge isn't efectivally shared when kept by information technology tools that involve static repositories like in an *intranet*, in which the knowledge can't transmit the richness of context in which it was applied.

An emerging collaborative tool is social computing, characterized as digital systems developed from informations of social interest and context to improve the activity and performance of people and organizations (Chua, 2004). Kwaifunip & Wagner (2008) believe that social computing's the trend for computing inside organizations that will represent an impacting turnaround for TI managing tools in today's enterprises.

When speaking of technological tools that depend on massive colaboration of the users' knowledge/informations, the success of implementing these kinds of systems, notwithstanding technological questions, is strongly related to (ambiental and human) factors like: quality of content, commitment to the use of a technology, users' satisfaction, profile of the ones involved and motivation to use that technology (AURELIE, BECHINA, & NDLELA, 2007).

The next topic will introduce the process of knowledge sharing, its concepts and dimensions, showing the aspects that motivate people's participation in a system based on knowledge.

3. KNOWLEDGE SHARING

Many terms found in the literature, like dissemination, distribution, partition, transference or knowledge sharing, are employed as synonyms to characterize the process in which knowledge migrates from one situation to another: between individuals and teams of people; from physical sources like data banks, documents, CDs, videos, books etc, to other recepting sources or to people and vice-versa (Tonet, 2005).

For Lin & Lee (2008), knowledge sharing equals a culture of social interaction involving the exchange of knowledge between the collaborators, experiences and competences. It can be individual, like, for exemple, talking with a colleague to help improve the performance in a given task, or organizational, like capture, organization, reutilization and transference of experiences based on knowledge existing in the organization.

There is a certain convergence in the authors about the aspects that influence knowledge sharing. Substancially, one can note in the literature two great groups or dimensions: individual factors and ambiental factors (cultural factors and technological factors) in which the individuals and technological artefacts are inserted.

The studies of Kaiser, Kansy, Mueller-Seitz, & Ringlstetter (2009); M.-J. J. Lin, Hung, & C.-J. Chen (2009); Nan (2008); Tonet (2005) and Ye, H. Chen, & Jin (2008) give prominence, specially, to the individual factors which, according to Vorakulpipat & Rezgui (2008) and Osterloh and Frey (2000), can be divided in: intrinsic (activities and behaviors that individuals naturally engage for their own benefit) and extrinsic (represented by the direct compensation that a person's work or actions exert).

In another study, the model of research developed by Ye, H. Chen, & Jin (2008), whose goal to examine specifically knowledge sharing in virtual communities, considers also the aspect of system usability.

A research about knowledge management in information technology organizations under the collaborator's perspective noticed that questions related to usability of technology like ease of accessing and finding informations were pointed by eighty per cent of the respondents as crucial to sharing knowledge (Han & Anantatmula, 2006). Nevertheless, another study pointed that students would like to use the collaborative system of annotations (PAMS 2.0) in instances of work groups due to the ease of use and stability (Su, S. J. H. Yang, Hwang, & Zhang, 2010).

The success of the Wikipedia tool can be explained by two values: the sociability and usability. The first, in a cybernetic context, allows establishing reward systems to stimulate the users' participation and the second, by being friendly and allowing an easy interaction between the users and the system (TSENG & HUANG, 2010).

Next, the systems usability content will be presented with the aim to analyse its intensity in knowledge sharing by the way of technological tools.

4. USABILITY OF SYSTEMS

In an all-encompassing way, usability refers to a collective of concepts like time for executing a task, performance, user satisfaction, ease of learning, not provoking errors, solving the tasks with efficiency and effectiveness (ISO 2011, 1998; Neves et al., 2006 ;NIELSEN, 1993).

Santos (2007) made a revision to the literature about usability by searching through the Brazilian data banks classed in the Qualis of CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) and obtained as results the standards adopted by authors Shackel, Jakob Nielsen, Bastien & Scapin, Jordan, Shneiderman and Quesenbery, in addition to the ISO usability norms. That revision was brought up to date and augmented with standards used by other authors. The similar attributes that are used by more than one author in ascending order are: memorization, error control, ease of learning, effectiveness, efficiency and satisfaction. Notice that the more recent works analysed are mostly based on the norm ISO 9241. The constitutive definition of each of these variables can be seen next:

- **Memorization / Ease of Remembering**: The system funcionalities should be easy to remember, even after not using it for some time (NIELSEN, 1993).
- **Error Control**: Mechanisms capable of preventing, reducing and facilitating the retrieval of errors generated by the system (BASTIEN & SCAPIN, 1993).
- **Ease of Learning**: Ease of learning to accomplish a given task in the system without help from supporters (QUESENBERY 2001).
- **Effectiveness**: Precision and completeness with which users attain specific objectives, accessing the correct information or generating expected results (ISO, 2010).
- **Efficiency**: Resources consumed to attain objectives in the system (Abran et al., 2007).
- **Satisfaction**: Convenience and acceptability of the product measured through subjective and/or objective methods (Feliphe & Lavor, 2008).

Among some authors, Blecken, Bruggemann & Marx (2010), Padilha (2004), and Santos et al. (2007, class Analytical and Empirical Evaluation as the two main ways of applying usability tests. In a complementary manner, Cybis (2000) adds another type entitled Prospective Evaluation, as do Winckler & Pimenta (2002) the Automatic Inspection one.
There's a solid theoric referential about the methods of usability evaluation, in compensation, the investigation about the individual traits which motivate the users to share knowledge through a system lacks validated evaluation tools.

5. RESEARCH METHOD

This study uses exploratory research, aiming to deepen the themes described in the previous chapter by means of bibliographical research in books and articles. It's highlighted that appart from the syncretism seen in the investigative methods outlined above, in the course of this work, the need was felt to employ another kind of research , the descriptive one, aiming to describe and evaluate knowledge sharing in a system through the application of a questionnaire. In spite of the quantitative framing of this study, a qualitative approach was made, through questions open to each dimension of the variables ascertained by the research tool.
To select the tool for this research, questionnaires already in the literature were analysed, preferencially the validated ones that were pertinent to the object of the proposed research. Among the ascertained works, two tools became the basis to the confection of the questionnaire adapted to this study.
After the questionnaire's elaboration, began the validation stage, which first was turned to a pilot-test, applied to three specialists, aiming an initial improvement, essencially related to the questions' comprehension and scale of answers, as well as the correction of grammatical errors. After the needed adjustements to the pilot-test, the questionnaire was applied to 10 users to verify its trustworthiness. With the confirmation of trustworthiness having being made, the tool was applied to the sample configuring the research's descriptive stage, where the collected data were analysed.
This study was applied to evaluate the web system called *PrecoPúblico*, a software with functionalities that cover some processes of knowledge managing, because, above all, it allows knowledge sharing by any individual through the publication of products and its prices, and also allowing the search of products through the internet.
The individuals' knowledge sharing analysis in this software uses the following hypotheses:
H1) The level of satisfaction in relation to the attitudes and behaviors of the system users is high.
H2) The system has a high level of usability.
H3) The individuals' knowledge sharing in the system is positively related to its usability.
Next, the variables used to measure the constructs of this research are demonstrated following the bibliographical survey that was developed, its concepts and measures of analysis of the investigation tool.

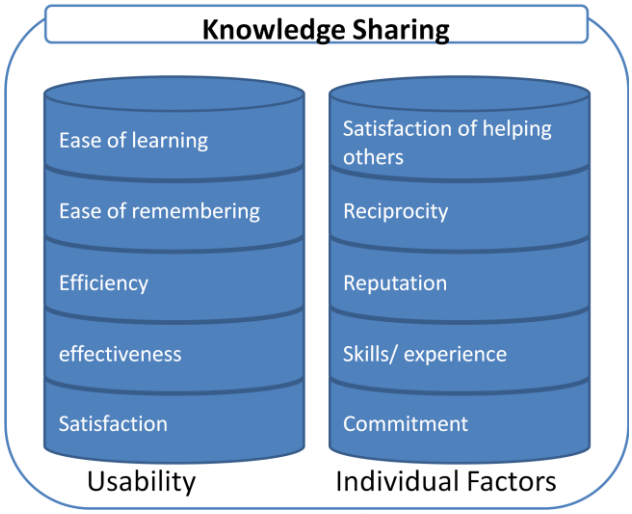


Image 2 – Constructs and variables of the research.
Source: From the author.

Among the various factors studied in the revision of the literature, the option was made to investigate the individual ones and those refering to usability, fundamented by the context of the research, which embraces a technological tool. Thus, the comprehension of the attitudes of the users who use it and of their traits are inherent aspects and can influence the process of knowledge sharing in this scenery. Image 2 ilustrates the composition of the constructs used in this research, including Usability (intervenient) and Individual factors (independent) and the dependent construct knowledge sharing in a system.
The synthesis of definitions, standards and their respective authors, extracted from the theoretical background of each construct in this research, is presented in table 1.

Factors	What it is (conceptual definition)	How to analyse (standard)	Authors (reference)
Usability	Ensemble of concepts like time of execution for a task, performance, user's satisfaction, ease of learning, not provoking errors, solving tasks with efficiency and effectiveness.	• By the ease of learning • By the efficiency • By the effectiveness • By the satisfaction	ISO 9241 (1998) Neves ET al. (2006) NIELSEN (1993)
Individual Factors	Concepts and personal traits related to the individuals' motivation to share knowledge.	• By the satisfaction of helping others • By the reciprocity • By the reputation • By the abilities and own knowledge • By the commitment	Lin, Lee, Wang (2008)

Table 1 – Summary of the constructs, their concepts and variables.
Source: From the author.

The universe of this research is constituted by any person who ever searched or bought goods or services through the internet, which, according to the Comitê Gestor da Internet no Brasil (2009), represents fourty five millions of the sixty two per cent users of internet in Brazil, as seen that the functionalities of the system analysed in this work are directly related to the action of searching and comparing product prices through the web.
In this work, the adding scale was applied to a form of Likert scale with five levels of answers about the respondents'

atittudes: a) I Strongly Disagree; b) I Disagree; c) I don't agree nor disagree; d) I Agree; e) I Strongly Agree.
Four of the twenty-three questions in the questionnaire are open, with the intention of getting closer to a qualitative approach, that in this tool, makes possible to collect any other perceptions different from the ones pre-formated in the questions. Each construct from the questionnaire contains an open space to other considerations and suggestions.

6. ANALYSIS AND INTERPRETATION OF THE RESULTS

Some analyses are interesting in the basic data from respondents like, for exemple, the age of respondents, where a little more than half (52,8%) is between fifteen and twenty-eight years old, being that the rest's between twenty-nine to sixty-four years old, putting in relief an even distribution of frequencies under this variable.
The major part of participants in the research has from twelve to sixteen years of experience in computer use and six to twenty years using the internet. Informations that indicate that the sample collected has a representative intimacy with the use of computers and good experience using the internet, which characterizes it as a sample that can be generalized with users apt to evaluate the sharing proposed for the PreçoPúblico system.
The scores reveal that the users, through their own traits and attitudes, are motivated to share knowledge in the system's context, confirming, thus, hypothesis H1 - The level of satisfaction in relation to the attitudes and behaviors of the system users is high. This is because all the variables analysed have an agreement result above 50%. On the other hand, the Recognition and Commitment dimensions attract attention to the software, as the percentuals of answers I don't agree nor disagree, were representative in comparison to the other questions of the analysed items. For this reason, the need to create resources that divulge and elevate the users' visibility is made evident, in the measure that they share their knowledge by means of the system. As an example, one could create a ranking with the images or names of the users who publish the most information or knowledge directly in the portal's home page.
In the same direction, the neutrality level (33%) about the pride to take part in the PreçoPúblico software, refers to the reflection about all the other variables in the Individual traits construct facing knowledge sharing. An example: if a user feels no satisfation in telling others he takes part in a project, that can be due to many other factors like lack of recognition of said attitude by other individuals who get this information, or even, due to the little visibility of the project itself, since it's still in the divulgation stage and possesses, by consequence, not many publications of product prices.
The interpretation of reached results about the publication and price research, which measure the system usability construct, is based on the percentual variation existing among that dimension's answers.
Confirming the H2 hypothesis - The system has a high degree of usability, both for publication and for price research, the scores show that the system's easy to use in all the studied variables. However, the task of Publishing prices was considered more difficult to learn, more time-consuming and complex than the task Research prices. This fact suggests that it's possible to create swifter mechanisms for the entry of data in the system as, for exemple, using a mobile device (cellphone, smartphone and others) with the bar code reader, capable of identifying a product's characteristics to register it directly in PreçoPúblico, avoiding the manual entry of data.
In the questions about the System's usability as a whole, the evaluation followed the already described results, confirming that users consider it as being easy to use, however, the scores of question seventeen (17), point to a representative percentage of neutrality in comparison to the other general questions about system usability, leaving a margin for reflection about what would justify this observation.
As in the research of Ye, H. Chen, & Jin (2008), it was made evident that system usability is also important to motivate knowledge sharing. It was discovered, also, that simplicity and quickness, respectively, are the variables that most influence knowledge sharing in the studied system, in spite of the little variation already explained in this section. Hypothesis H3 is, thus, confirmed - The individuals' knowledge sharing in the system is positively related to its usability.
Individually interpreted, four main types of information were identified in the open answers: congratulations, inexperience/familiarization, outdated content and error reports or improvement.
Another perception refers to the necessity to create new resources that make the process of price publication easier, in this case, also, some are already in the process of development. Integration with social networks and mobile devices; and to solicit other informations besides the product's price, are some examples that should potentialize the "sharing of the experience acquired by the individual in the buying decision process which involves the search and evaluation of alternatives and the decision making, evidencing more completeness in the knowledge to be shared" (Juliani, 2008).

7. FINAL CONSIDERATIONS

The evaluation of the knowledge sharing process in a system based on knowledge has been reached by means of the research tool's application, which made possible to collect the participants opinions, evidencing a high level of usability and predisposition by the users' individual traits to share knowledge, as well as confirming the influence of usability in this process, specially over the attributes Simplicity and Quickness to accomplish tasks in the software.
The evaluation also demonstrates that, as shown through the representative neutral score in the individual factors Recognition and Commitment, it's necessary to create resources capable of offering visibility to the individuals who share their knowledge with others through the system.
The open questions made it possible to collect suggestions of improvement, encouragements, errors, and showed that due to the major part of the participants having answered the questionnaire based on their first access to the system, some doubts related to the inexperience of using the software affected their answers.
Based on this study related to the high importance of individual factors and usability in the studied system, it's recommended to developers the conception of collaborative softwares directed to the characteristics of people who will use it, being easy to use, as well as making cyclical evaluations of the knowledge sharing process as a way to spot the deficiencies and identifying its real needs to potentialize the contribution of the users' knowledge.

8. BIBLIOGRAPHICAL REFERENCES

ABRAN, A., KHELIFI, A., SURYN, W., & Seffah, A. **Consolidating the ISO Usability Models**. Evaluation, 2007.

BASTIEN, J.M.C.; SCAPIN, D.L. **Human factors criteria, principles, and recommendations for HCI:** methodological and standardization issues. França: INRIA, 1993.

BLECKEN, A., BRUGGEMANN, D., & MARX, W. **Usability Evaluation of a Learning Management System**. Proceedings of the 43rd Hawaii International Conference on System Sciences.2010, pp.1-9.

CHUA, a. **Knowledge management system architecture:** a bridge between KM consultants and technologists. International Journal of Information Management, 2004, vol.24, no.1, p. 87-98. doi: 10.1016/j.ijinfomgt.2003.10.003.

CURRIE, G., & KERRIN, M. **The limits of a technological fix to knowledge management:** epistemological, political and cultural issues in the case of intranet implementation. Management Learning, 2004, vol.35, no.1, p.9–29.

CYBIS, W. de A. **Ergonomia de Interfaces Homem-Computador**. Apostila para o Curso de Pós-Graduation em Engenharia de Prodution, UFSC, 2000. Disponível em: <http://www.labiutil.inf.ufsc.br/apostila.htm>.

FELIPHE, L., & LAVOR, F. **Central de Projetos Calandra :** Usability em systems corporativos and a perspectiva da Engenharia de Prodution, 3º EBAI – Encontro Brasileiro de Arquitetura de Information 2008, p.1-20.

HAN, B. M., & ANANTATMULA, V. S. **Knowledge Management in IT Organizations From Employee's Perspective**. Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06), 2006. doi: 10.1109/HICSS.2006.243.

HENDRIKS, P. **Knowledge-based systems and knowledge management:** Friends or foes? Information & Management, 1999, vol.35, no.2, p.113-125. doi: 10.1016/S0378-7206(98)00080-9.

HAYES, N., & Walsham, G. **Safe enclaves, political enclaves and knowledge working**, Managing knowledge: Critical investigations of work and learning. 2000, Basingstoke: MacMillan.

ISO. **Normas ISO para Usability**. Disponivel em: http://www.iso.org/iso/catalogue_detail.htm?csnumber=52075. Data de acesso: 29/01/2011.

ISO 9241. **Ergonomic requirements for office work with visual display terminals** (VDTs). ISO 9241. Switzerland: ISO, 1998.

JORDAN, Patrick W. **An Introduction to Usability**. London, UK: Taylor & Francis, 1998.

JULIANI, Jordan. P. **A socialization de knowledge entre consumidores na busca de melhores alternativas de compra. Um modelo technological.** Tese (Doutorado em Engenharia and Gestão do Knowledge) – Programa de Pós-graduation em Engenharia and Gestão do Knowledge, University Federal de Santa Catarina, Florianópolis, 2008.

KAISER, S., KANSY, S., MUELLER, G., & RINGLSTETTER, M. **Weblogs for organizational knowledge sharing and creation:** a comparative case study**.** Knowledge Management Research & Practice, 2009, vol.7, no.2, p.120-130. doi: 10.1057/kmrp.2008.35.

KING, W., & MARKSJR, P. **Motivating knowledge sharing through a knowledge management system.** 2008, Omega, vol.36, no.1, p.131-146. doi: 10.1016/j.omega.2005.10.006.

KWAIFUNIP, R., & WAGNER, C. **Weblogging:** A study of social computing and its impact on organizations. Decision Support Systems, 2008, vol.45, no.2, p.242-250. doi: 10.1016/j.dss.2007.02.004.

LIAO, S. **Knowledge management technologies and applications—**literature review from 1995 to 2002. Expert Systems with Applications, 2003, vol.25, p.155-164. doi: 10.1016/S0957-4174(03)00043-5.

Lin, HSIU-FEN & Lee, H.-S. **Evaluation of factors influencing knowledge sharing based on a fuzzy AHP approach**. Journal of Information Science, 2008, vol.35, no.1, p.25-44. doi: 10.1177/0165551508091310.

LIN, HSIU-FEN. **Knowledge sharing and firm innovation capability:** an empirical study. International Journal of Manpower, 2007, vol.28, n.3/4, p.315-332.

LIN, M.-J. J., HUNG, S.-W., & CHEN, C.-J. **Fostering the determinants of knowledge sharing in professional virtual communities.** Computers in Human Behavior, 2009, vol.25, no.4, p.929-939. Elsevier Ltd. doi: 10.1016/j.chb.2009.03.008.

MACKINLAY, A. **The limits of knowledge management.** New Technology, Work and Employment, 2002, vol.17, no.2, p.76–88.

NAN, N. **A principal-agent model for incentive design in knowledge sharing.** Journal of Knowledge Management, 2008, vol.12, no.3, 101-113. doi: 10.1108/13673270810875895.

NEVES, R. M., Castanheira, D., Franco, C., Costa, R., Augusta, M., & Machado, S. **Uma Aplication da Matemática Nebulosa na Usability do Orkut.** 2006

NIELSEN, J. **Usability Engineering**. Boston, MA: Academic Press, 1993.

OSTERLOH, M. and FREY, B. **Motivation, knowledge transfer and organizational forms**. Organization Science, v. 11, n. 5, p. 538-550, 2000.

ORDAZ, C., CRUZ, G., & GINEL, E. "**Facilitadores de los processos de compartir conocimiento y su influencia sobre la innovación**" Cuadernos de Economía y Dirección de la Empresa. Núm. 42, marzo 2010, p.113-150, ISSN: 1138-5758.

QUESENBERY, W. **What does usability mean:** Looking beyond 'ease of use'. In: 48th Annual Conference Society for Technical Communication. Chicago, 2001.

SANTOS, B. S., Teixeira, L., Teixeira, B., Lucas, C., Pacheco, D., Cunha, J., et al. **Evaluation de usability de alguns aspectos de um serviço integrado no portal ua . pt : My . UA**. Revista do DETUA, 2007, vol.15, p.15-20.

SANTOS, R. C. D. **Desenvolvimento de uma metodologia para evaluation de usability de systems utilizando a lógica fuzzy baseado na iso**. Dissertation (Mestrado Profissionalizante em Administration) – Faculty De Economia and Finanças Ibmec Programa de Pós-Graduation and Research em Administration and Economia, Rio de Janeiro, 2007.

SHACKEL, B. **Ergonomics in design for usability**. In: Harrison, M. D., & Monk, A. F. People and computers: Designing for usability, In: HCI 86 Conference on People and Computer. New York: Cambridge University Press, p. 44-64, 1986.

SHNEIDERMAN, B. **Designing the User Interface**: Strategies for Effective Human- Computer Interaction. 3. ed. EUA: Addison Wesley, 1998.

SU, A. Y. S., Yang, S. J. H., HWANG, W.-Y., & ZHANG, J. **A Web 2.0-based collaborative annotation system for enhancing knowledge sharing in collaborative learning environments**. Computers & Education, 2010, vol.55, no.2, p.752-766. Elsevier Ltd. doi: 10.1016/j.compedu.2010.03.008.

TERRA, José Claudio. **Gestão 2.0:** como integrar a colaboration and a participation em massa para o sucesso nos negócios. Rio de janeiro: Elsevier, 2009.

TONET, H. C**. Knowledge sharing no trabalho :** o impacto das atitudes and da cultura organizacional, 2005.

TONET, Helena Correa; PAZ, Maria das Graças Torres da. **Um modelo para o knowledge sharing no trabalho**. *Rev. adm. contemp.* [online]. 2006, vol.10, n.2, pp. 75-94. ISSN 1982-7849.

TSENG, S. **The effects of information technology on knowledge management systems**. Expert Systems with Applications, 2008, vol.35, no.1-2, p.150-160. doi: 10.1016/j.eswa.2007.06.011.

TSENG, S.-M., & HUANG, J.-S. **The correlation between Wikipedia and knowledge sharing on job performance**. Expert Systems with Applications. 2010 Elsevier Ltd. doi: 10.1016/j.eswa.2010.11.009.

VORAKULPIPAT, C., & REZGUI, Y. **An evolutionary and interpretive perspective to knowledge management.** Journal of Knowledge Management, 2008, vol.12, no.3, p.17-34. doi: 10.1108/13673270810875831.

WANG, S., & NOE, R. A. **Knowledge sharing: A review and directions for future research**. Human Resource Management Review, 2010, vol.20, no.2, 115-131. Elsevier Inc. doi: 10.1016/j.hrmr.2009.10.001.

WINCKLER, M., & PIMENTA, M. S. **Evaluation de Usability de Sites Web**. World Wide Web Internet And Web Information Systems, 2002, p.1-54.

YE, S., CHEN, H., & JIN, X. **An Empirical Study of What Drives Users to Share Knowledge in Virtual Communities**, 2008, p.563-575.

Optimum Controller Design of an Overhead Crane

Kamal A.F. Moustafa
kamalam@uaeu.ac.ae

Khalifa H. Harib

and

Farag Omar

Department of Mechanical Engineering, United Arab Emirates University
Al-Ain, United Arab Emirates

ABSTRACT

The optimum control problem of an overhead crane with uncertain parameters is considered. The crane parameters, e.g., the payload mass, the cable length, hoisting velocity, hoisting acceleration are assumed to be interval quantities in the sense that they are known only within specified lower and upper bounds. The Monte Carlo technique is applied to guarantee the desired performance of the considered overhead crane model. The proposed method is used to find optimum values of the controller gains based on minimizing a least square error function. Simulation results are provided to show the usefulness of the suggested approach to control overhead cranes.

Keyword: Overhead crane, Optimum Control, Monte Carlo Simulation

1. INTRODUCTION

The design problem of an optimum controller for dynamic systems with uncertain parameters has recently attracted considerable attention of researchers [1-4]. It is required to incorporate robustness in the controller design of such systems to overcome the uncertainties inherent in the model parameters. In overhead crane industry, the underlying model is based on the nominal operation conditions. However, in actual operation, the mass of the payload, the hoisting cable length, hoisting velocity and acceleration are interval quantities in the sense that they are only known within a priori defined lower and upper bounds.

In a recent work [2], the authors developed a Monte Carlo based technique to synthesize an interval controller for an overhead crane with only one interval parameter; namely, the cable length. In this work, the developed technique will be extended to a wider class of overhead cranes where four parameters are uncertain. These parameters are the payload mass, the cable length, the hoisting velocity and the hoisting acceleration. The values of these parameters are assumed to be known only within specified lower and upper bounds that depend on a given operating conditions of the underlying crane. The design objective is to synthesize a Proportional-Derivative (PD) feedback controller to minimize the overshoot of the swing angle response of the payload regardless of the inherent uncertainties of the crane parameters. Simulation results are provided to show the usefulness of the suggested approach to control overhead cranes.

The Monte Carlo approach [5-7] is utilized to find optimum values of the controller gains based on minimizing a least square error function. In this method, a large number of simulation experiments are run by using randomly chosen values of the controller gains generated from a uniform distribution of specified lower and upper bounds as defined by the operating conditions. For each randomly

generated gain, the corresponding system variables of interest are simulated and the corresponding deviation from a desired trajectory is computed at each time instant. The optimum controller gain is then calculated as the value that minimizes the mean square error for all trajectories. Simulation results are provided to compare to illustrate the results.

2. MATHEMATICAL PRELIMINARIES

Interval analysis [8,9] is a method developed by mathematicians as an approach to putting bounds on mathematical computation and thus developing numerical methods that yield reliable results. Treatment is typically limited to real interval quantities defined as

$$[a, b] = \{x \in R, a \le x \le b\} \quad (1)$$

Instead of working with an uncertain real x we work with the two ends of the interval [a,b] which contains x: x lies between a and b, or could be one of them. Similarly, a function f when applied to x is also uncertain. In interval analysis, f produces an interval quantity which is all the possible values of f(x) for all $x \in [a, b]$.

To investigate the stability of an interval closed loop system, one needs to consider its interval characteristic polynomial, which is generally given by

$$\Delta(s) = [a_o, b_o] + [a_1, b_1]s + \cdots + [a_n, b_n]s^n + [1, 1]s^{n+1} \quad (2)$$

Two theorems that are recently proved in [1] give a necessary condition and a sufficient condition for stability of the interval polynomial (2). These theorems are reproduced below for the convenience of the reader.

Theorem 1. The interval polynomial defined in (2) is stable if the following necessary conditions are satisfied:

$$b_i \ge a_i > 0, \ \ i = 0, 1, 2, \dots, n$$
$$a_i a_{i+1} \ge b_{i-1} b_{i+2} > 0, \ \ i = 1, 2, \dots, n-2$$

Theorem 2. The interval polynomial defined in (2) is stable if the following sufficient conditions are satisfied:

$$b_i \ge a_i > 0, \ \ i = 0, 1, 2, \dots, n$$
$$0.4655 a_i a_{i+1} \ge b_{i-1} b_{i+2} > 0, i = 1, 2, \dots, n-2$$

The stability conditions of Theorems 1 and 2 can be applied to the closed loop polynomial of Equation (2). This produces inequalities in terms of the controller parameters that can be solved to obtain their lower and upper bounds so that stability of the underlying system is ensured. In the next section, an overhead crane model with interval parameters is considered to illustrate the results.

3. REDUCED ORDER INTERVAL MODEL

The model considered here is shown schematically in Fig. 1. It is a 3-degrees-of-freedom overhead crane model with a trolley mass m, and a payload material point with mass M hanging from a cable with length l [10-12]. The crane is driven by two actuating forces: the trolley force F_x and the hoisting force F_l. The generalized coordinates of the model are the trolley displacement x; the cable length l; and the payload swing angle ϑ.

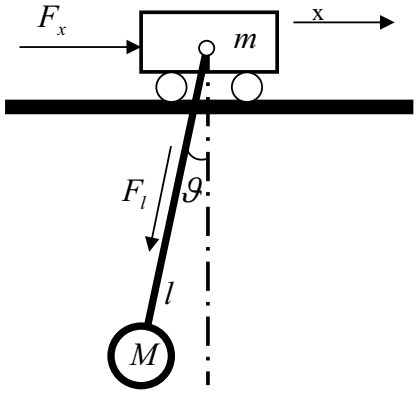


Fig.1. Schematic of the overhead crane

It can be shown that the completely controllable and observable part of the considered overhead crane model is the subsystem related to the dynamic behavior of the generalized coordinate $\delta\vartheta$ [13]. This subsystem is a 2nd order ordinary differential equation that can be written as

$$\begin{bmatrix} \delta\dot{\theta} \\ \delta\ddot{\vartheta} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\beta & -\alpha \end{bmatrix} \begin{bmatrix} \delta\vartheta \\ \delta\dot{\vartheta} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ -1 & \vartheta_o \end{bmatrix} \begin{bmatrix} \delta F_x^* \\ \delta F_h^* \end{bmatrix} \quad (3)$$

Where,

$$\beta = \frac{M}{m}\left[\frac{g}{l_o} - \frac{\ddot{l_o}}{l_o}\right] + \frac{g}{\ddot{l_o}} \qquad , \alpha = \frac{2\dot{l_o}}{l_o} \tag{4}$$

$$\delta F_x^* = \delta F_x/(ml_o),\ F_h^* = \delta F_h/(ml_o) \tag{5}$$

It is assumed that the parameters of the crane as well as the nominal operating conditions are interval quantities with known upper and lower bounds. In particular, the following interval quantities are defined, respectively, for the nominal operating cable length, velocity and acceleration

$$l_o \in [\underline{L},\ \overline{L}]\ ,\ \dot{l_o} \in \left[\underline{v},\ \overline{v}\right],\ \ddot{l_o} \in [\underline{a},\ \overline{a}] \tag{6}$$

where $\underline{x}$ and $\overline{x}$ are, respectively, the lower and upper bounds of the corresponding interval quantity. The payload mass is similarly defined as

$$M \in [\underline{M},\ \overline{M}\ ] \tag{7}$$

It should be noted that the bounds $\underline{v}$, $\underline{a}$ and $\underline{a_x}$ are negative real numbers and all other bounds are, on the other hand, positive real numbers.

Applying Theorems 1 and 2 of Section 2 to the characteristic polynomial of Equation (3) yields that the necessary and sufficient conditions for the reduced order overhead crane model to be stable is that the coefficients α and β must be positive values. However, it is obvious that there is no guarantee that they are positive during all operating conditions of the crane. Coefficient α is clearly negative when $\dot{l_o}$ is negative, i.e., when the load is being hoisted. The same argument is also valid for the coefficient β that could be negative if the hoisting acceleration is higher than the gravitational one. It is, therefore, necessary to employ an appropriate feedback control to guarantee the stability of the above model for all operating conditions. Such a controller should be robust even if the nominal operating cable length, hoisting/lowering velocity and acceleration are uncertain and known within given known intervals as discussed above. Moreover, the controller gains are to be selected in such a way to satisfy the desired performance characteristics. One main interest of this paper is to achieve this task by applying a Monte Carlo based approach to select the gains to minimize a function of the least square error. This method is a generalization of the technique recently developed by the authors in [2] where only the cable length is considered an interval quantity.

4. MONTE CARLO OPTIMUM CONTROL

In this section, an optimum controller is designed by applying the Monte Carlo technique [2,5-7] where a large number of simulation experiments are run by using randomly chosen values for the controller gain. These values are generated from a uniform distribution of specified lower and upper bounds as defined by the operating conditions. For each randomly generated gain, the corresponding swing angle trajectory is simulated and the corresponding deviation from a desired trajectory is computed at each time instant. The optimum controller gain is then calculated as the value that minimizes the mean square error of the swing angle for all trajectories. Therefore, the optimum controller gain is calculated as

$$k_{d-opt} = \min_k\{MSE\} \tag{8}$$

Where the mean square error is defined as

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\int_{t_o}^{t_f}[(\theta_i(t,k_{di}) - \theta_d(t)]^2 dt\ ] \tag{9}$$

In the above equation, $\theta_i(t,k_{di})$ is the swing angle trajectory corresponding to the i^{th} randomly generated controller gain k_{di}; $\theta_d(t)$ is the desired swing angle trajectory; n is the number of Monte Carlo experiments; and t_o and t_f are the initial and final simulation time, respectively.

It should be noted that, for nominal operation, the swing angle, ϑ_o, is usually kept small and the trolley mass, m, is relatively large. As a result, the contribution of δF_h to the control of $\delta\vartheta$, as seen from Equations (3) and (5) is almost negligible. The anti swing controller proposed in this paper will be based, therefore, on the actuation of the trolley force δF_x. Toward that end, the following feedback control scheme is proposed to eliminate the payload swing, $\delta\vartheta$, as fast as possible and without overshoot.

$$\delta F_x^* = \left[-\frac{Mg}{ml_o} + k_P\right]\delta\vartheta + k_D\delta\dot{\vartheta} \tag{10}$$

$$\delta F_h^* = 0 \tag{11}$$

Where k_P and k_D are, respectively, the proportional and derivative controller gains. These gains will be selected in such a way that the desired characteristics for the decay of the swing angle are achieved. Substituting Equations (10) and (11) in (3), one can obtain

$$\delta\ddot{\vartheta} + \left[\frac{2\dot{l_o}}{l_o} + k_D\right]\delta\dot{\vartheta} + \left[\frac{g}{l_o} - \frac{M\ddot{l_o}}{ml_o} + k_P\right]\delta\vartheta = 0 \tag{12}$$

Let the desired payload swing characteristics be define by

$$\delta\ddot{\vartheta} + 2\xi\omega_n\delta\dot{\vartheta} + \omega_n^2\delta\vartheta = 0 \tag{13}$$

Where ξ and ω_n are, respectively, the desired interval damping factor and interval natural frequency for the payload swing dynamics. Comparing Equations (12) and (13), the following interval equations must be satisfied to achieve the desired payload sway performance.

$$2\xi\omega_n = \frac{2\dot{l_o}}{l_o} + k_D \tag{14}$$

$$\omega_n^2 = \frac{g}{l_o} - \frac{M\ddot{l_o}}{ml_o} + k_P \tag{15}$$

It can be shown that $k_P = 0$ and $\xi = 1$ give the optimum performance with no overshoot [1]. Using Equations (12) and (13) above, one can derive the required interval derivative gain as

$$k_D = \left[\underline{k_D},\ \overline{k_D}\right] \tag{16}$$

Where the lower and upper bounds are

$$\underline{k_D} = 2\left[\left(\frac{\underline{g}}{\overline{L}} - \frac{\overline{M}\underline{a}}{m\,\underline{L}}\right)^{0.5} - \frac{\underline{v}}{\underline{L}}\right] \tag{17}$$

$$\overline{k_D} = 2\left[\left(\frac{\underline{g}}{\underline{L}} + \frac{\overline{M}|\underline{a}|}{m\,\underline{L}}\right)^{0.5} + \frac{|\underline{v}|}{\underline{L}}\right] \tag{18}$$

Practical considerations for crane operations require that the maximum lowering acceleration of the cable to be much less than the gravitational acceleration. Thus, the lower bound of the derivative gain of the controller as given by Equation (18) is guaranteed to be a real number in practice.

5. SIMULATION RESULTS

We consider an overhead crane with the lower and upper bounds of its parameters are given in Table 1.

Table 1. Crane Interval bounds

Variable	Lower Bound	Upper Bound
Cable Acceleration	-0.2 m/s2	0.2 m/s2
Cable Velocity	-0.5 m/s	1.2 m/s
Cable Length	1 m	20 m
Payload Mass	0 kg	500 kg
Trolley Mass	1500 kg	1500 kg

For the parameter values given in Table 1, a simulation run is carried out, for crane parameters and gain values randomly generated from a uniformly distributed population. The responses due to step initial disturbance for 50 runs are shown in Fig. 2. It can be observed that all runs are stable as expected. However, the overshoot of many of them is relatively large.

It is suggested, in this work, to use an optimum controller gain that corresponds to the minimum value of the mean square error of the swing angle. This optimum gain is found off-line by using a large number of runs, simulated using randomly generated crane parameters that cover the possible operating conditions. The error from the desired swing angle (zero) is then calculated for each run and the optimum gain is determined as 5.33. This optimum gain is then used on-line to control the crane under any operating conditions within the given lower and upper bounds of the parameter values. The simulation for the case of using the optimum gain is illustrated in Fig. 3. It is clear that all responses are stable with no overshoot as desired.

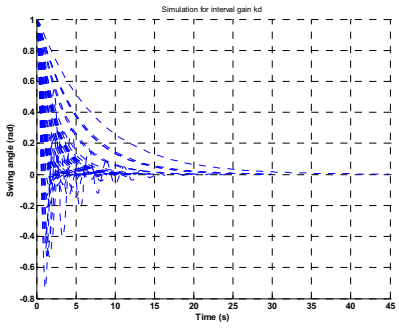


Fig. 2 Response for random crane parameters and controller gain

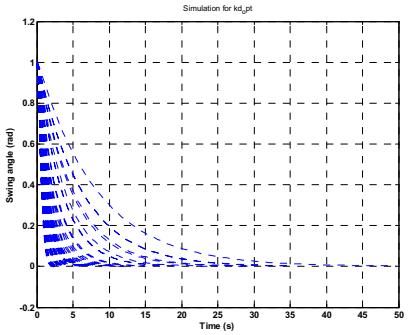


Fig. 3 Response for random crane parameters and optimum gain (k_d = 5.33)

6. CONCLUSIONS

The control problem of an overhead crane with uncertain parameter is considered in this paper. The parameters of the considered model are assumed to be interval quantities rather single valued ones. The Monte Carlo approach is applied to control the considered overhead crane model. The design objective is to synthesize a Proportional-Derivative (PD) feedback controller for a reduced order interval crane model such that the desired performance can be ensured despite of the parameter uncertainty. A simulation study is performed to illustrate the results. It is shown that the control performance of the payload swing angle is reasonable despite the presence of uncertain parameters.

7. REFERENCES

[1] Patre, B.M., and Deore, P.J., "Robust Stability and Performance for Interval Process Plants," **ISA Transactions**, Vol. 46, 2007, pp. 343–349.

[2] Moustafa, K.A., Omar, F., and Harib, K., "Control of Interval Parameter Overhead Cranes via Monte Carlo Simulation," **Transactions of the Institute of Measurement and Control**, Vol. 33, No. 2, 2011, P. 260-273.

[3] Chen, C.T., and Wang, M.D., "Robust Controller Design for Interval Process Systems," **Computers and Chemical Engineering**, Vol. 21, 1997, pp. 739-750.

[4] Chen, C.T., and Wang, M.D., "A Two Degrees of freedom Design Methodology for Interval Process Systems," **Computers and Chemical Engineering**, Vol. 23, 2000, pp. 1745-1751.

[5] Manno, I, **Introduction to the Monte Carlo Method**, Budapest, Hungary: Akadémiai Kiadó, 1999.

[6] Robert, C.P., and Casella, G., Monte Carlo Statistical Methods, 2nd ed, Springer Texts in Statistics, Springer Science & Business Media Inc., 2004.

[7] Gazi, E., Seider, W.D., and Ungar, L.H., "A Non-Parametric Monte Carlo Technique for Controller Verification," **Automatica**, Vol. 33, No. 5, 1997, pp. 901-906.

[8] Moore, R.E., Kearfott, R.B., and Cloud, M.J., "Introduction to Interval Analysis," **SIAM**, Vol. 10, 2009.

[9] L. Jaulin, L., Kieffer, M., Didrit, O., and Walter, E., **Applied Interval Analysis-with Examples in Parameter Estimation, Robust Control, and Robotics**, Spriger, London, 2001.

[10] Moustafa, K. A., and Ebeid, A. M., "Nonlinear Modeling and Control of Overhead Crane Load Sway," ASME Trans., **Journal of Dynamic Systems, Measurements and Control**, Vol. 110, No. 3, P. 266-271, 1988.

[11] Moustafa, K. A., and Abo-El Yazid, T., "Load Sway Control of Overhead Cranes with Load Hoisting Via Stability Analysis," Japan **Society of**

Mechanical Engineers Int. Journal (JSME), Series C, Vol. 39, No. 1. P. 34-41, March 1996.

[12] Moustafa, K.A., "Reference Trajectory Tracking of Overhead Cranes," **ASME Trans., Journal of Dynamic Systems, Measurement and Control**, Vol. 128, No. 1, P. 139-141, March 2001.

[13] Moustafa, K. A., "Feedback Control of Overhead Cranes Swing with Variable Rope Length," **Proc. 1994 American Control Conference (ACC)**, Maryland, USA, P. 691-695, June 1994.

COMPUTATIONAL SIMULATION OF THE FLOW PAST AN AIRFOIL FOR AN UNMANNED AERIAL VEHICLE

L. Velázquez-Araque[1] and J. Nožička[2]

[1] Division of Thermal fluids, Department of Mechanical Engineering, National University of Táchira, Venezuela.

[2] Department of Fluid Dynamics and Power Engineering, Faculty of Mechanical Engineering
Czech Technical University in Prague, Czech Republic.

ABSTRACT

This paper deals with the numerical simulation of the two-dimensional, incompressible, steady air flow past a NACA 2415 airfoil and four modifications of this one. The modification of this airfoil was made in order to create a blowing outlet with the shape of a step on the suction surface. Therefore, five different locations along the cord line for this blowing outlet were analyzed. This analysis involved the aerodynamic performance which meant obtaining lift, drag and pitching moment coefficients curves as a function of the angle of attack for the situation where the engine of the aerial vehicle is turned off called the no blowing condition by means computational fluid dynamics. The RNG k-ε model is utilized to describe the turbulent flow process. The simulations were held at a Reynolds number of 10^5. Results allowed obtaining lift and drag forces and pitching moment coefficient and also the location of the separation and reattachment point in some cases for different angles of attack, from 0 to 16 degrees with the smallest increment of 4 degrees. Finally, numerical results were compared with results obtained from wind tunnel tests by means of an aerodynamic balance and also oil and smoke visualization techniques and found to be in very good agreement.

INTRODUCTION

In the Laboratory of Fluid Mechanics and Thermodynamics of the CTU in Prague, an unmanned aerial vehicle (UAV) with an internal propulsion system is being developed. In order to acomplish this main objective, all the components must be designed. This paper is part of the development of an airfoil for a UAV with internal blowing propulsion system for the gliding condition.

The main motivation of this research is the validation of experimental results obtained in wind tunnel tests of the aerodynamic characteristics by means of an aerodynamic balance as well as the flow field by oil and smoke flow visualization techniques. The analysis of the air fluid flow past an airfoil from the NACA 4 digits family and four modified models is performed by means of obtaining lift and drag forces and pitching moment coefficient and also the location of the separation and reattachment point in some cases for different angles of attack. Then an exhaustive comparison to the experimental results is performed. The whole process is described in the following sections.

AIRFOILS TESTED

A NACA 2415 airfoil (Figure 1), which has become increasingly popular on ¼ scale pylon racers [1] was tested and also four modifications of this one. The modification is based mainly on the creation of an abrupt step on the suction side of the original NACA 2415 airfoil.

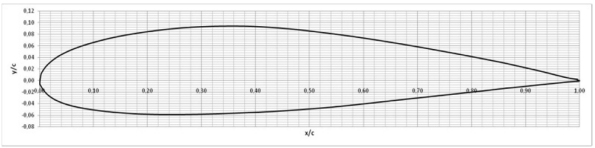


Figure 1: *NACA 2415 airfoil*

This step simulates a blowing propulsive outlet of the wing in normal flight conditions. Four different configurations where designed which involved the location of the step at different strategic points chordwise (Figure 2). These points are:

- At the location of the maximum thickness: 30% of the chord. (2415-3).
- At the location of the maximum camber: 40% of the chord. (2415-4).
- Before the transition point (at 0 AOA): 50% of the chord. (2415-5).
- Passed the transition point (at 0 AOA): 60% of the chord. (2415-6).

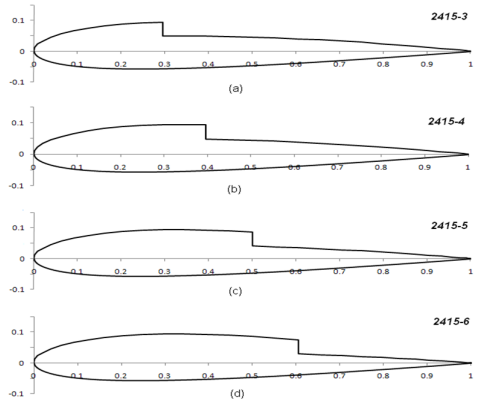


Figure 2: *Airfoils developed for testing (a) 2415-3, (b) 2415-4, (c) 2415-5, (c) 2415-6.*

COMPUTATIONAL DOMAIN

Something very important in this part is the choice of the domain, because it is formed by real borders such as the upper and lower surfaces of the airfoil and also by imaginary borders which enclose the external environment. The domain extends from 8 chords lengths upstream to 20 chord lengths downstream according to [2] an also 8 chord lengths for the upper and lower heights. The fluid flow which is simulated is air past five different airfoils with a Reynolds number of 10^5. These five airfoils correspond to the NACA 2415 and the four modifications with the step at 30, 40, 50 and 60 percent of the chord length. In Figure 3 it is possible to see the geometry of the domain for the airfoil 2415-3 as an example.

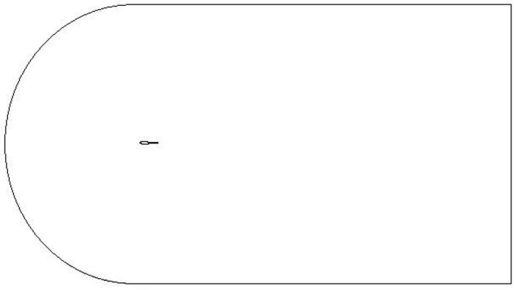


Figure 3: *Computational domain for the numerical simulations*

DISCRETIZATON OF THE DOMAIN

The geometry shown in Figure 3 is discretized using a structured mesh of 188 x 200 tetrahedral elements, this mesh has been also supplemented with very small elements in the vicinity of the surface of the airfoil forming a boundary layer with a grow factor of 1,2. References when creating the mesh were followed in [3], therefore the created mesh had a size change of 2,66 and an equisize skew of 0,348.

The domain and the mesh were created using the commercial software GAMBIT, version 2.3. In order to obtain the lift and drag as a function angle of attack, single meshes were created for 0, 4, 12 and 16 degrees and for every airfoil, thus there were created a total of 20 meshes (Figure 4).

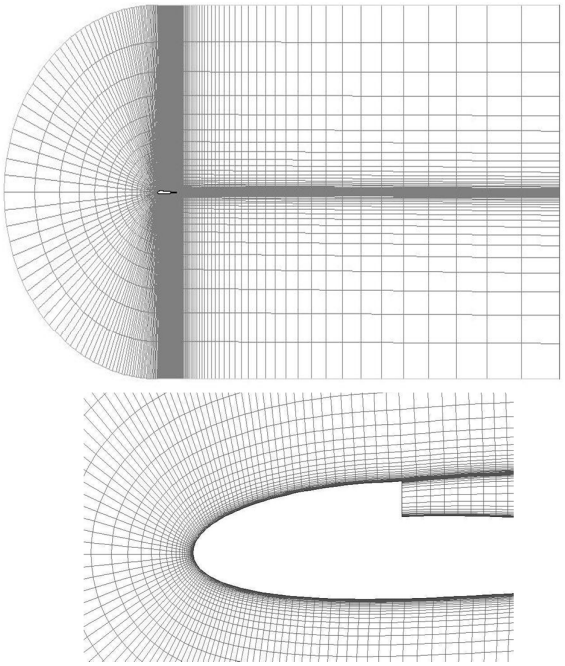


Figure 4: *A mesh used for the numerical simulation.*

Then, from the governing equations, the discretization of the domain and using the finite volume method based on finite elements, a discrete set of algebraic equations is set which solution is obtained as coupled, iteratively, using the commercial solver ANSYS FLUENT, version 12.0 using a scheme of second order upwind.

TURBULENCE MODEL

The *k-ε* model is derived from the Navier-Stokes equations and it is one of the simplest complete models of turbulence with two-equation models in which the solution of two separate transport equations allows the turbulent velocity and length scales to be independently determined. The standard *k-ε* model in ANSYS FLUENT falls within this class of models and has become very used for practical engineering flow calculations. It is a semi-empirical model. It is robust, economic, and presents reasonable accuracy for a wide range of turbulent flows.

The chosen turbulence model was the RNG *k-ε*. The RNG (renormalization group theory) is an improvement of this model of turbulence because it provides an analytically derived differential formula for effective viscosity that accounts for low-Reynolds-number effects. Therefore it is more accurate and reliable for a wider class of flows.

BOUNDARY CONDITIONS

At the inlet it is specified the air absolute velocity magnitude and also its components; in this case the velocity is parallel to the horizontal axis, therefore it does not have any component in the ordinates. Concerning turbulence, it was also specified the turbulence intensity of 1,3 % in accordance to [4] and also the turbulent length scale. The upper and lower surfaces of the airfoil are set as walls. At the outlet it is specified the pressure as the atmospheric pressure. For the lateral walls of the domain they are set as symmetry.

GOVERNING EQUATIONS

Since this problem does not involve heat transfer nor compressibility the equation for energy conservation is not required, therefore the most important equations such as conservation of mass and momentum used by the software's solver are listed as follows:

Continuity equation:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{v}) = 0 \qquad (1)$$

Conservation of momentum in a non-accelerating reference frame:

$$\frac{\partial (\rho \vec{v})}{\partial t} + \nabla \cdot (\rho \vec{v} \vec{v}) = -\nabla p + \nabla \cdot (\bar{\bar{\tau}}) + \rho \vec{g} + \vec{F} \qquad (2)$$

where p is the static pressure, $\rho \vec{g}$ and $\vec{F}$ are the gravitational and external body forces and $\bar{\bar{\tau}}$ is the stress tensor which is described as:

$$\bar{\bar{\tau}} = \mu \left[(\nabla \vec{v} + \nabla \vec{v}^T) - \frac{2}{3} \nabla \cdot \vec{v} I \right] \qquad (3)$$

where μ is the dynamic viscosity, I is the unit tensor, and the second term on the right hand side is the effect of volume dilation.

Due to the RNG *k-ε* model was selected for the problem, the transport equations for *k* and *ε* are described.

$$\frac{\partial(\rho k)}{\partial t}+\frac{\partial(\rho k u_i)}{\partial x_i}=\frac{\partial}{\partial x_j}\left(\alpha_k \mu_{eff}\frac{\partial k}{\partial x_j}\right)+G_k+G_b-\rho\varepsilon-Y_M+S_k \quad (4)$$

$$\frac{\partial(\rho\varepsilon)}{\partial t}+\frac{\partial(\rho\varepsilon u_i)}{\partial x_i}=\frac{\partial}{\partial x_j}\left(\alpha_\varepsilon \mu_{eff}\frac{\partial \varepsilon}{\partial x_j}\right)+C_{1\varepsilon}\frac{\varepsilon}{k}(G_k+C_{3\varepsilon}G_b)-C_{2\varepsilon}\rho\frac{\varepsilon^2}{k}-R_\varepsilon+S_\varepsilon \quad (5)$$

where *k* is the specific turbulence kinetic energy and it is defined as the variation in the velocity fluctuations; it has units m^2/s^2. *ε* is the turbulence dissipation of small vortices (eddies), in other words, the rate at which the velocity fluctuations are dissipated, its units are m^2/s^3.

Likewise, G_k represents the generation of turbulence kinetic energy due to the mean velocity gradients. G_b is the generation of turbulence kinetic energy due to buoyancy, Y_M represents the contribution of the fluctuating dilatation in compressible turbulence to the overall dissipation rate. α_k and α_ε are the inverse effective Prandtl numbers for *k* and *ε* respectively. S_k and S_ε are user-defined source terms.

CONVERGENCE CRITERIA

The convergence criteria selected for this problem was the recommended by the software, it is 10^{-3} for all the scaled residuals, however the convergence checking was deactivated because the drag and lift monitors were activated, therefore the convergence was achieved when the values of C_D and C_L remained constant for a minimum of 1000 iterations.

RESULTS AND ANALYSIS

Figures 5 - 8 show numerical C_L versus AOA, C_D versus AOA, C_L versus C_D and C_M versus AOA for all models tested experimentally, including the original NACA 2415 airfoil.

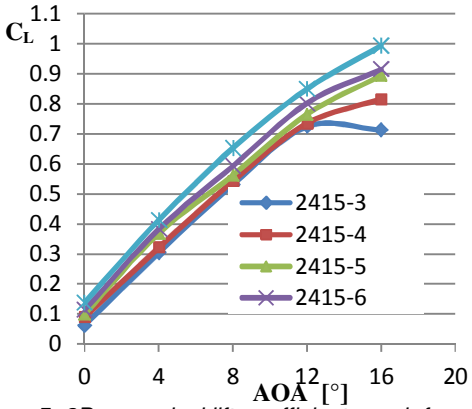


Figure 5: *2D numerical lift coefficient graph for all airfoil models tested.*

In Figure 5 it is possible to see the lift coefficient as a function of AOA, the values of C_L for all AOA were obtained with CFD software. As expected the highest lift slope corresponds to the original NACA 2415 airfoil and then, it is decreasing as the position of the step moves towards the leading edge. All slopes seem approximately straight up to 12 degrees of AOA because the minimum AOA displacement was 4 degrees. The stall point is only clear for the 2415-3 airfoil at 13 degrees of AOA; the other airfoils present a soft decreasing of the slope from 12 degrees of AOA.

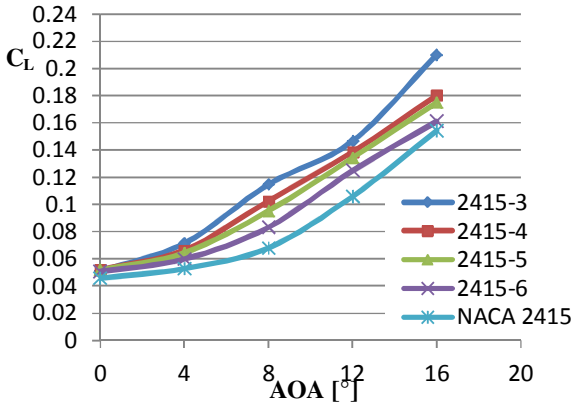


Figure 6: *2D numerical drag coefficient graph for all airfoil models tested.*

In Figure 6 it is possible to see the drag coefficient as a function of AOA; the values of C_D for all AOA were obtained with CFD software. All curves begin at a common point for zero AOA approximately of 0,05 C_D. After this point each curve follows its pattern and the values of C_D increase with the increment of the AOA. Among the curves, the NACA 2415 airfoil presented the lowest values of drag as expected followed by the 2415-6 and so on. This shows that the drag increases as the position of the step moves towards leading edge.

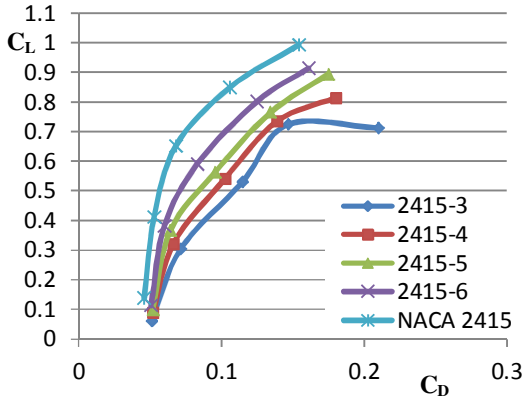


Figure 7*: 2D numerical polar graph for all airfoils model tested.*

Figure 7 shows the polar graph, it was also possible to obtain the Optimum Glide Ratio based on the numerical results:

- NACA 2415: OGR = 9,429.
- 2415-6: OGR = 6,944.

- 2415-5: OGR = 6,133.
- 2415-4: OGR = 5,580.
- 2415-3: OGR = 5,090.

According to this the curves it is possible to notice that the minimum drag is similar for all airfoils at 0 degrees of AOA of approximately 0,05, however, the maximum lift is obtained by the NACA 2415 of 0,993 as expected followed by the 2415-6 airfoil which presented a maximum lift of 0.914 the other maximum values of C_L can be seen in detail in Figure 7 and these ones decrease as the step moves towards the leading edge.

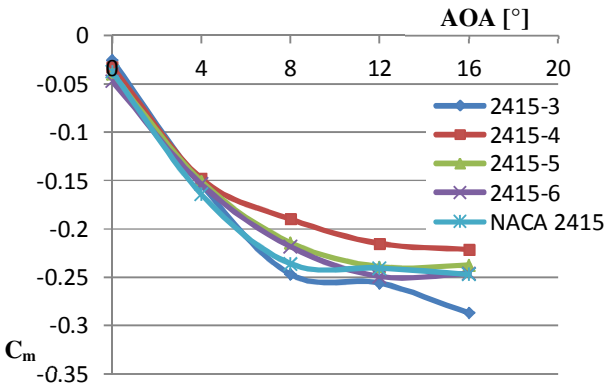
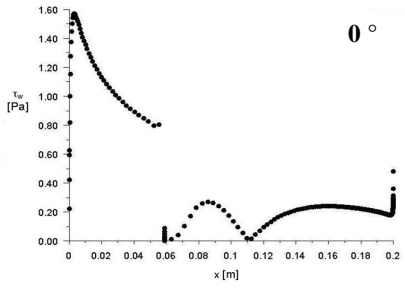


Figure 8: *2D numerical pitching moment graph for all airfoils models tested.*

The numerical pitching moment coefficient was obtained with CFD software; it is computed with respect to the leading edge for several values of AOA (Figure 8). In this graph it is possible to observe that all airfoils tested presented a very similar behavior between 0 and 4 degrees of AOA, from this point the 2415-4 airfoil presents the lowest values, followed by 2415-5, 2415-6 and NACA 2415 which are very small. The 2415-3 airfoil presents the highest values of pitching moment, however this values are not so high compared to the other airfoils.

In Figure 9 it is possible to see the numerical wall shear stress on the suction surface along the chord line for the NACA 2415-3 airfoil tested from 0 to 16 degrees of AOA which allows observing points of separation and reattachment of the flow. Likewise it was possible to obtain pictures for all of the other models.



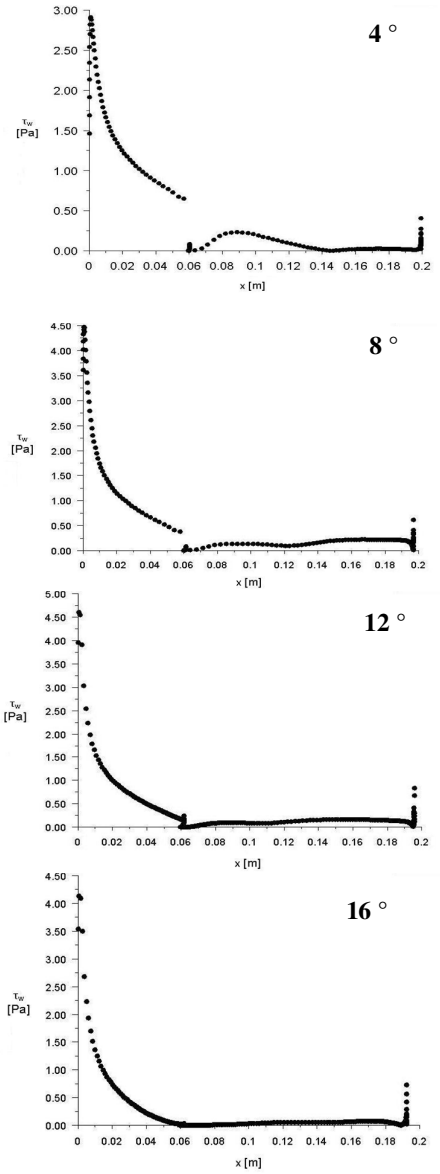


Figure 9: *Wall shear stress of the 2415-3 airfoil for different AOA.*

A shear stress is applied parallel or tangential to a face of a material. Any real fluids (liquids and gases included) moving along a solid surface will incur a shear stress on that surface. That is the reason why the wall shear stress is considered an indicative of separation of flow because when it is equal to zero, it means that the flow is not attached to the surface of the airfoil. After this point, values of shear stress are different of zero and the separation region begins. In the case of reattachment of flow, it is noticed when the values of wall shear stress reach zero again, and the area between these two points is the separation region, in this region, the values of wall shear stress are negative, this can be seen if only the x-component of the wall shear stress is plotted but for a better observation, it was decided to plot the resultant wall shear stress, where all values are always positive.

In Figure 9 it can be seen that the flow detaches at the location of the step for the 2415-3 airfoil for all AOA. It is presented as an abrupt fall in the wall shear stress curve until zero; however, for all other modified airfoils analyzed, this behavior is present until an AOA of 12 degrees; at 16 degrees of AOA the separation point is located before the step. Concerning reattachment of flow, it is observed for AOA between 0 and 4 degrees, after the step. At higher AOA there is not reattachment of flow. The NACA 2415 airfoil presented a separation point for an AOA of 4 degrees located at 95% of the chord and while the AOA was increasing, this separation point was moving towards the leading edge until reaching 30% of the chord for 16 degrees of AOA, this is important because it explains why the separation of flow for the modified airfoils begins to be present before the step for an AOA of 16 degrees.

In Figure 10, it is possible to observe the flow field as velocity contours of the air flow past the NACA 2415-6 airfoil tested from 0 to 16 degrees of AOA.

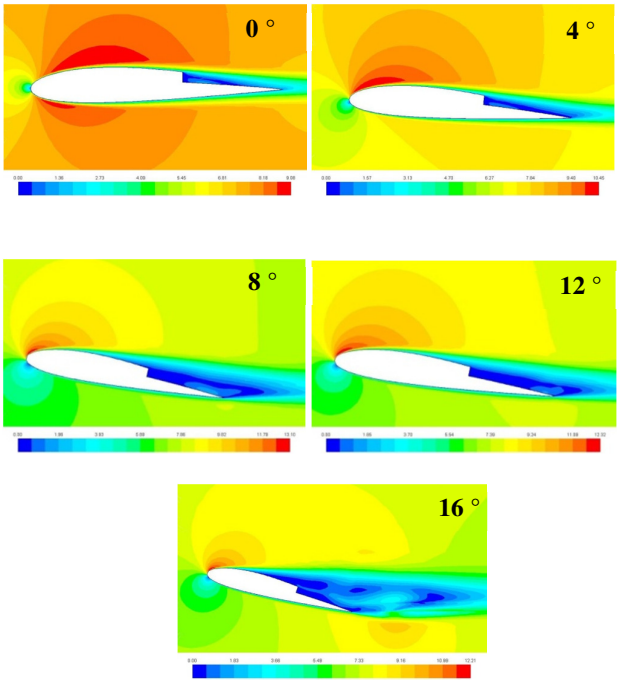


Figure 10: *Wall shear stress of the 2415-6 airfoil for different AOA.*

In Figure 10 it is possible to see the first numerical graphical approach to the behavior of the air flow past the tested airfoils. Here we can observe how the velocity changes in the selected domain; in this case the most important is to observe this phenomenon near the surface of the model. However these pictures do not show clearly the separation and reattachment points. The 2415-6 airfoil presents the biggest regions of high speed for all AOA and the reason is because the step is located closer to the leading edge so that the flow is attached to the airfoil's surface for a longer distance than the other modified models. On the contrary, the 2415-3 airfoil presents the smallest regions of high velocity for all AOA this and therefore the biggest regions of separation of flow for all AOA this incurs a higher drag compared to other tested models.

In Figure 11 it is possible to see the streamlines of the flow past the 2415-4 airfoil.

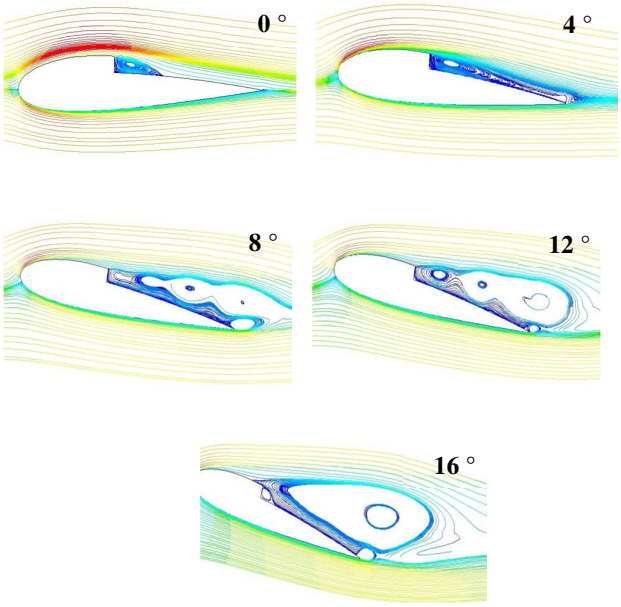


Figure 11: *Streamlines of the 2415-4 airfoil for different AOA.*

It is observed that the flow is fully attached to the suction surface of the airfoil until the step where separation of flow occurs, this phenomenon occurs for all AOA, the spatial extension of the separation region can be detected by exploring the wall shear stress along the surface of the airfoil (Figure 9). Inside this region, it is possible to observe that the adverse pressure gradient causes a reversed flow and this becomes into a counter-rotating vortex. Then the flow reattaches again and remains in contact with the surface until the trailing edge, this reattachment was observed in all modified airfoils from 0 to 4 degrees of AOA.

In Figure 11 for the 2415-4 airfoil, a very interesting phenomenon occurs at 16 degrees of AOA, a small induced vortex appears just next to the step inside the big separation region which begins upstream.

For the NACA 2415 airfoil, the streamlines remain attached along the whole surface of the airfoil until 8 degrees of AOA where a small detachment is observed very close to the trailing edge. As the AOA increases, this separation region begins more upstream. For the highest AOA, a big counter-rotating vortex is observed within the separation region.

ANALYSIS AND DISCUSSION

This section is devoted to different comparisons between obtained experimental and numerical results.

Comparison to force and moment coefficients obtained by wind tunnel tests.

Concerning the lift coefficient, experimental and numerical results are in good agreement; it is possible to see in Figure 12, where only two airfoils have been included on purpose for a better appreciation, that the differences are very small, for the case of the 2415-3 airfoil, the stall point could be seen clearer in the experimental results because this method had a smaller increment of the angle of attack [5].

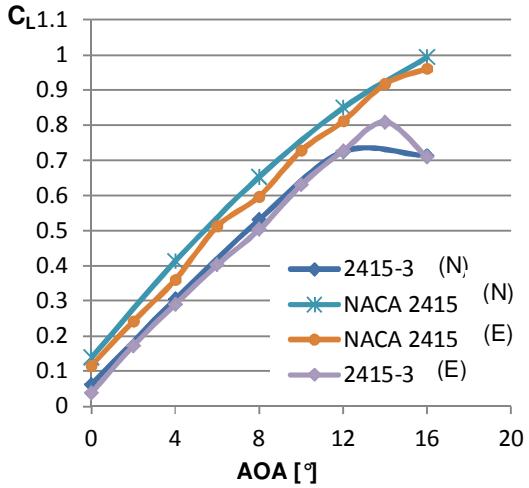


Figure 12: *Numerical and experimental lift coefficient for two airfoils tested.*

The behavior of the other airfoils is pretty similar and for that reason the curves were omitted.

Concerning the drag coefficient, experimental and numerical results are similar, however some discrepancies are present. In Figure 13 are shown the most representative cases of those discrepancies, only two airfoils have been included on purpose for a better appreciation. It is possible to see that in general, the numerical values for drag coefficients resulted slightly lower than the experimental ones. Since the method of computing forces used by the software consists in summing the dot product of the pressure and viscous forces on each face with the specified force vector, in this case the force is parallel to the flow direction, only abscissas, therefore, the theory of the software which predicts the force and then the coefficient does not seem very accurate.

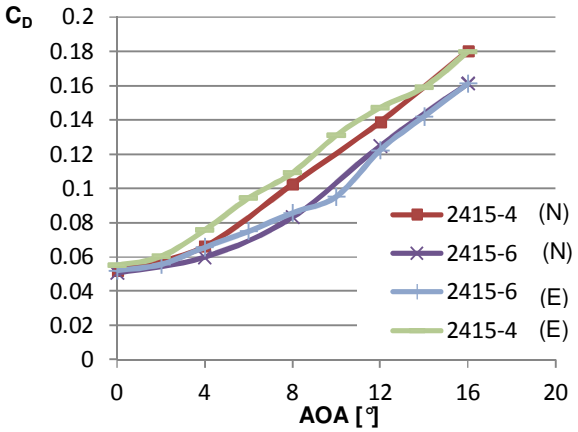


Figure 13: *Numerical and experimental drag coefficient for two airfoils tested.*

Concerning the pitching moment coefficient it is possible to observe in Figure 14 that experimental and numerical results present significant differences, for instance in these two cases experimental results are lower than numerical ones and so on for the rest of the models tested.

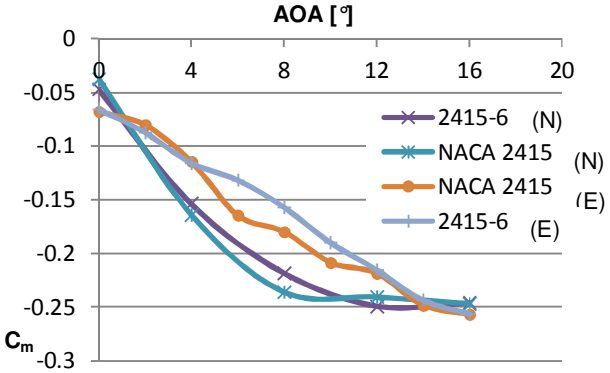


Figure 14: *Numerical and experimental pitching moment coefficient for two airfoils tested.*

The theory of the software which predicts the pitching moment and then its coefficient consists in summing the cross products of the pressure and viscous force vectors for each face with the moment vector, which is the vector from the specified moment center to the force origin. Based on this it can be said that these differences could be due to possible inaccuracy in the measurements with the wire balance. For a clearer determination of these discrepancies it would be necessary performing these measurements with another type of balance and compare the results [5].

Comparison to results obtained by experimental oil and smoke visualization of flow.

According to experimental results reviewed in [6], numerical results are quite in good agreement. For instance, in Figure 15 it is possible to see a larger view of the oil visualization for the 2415-3 airfoil at 0 degrees

AOA, the values of points of separation an attachment are in very good agreement with numerical ones which are shown in Figure 16 for the same airfoil at 0 degrees of AOA.

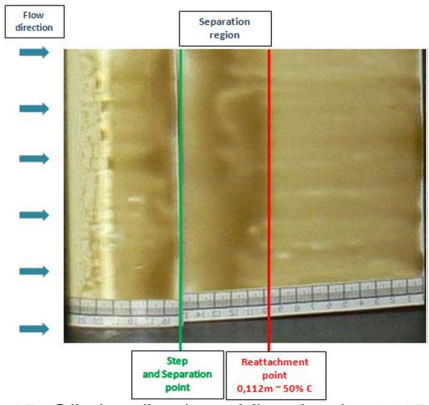


Figure 15: *Oil visualization of flow for the 2415-3 airfoil at 0 ° of AOA.*

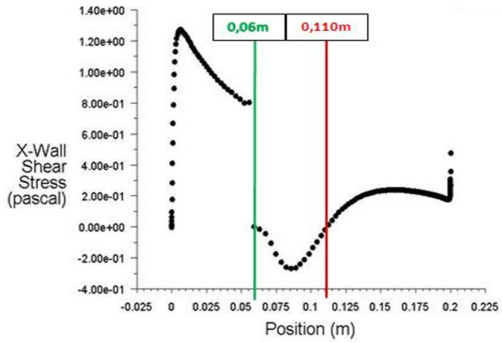


Figure 16: *X-Wall shear stress for the 2415-3 airfoil at 0 ° of AOA.*

Likewise streamlines obtained numerically were compared to smoke visualization pictures reviewed in [6]. Figures 17 and 18 are shown as an example to confirm the good agreement between numerical and experimental results.

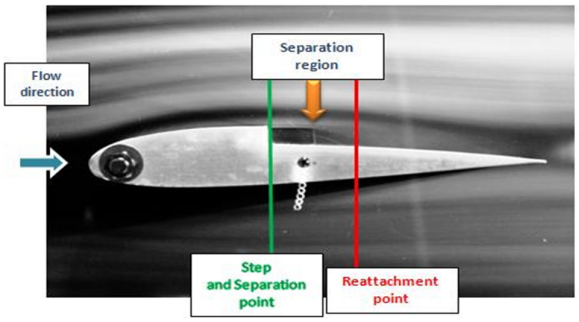


Figure 17: *Smoke visualization of flow for the 2415-4 airfoil at 0 ° of AOA.*

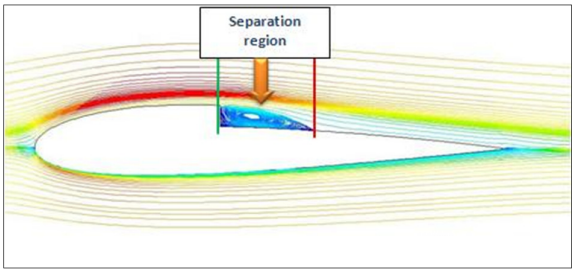


Figure 18: *Streamlines for the 2415-4 airfoil at 0 ° of AOA.*

CONCLUSION

By means of the use of CFD it has been possible to obtain lift, drag and pitching moment coefficients and also the flow field of air past an original NACA 2415 airfoil and four modifications of this one. It was also possible to obtain the location of separation and reattachment points in some cases for different angles of attack which made possible the analysis of the influence of the location of the propulsing outlet along the chord line, turning out that for the non-blowing condition the aerodynamic performance of the airfoil increases as the propulsing outlet moves towards the trailing edge. The validation of the results has been performed through an exhaustive comparison to experimental obtained results for forces and moments by means of wind tunnel tests and separation and reattachment points by means of oil and smoke visualization having found them in good agreement.

REFERENCES

[1] Selig M.S., Lyon C.A., Giguere P., Ninham C., Guglielmo J.J.: *Summary of Low- Speed Airfoil Data*, Vol. 2, Viginia Beach, SoarTech Publications, 1996.

[2] Malan P., Suluksna K., Juntasaro E.: *Calibrating the γ-Re_θ Transition Model for Commercial CFD.* Proceedings of The 47th AIAA Aerospace Sciences Meeting, 2009.

[3] Rhie C.M., Chow W.L.: *Numerical Study of the Turbulent Flow Past an Airfoil with Trailing Edge Separation,* AIAA Journal, Vol. 21, No. 11, 1983.

[4] http://profily.fs.cvut.cz/, Laboratory of Department of Fluid Dynamics and Power Engineering of the CTU in Prague. Section Airfoils and Straight Blade Cascades.

[5] Velazquez L., Nožička J., Vavřín J.: *Experimental Measurement of the Aerodynamic Characteristics of Two-Dimensional Airfoils for an Unmanned Aerial Vehicle,* Proceedings of Experimental Fluid Mechanics 2010. Liberec, 2010.

[6] Velazquez L., Nožička J, Kulhanek R.: *Oil and Smoke Flow Visualization past Two-Dimensional Airfoils for an Unmanned Aerial Vehicle,* Proceedings of The 11th Asian Symposium of Visualization. Niigata, Japan. 2011

Barriers to Collaboration Systems Effectiveness

Dr. T. Herring, Dr. R. Cherinka, and J. Prezzama
The MITRE Corporation
4830 W. Kennedy Blvd., Tampa, FL 33609
Phone: 813-287-9457, Fax: 813-287-9540
therring@mitre.org, rdc@mitre.org, prezzama@mitre.org

ABSTRACT

A major transformation within the Department of Defense (DoD) occurred at the onset of combat operations in the Middle East. The premise of the transformation was to restructure and equip military forces with technological capabilities to enhance information sharing and situational awareness toward the goal of achieving information superiority. To date, few studies have been conducted to determine the level of effectiveness of collaboration systems employed within the operational forces of the DoD. As a result, a case study was conducted consisting of exploratory research to identify the primary components of communication and how they affect collaboration systems employed in support of military operations. Specifically, the focus of my corporation's research was to better understand the factors that contribute to and the barriers that prohibit collaboration systems effectiveness.

Keywords: Transformation, collaboration, information sharing, communication theory.

1. INTRODUCTION

In light of the modern warfare tactics in Iraq and Afghanistan, elements of the Department of Defense underwent a major transformation to create a modular, scalable, and modernized military force. This initiative involved incorporating new concepts, organizational structures, and technologies to modernize military forces. However, after 6 years of sustained combat operations in the Middle East many of the collaboration systems employed are estimated to be only 75% effective and fall short of the required 99% level of service [1]. Collaboration systems are defined as information sharing tools specifically designed to access, store, and disseminate voice, data, imagery, and video [2]. In this paper, we discuss efforts to assist our DoD sponsors to understand and address many of the challenges associated with rapid integration of new collaboration systems employed in support of sustained combat operations in remote austere environments. In particular, we address barriers to collaboration systems effectiveness as it relates to command and control systems employed within the DoD's Operational Forces. In addition, we intend to provide a deeper understanding of communications theory through the evaluation of data collected from subject matter experts.

2. BACKGROUND

Immediately following the attacks on September 11, 2001, the most significant reorganization of military forces since the Cold War began within the DoD. This initiative involved incorporating new concepts, organizational structures, and technologies to modernize military forces. The premise for transformation was to restructure and equip military forces to account for the significant increase in humanitarian relief efforts and to effectively engage in a prolonged war against an irreconcilable and adaptive adversary. The key emphasis during the transformation process was the reorganization of military forces and the fielding of new communications equipment. The fundamental concept behind the transformation within the DoD was based on the emerging theory of network centric warfare. This concept is aligned with the belief that by leveraging modern technology military commanders can gain a competitive advantage over an adversary.

Vice Admiral (Ret.) Arthur K. Cebrowski, Director of the Office of Force Transformation, stated that in the information age military power comes from the ability to access and disseminate information (Office of Force Transformation, 2005). This statement was the premise for the DoD's transformation to restructure and equip organizations with new collaboration systems to enhance information sharing (Office of Force Transformation, 2005). However, as previously stated, many of the collaboration systems employed are estimated to be only 75% effective and fall short of the required 99% level of service. Thus, the problem addressed in this study was the lack of collaboration system effectiveness employed in support of combat operations.

3. THEORETICAL FRAMEWORK

The theoretical foundation for this study is focused on communications theory. Communications theory provides a means for understanding the basic transmission of information between two entities [3]. In the simplest form, the communications process consists of the following components: sender, message, and receiver [4]. The major source of contention found in the field of communications theory stems from whether or not the available models address all the components and influential factors identified with the communications process. Another challenge associated with communications theory is that no one model provides a holistic perspective of the various types of communications. Early models portray the communications process in a simple form which is easily understood. However, modern models are tailored to specific forms of communication and are displayed in a complex array of processes which may be difficult to comprehend [5]. Thus, one of the goals of this study was to provide a deeper understanding of communications theory by examining how communication is effected by the integration of complex collaboration systems.

4. Research Methodology

This study was aligned with a phenomenological qualitative research design to explore specific components of communication and how they affect collaboration systems. Phenomenological research involves collecting data on human experiences based on a specific phenomenon in order to develop an understanding and derive deeper meaning [6]. Thus, the purpose of a qualitative study was to develop a more thorough understanding of a phenomenon based on collecting data on human experiences in a prescribed social environment [7]. Furthermore, exploratory research was incorporated into the design to provide clarity and define the nature of the problem being studied [8]. In this case, the research design facilitated the collection data from 15 subject matter experts to develop a deeper understanding of the effectiveness of collaboration systems. Purposeful sampling was used to select participants who have operational experience using collaboration systems in a remote/austere environment. This approach was necessary since no foundational data exists on the topic. The data collection strategy consisted of utilizing semi-structured interview questions designed to extract data from the participants. Individual interviews were conducted to collect data on the experiences of select individuals reference a phenomenon of interest. The measurement of the qualitative data was portrayed through the analysis of text and was further refined by the categorization of major topics or concepts.

5. CASE STUDY HIGHLIGHTS

In the analysis of the results of the study, several key observations can be made. Subject matters experts identified the following communication barriers with employing collaboration systems in combat operations: system interoperability, security domains, bandwidth limitations, duplication, austere environments, power, and training.

- Systems interoperability challenges were reported with both hardware and software configurations. Specifically, specialty cables had to be manufactured to enable physical connectivity between critical systems. Software interoperability stemmed from the use of incompatible software versions and implementation of proprietary applications.

- The use of multiple domains was identified as a significant hindrance to collaborative efforts by force the manual transfer of data between collaboration systems.

- Bandwidth limitations were encountered due to lack of satellite availability, competing requirements, and network infrastructures. As a result, mitigation measures were emplaced which limited the size of email attachments, prioritized information sharing, limited access to external websites/portals.

- Duplication in systems, supporting network architectures, and data storage degraded information sharing between forces. In each case, unintentional latency was encounter in determining the appropriate collaborative system to use or the location in which data was stored.

- Operating in an austere environment led to partial and complete outages due to extreme heat and high winds.

- Power issues primarily resulted from a lack of backup power source and/or inadequate commercial power availability.

- Training deficiencies were also cited as contributors to collaboration systems effectiveness.

These barriers are consistent with observed challenges found within our current customer base. For example, as part of this study, participants cited that bandwidth limitations prevented them from sharing critical

information with external entities in a timely manner. These limitations were caused by a myriad of factors such as lack of satellite availability, competing requirements, and costs. Outside of this study, these same limitations have been observed to varying degrees within our current customer base that operate in remote/tactical communications environments.

This study was also used to collect data on mitigation measures implemented to overcome known barriers to collaboration systems effectiveness. Based on the data collected, the following mitigation measures were identified: training, network management, redundancy, and standard operating procedures.

- Training was conducted as part of pre-deployment operations and for sustainment.
- Network management included regular scans, monitoring, and bandwidth management
- Redundancy systems and networks were employed to mitigate the potential for a complete outage.
- Standard operating procedures were documented and enforced.

The findings with this study were in line with the other studies conducted on specific combat operations. For example, one study found collaboration challenges as a result of the employment of collaboration systems using proprietary solutions and multiple security domains [9]. Furthermore, in a 2007 report to congress, collaboration challenges such as bandwidth limitations, systems interoperability, information management and use of multiple security domains was cited as having adverse effects on collaboration [10]. In contrast, this study led to the discovery of additional communication barriers such as duplication, austere environment, power outages and lack of training. Furthermore, the data collected as part of this study provided further insight into criticisms cited with existing communications models, specifically as they relate to communications barriers.

6. IMPLICATIONS

The implications associated with this study revealed a common theme: lack of understanding of the adverse effects of collaboration barriers.

The aggregate responses from participants resulted in the identification of critical barriers to collaboration systems effectiveness. However, the magnitude of the adverse effects of each of these barriers have on collaboration systems effectiveness is unknown. In addition, some of the mitigation measures identified in this study are cited as barriers. Thus, further research is warranted to better understand barriers to collaboration systems effectiveness as well as adverse any effects of mitigation measures.

7. RECOMMENDATIONS

The findings of this study have brought to light several practical applications which could be utilized by other organizations to enhance collaboration system effectiveness. Additionally, there were two specific topics which emerged from an evaluation of the participant responses which warrant further research.

Practical application 1. Evaluate the adverse effects of mitigation measures for adverse effects on collaboration systems effectiveness. Mitigation measures employed within any organization to enhance collaboration should be fully evaluated for any potential adverse effects.

Practical application 2. Reduce unnecessary redundancy in systems and networks. Employing multiple systems with redundant or overlapping capabilities should be evaluated for efficiency and adverse effects.

Future research 1. Based on the findings of this study, a subsequent quantitative study should be conducted to determine the actual level of effect each challenge identified in this study has on collaboration system effectiveness.

Future research 2. Based on the evaluation of collaboration barriers and mitigation measures, an additional study should be conducted to determine the effects of mitigation measures on collaboration system effectiveness.

8. CONCLUSIONS

This paper presents the results of a study to identify barriers to collaboration systems effectiveness through the use of a phenomenological qualitative study. Data was collected from subject matter experts.

Barriers to collaboration systems effectiveness identified in this study include:

- system interoperability
- security domains
- bandwidth limitations
- duplication
- austere environment
- power
- training

Mitigation measures employed based on known barriers to collaboration systems effectiveness include:

- Training
- Network management
- Redundancy
- Standard operating procedures

Based on the findings, implications were made referencing each research question. Four of the five implications focused on the need to better understand the adverse effects of barriers on collaboration system effectiveness. The final implications involved evaluation the effects of mitigation measures used to enhance collaboration.

Practical applications derived from this study consist of the need to evaluate the adverse effects of mitigation measures for adverse effects on collaboration systems effectiveness and reduce unnecessary redundancy in systems and networks. Lastly, recommendations for future research include a subsequent quantitative study to determine the actual level of effect of known collaboration challenges on collaboration system effectiveness and a separate study to determine the effects mitigation measures have on collaboration system effectiveness.

9. REFERENCES

[1] Office of Force Transformation. (2005). *The implementation of network-centric warfare*. Washington D.C.: U.S. Government Printing Office.

[2] Defense Information Systems Agency. (2010). Protecting critical infrastructure in DISA and the global information grid. Retrieved from http://www.disa.mil/conferences /2010/files/.../Protecting_Critical_Infrast_GIG.pdf

[3] Craig, R. (2009). *Traditions of Communication Theory*. Los Angeles, CA: Sage Publications, Inc.

[4] Russell, E. (2009). *The fundamentals of marketing*. La Vergne, TN: Ingram Publisher Services, Inc.

[5] Croft, R. (2004). Communications theory. Retrieved from http://www.eou.edu/~rcroft/MM350/ CommunicationModels.pdf

[6] Moustakas, C. (1994). *Phenomenological research methods*. Thousand Oaks, CA: Sage Publications, Inc.

[7] Creswell, J. (2009). *Research design: qualitative, quantitative, and mixed methods approaches (3rd ed.).* Thousand Oaks, CA: Sage Publications, Inc.

[8] Zikmund, W. (2003). *Business research methods*. Mason, OH:Thomson/South-Western.

[9] Tisserand, J. (2006a). *Network centric warfare case study volume II: A view of C4 architectures at the dawn of network centric warfare*. Retrieved from the United States Army War College Library website: http://www.carlisle.army.mil/library

[10] Wilson, C. (2007). *Network centric operations: Background and oversight issues for congress.* Retrieved from http://www.fas.org/sgp/crs/natsec /RL32411.pdf

Conservativeness Judgement Of The Controller For Systems With Time-varying Delay

Jiuying Deng
Guangdong University of Petrochemical Technology ,
Maoming, 525000, P.R. China

and

Huifei Deng, Qinruo Wang
Guangdong University of Technology, Guangzhou,
310006, P.R. China

ABSTRACT

The conservatism of asymptotic stability conditions is considered in terms of linear matrix inequalities for time-varying delay systems. The conservative index is defined to evaluate the conservativeness for both delay-dependent and delay-independent stability conditions. The general results on H_∞ performance analysis are presented based on descriptor system approach. The optimization approach is given to obtain the upper delay and rational performances for the state-feedback controller of time-delay systems. Experimental results verify the effectiveness of the new method.

Keywords: time-varying delay systems, conservative index, descriptor system

1. INTRODUCTION

Time-delay existing in many control systems is often a source if instability. For the stability analysis of time-delay systems, there are mainly two type of stability conditions proposed: delay-dependent and delay-independent stability criteria[1-3]. The delay-dependent conditions are generally less conservative than delay-independent ones which do not include any information on the size of delays. The choice of an appropriate Lyapunov-Krasovskii functional is crucial for obtaining a solution to various H_∞ control problems[4-7]. In the state-feedback H_∞ controller design, special forms of Lyapunov-Krasovkii functionals lead to simpler delay-independent and delay-dependent linear matrix inequalities(LMIs)[8-12]. Concerning the H_∞ control problem for some time-delay systems, only a delay-independent state-feedback solution has been achieved[13]. In order to reduce the conservatism of these stability conditions, the descriptor system approach has been proposed in [1-2,5]. The improved solutions are delay-dependent, however delay-independent results can be obtained, for certain values of the design parameters in [1]. We integrate the delay-dependent conditions with time-independent ones by descriptor system approach. A conservativeness index is introduced to evaluate the conservatism for the state-feedback controller of time-varying delay systems. The general asymptotic stability results are presented on H_∞ performance analysis. The optimization problem is given to compute the upper value of delay and minimum disturbance attenuation performance for system with various size of delay. At last, the given numeral experiments show that the new method is effective.

Notation: Throughout the paper the superscript "T" stands for matrix transposition, R^n denotes the n dimensional Euclidean space with vector norm $|\cdot|$. The space of functions that are square integrable $[0\ \infty)$ is denoted by $L_2[0,\infty)$. For real symmetric matrices X and Y, the notation $X \geq Y$ (respectively, $X > Y$) means that the matrix $X - Y$ is positive semi-definite

(respectively, positive definite). I is an identity matrix with appropriate dimension. In symmetric block matrices or long matrix expressions, an asterisk (*) is used to represent a term that is induced by Symmetry. Matrices, if not explicitly stated, are assumed to have compatible dimension.

2. ELEMENTARY KNOWLEDGE

We consider the following time-delay system[2]:

$$\begin{aligned}\dot{x}(t) &= \sum_{i=0}^{2} A_i x(t-\tau_i(t)) + B_1\omega(t) + B_2 u(t)\\ x(t) &= \phi(t),\ \forall t \in [-h, 0]\\ z(t) &= Cx(t) + D_1\omega(t) + D_2 u(t)\end{aligned} \tag{1}$$

Where $x(t) \in R^n$ is the state, $\tau_0 \equiv 0, \phi(t)$ is the initial condition, the scalar $h > 0$ is an upper bound on the time delays $\tau_i(t), i = 1, 2$, and $A_i, i = 0, 1, 2, B_1, B_2, C, D_1, D_2$ are known real constant matrices. $\omega(t) \in R^q$ is the noise signal which is assumed to be in $L_2[0,\infty)$; $z(t) \in R^p$ is the output. However, the results in this paper can be extended to the case of multiple delays.

As in [3], we consider two different cases for time-varying delays:

(1) $\tau_i(t)$ are differentiable functions, satisfying for all $t \geq 0$:

$$0 \leq \tau_i(t) \leq h_i, \quad \dot{\tau}_i(t) \leq d_i < 1, \quad i = 1, 2. \tag{2}$$

(2) $\tau_i(t)$ are continuous functions, satisfying for all $t \geq 0, 0 \leq \tau_i(t) \leq h_i, i = 1, 2$.

Now set $\dot{x}(t) = y(t)$, then as in [2], the delay system (1) can be transformed into an equivalent descriptor form

$$\begin{aligned}&\dot{x}(t) = y(t)\\ &0 = -y(t) + \{\sum_{i=0}^{2} A_i\} x(t) - \sum_{i=1}^{2} A_i \int_{t-\tau_i(t)}^{t} y(s)ds + B_1\omega(t) + B_2 u(t)\end{aligned} \tag{3}$$

or

$$E\dot{\bar{x}}(t) = \begin{bmatrix} 0 & I \\ \sum_{i=0}^{2} A_i & -I \end{bmatrix} \bar{x}(t) - \sum_{i=1}^{2} \begin{bmatrix} 0 \\ A_i \end{bmatrix} \int_{t-\tau_i(t)}^{t} y(s)ds + \begin{bmatrix} 0 \\ B_1 \end{bmatrix} \omega(t) + \begin{bmatrix} 0 \\ B_2 \end{bmatrix} u(t) \tag{4}$$

$$z(t) = \begin{bmatrix} C & 0 \end{bmatrix} \bar{x}(t) + D_1\omega(t) + D_2 u(t) \tag{5}$$

Where

$$\bar{x}(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}, \quad E = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}.$$

The following lemma is delay-dependent result on H_∞ performance analysis as theorem 5 in [2].

Lemma 1 Suppose $\gamma > 0$ is a given scalar. Then, under zero input $u(t) = 0$, for all $\tau_i(t), i = 1, 2$, satisfying (2), the time-delay system (4),(5) is asymptotically stable and satisfies

$$\| z \|_2 < \gamma \| \omega \|_2 \tag{6}$$

Under zero initial condition for all non-zero $\omega \in L_2[0,\infty)$ if there exist matrices $P_1 > 0, P_2, P_3, R_i > 0, S_i$, and $Y_i, i = 1, 2$, such that the following LMI holds:

$$\begin{bmatrix} \Omega & P^T\begin{bmatrix} 0 \\ A_1 \end{bmatrix} - Y_1^T & P^T\begin{bmatrix} 0 \\ A_2 \end{bmatrix} - Y_2^T & -h_1 Y_1^T & -h_2 Y_2^T & P^T\begin{bmatrix} 0 \\ B_1 \end{bmatrix} & \begin{bmatrix} C^T \\ 0 \end{bmatrix} \\ * & -(1-d_1)S_1 & 0 & 0 & 0 & 0 & 0 \\ * & * & -(1-d_2)S_2 & 0 & 0 & 0 & 0 \\ * & * & * & -h_1 R_1 & 0 & 0 & 0 \\ * & * & * & * & -h_2 R_2 & 0 & 0 \\ * & * & * & * & * & -\gamma^2 I & D_1^T \\ * & * & * & * & * & * & -I \end{bmatrix} < 0 \tag{7}$$

where
$$\Omega = P^T\begin{bmatrix} 0 & I \\ A_0 & -I \end{bmatrix}+\begin{bmatrix} 0 & I \\ A_0 & -I \end{bmatrix}^T P+\begin{bmatrix} \sum_{i=1}^{2} S_i & 0 \\ 0 & \sum_{i=1}^{2} h_i R_i \end{bmatrix}+\sum_{i=1}^{2}\begin{bmatrix} Y_i \\ 0 \end{bmatrix}+\sum_{i=1}^{2}\begin{bmatrix} Y_i \\ 0 \end{bmatrix}^T \tag{8}$$

$$P=\begin{bmatrix} P_1 & 0 \\ P_2 & P_3 \end{bmatrix} \tag{9}$$

There is the proof in [2].

A Lyapunov functional candidate is

$$V(t) = V_1(t) + V_2(t) + V_3(t)$$

Where

$$V_1(t) = \bar{x}(t)^T EP\bar{x}(t)$$

$$V_2(t) = \sum_{i=1}^{2}\int_{-h_i}^{0}\int_{t+\theta}^{t} y(\alpha)^T R_i y(\alpha)\,d\alpha d\theta$$

$$V_3(t) = \sum_{i=1}^{2}\int_{t-\tau_i(t)}^{t} x(\alpha)^T S_i x(\alpha)\,d\alpha$$

Similarly, for all constants $\tau_i(t), i = 1, 2$, the delay-dependent result on H_∞ performance analysis was given by Theorem 6 in [2], corresponding to (7) with $S_i = 0, i = 1, 2$.

3. CONSERVATISM JUDGEMENT OF TIME-DELAY SYSTEM

The conditions of delay-dependent stability criteria take into account of the size of delays. The delay-dependent approach often has a limit of upper delay the time-delay system. The delay-independent approach is developed to stabilize the systems with long delay.

Set $R_i = 0, i = 1, 2$, the condition given by Lemma is transformed into the delay-dependent result on H_∞ performance analysis. It is known that the delay-dependent stability conditions are generally less conservative than delay-independent ones which do not include any information on the size of delays. So, we define an index to evaluate the conservativeness of stability conditions[1].

Now, choose a new $V_2(t)$:

$$V_2(t) = \sum_{i=1}^{2}\int_{-h_i}^{0}\int_{t+\theta}^{t} y(\alpha)^T (1-\varepsilon) R_i y(\alpha)\,d\alpha d\theta$$

Where the conservativeness index $\varepsilon \in [0,1]$. For $\varepsilon = 0$, the stability condition is derived as the delay-dependent result in Lemma 1; for $\varepsilon = 0$, the stability condition is derived as the delay-independent result.

Then, we have the following stability condition on H_∞ performance analysis.

Theorem 1 Suppose $\gamma > 0$ is a given scalar, $\varepsilon \in [0,1]$. Then, under zero input $u(t) = 0$, for all $\tau_i(t), i = 1, 2$, satisfying (2), the time-delay system (4),(5) is asymptotically stable and satisfies (6), Under zero initial condition for all non-zero $\omega \in L_2[0,\infty)$ if there exist matrices $P_1 > 0, P_2, P_3, R_i > 0, S_i$, and $Y_i, i = 1, 2$, such that the following LMI holds:

$$\begin{bmatrix} \Omega_1 & P^T\begin{bmatrix} 0 \\ A_1 \end{bmatrix} - Y_1^T & P^T\begin{bmatrix} 0 \\ A_2 \end{bmatrix} - Y_2^T & -h_1 Y_1^T & -h_2 Y_2^T & P^T\begin{bmatrix} 0 \\ B_1 \end{bmatrix} & \begin{bmatrix} C^T \\ 0 \end{bmatrix} \\ * & -(1-d_1)S_1 & 0 & 0 & 0 & 0 & 0 \\ * & * & -(1-d_2)S_2 & 0 & 0 & 0 & 0 \\ * & * & * & -h_1(1-\varepsilon)R_1 & 0 & 0 & 0 \\ * & * & * & * & -h_2(1-\varepsilon)R_2 & 0 & 0 \\ * & * & * & * & * & -\gamma^2 I & D_1^T \\ * & * & * & * & * & * & -I \end{bmatrix} < 0 \tag{10}$$

where $\Omega_1 = P^T\begin{bmatrix} 0 & I \\ A_0 & -I \end{bmatrix} + \begin{bmatrix} 0 & I \\ A_0 & -I \end{bmatrix}^T P + \begin{bmatrix} \sum_{i=1}^{2} S_i & 0 \\ 0 & \sum_{i=1}^{2} h_i(1-\varepsilon)R_i \end{bmatrix} + \sum_{i=1}^{2}\begin{bmatrix} Y_i \\ 0 \end{bmatrix} + \sum_{i=1}^{2}\begin{bmatrix} Y_i \\ 0 \end{bmatrix}^T$, P is given in (9).

We consider the state-feedback control law:

$$u(t) = Kx(t) \tag{11}$$

Substituting (9) into (4),(5), wo obtain the structure of (4), (5) with

$$\overline{A}_0 = A_0 + B_2K, \quad \overline{C} = C + D_2K,$$

$A_i, i=1,2, B_1, C, D_1$ are the same. It is obvious that the stability condition of the state-feedback control system on H_∞ performance analysis is generated for Theorem.

Theorem 2 Suppose $\gamma > 0$ is a given scalar, $\varepsilon \in [0,1]$. For all $\tau_i(t), i=1,2$, satisfying (2), the time-delay system (4),(5) with control law (11) is asymptotically stable and satisfies (6), Under zero initial condition for all non-zero $\omega \in L_2[0,\infty)$ if there exist matrices $P_1 > 0, P_2, P_3, R_i > 0, S_i$, and $Y_i, i=1,2$, such that the following inequality holds:

$$\begin{bmatrix} \Omega_2 & P^T\begin{bmatrix} 0 \\ A_1 \end{bmatrix} - Y_1^T & P^T\begin{bmatrix} 0 \\ A_2 \end{bmatrix} - Y_2^T & -h_1Y_1^T & -h_2Y_2^T & P^T\begin{bmatrix} 0 \\ B_1 \end{bmatrix} & \begin{bmatrix} \overline{C}^T \\ 0 \end{bmatrix} \\ * & -(1-d_1)S_1 & 0 & 0 & 0 & 0 & 0 \\ * & * & -(1-d_2)S_2 & 0 & 0 & 0 & 0 \\ * & * & * & -h_1(1-\varepsilon)R_1 & 0 & 0 & 0 \\ * & * & * & * & -h_2(1-\varepsilon)R_2 & 0 & 0 \\ * & * & * & * & * & -\gamma^2 I & D_1^T \\ * & * & * & * & * & * & -I \end{bmatrix} < 0 \tag{12}$$

where $\Omega_2 = P^T\begin{bmatrix} 0 & I \\ \overline{A}_0 & -I \end{bmatrix} + \begin{bmatrix} 0 & I \\ A_0 & -I \end{bmatrix}^T P + \begin{bmatrix} \sum_{i=1}^{2} S_i & 0 \\ 0 & \sum_{i=1}^{2} h_i(1-\varepsilon)R_i \end{bmatrix} + \sum_{i=1}^{2}\begin{bmatrix} Y_i \\ 0 \end{bmatrix} + \sum_{i=1}^{2}\begin{bmatrix} Y_i \\ 0 \end{bmatrix}^T$, P is given in (9).

4. CONSERVATIVENESS EVALUATION AND EXPERIMENTAL RESULTS

The inequality (12) in Theorem is affine in the system matrices and feedback gain *K*. The solver of Matlab is not directly applied to solve the non-linear matrix inequality. We decompose the condition (12) as the followings (13.1)+(13.2).

$$\begin{bmatrix} \Omega_1 & P^T\begin{bmatrix} 0 \\ A_1 \end{bmatrix} - Y_1^T & P^T\begin{bmatrix} 0 \\ A_2 \end{bmatrix} - Y_2^T & -h_1Y_1^T & -h_2Y_2^T & P^T\begin{bmatrix} 0 \\ B_1 \end{bmatrix} & \begin{bmatrix} \overline{C}^T \\ 0 \end{bmatrix} \\ * & -(1-d_1)S_1 & 0 & 0 & 0 & 0 & 0 \\ * & * & -(1-d_2)S_2 & 0 & 0 & 0 & 0 \\ * & * & * & -h_1(1-\varepsilon)R_1 & 0 & 0 & 0 \\ * & * & * & * & -h_2(1-\varepsilon)R_2 & 0 & 0 \\ * & * & * & * & * & -\gamma^2 I & D_1^T \\ * & * & * & * & * & * & -I \end{bmatrix} < 0 \tag{13.1}$$

$$\Omega_3 = P^T\begin{bmatrix} 0 & 0 \\ B_2K & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ B_2K & 0 \end{bmatrix}^T P < 0 \tag{13.2}$$

The LMI (13.1) is solved by the solver of Matlab under the constraint (13.2). For the prescribed scalar $\gamma > 0$, and a conservativeness index $\varepsilon \in [0,1]$, we applied the Theorem 2 to obtain the maximum value of the delay *h*.

For convenient computation, the optimization problem is formulated as

$$\begin{aligned} &\min_{\gamma,\varepsilon\in[0,1]} 1/h \\ &s.t.\ (13.1), (13.2) \\ &\quad P>0 \\ &\quad R>0 \end{aligned} \tag{14}$$

Example 1. We consider the following system[1]:

$$\begin{aligned} \dot{x}(t) &= A_0 x(t) + A_1 x(t-\tau(t)) + B_1\omega + B_2 u(t) \\ z(t) &= C_0 x(t) + D_1\omega + D_2 u(t) \end{aligned} \tag{15}$$

where $D_1 = 0$, $A_0 = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}$, $A_1 = \begin{bmatrix} 0 & 0.9 \\ -1.3 & -1.9 \end{bmatrix}$,

$B_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $B_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $C_0 = \begin{bmatrix} 1 & 0 \end{bmatrix}$, $D_2 = 0.1$,

$0 \le \tau(t) \le h, \quad \dot{\tau}(t) \le d = 0.5$.

The LMI of Theorem 2 can be used to find the maximum value of *h* for which a state-feedback controller stabilizes the system. For $\varepsilon = 0.31$, we obtained that $h = 1$, and a minimum value of $\gamma = 4.12$ with a corresponding gain $K = [-10\ 0]$. For another $\varepsilon = 0.1$, a value of $h = 0.9$, $\gamma = 2.1$ was achieved. For $\varepsilon = 0.95$, an upper delay $h = 0.9$, and a minimum value of $\gamma = 1.86$ were achieved. We will have better results in cost of conservatism raising.

5. CONCLUSION

In this paper, we utilize the descriptor system approach for time-varying systems. The conservatism of asymptotic stability condition is analysed on both delay-dependent and delay-independent methods. The delay-dependent stability condition is less conservative than delay-independent stability condition. We introduce the conservativeness index to denote the conservatism between delay-dependent and delay-independent condition. The general results with the conservativeness index on H_∞ performance analysis are presented for time-delay systems. This method links delay-dependent condition with delay-independent condition. So, it is convenient to compute the maximum value of delay and optimal performances for time-varying delay systems, or systems with various size of delay. The experimental results of numeral example show that the new method is effective.

REFERENCES

[1] E. Fridman and Uri Shaked. A Descriptor System Approach to H_∞ Control of Linear Time-Delay Systems. **IEEE TRANSACTIONS ON AUTOMATIC CONTROL**, Vol.47, 2002(2): 253-270.

[2] S. Xu, J. Lam and Y. Zou. Simplified descriptor system approach to delay-dependent stability and performance analyses for time-delay systems. **IEE Proc.-Control Theory Appl.**, Vol.152, 2005(2): 147-151.

[3] Emilia Fridman and Uri Shaked. An Improved Stabilization Method for Linear Time-Delay System. **IEEE TRANSACTIONS ON AUTOMATIC CONTROL**, Vol.47, 2002(11): 1931-1937.

[4] Yong He, Qing-Guo Wang, Lihua Xie, and Chong Lin. Further Improvement of Free-Weighting Matrices Technique for Systems With Time-Varying Delay. **IEEE TRANSACTIONS ON AUTOMATIC CONTROL**, Vol.52, 2007(2): 293-299.

[5] Huijun Gao and Changhong Wang. Comments and Further Results on "A Descriptor System Approach to H_∞ Control of Linear Time-Delay Systems. **IEEE TRANSACTIONS ON AUTOMATIC CONTROL**, Vol.48, 2003(3): 520-525.

[6] PooGyeon Park. A Delay-Dependent Stability Criterion for Systems with Uncertain Time-Invariant Delays. **IEEE TRANSACTIONS ON AUTOMATIC CONTROL**, Vol.44, 1999(4): 876-877.

[7] Yong He, Qing-Guo Wang, Chong Lin, Min Wu. Delay-range-dependent stability for systems with time-varying delay. **Automatica**, Vol.43 2007: 371–376.

[8] Wu-Hua Chen, Zhi-Hong Guan, Xiaomei Lu. Delay-dependent output feedback guaranteed cost control for uncertain time-delay systems. **Automatica**, Vol.40, 2004: 1263–1268.

[9] J.H. Park. Design of a dynamic output feedback controller for a class of neutral systems with discrete and distributed delays. **IEE Proc.-Control Theory Appl.**, Vol.151, 2004(5): 610-614.

[10] Chong Lin, Qing-Guo Wang, and Tong Heng Lee. A Less Conservative Robust Stability Test for Linear Uncertain Time-Delay Systems. **IEEE TRANSACTIONS ON AUTOMATIC CONTROL**, Vol.51, 2006(1): 87-91.

[11] D. Ye and G.-H. Yang. Adaptive reliable H_∞ control for linear time-delay systems via memory state feedback. **IET Control Theory Appl.**, 2007, 1, (3), pp. 713-720.

[12] Guang-Hong Yang and Dan Ye. Adaptive Fault-tolerant H_∞ Control via Dynamic Output Feedback for Linear Systems against Actuator Faults. **Proceedings of the 45th IEEE Conference on Decision & Control,** San Diego, CA, USA, December, 2006: 3524-3529.

[13] M. Mahmoud. Robust H_∞ control of linear neutral systems. **Automatica**, vol.36, 2000, pp. 757-764.

Technological Innovation in Greece: Trends, Challenges and Opportunities for a Sustainable Future for IMETI 2012

Maria MARKATOU
Department of Economics- National and Kapodistrian University of Athens
Larissa, Greece

ABSTRACT

This paper examines the development of technological innovation in Greece based on patent data. It is based on the description and measurement of the new and original technologies that were first developed by Greek inventors and then been granted by the Greek patent office between 1988 and 2010. The study relies on the theoretical and empirical framework of the importance of innovation in technological development and on its description and measurement through relative patent records, which have been considered and accepted to be good and reliable innovation indicators. This paper starts by discussing the existing theoretical and empirical evidence of both technology and innovation in relation to technological development and continues by examining the Greek case, presenting both an historical evolution of the new developed and protected technologies and their weight based on their importance in quantitative and qualitative terms. The total period of analysis has been divided in two sub-periods, 1988-2000 and 2000-2010 in order to trace changes in trends and behaviour in both the introduction or withdraw of technologies and their relative importance. The paper also provides evidence of the high or low concentration of the developed technological innovation inside relative technology and industrial sectors.

Keywords: Greece, Intellectual property rights, Measurement, Patents, Technological innovation

1. INTRODUCTION

Innovation has become a very important factor to corporate success, technological leadership and economic development, both at national and regional level [1], [2]. Schumpeter [3], with his "gales of creative destruction", gave a vivid description of the effects of the introduction and diffusion of major technological discoveries and inventions in industry and the world economy and Romer [4] emphasized the role of innovation. As a production factor innovation affects growth and contributes to the development of nations. According to Porter [5], innovation, whether it relates to processes, products or organizations, determines the competitiveness of a nation, which depends ultimately on the companies' ability to innovate and improve.

Since the pioneering study on the nature of innovation in the 1970s [6], [7], many research works have been presented regarding innovation analysis. Innovation is the output of innovation process. Innovation process is considered to be a highly systemic and complex process, which varies across industry, technology and firm size. Particularly for firms, firms develop innovations responding to their particular markets and technological challenges. All these factors make innovation difficult to be measured in complete and standardized ways. However, given the importance of innovation for national and firm wealth and welfare, the issue of measurement has become even more demanding. Nowadays, the literature on measurement of innovation is abundant, every day being improved and increased, and focused on measuring innovation at both firm, sector, regional and national level, as well as in combination to other economic and managerial parameters. The main methodological conclusion that can be derived from the study of literature is that innovation can be measured only through its products and only indirectly, with the contribution of relative indicators. One very important category of such indicators is patent indicators.

The objective of this paper is to study technological innovation in Greece, through the examination of patent records. Two points have to be noted: First, It is of particular importance, when examining a country's technological activities and identifying new trends, to highlight major technology fields, locate fields of dynamism and indicate fields for further development and specialization. Second, the reference to the historical evolution or long-tern perspective is necessary or even inevitable for such an analysis. Therefore, this paper examines Greek granted patents since the establishment of the Greek Patent Office, to study technological innovation, aiming at providing a deep understanding of the existing situation and an objective statistic reference for future policy in this field. The paper is structured as follows: Section one discusses the theoretical and empirical framework of innovation and technological development in relation to patent records and the construction of relevant indicators. Section two describes the data that was used and the methodology that was followed in the study. Section three focuses on the empirical results of the study based on the Greek case. Section four synthesizes, further discusses the results, presenting at the same time some concluding remarks.

2. BIBLIOGRAPHY REVIEW

Joseph Schumpeter is considered to be the first economist to focus on innovation and its importance. According to him,

innovation is a product (new one or change in an existing), a process new to an industry, an opening to a new market, the development of new sources of supply for raw materials or other inputs and changes in industrial organization [8]. Thus, innovation involves many and different things and this means that its measurement is difficult due to the broad nature, scope and products of innovation activities. This is the reason that innovation can only be measured indirectly and mainly with the use of indicators.

The existing bibliography has proposed several indicators for the description and measurement of innovation. The most common of them are the indicators that derive from R&D, patents and new products [9], [10], [11], [12]. A patent is a document, which contains structured and detail information regarding the hidden invention and the probable or future technological innovation, and is accessible to the general public through the dissemination of such documents by a national or international authorized government agency [13]. Each patent document is issued by a patent office and grants the owner a monopoly over the exploitation of a precisely defined technological advancement or incremental improvement (e.g. new device, apparatus, or process) over a stated period of time, which is usually 20 years. Generally an invention can be eligible for a patent, only if the innovation that this invention 'hides' is novel, involves a non-obvious inventive step, and could be commercially viable [14], [15].

Among the advantages of using patent data at the study of technological innovation are first the proximity of patents to the inventive and innovative activities, the wide range of fields covered by patents and the geographical scope of patents. Second, patents are characterized by their easy accessibility, high reliability and precise definition [16], [17]. Third, patent data are accurately recorded and easily elaborated, while they can be used to examine and study different levels and kinds of analysis (e.g. technological, sectoral-industrial, national). Four, patent data are rather 'objective' indicators, as patent documents are examined and eventually granted by a single national patent office, based on specific criteria, while every patent is then classified to sectors, classes and main groups according to the same classification system (e.g. International Patent Classification) [18]). Finally, in comparison with or in contrast to other sources, patents are often the only timely measure of rapid technological change, particularly in the context of global competition and an important tool for assessing the performance of technological systems. For example patents permit the study of technological change since they represent inventive activity and output from applied research over different fields, countries, and time [19], [20].

However, as every tool of analysis, patent data exhibit also limitations. First, obviously not all inventions become patents and not all patents become innovations. Especially for firms this has a dual meaning. Firms use the patent protection in order to be protected from competitors. However, they may choose to use other ways of protection, such as secrecy, very fast introduction to the market and so on. Second, every patent office treats patents equally, while they are not and nor do all patents exert the same economic impact and the same technological and economic value [21], [22], [23], [24]. Third, the propensity to patent differs across countries, sectors, firms, fields and technologies and this difference overestimates or underestimates the results in terms of performance [25], [26]. Meanwhile this difference is due in part to the level of protection afforded by the patent, but also to the possibility of protecting monopoly rights by other means depending upon market conditions. Four, there are differences in patent regimes across countries and this means that it is difficult to be certain that one is comparing 'like with like'. In fact, it is widely accepted that there are differences among the various patent systems (e.g., United States Patent and Trademark Office, European Patent Office, etc.), due to variations in legal, geographic, economic, and cultural factors. For instance, some countries would require multiple patents for the same innovation which could be covered by a single patent in other countries.

As already mentioned, the bibliography on innovation measurement is abundant and there are many research works that have studied innovation and related fields at national level based on patent records. For example, the Norwegian technological changes were analysed through patents [27]. The Germany's long term evolution in chemistry was examined by patent statistics [28]. Patent growth was also used to evaluate the India's patent policy during 1979-2002 [29]. The Taiwan's innovative potential between 1998 and 2002 was reviewed based on the number of patents [30]. Characteristics of patent application were used to compare France, Germany, and England in the arena of biotechnology [31]. Obviously, the above list in only indicative of the bibliography in this field and it was recorded to show that first the description and measurement of innovation concerns all countries, developing and developed and second patent records have been used to investigate different aspects of innovation and related parameters at different levels of scale and aggregation.

This study uses patent data as a proxy and measurement indicator to elucidate technological innovation in Greece, taking advantage of their positives and also considering their negatives. To the best of our knowledge, this paper is the first research effort to study technological innovation in Greece based on patent data.

3. METHODOLOGY AND DATA

The data for this study is based on patent records during the period 1988-2010. The rationale behind the construction of the patent database and its elaboration is the result of the selected methodology and the methodological choices that had to be done for this study. Four methodological choices have been made: The first concerns the data source, the second the decision on working with patent grants instead of using simple patent applications, the third the selection between the different technological classifications and the fourth the level of aggregation.

Regarding the first methodological choice it has been decided to use patent data from the Greek Patent Office. The choice of

using data only from the Greek Patent Office has been made for two reasons: First, the main interest of this paper is the examination of the Greek case. Second, the researcher's previous experience [32] has shown that Greek inventors usually protect their patents through the national way, as it is cheaper, easier and a convenient procedure with which they are much more familiar than applying to a broader patent office. The choice of using patent grants instead of using mere patent applications (second methodological choice) is related to a choice of assigning higher value to this study and so to results. Based, therefore, on the fact that a patent is only granted when it contains a technological innovation which exceeds a certain level of newness, only patent grants can guarantee that. The choice of using the IPC system of patent classification (third methodological choice) relies on its main advantage, namely the fact that it is application-based and thus facilitates both the identification of technological innovations and their assignment to different industrial sectors. The IPC is a hierarchy of codes, structured into different levels with several levels of breakdown primarily concerned with the technological characteristics of the invention. The last methodological choice concerns the aggregation level. IPC allows linking patents to one or more economic areas, but only when examining the class level. This is due to the fact that there is no natural or perfect correspondence between technological classifications and economic areas. This problem of classification refers primarily to the difficulties in allocating patent data, organised by technological classes, into economically relevant industries or product groupings. In this study the class level of aggregation has been used (>130 classes and >850 groups of technologies).

Based on the above methodological choices, the analysis is based on the elaboration of a patent database, which has been constructed and elaborated especially for this study and according to the following steps: Step one, patent documents in paper sheets have been collected from the Greek Patent Office for the period of 1988-2010. Based on these patent documents a patent database has been constructed, which contains all patents (Greek and foreign) that have been granted by the Greek Patent Office, a total of 7187 patents. Second step, patents by Greek owners (Greek patents) have been separated by patents of other nationalities (foreign patents) based on the address of the assignee or patent owner. This part of patents composes the sample of analysis, which includes 5339 patents. Third step, each Greek patent has been classified to one or more technology sectors, sub- sectors, classes- subclasses and main groups (five levels of technology analysis) based on the number of patent codes that the Greek Patent Office has attributed to the referring patent. This means that if a patent has one patent code, it is described based on five levels of analysis, if it has two patent codes, it is described based on ten levels of analysis and so on. Forth step, each patent has been related and corresponded to one or more industrial sectors, indicating this way its potential application or industrial use. Fifth, indexes of technological specialization have been constructed, based on the form of Herfindhal index of concentration.

Some further remarks have to be clarified regarding methodology and elaboration and data: First, the total period of analysis has been divided to two sub-periods: 1988-2010 (period 1), and 2000-2010 (period 2). This enables the historical evolution of technological innovation in Greece and major highlights technological fields in time and in relative importance. Second, the technological content of patents is described at five levels of analysis, from the more general (sector, sub-sector) to the more detailed (class, main group), thus ending up to group of products or simple products. Third, the correspondence of patents into relevant industrial sectors is based on their main technological content, thus the first patent code assigned to each patent. In this way it is also avoided the problem of overlapping, while being focused on its main potential application or industrial use. Four, four different kinds of indicators have been constructed and presented in this paper and more specific indicators of patent counts per sub-period and total period, indicators of technological content at class technology level, indicators of technological specialization based on the size of the respective index of concentration and indicators of application or industrial use at NACE 2-digit level.

4. MAIN RESULTS

This study shows that technological innovation in Greece is widely dispersed among the eight broad technology sectors. Shares higher than 10% have been recorded to the sectors of 'human necessities', which ranks first, 'performing operations-transporting', 'fixed constructions', 'mechanical engineering-lighting- heating weapons- blasting' and 'physics' which barely surpasses the 10% share. 'Chemistry- metallurgy' and 'electricity' range between 5-10%, while only 'textiles- paper' accounts for very low absolute numbers. 'Human necessities' are further focused on technologies of 'agriculture', 'foodstuffs-tobacco', 'personal or domestic articles' and 'health, lifesaving, amusement'. 'Performing operations- transporting' is mainly related to technologies of 'transporting', while 'fixed constructions' to those of 'building'. 'Mechanical engineering-lighting- heating weapons- blasting' is further specialized to technologies of 'engines or pumps' and 'engineering in general', while 'electricity' leads to 'electric communication technique'.

The comparison between the two sub-periods of analysis (1988-2000 and 2000-2010) shows that there is no change in the two sub-rankings. However, there is change in the recorded shares between sectors. The shares of 'human necessities', 'performing operations- transporting' and 'electricity' have been decreased and this reduction ranges between 10-15%. On the contrary the shares of 'chemistry- metallurgy', 'mechanical engineering-lighting- heating weapons- blasting' and 'physics' have been increased and this increase ranges between 18-41%, with 'chemistry- metallurgy' being characterized by the higher (42%). The sector of 'fixed constructions' exhibits a remarkable stability, accounting for the 14% of Greek patents and technological innovations during the whole period of analysis (1988- 2010). The calculation of concentration indexes shows that concentration based on sectors is 0.186 for the whole period of analysis (1988-2010), 0.192 for 1988-2000 and 0.176 for 2000-2010, thus is being decreased gradually leading to more technological dispersion and variety. Inside sectors and at class level, concentration varies and ranges from 0.381 in 'textiles-paper', a sector with few patents, and 0.009 in 'chemistry-

metallurgy'. 'Fixed constructions' and 'electricity' are relatively sectors of high concentration, while 'performing operations-transporting' is a rather sector dispersed sector with many important classes.

Table 1 presents the 15 most important technology classes of the classifying Greek patents according to the IPC standards. A closer examination of the table shows that core technologies in 'human necessities' are those of (1) 'medical or veterinary science- hygiene', (2) 'agriculture, forestry, animal husbandry, hunting, trapping, fishing', (3) 'furniture, domestic appliances, coffee mills, spice mills, suction cleaners in general', (4) 'foods or foodstuffs, their treatment, not covered by other classes'. These four technologies account for more than the 80% of the respective sector and are further specialized to 'planting; sowing; fertilising; harvesting; mowing; animal husbandry' and 'preparations for medical, dental or toilet purposes' at technology main group level. Main technologies in 'performing operations- transporting' are those of (1) 'conveying, packing, storing, handling thin or filamentary material', (2) 'vehicles in general' and (3) 'ships or other waterborne vessels, related equipment'. Core technologies in 'fixed constructions' are those of different aspects of building, such as (1) 'doors, windows, shutters or roller blinds in general, ladders' and (2) 'locks, keys, window or door fittings, safes'. In total technologies focused on building account for more than the 80% of the respective sector and are further specialized to 'general building constructions; finishing works on buildings'. Main technologies in 'mechanical engineering- lighting- heating weapons- blasting' are those of (1) 'machines or engines for liquids and various motors', (2) 'heating, ranges, ventilating' and (3) 'engineering elements or units, thermal insulation in general'. Core technologies in 'physics' are those of different kinds of instruments, such as (1) 'computing, calculating, counting', (2) 'Educating and display' and (3) 'measuring, testing'.

The comparison between the two sub-periods of analysis at class level shows the withdrawal of some technology classes and the entrance of some others. Among the former, which account for 18 in total, the most characterizing cases are those of 'paper making- production of cellulose' and 'earth drilling-mining', while among the latter the most important cases are those of 'fabrics' and 'organic macromolecular compounds'**.** In addition, the further examination of classes highlights the very important increase in absolute and relative numbers of some classes, when moving from the first period to the second. These classes are related to the 'performing operations-transporting' ('spraying-atomising' and 'mechanical metal-working'), chemistry-metallurgy (e.g. 'treatment of water, waste, sewage or sludge', 'organic chemistry, 'biochemistry') and 'mechanical engineering- lighting- heating weapons- blasting' (classes 6 and 10b).

Table 1. The 10% of the most important classes in Greece

Ranking	Technological classes	1988- 2010	%[1]	%[2]
1	Medical or veterinary science, hygiene	486	9.20	30.07
2	Agriculture, forestry, animal husbandry, hunting, trapping, fishing	414	7.84	25.62
3	Building	338	6.40	44.89
4	Furniture, domestic appliances, coffee mills, spice mills, suction cleaners in general	238	4.51	14.73
5	Conveying, packing, storing, handling thin or filamentary material	196	3.71	18.88
6	Machines or engines for liquids; wind, spring or weight motors; producing mechanical power or a reactive propulsive thrust	172	3.26	25.79
7	Foods or foodstuffs, their treatment	167	3.16	10.33
8	Vehicles in general	153	2.90	14.74
9	Measuring; testing	152	2.88	19.07
10	-Doors, windows, shutters or roller blinds in general, ladders -Heating; ranges; ventilating	147	2.78	19.52 22.04
11	Locks, keys, window or door fittings, safes	114	2.16	15.14
12	Educating, cryptography; display; advertising; seals	112	2.12	14.05
13	-Computing; calculating; counting -Electric communication technique	107	2.03	13.43 13.43
14	Ships or other waterborne vessels, related equipment	87	1.65	8.38
15	Engineering elements or units; general measures for producing and maintaining effective functioning of machines or installations; thermal insulation	85	1.61	12.74

[1]Share of technology class based on total technology taxonomy. [2]Share of technology class based on the technology sector it belongs to.

Table 2 presents the economic direction of the 15 most important technology classes, after relating them with industrial sectors of application and use. A careful reading of this table shows that the Greek technological innovation is related to the following industrial activities based on NACE codes: (1) chemicals (e.g. pharmaceuticals and pesticides-agrochemical products), (2) different kinds of machinery (e.g. agricultural-forestry, energy, office and computers, special purpose machinery, non-specific purpose machinery), (3) different kinds of equipment (e.g. medical, other transport, other electrical,

industrial process control), (4) different kinds of instruments (e.g. measuring, optical), (5) products and final products from different materials (e.g. non-metallic mineral, fabricated metal, rubber-plastic, basic metal, wood), (7) general basic and consumer goods (e.g. food-beverages, domestic appliances, furniture-consumer goods), (8) electric and electronic equipment (e.g. signal transmission-telecommunications, television, radio receivers-audiovisual electronics) and (9) particular forms of vehicles. More than 60% of all Greek patents are related to the above activities. The table also shows that the majority of technology classes are directed to more than one industrial activities of different sectors, while a particular industrial sector may be related to more than one technology classes. The former is more representative in classes (5), (8), (9) and (15). The latter is clearer in the industrial sectors of 'fabricated metal products' and 'non-specific purpose machinery'. However there are also technology classes and industrial sectors of application and use that are characterized by 100% '1 to 1' correspondence. This is the case of class (6) to 'energy machinery', both classes (10) and (11) to 'fabricated metal products', class (13) to 'office machinery and computers' and class (14) to 'other transport equipment'.

Table 2. The 10% of the most important industrial sectors of application and use in Greece

Classes[1]	Industrial sectors of application and use
1	Medical equipment, pharmaceuticals
2	Agricultural-forestry machinery, pesticides-agrochemical products
3	Non-metallic mineral products, basic metals, wood products
4	Domestic appliances, furniture- consumer goods, fabricated metal products
5	Rubber and plastic products, Fabricated metal products, special purpose machinery, non-specific purpose machinery
6	Energy machinery
7	Food-beverages, special purpose machinery
8	Motor vehicles, other transport equipment, rubber-plastic products, other electrical equipment
9	Measuring instruments, non-specific purpose machinery, industrial process control equipment, medical equipment, motor vehicles
10	Fabricated metal products (100%) Fabricated metal products, domestic appliances, non-specific purpose machinery
11	Fabricated metal products
12	Optical instruments, signal transmission-telecommunications
13	Office machinery and computers (100%) Signal transmission-telecommunications, television, radio receivers-audiovisual electronics
14	Other transport equipment (100%)
15	Energy machinery, rubber-plastic products, fabricated metal products, motor vehicles, non-specific purpose machinery

[1]Classes 1-15 based on the taxonomy of table 1

5. CONCLUSIONS AND IMPLICATION FOR POLICY

Innovation plays a key role in economic development and is therefore a primary concern for practitioners, policy makers, and researchers. Innovation description and measurement led, among other things to the theoretical and empirical analysis of patent value, which has attracted the attention of all related private-public agents for years and contributed to the better examination and interpretation of innovation. In this context, this paper aims at studying the production and the development of technological innovation in Greece based on patent data.

Results show that technological innovations related to specific technologies of 'human necessities', 'fixed constructions' and 'performing operations- transporting' are the most important in quantitative terms. This is the Greek pattern but also the dominant pattern of other countries, with the exception of 'fixed constructions', like Denmark, Hungary, Poland, Portugal, and Slovakia (based on their patents grants at the U.S.A. patent office). On the contrary, in Germany and France different patenting trends can be recorded: The production of technological innovations is mainly concentrated in 'performing operations-transporting', 'electricity' and 'physics'. Comparing to other countries Greece is both characterized by an unusually large share of patents related to 'fixed constructions' and an insufficiency of technologies related to 'chemistry-metallurgy', 'physics' and 'electricity'.

This paper may have some implications for government policy. Greece is a country with very severe fiscal and structural problems. Whatever Greek government does, it is necessary to 'build' a new development agenda and policy and innovation plays a central role in this procedure. Therefore, focusing on technological innovations is the main challenge, but to what direction? The country needs to balance between 'world' and 'endogenous' important innovations'. This means that Greece has to do two things: First, look for and then develop competitive innovations, investing and expanding its relative strengths in line with its industrial structure (e.g. innovations in 'fixed constructions'). The existing pattern shows that technological innovations in industrial sectors related to drugs-sanitary products and both the agricultural and construction industry are the most important. However, based on its geographical position and physical environment and in

correspondence to future global needs, we argue that only innovations related to the agricultural and construction industry could be a viable and technological direction. Thus and in accordance with the paper results Greek innovation should focus on ecological building, sustainable forms of cultivations, renewable resources and saving in energy.

6. REFERENCES

[1] R.J. Buswell, "Research and development and regional development: a review", In: A. Gillespie (Ed.) **Technological Change and Regional Development**, London: Pion, 1987.

[2] E.J. Malecki, **Technology and Economic Development: the Dynamics of Local, Regional, and National Change**. New York: Longman Scientific & Technical, 1991.

[3] J.A. Schumpeter, **The theory of Economic development**, Cambridge: Harvard University Press, 1934.

[4] P.M. Romer, "The origins of endogenous growth", **Journal of Economic Perspectives**, Vol. 8, 1994, pp. 3–22.

[5] M. Porter, **The competitive advantage of nations**, New York: The Free Press, 1990.

[6] M. Gibbons and R. Johnston, "The Roles of Science in Technological Innovation", **Research Policy**, Vol. 3, 1974, pp. 220-242.

[7] C. Freeman, **The Economics of Industrial Innovation,** London: Pinter, 1974.

[8] OECD, **The Oslo Manual: Proposed guidelines for collecting and interpreting technological innovation data**, Paris: OECD, 1997.

[9] B.L. Basberg, "Patents and the measurement of technological change: a survey of the literature", **Research Policy**, Vol. 16, 1987, pp. 131-141.

[10] K. Pavitt, "Uses and abuses of patent statistics", In: A.G.J. Raan, (Ed.) **Handbook of quantitative studies of science and technology**, Amsterdam: North Holland, 1988.

[11] Z. Griliches, "Patent Statistics as Economic Indicators: A Survey", **Journal of Economic Literature**, Vol. 28, No. 4, 1990, pp. 1661-1707.

[12] D. Archibugi, "Patenting as an indicator of technological innovation: a review", Science and **Public Policy**, Vol. 19, No. 6, 1992, pp. 357-368.

[13] Z. Huang, H. Chen, A. Yip, G. Ng, F. Guo and Z.K. Chen, "Longitudinal patent analysis for nanoscale science and engineering: Country, institution and technology field", **Journal of Nanoparticle Research**, Vol. 5, 2003, pp. 333–363.

[14] H. Dernis and D. Guellec, "Using Patent Counts for Cross-Country Comparisons of Technology Output", STI Review, No. 27, Paris: OECD, 2001.

[15] H. Dernis & M. Khan, "Triadic patent methodology", STI Working Paper, No. 2004/2, Paris: OECD, 2004.

[16] H. Ernst, "Patent Applications and Subsequent Changes of Performance: Evidence from Time-Series Cross-Section Analyses on the Firm Level", **Research Policy**, Vol. 30, No. 1, 2001, pp. 143-157.

[17] K. Debackere, A. Verbeek, M. Luwel and E. Zimmermann, "Measuring progress and evolution in science and technology-II: The multiple uses of technometric indicators", **International Journal of Management Reviews**, Vol. 4, 2002, pp. 213–231.

[18] OECD, **Patent Manual**, Paris: OECD, 2009.

[19] M. Trajtenberg, "A penny for your quotes: Patent citations and the value of innovations", **RAND Journal of Economics**, Vol. 21, 1990, pp. 172–187.

[20] A. Hullmann and M. Meyer. "Publications and nanotechnology", **Scientometrics**, Vol. 58, No. 3, 2003, pp. 507-527.

[21] A.B. Jaffe, M.S. Fogarty and B.A. Banks, "Evidence from Patents and Patent Citations on the Impact of Nasa and Other Federal Labs on Commercial Innovation", **Journal of Industrial Economics**, Vol. 46, No. 2, 1998, pp. 183-205.

[22] C. Gay and C. Le Bas, "Uses without too many abuses of patent citations or the simple economics of patent citations as a measure of value and flows of knowledge 1", **Economics of Innovation and New Technology**, Vol. 14, 2005, pp. 333-338.

[23] S. Wang, "Factors to evaluate a patent in addition to citations", **Scientometrics**, Vol. 71, 2007, pp. 509–522.

[24] Y.G. Lee, "What affects a patent's value? An analysis of variables that affect technological, direct economic and indirect economic value: An exploratory conceptual approach", **Scientometrics**, Vol. 79, No. 3, 2009, pp. 623-633

[25] A. Arundel and I. Kabla. "What percentage of innovations are patented? Empirical estimates for European firms**",** **Research Policy**, Vol. 28, 1998, pp. 127-141.

[26] I. Mäkinen, "To Patent or Not to Patent? An Innovation-Level Investigation of the Propensity to Patent", Technical Research Centre of Finland, No. 646, Finland: VTT Publications, 2007.

[27] B.L. Basberg, "Patent Statistics and the measurement of technological change: An assessment of the Norwegian patent data, 1840-1980", **World Patent Information**, Vol. 6, No. 4, 1984, pp. 158-164.

[28] H. Grupp, I.D. Lacasa, and U. Schmoch, "Tracing technological change over long periods in Germany in chemicals using patent statistics", **Scientometrics**, Vol. 57, No. 2, 2003, pp. 175-195.

[29] P. Ganguli, "Patents and patent information in 1979 and 2004: A perspective from India" **World Patent Information**, Vol. 26, 2004, pp. 61-62.

[30] D.Z. Chen, H.W. Chang, M.H. Huang and F.C. Fu, "Core technologies and key industries in Taiwan from 1978 to 2002: A perspective from patent analysis", **Scientometrics**, Vol. 64, No. 1, 2005, pp. 31-53.

[31] S.V. Ramani and M.A.D. Looze, "Country-specific characteristics of patent applications in France, Germany and the UK in the biotechnology sectors", **Technology Analysis & Strategic Management**, Vol. 14, No. 4, 2002, pp. 457-480.

[32] M. Markatou, **Patents and technological change in Greece**, PhD dissertation, Athens: National and Kapodistrian University of Athens, 2008.

Cash for Clunkers: Does it Stimulate Car Markets?

Pavel ŠTRACH
Institute of Management and Marketing, ŠKODA AUTO University,
Mladá Boleslav 293 60, Czech Republic

and

Tomáš KINCL
Department of Information Management, University of Economics, Prague
Jindřichův Hradec 377 01, Czech Republic

ABSTRACT

Automotive industry in several developed countries has been the back bone of their economies employing a large number of workers and contributing significantly to the overall industrial output. Sudden decline in demand for new cars caused by reduced finance accessibility and by bleak consumer outlook led many governments to the introduction of special car scrapping schemes – special programs designed to stimulate new car purchases in search for lower emissions and increased road safety. The question of that time as well as today remains the same: do car scrapping schemes unlock the demand for new cars, alleviate consumer fears and contribute to the long term prospect of car markets or did they simply bring the demand forward, exhaust the market capacity, and immediate gains would be offset later on by lower car sales?

Keywords: Automotive Industry, Market, Strategy, Cash for clunkers, Car scrappage schemes, Demand shift, Crisis

1. INTRODUCTION

At the time of economic downturn, governments typically incentivize frozen financial markets, consumer spending and become more active players in deterring threat of bankruptcies – from direct subventions to certain companies, through government-backed loans to direct privatization or direct stimuli to consumers of goods, producers of which have been affected the most. The recent economic recession has been marked in a number of countries by car scrapping schemes. Consumers, initially stimulated through ever increasing fuel prices, turned their back on new cars in the fear of job security, decreased personal wealth and limited finance options. On the other hand, governments have been seeing prospects of lower tax collection, higher unemployment and more aged vehicles on the road yet pursuant to meet obligations of global anti-warming initiatives.

Car scrapping schemes offered owners of well used cars an opportunity to receive a special bonus if they dispose of their old vehicle indefinitely losing the opportunity to trade-in. Suddenly incentivized demand for new cars through scrapping schemes offers market opportunity for many producers and dealers. However, the opportunity goes hand in hand with uncertainty: would the sudden consumer interest be followed by even harder landing on the tarmac of economic crisis, would the market capacity be simply exhausted more rapidly, would there be a slump in demand after governmental money runs out?

2. SCRAPPING SCHEMES

Car scrapping schemes have been introduced in most top-10 car producing countries around the world with the exception of Brazil and India [19]. Car scrapping programs were perhaps first introduced in Europe throughout 1970s. Countries such as Spain or Greece tried to reduce the age of passenger cars on roads, more in line with other European countries. Since 1970s, scrappage schemed have also been run by other countries such as Sweden, Norway and Iceland, although they targeted environmental aspects concerning end-of-life vehicles with no further obligation to buy a new car [8]. Schemes introduced during the recent economic crisis should be labeled as cash-for-replacement schemes [5] as they offered a financial bonus to everyone who brought an old junk to a scrap yard and purchased a brand new vehicle. Those kinds of accelerated replacement schemes came into existence in 1990s. In 1990, California's Unocal oil company implemented the SCRAP – South Coast Recycled Auto Programme. 700 US dollars were given to retire pre-1971 vehicles and some local dealers as well as Ford Motor Company threw in some additional funds to allure buyers [5]. Greece is believed the first European country to launch a replacement scheme as we know it today in January 1991. Cars older than 10 years in Athens area could be replaced with more modern catalysator-equipped vehicles while being exempted from 40 to 60 % exercise duty and some registration charges. Similar although lower reduction in registration and road charges was applied to all new car purchases (with catalytic device) across Greece[5]. First European scheme driven primarily by economic objectives could be the Spanish Plan Renove program implemented in 1994 [9]. The scheme provided less than 500 euro to any new car buyer if they scrapped at least 10 year old car. Economic objectives resurfaced just recently. The collapse of financial markets introduced the scrapping scheme idea to many other countries, including those with relatively young car fleets such as Germany or United Kingdom (see Table 2).

Magnitude, technology and determinants of scrappage schemes were clearly different (see Table 1). Usually, the bonus was given directly to the individual consumer, payee of value-added (sales) tax, who owned an old vehicle for a certain period of time (commonly 18 to 24 months). Their old car had to meet

certain age criteria and/or score badly on environmental performance in terms of fuel consumption or air pollution. Most countries running scrapping schemes required junks to be between 10 and 13 years of age, with the exception of Germany with its relatively young fleet enabling scrappage of 9 years old vehicles.

Table 1 Scrapping Bonus

Country	Minimum Age of Cars	Scrapping Bonus (EUR)
Austria	13	1,500
Denmark	10	234
France	10 (emission based)	2,000
Germany	9	2,500
Great Britain	10	2,300
Greece	10	3,400
Italy	15	1,500
Japan	13	1,800
Netherlands	17	1,000
Portugal	15	1,250
Romania	15	790
Slovakia	10	2,000
USA	(15.8 mpg or lower)	3,150

Adapted and updated from [11]

The average scrapping incentive (in some countries there was a direct governmental incentive plus compulsory dealer participation) across the European Union fluctuated around 1.500 euros per vehicle. Extraordinarily high subsidies hawked buyers in Greece (3,400 Euros), Germany (2,500 Euros) and the Great Britain (2,300 Euros). In most countries, drivers had to take their four-wheelers to a junk yard with registration plates and in a drivable condition. Disposal of old cars was free. Owners were issued a confirmation which then could be exchanged with local transport authorities or directly used at dealers. The incentive was paid to dealers once they delivered the new car – a couple of weeks after order in Europe, almost immediately in the United States where cars have been sold directly off the lot. Vouchers issued for a junk had usually a limited acceptance not much beyond a few months or within the particular fiscal year. In most countries, a few vouchers remained unused.

Scrapping scheme does not affect only auto manufacturers, their suppliers across the value chain and dealers. It provides subsequent stimuli for other sectors such as the car finance companies or even local garages. For instance, the Slovakian scrapping scheme temporarily increased demand for repair workshops [12]. Many junks were growing rust in gardens and bringing them to shape in which they could be driven to a yard required some mechanical ingenuity. Similarly, second hand parts retailers were circling around queues in front of car yards dismounting some resalable components, price of which were decreasing [17].

Table 2 Average Age of Passenger Cars

Country	Scrapping Bonus (EUR)	Average Age (2001)*	Share of Cars Older Than 10 yrs (2004)**
Austria	1,500	7.4	33.5
Denmark	234	8.2	31.9
France	2,000	7.5	32.2
Germany	2,500	6.8	30.6
Great Britain	2,300	5.9	20.4
Italy	1,500	8.1	38.9
Japan***	1,800	5.8	29.9
Netherlands	1,000	7.0	31.1
Portugal	1,250	11.4	-
Romania	790	11.5	-
Slovakia	2,000	14.8	55.8
USA^	3,150	9.0	34.8

Source: *European Environment Agency (2009), **Eurostat (2009), *** JAMA (2010) – 2009 figures, „-“ data unavailable, ^ 11 years and older, median 2009, share 2005

Governmental incentives programs were limited also in the absolute monetary allocation which could be transferred through the scheme. This measure indirectly abridged the number of incentivized new vehicles. Limited programs allocation forced consumers to act rather swiftly and created sudden demand. Scrapping schemes were exhausted within a month or two or even in weeks in countries such Austria, Slovakia, and Germany. Some schemes limited the maximum price of the new vehicle under the umbrella of avoiding luxury or large car purchases, others such as France or Netherlands set levels of environmental acceptability in terms of CO_2 emissions per kilometer. Last but not least, some programs did not allow consumers to lease new cars or take other consumer finance options in the effort to shelter buyers from irresponsible purchasing decisions.

Media referred to the U.S. Car Allowance Rebate System (CARS) as cash-for-clunkers. The program came to implementation in July 2009 under the assumption that 3 billion dollar allocation would last until November. However, the program was saturated by 24 August 2009. Clunkers had to be younger than 25 years and their fuel economy had to be below 15.8 mpg (over 13 l/100 km). Most junkies returned with dealers were pick-up trucks (such as Ford F150) or large SUVs (Ford Explorer, Jeep Grand Cherokee). Maximum rebate of 4,500 US dollars received only buyers of cars with fuel economy exceeding 10 l/100 km. Others were eligible for 3,500 dollars only. 700,000 new vehicles purchased under CARS were mostly smaller sedans and hatchbacks with badges such as Toyota, Honda, and Ford. Winning brands had enough stock at the time of the scheme while other traditional Detroit automakers (General Motors and Chrysler) did not have enough inventories or the right product mix.

Japan, another well developed and mature car market, launched their incentive scheme in June 2009 for cars above 13 years of age. The incentive of 250,000 yen (about 1,800 Euros) should increase the demand for new vehicles by 20 % and put new 900,000 vehicles on the road [13].

In 2009, China has become the largest car market in the world. Recognizing the potential and impact of the automotive industry, Chinese government has decided to fuel car market growth. For 2009 and 2009, the luxury tax was lowered from 10 to 5 % on cars with engines smaller than 1,600 ccm as there was a relatively small amount of old vehicles. Outside large agglomerations, three-wheelers could be scrapped and traded for an automobile or a minivan with engines up to 1,300 ccm. Only between January and June 2009, the demand for cars between 1,000 and 1,600 ccm grew by 40 % [22]. South Korea took a similar tax-break approach – from May til December 2009, owners of cars registered in 1999 and earlier were given a 70% cut in taxes when buying a replacement.

It was not until March 2010 when Russia adopted an incentive policy of its own kind. Buyers interested in disposing of their vehicles aged 10 years and more could opt only for vehicles assembled or manufactured in Russia if they intended to claim the governmental subsidy of 1,300 Euros (50,000 rubles). 66 models of Lada, UAZ, Gaz, Chevrolet, Fiat, Ford, Renault, Volkswagen, and Skoda brands were manufactured in the country. However, 70 % of all recycling certificates were used in favor of local Lada. 10 billion rubles allotment (200,000 vehicles) ran out shortly and government decided to put additional 10 billion rubles on the table in June 2010.

3. ISSUES WITH SCRAPPING SCHEMES

Scrappage schemes are distortive by nature and target one particular sector and only a portion of the market, just like most other economic measures introduced by political representations through parliaments or governments. Controversial informed as well as popularly shallow or politically colored discussions have accompanied car scrappage schemes since their inception. All economic measures ever implemented by states – grant schemes, direct subsidies or tax breaks – always assist or suppress selective sectors or segments, sometimes for limited periods only. A stimulus such as the scrapping program would immanently have only a short kick-in effect (if designed wisely; although its magnitude and duration may clearly differ). Guardians of budgetary discipline voiced doubts over income neutrality and feared a scrapping bonus could be yet another net public expense. However, most countries concluded the scheme was really expense neutral or even income positive [9]. Sales tax collection from stimulated car sales offset the stimuli.

Scrapping incentives such as tax breaks, reductions, direct (to consumers) or indirect (through dealers) bonuses were based on car's age, its fuel economy, air pollution, length of ownership and on price and specs of the new car. Programs effectively discriminated those who bought their (aged) cars just recently (short ownership) and supported primarily lower income brackets who might recipients of other social welfare. In some countries (such as Great Britain), most of the passenger car market recruits from business side and businesses were ineligible to take piece of the scrappage pie.

Car scrapping schemes have been fundamentally discarded for supporting the least environmentally-cautious mean of transportation, although the green argument was present in all programs [19]. New cars emit fewer pollutants and consume less fuel, even might be safer for passengers and pedestrians being loaded with numerous safety features, electronic and restraint systems, and undergoing more crash approvals. Older cars become gradually less rigid and do not shelter passengers from impact as well as new cars do.

In spite of scrapping schemes being implemented primarily in car producing countries (governments might have been concerned with local employment at the time of crisis), except for Russia mainly imported vehicles and foreign-based manufacturers gained the most new market presence. In Germany, local makes such as Mercedes, BMW or Audi were not flying from dealerships as scrapping schemes stimulate consumers to purchase generally smaller and less expensive cars. The scrapping bonus is relatively the highest for small/cheapest cars, which provides greater motivation to buyers. Even in countries, in which smaller cars have been produced (such as Slovakia), the model which were sought after came from abroad (e.g. Renault Thalia or Dacia Logan small sedans – [16]).

The demand for new cars increases at the time of scrapping scheme and not all brands gain. In Europe, the demand for certain makes and models increased disproportionately (Dacia, Hyundai, Kia, Skoda, Volkswagen). In Germany, most successful models included VW Golf and Polo, Opel Astra, Skoda Fabia, Opel Corsa, and Ford Fiesta ([14]. Austrian buyers chose Seat Ibiza, VW Golf, Opel Corsa or Peugeot 207 [20]. In the United States, Toyota Corolla, Honda Civic, Ford Focus, Toyota Camry and Hyundai Elantra were popular 0. All models mentioned above (perhaps besides Camry) belong to what can be considered a small car for a given market. In Europe, where about half of all passenger cars had diesel powered engines (commonly purchased by companies), scrappage stimulated disproportionate demand for gasoline powered cars.

Owners of older vehicles usually do not belong to most affluent buyers. Stimulating generally poorer consumers to make long-term purchasing decisions at the time of economic uncertainty was another concern raised by scrapping scheme opponents. Buyers might have been thrown even deeper into financial troubles and more mildly used cars might have appeared on the market later on. Or some consumers may have preferred a new vehicle before buying a second hand. Buyers of new cars might have spent their money elsewhere; hence crisis might have just been shifted away from automotive industry to other parts of national economies which are traditionally reliant on discretionary spending (such as restaurants or entertainment establishments), offering other durables (such as furniture), or even away from retail sector in general [7]. In Slovakia and Austria, scrappage schemes assisted to second hand car dealers in spite of not providing that many mildly used cars which were purchased with a bonus. Throughout 2010 a different trend prevailed: all Europe have suffered lack of mildly used cars (jahreswagen) as private ownership takes usually longer than corporate cycle of fleet renewal.

A question mark appeared after some countries (such as Great Britain) required compulsory dealer participation in the bonus. Cash bonus offered in the program was evenly split between government and dealers. However, dealers' discounts have been obvious ever since car salesmen became a symbol of elaborate selling techniques (e.g. [23]) and any participation beyond slimmer margins was a wishful thinking of policy makers. There is no evidence that dealers' or manufacturer's profits would be sacrificed: in the contrary, dealers and manufacturers, who were scrappage scheme winners, reported strong financial results (e.g. [21]; [18]). In France, Netherlands or the United States, the actual bonus varied based on attributes of the old scrapped and the brand new car. The average U.S. bonus reached 2,877 US dollars per car [10].

It would be misleading to believe that scrapping bonuses anyhow hinder competition between brands. Similarly, it must be noted that vehicles exchanged for a bonus are not worthless and would have been normally traded-in with dealerships. Therefore, it is not just the bonus but the resale value which impact on consumer decision making process. Market literacy of some buyers might be questioned though, as scrapped vehicles included more than 5,000 Mercedes cars, more than 1,000 Jaguars and even 11 Porsches, 6 Maseratis and 1 Aston Martin in the United States for instance [15]. Resale value is

perhaps to blame for much older and much less fuel efficient cars than required were disposed of. Large bonuses in combination with a younger fleet and weak administrative controls could be blamed for recorded exports of 50,000 presumably scrapped junks to third countries out of Germany [3].

On the other hand, capacity of junk yards was not unlimited and some collection points in Slovakia stopped accepting vehicles after utilizing all available space. In the United States, it was even believed that the program might cause significant future expenses related to indefinite environmentally considerate disposal [4]. The price of scrap metal had been decreasing throughout 2009 [2] possibly due to sudden influx of used cars.

Proponents of scrapping schemes remarked that the automotive industry is labor intensive and provide lot of job opportunities. Only in Germany, one in seven jobs, i.e. 5.3 million jobs depend on the automotive sector [6]. A car has been one of the most complex products consumed and purchased on a regular basis. Supporting the auto industry may, therefore, help several other sectors from suppliers, dealers to local communities, in which these companies operate. Scrappage schemes were in line with official environmental policies, even on the international scale (Kyoto Protocol). Whereas oil resources have been scarce and most auto nations rely on imports, potentially lower fuel consumption might contribute to environmental as well as national security issues

.

The major issue with scrapping scheme design and acceptability has been the potential overutilization of market capacity: cars purchased now will not be sold later and scrapping scheme may just pre-sell what would materialize later anyhow. A car is driven, has to be maintained and the maintenance costs mounts the older the vehicle gets and the more miles it features on the odometer. It is certainly possible to extend vehicle life but through time it may become more costly to run an oldie than to purchase a new car. The discussion was tense but evidence was lacking and it is not until about a year at least until scrapping schemes might be assessed for mid-term impact.

4. STIMULATION OR SUPPRESSION

Both manufacturers and car dealers raised concerns about sales sustainability. Do scrapping schemes stimulate or suppress car markets in the mid-term? A stimulus program means that the demand would ease up instantly once the scheme runs out (if now phased out gradually which was more or less the case in France). However, if the stimulus package was designed wisely, the new demand levels should exceed those before the program had been launched. Looking one year in the program, there might be a more accurate answer to the problem.

Formula 1 Future Loss Index (FLI)

$$FLI = \frac{Vehicles\,on\,Bonus + \frac{Bonus\,in\,Euro}{1{,}000}}{Passenger\,Car\,Sales\,2008} \qquad (1)$$

The scheme distorts the market the more cars it includes (as a portion of the market) and the higher is the bonus. To address the issue, we have constructed a future loss index as a measure of scrapping scheme appropriateness. Future loss index (see Formula 1) multiplies the number of vehicles purchased on bonus by multiples of thousands of euro under the assumption that the stimulus increases with dollar amount. As discussed previously, typically purchased vehicles were smaller cars for which the rebate was relatively more substantial. Future loss index is then divided by a number of passenger cars sold in a given market in 2008 (a year before the scrapping scheme). Broadly speaking, the index might be perceived as a percentage of 2008 market affected by the scheme. The higher the index is the more it should impact on future sales after the subsidy runs out. A number of counter-arguments might be voiced against such measure: it does not consider specifically the average age of registered cars (the effect on future sales might be smaller if the current fleet is aged), usual portion of the market belonging to fleet sales (the future loss would be smaller if most usual sales were fleet) or local variations in car pricing and taxes. Furthermore, consumer's appetite measured by consumer confidence, GDP per capita, unemployment rate, savings rate or availability of consumer finance might play their part in future loss assessment.

Table 3 Future Loss Index: One Year in Scrapping Scheme

Country	Bonus (EUR max.)	Scheme Introduced	Vehicles on Bonus	FLI	y/y Growth Rate (year in bonus)	y/y Growth Rate (two years in bonus)
Austria	1,500	4/2009	30,000	15.3	25.7	5.0
Portugal	1,250	1/2009	32,500	19.0	61.9	-9.2
Spain	1,000	6/2009	240,000	20.7	25.6	-9,6
UK	2,300	5/2009	200,000	21.6	13.5	-1.7
France	2,000	1/2009	600,000	58.5	14.3	8.2
Italy	1,500	2/2009	856,000	59.0	20.6	-20.5
Japan	1,800	6/2009	900,000	23.5	20.6	-21,6
Slovakia	2,000	3/2009	40,000	114.2	-16.8	20.1
Germany	2,500	1/2009	2,000,000	161.8	-29.8	16.5
USA	3,150	7/2009	670,000	22.1	5.0	10,5

Source: ACEA, www.inautonews.com, The Wall Street Journal

Having those limitations in mind, the future loss index seems to be a solid indicator whether scrapping schemes were designed appropriately and whether they succeeded in jump starting local markets. Looking at the market annual growth rate 12 months after the scheme was introduced, it is apparent that only Slovakia and Germany, schemes of which offered substantial incentives and affected a big chunk of the market pie, recorded a decline. Even with limited number of observations, Spearman's correlation between the future loss index and the market growth rate a year in the scheme is – 0.784 (p-value = 0.01, 2-tailed) implying that higher future loss index leads to lower market growth: with high future loss index the schemes were outselling car market future. Beta coefficient of the estimated linear regression (see Formula 2) is estimated with 95 % significance and the model R Square was calculated at 0.649.

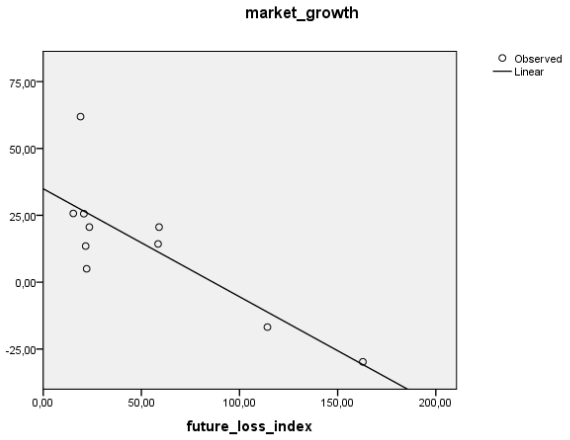
Formula 2 Future Loss Index (FLI)

$$Market\,Growt\,Rate = 34.925 - 0.404 \times FLI \qquad (2)$$

It remains a question whether the relatively optimistic outcomes would persist in year two and beyond. It appears that linear regression (as documented in Figure 1) would be able to explain

a significant portion of market growth rate variation (p-value = 0.05). Would the demand have been substantially brought forward due to scrapping schemes, there would have been at least anecdotal evidence of junk owners waiting for the incentive plan introduction in various countries. Long term effects might be difficult to capture as many other external and internal factors affecting national economies would come to play.

Figure 1 Linear Regression: Market Growth vis-à-vis Future Loss Index



We have also attempted to run similar analyses with growth rates two years in the scheme (Table 3 – last column). However, there seem to be no relationships among variables. It may well be that effects of car scrapping schemes were covered by fiscal and macroeconomic issues in multiple countries of the European Union in 2011, which implemented auto-industry support in 2009.

5. CONCLUSION

Scrapping schemes have been special measures implemented by governments at the time of economic downturn not only to support car markets but to contribute to environmental sustainability of individual motoring or to reduce dependency on imported oil. Scrapping schemes are not systemic by definition and skew usual patterns of car markets. Demand shifts towards smaller cars. Early post-scrappage evidence suggests there could be the right way to introduce scrapping schemes if they should vitalize car markets and not just outsell the future. The future loss index might be a good proxy in making initial judgments and assisting to policy development.

6. REFERENCES

[1] Bartlett, J., **Cash for clunkers: Top 10 most popular new cars and trade ins**, Consumerreports.org, 25 August 2009, available at: http://blogs.consumerreports.org/cars/2009/08/cash-for-clunkers-top-10-most-popular-new-cars-and-trade-ins.html (accessed 15 September 2009), 2009.

[2] Beck, B., **Scrap Prices Riding Growth of Auto Sales**, Scrap Price Bulletin, 18 January 2011, available at: http://www.scrappricebulletin.com/FeaturesColumn.aspx?marketcode=WK&subcategoryID=TA (accessed 30 January 2011), 2011.

[3] cis, **Wreck Rebate Racketeering: Car Recyclers Betray German Government**, Spiegel.de, 8 May 2009, available at: http://www.spiegel.de/international/germany/0,1518,640572,00.html (accessed 15 November 2010), 2009.

[4] de Paula, M., **The Cash for Clunkers Junkyard Glut**, Portfolio.com, 19 August 2009, available at: http://www.portfolio.com/business-news/portfolio/2009/08/19/cash-for-clunkers-creates-salvage-yard-glut/ (accessed 20 September 2010), 2009.

[5] ECMT, **Cleaner Cars: Fleet Renewal and Scrappage Schemes, Guide to Good Practice Series**. European Conference of Ministers of Transport, OECD Publications, Paris, 1999.

[6] European Commission**, Responding to the Crisis in the European Automotive Industry**, 25 February 2009, European Commission, DG Enterprise and Industry, Brussels, 2009.

[7] Ewing, J., **Car-Scrapping Plans: Germany's Lessons**, Businessweek, 6 May 2009, available at: http://www.businessweek.com/globalbiz/content/may2009/gb2009056_301566.htm (accessed 10 September 2010), 2009.

[8] Forslind, K. H., Implementing Extended Producer Responsibility: The Case of Sweden's Car Scrapping Scheme, **Journal for Cleaner Production**, Vol. 13, No. 6, , 2005, pp. 619–629.

[9] Global Insight, **Assessment of the Effectiveness of Scrapping Schemes for Vehicles: Economic, Environmental, and Safety Impacts, European Commission**, DG Enterprise and Industry, Automotive Industry, available at: http://ec.europa.eu/enterprise/sectors/automotive/files/projects/report_scrapping_schemes_en.pdf (accessed 5 August 2010), 2010.

[10] Hicks, M., **Some Preliminary Evaluation of the Cash for Clunkers Program**, CBER Policy Brief. Ball State University, November, 2009, pp. 1–4.

[11] Kaláb, V., Vašek, P., **Česko chce eurošrotovné**. iHNed.cz, 5 February 2009, available at: http://ihned.cz/c3-33901210-00000_d-cesko-chce-eurosrotovne (accessed 6 February 2009), 2009.

[12] Krajanová, D., **Čo treba urobiť, ak chcete získať šrotovné**, Sme.sk, 6 March 2009, available at: http://ekonomika.sme.sk/c/4337492/co-treba-urobit-ak-chcete-ziskat-srotovne.html (accessed 2 February 2011), 2009.

[13] Krebs, M., **Japan Adopts Scrappage Program in Its Economic Stimulus Plan**, Edmuds.com – Autoobserver, available at: http://www.autoobserver.com/2009/04/japan-adopts-scrappage-program-in-its-economic-stimulus-plan.html (accessed 16 September 2009), 2009.

[14] Nacken, H.P., **Kompakt-Pkw und Kleinwagen bauen Führung aus - Zulassungsplus auch nach Abwrackprämie**, DDP – Nachrichtagentur, available at: http://de.finance.yahoo.com/nachrichten/kompakt-pkw-und-kleinwagen-bauen-fhrung-aus-zulassungsplus-auch-nach-abwrackprmie-von-ddpdjn-korrespondent-hans-peter-nacken-ddpnews-6be9cadb00a0.html (accessed 15 October 2009), 2009.

[15] NHTSA, **Consumer Assistance to Recycle and Save Act of 2009**, Report to Congress, National Highway Traffic Safety Administration, Washington, 2009.

[16] Němcová, V., **Slováci za šrotovné kupují nejvíc renaulty a dacie**, IDNES.cz, 31 March 2009, available at:

<http://ekonomika.idnes.cz/slovaci-za-srotovne-kupuji-nejvic-renaulty-a-dacie-f8k-/eko-zahranicni.asp?c=A090331_145001_eko-zahranicni_vem (accessed 15 September 2009), 2009.

[17] Novotný, R., **Šrotovné: Co je a co není vidět?**, Mesec.cz, 15 April 2009, available at: http://www.mesec.cz/clanky/srotovne-co-je-a-co-neni-videt/ (accessed 1 February 2011), 2009.

[18] Renault, **Renault – 2009 Annual Report**, Renault: Boulogne-Billancourt, 2010.

[19] Schweinfurth, A., Car-Scrapping Schemes: An Effective Economic Rescue Policy?, **The Global Subsidies Initiative: Policy Brief**. December, 2009, pp. 1–8.

[20] Statistik Austria, **April 2009: Ökoprämie lässt Pkw-Neuzulassungen um 12,8% zulegen**, Statistik Austria, 12 May 2009. Available at: <http://www.statistik.at/web_de/presse/pressemitteilungen_laufendes_jahr/5/036477?year=2009&month=5 (accessed 10 September 2009), 2009.

[21] Vokswagen, **Volkswagen AG Annual Report 2009**, Volskwagen AG: Wolfsburg, 2010.

[22] Weerts, S., **A Close Look at China Auto Sales Data. Seeking Alpha**,18 August 2009, available at: http://seekingalpha.com/article/156693-a-close-look-at-china-auto-sales-data (accessed 14 September 2009), 2009.

[23] Wise, D. H., Differential Pricing and Treatment by New-Car Salesman: The Effect of the Prospect's Race, Sex, and Dress, **Journal of Business**, Vol. 47, No. 2, 1974, pp. 218–230.

Developing and launching the INNOPENA Internet platform to unblock the flow of innovative solutions between R & D and industry in Poland

Romuald I. ZALEWSKI
Poznań University of Economics, Poznań, Poland
and
Eulalia SKAWIŃSKA
Poznań University of Technology, Poznań, Poland

Summary
This paper presents actual, unsatisfactory innovative activity of processing industry in Poland and one of its regions named Wielkopolska (Greater Poland). The analysis uses data from international and internal sources, as well as results of special research programmes. Various barriers were identified and particularly very low level of cooperation between enterprises and R&D sector represented by universities, institutes of Polish Academy of Science and others. The aim of this paper is to present newly established Internet platform INNOPENA®, which was built relying on the 'open innovation' concept. There is a hope, that this new instrument will improve cooperation for better innovation activity, especially in favour of small and medium enterprises.

Key words: innovation, knowledge economy, open innovation, innovative platform, INNOPENA®

Introduction

Research and development activity refers to the transformation of money into knowledge.
Innovations refer to the transformation of knowledge into money.

Esko Aho, former Prime Minister of Finland

The most important milestones in the progress of civilization are occupied by broadly understood technological innovations and achievements in technical sciences. They allow for production and refinement of new products and services.
Present key technologies are characterized by high expenditures on R & D, the complexity of solutions, decreased time to enter the market and shorter product's life cycle, rapid diffusion of innovation, rapid aging of investment and technology, and in the initial period by high financial and market risk. According to the global statistics, a given branch of industry is considered high technology if the expenditure on R & D exceeds the value of 5 % and even much more as in the case of biotechnology, nanotechnology, and pharmaceuticals. Companies specializing in such industries currently spend up to 20 % of the revenue coming from their use on R & D. Research and development activities last even several years. These conditions create a technological gap between the leading countries and those catching-up. In Poland, compared to most developed economies in the world, including some European Union countries, this gap is quite significant. Changing this state should be one of the most important objectives of the Polish economic policy as well as the scientific and technological one.
Poland is currently doing well with the economic crisis and challenges of new markets. The country's economy is more modern, innovative and competitive than a few years ago. But still the resources of research and development facilities and scientific institutes are not being utilized. Expenditures on R & D are small, which can lead to a trap of median income (approximately 70% of the average GDP *per capita* in the EU), or medium innovativeness and competitiveness [Rate]. The article attempts to present the current situation in this regard, especially the intensity of cooperation between the processing industries with the sphere of science, as the determinants of improvement. Data analysis involves several variables for Poland and in meso-scale for Greater Poland Voivodeship. The analysis used data based on GUS statistics (Science and Technology 2010, PNT-02 reports for the year 2010), survey data (Talaga, Zalewski, in press) and the international rankings. However, the main aim is to present a new and unique platform - INNOPENA ® which aims to facilitate cooperation between small and medium enterprises (SME) and the sphere of R & D based on the model of 'open innovations'.

1. A few comparisons

Poland's innovativeness in the macro scale has not been improving over the years. Poland belongs to a group of catching-up states among the other EU countries. Competitiveness, however, has been improving steadily and in 2011 Poland ranked 34th among 58 countries (IMD 2010). Also the entrepreneurship of managers increased in 2010 (from 33rd position in 2008) to 6th position most probably due to good growth of GDP in 2009 and 2010 during the current economic crisis.
There is a widespread belief that the innovative activity of enterprises depends on available resources, among which financial ones play a major role. According to GUS (Central Statistical Office) data, internal expenditure on research and development in 2009 totaled 9.07 billion zł. This represented 0.74% of GDP or about 238 zł per Polish citizen (Year 2011). These expenditures situate Poland on the 22nd place in the EU. A very low share of expenditure on R & D in the structure of budget expenditure has been quite disturbing over the period of transition. It is several times lower than the Lisbon target for the EU (3%) and three times smaller than the average for the EU-27, which equals to 2.0% (Eurostat).
According to Bukowski and Śniegocki (2012) ... 'there, unfortunately, prevails a malicious conviction that the state does not have to spend more on research than it does

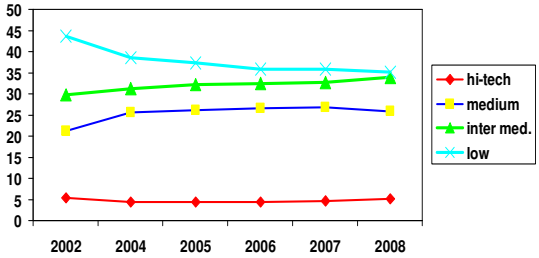
today, because it is the role of the private sector ...'. Representatives of this view do not realize that public money serves as a point of ignition for private innovativeness through the development of technical infrastructure and research. When public funding of R & D exceeds the critical threshold of approximately 0.7% GDP, which is about 0.3 -0.4 %age points more than in Poland today, the tendency of companies to undertake risks linked to innovativeness increases greatly (Bukowski, Śniegocki 2012). It becomes necessary to increase investments in R & D in order to slowly move towards the average EU level.

The responsible management of innovativeness in Poland requires, in the next few years, to double the public expenditure on education and R & D. Just to ensure that every year it rises by nearly 1.5 billion zł which is about 0.1 % GDP. So, it is as much as it has been spent on preparations for the EURO 2012 in the last four years. Causes of low expenditures on R & D in Poland are believed to stem also from the absence of fiscal mechanisms to support research and development and the little importance of the state in creating innovation-friendly environment.

Not only the level but also the structure of expenditure regarding innovativeness is incorrect. The budget is dominated by 60.4%, while enterprises are only involved in 27.1%. The other shareholders are composed of foreign funds - 5.5%, higher education - 6.7% and private non-profit institutions - 03%. International statistics shows that highly innovative countries have an inverse investment structure with the dominance of private entities and only a few dozen % share the state budget. This is the case of Japan, the United States and many other countries (Zalewski, Skawińska 2011).

The lack of funding of the R & D sphere affects the number of professionals employed in this sector. In 2008 only 4.7 persons out of every 1,000 employees worked in research and development, while in Germany it was 13.4, in the UK - 11.7 and in Finland even 22.4. There has been a reduction in the number of individuals employed in research and development facilities in recent years. Krzysztof Kurzydłowski, director of the National Research and Development Center believes that '... the number of scientists in Polish industry is dramatically low. You have to associate the leading companies with research centers ... '(2012). More and more companies invest in ambitious projects.

Figure 1. Structure of industrial output by level of technology



Source: based on Science and Technology, 2010.

The weak financial situation of the research and development sector is reflected in the structure of sales in industrial processing. In 2009 the share of high-tech products amounted to 5.36% (slight increase compared to 2002), of medium high technology products - 26.3% (stagnation), of medium low technology - 33.75% (slight increase) and of low technology 34.6% (slight decrease). Favorable evolution of high-tech and inter-med is very slow (Fig. 1). Still more than 68% of all goods sold belong to a group of not modern goods, which indicates a high proportion of raw materials and low processed products in the volume of production.

Polish industrial enterprises purchased 463 licenses in 2009 but sold only 82. The Patent Office reported 2899 inventions and 1536 patents issued along with 431 protection rights. Per 1 million inhabitants in Poland there are only 3.61 of European patents, while in Sweden this number amounts to 280 and in Germany to 283. Poland occupies the last position in this ranking (Year 2010). The reason for this is the lack of culture regarding protection of intellectual and industrial property rights - proven by several years of follow-up reports, PNT-02 (Zalewski, Talaga, 2011).

Greater Poland Voivodeship occupies places above the average in different rankings of innovation and investment attractiveness of Poland (e.g. Analysis of 2008; Zalewski 2011; Brodzicki 2010). It can therefore be accepted as a good regional representative of the whole country. Reports of the Central Statistical Office on the innovative activity for the years 2008-2010 were filed by 3479 manufacturing companies from Greater Poland. Out of them 559 subjects identified themselves as actively innovative. This shows an increase in the number of entities to fill the PNT-02 form as compared to 2006-2008, while there has been a reduction by 102 entities in the number of innovative enterprises.

Private companies dominate the group of innovative enterprises - 551. The numbers in terms of employment are nearly equal in the case of small and medium-sized entities (230 and 243 respectively).

The share of high technology companies in product and process innovations remains very low in relation to the years 2006-2008 - over 4%. Low technology enterprises dominate in the number of 258 (51%), while the number of medium-high and medium-low technology enterprises remains 26% and 19% respectively.

In light of the statements included in the report the information about the importance of various sources of information for innovative activity is very disturbing. For the biggest number of companies they themselves form the most valuable source of information. That means closing to the inside and insulation from the environment. Market and institutional sources are rated as 'irrelevant' by most companies. Institutes of the Polish Academy of Sciences, research and development units and universities are rated as an 'irrelevant' source of innovation by about 85% of companies. Even the scientific and technical societies are not considered by companies as partners and do not have any stimulating effect.

The level of protection of intellectual property remains an important factor in determining the degree of innovativeness. According to the data obtained from the Central Statistical Office (for 2008-2010) - 229 representatives of manufacturing companies in Greater Poland filed trademark applications in the Patent Office - 113 companies; industrial designs - 51 companies; designs - 29; inventions – 36 companies. A total of 39% of all innovatively active enterprises protect their technical ideas but only five companies (2%) belong to high technology sector.

Independent survey carried out in the summer 2011 on a sample of companies in Greater Poland (Talaga, Zalewski 2011) confirms the trends described above. Most companies (55 to 83%) after 2008 incurred no expenditure on various forms of innovative activity. The exception is the purchase of new machinery, equipment and software, in which 70% of enterprises invested. In the same period at least 66% of companies were not strengthened by the activities that shape the innovative potential of employees (14 criteria), which is evaluated as average. Many respondents believe that they do not have to cooperate with similar entities in the region or with the science sector. For most of them the major sources of competitive advantage in the market will continue to be cost optimization, improvement of processes and products or introduction of new ones. Although these are the important sources of competitive advantage, entrepreneurs fail to acknowledge that the cost advantage loses its meaning. Only about 15% of respondents believe that they will need to change their business model in the future (see Chesbrough 2003). There is a lot of fondness for organic produce, conserving energy and other resources. But only 5% of the representatives of businesses acknowledge opportunities in product and service innovation for the elderly who constitute an increasing proportion of the population.

2. Position of Poland

The above reasons form one of the main causes for Poland scoring 22nd in the European ranking for innovation - the so called Cumulative Index of Innovation in 2010 and 23rd in 2011 among 23 EU countries (EIS 2010). This result allows one to place Poland among the states catching-up more innovative countries. Low innovativeness confirmed by the SII index, the structure of industrial output and the participation of high technology products in exports all affect the buggy position of Poland in terms of competitiveness of its economy. Poland ranked 34th in 2011 in the ranking of 59 world countries, which gives it a 14th place among the EU member states (World 2011).

The sources of innovation for companies are varied and it is worth paying attention to some of them. For most companies the important sources of innovation lay inside themselves, but they exist within the same group of enterprises only for a few firms. Unlike in many other countries, Polish companies do not utilize the achievements of science (universities, Academy of Sciences) and of research and development (R & D) units. The vast majority believes that PAN institutes, research institutes and universities are of no importance as a source of innovation (Zalewski, Talaga 2011). Only a few entities have recognized that these units have high or medium impact on innovation. Many companies also raise the important problem of the lack of qualified staff, information on markets and finding a business partner or a technology.

All the above confirm the existence of barriers that impede innovation and inhibit the activity of almost 84% of enterprises. Without going into intricate and detailed analyses, it can be concluded that companies mostly rely on their own strength and resources in the development of inspiration. Thus, they remain closed for other companies of their group, as well as for competitors from other industries. Innovative activity is costly and often risky (also in financial terms). Approximately 25% of companies indicate that lack of their own or external capital, and the high cost of innovation form the most important obstacles to innovative activity. Also, about 25% of respondents indicated quite the opposite.

3. The barriers between science, R & D and businesses

There is a "wall" between enterprises and the sphere of science which often prevents collaboration, sharing of knowledge, ideas, concepts, information about technology markets and the demand for trained personnel. Documenting the gap in the science-industry co-operation should not be based solely on scientific and statistical studies. This problem is "very hot" as evidenced by the speeches of politicians and especially entrepreneurs in recent months. On April 24, 2012 a list of 500 largest Polish firms was published (Rzeczpospolita 24.04.2012). In an interview on the key stakeholders who have a significant impact on the strategy of these companies none of the respondents mentioned universities, research institutes, NGOs, employers' organizations. We must remember that "the sphere of science possesses knowledge but the key to its use lies in the realm of industry and in Poland there is a deep chasm between these spheres because businesses speak a completely different language than scientists. Cooperation does not belong to the strengths of the Polish society. Therefore, often the potential of an individual does not translate into innovativeness of the public. In addition, companies also rarely cooperate with each other" (Firmy 2011).

There are various causes for the existence of this barrier separating the two areas. We express the view that the faster it falls, the better for the economy. So an ongoing search is taking place for appropriate tools that can be used to weaken and dismantle the barrier. It should be noted that these barriers also exist between business, academia and the third link in the triple helix of innovation (Etzkowitz, Leydesdorff, 1997): national and local authorities. It is difficult to talk about a clear and effective scientific and industrial policy. The belief that "modernization aims to reduce the quantitative differences between us and the West and that concrete and steel are enough to turn Poland into a rich country" (Course 2012) is not enough. Simple reserves of economic growth and competition in global markets are becoming exhausted.

Changing the rules of law in the sphere of economy, finance, education and the science sector is and will be essential to boost this cooperation. State, local government authorities, educational system should all work to improve education and the social capital in Poland. High social capital equals high level of trust, cooperation, integrity, loyalty, respect for norms, solidarity and participation (Study 2010). Without these qualities it is difficult to talk about common construction of relationships and cooperation for the future. They form a source of information exchange needed for the creation of new products and innovations.

A lot has already been achieved to strengthen the Polish innovativeness. For example, Regional Innovation Strategies have been developed. About 61 institutions for technology transfer have been founded in Poland (the so-called science parks, technology, innovation and entrepreneurship incubators, centers of excellence) (Centers 2011). However, their effects for the growth of innovation are staggering (Benchmarking 2010, Józefiak 2006). A similar situation is taking place in Greater

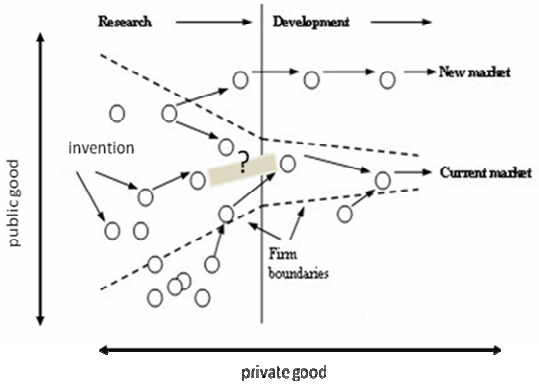
Poland according to recent research (Skawińska, Zalewski, 2012).

According to Jerzy Hausner (2012) "Poland is developing, but it is not using all the possibilities available. This is an extensive development. We are using the available resources obtained from the EU but we are not multiplying them. 'Poland 2030' report, just like the report of the European Union's development, is a vision for the future, but does not provide any means to achieve strategic goals. Jerzy Hausner holds a view 'that the EU money spent on innovation used to spoil the market and companies. We showed an extremely bureaucratic system apparently subordinate to the logic of effectiveness, the system of bureaucratic safeguarding. Money for innovative projects is spent on not innovative ones because they are preferred as those that do not generate any risk. "The creation of e-government system, which is to assist the decision making process constantly, is far from being accomplished.

The Internet as a tool for communication and information in the field of innovation has also been used. There are various online platforms run by central government ministries and agencies. At the local level such platforms have been established by city offices, marshals' offices, technology transfer units, etc. They conduct diverse and rich activity mainly in the field of information, training provision and advice, organization of competitions, trainings, databases etc.

Poland, however, lacks a web platform based on the concept of open innovation (Chesbrough 2003). The idea of open innovation is presented in Figure 2. A few platforms around the world are based on such a scheme - Imocentivie (USA), Innoget (Spain), Prosans (France), Nine Sigma (Japan). The principle of their operation is based on the following: A company reports a specific need for an innovative product, technology, process. Representatives of science and even certain individuals create innovative solutions by strictly defined guidelines. It is even better if such a solution is ready and waiting for the possibility of commercialization. This would speed up the implementation of innovative ideas into the economy.

Figure 2. The idea of open innovation



This gap is to be filled by the INNOPENA ® platform. It was built and launched at the University of Economics in Poznan, the project funded by the National Centre for Research and Development (NCBiR).

4. INNOPENA

The word (acronym) INNOPENA ® comes from the fragments of words in the name of the project. Its title is: „Developing and launching INNOPENA® Internet platform – INNOVATIONS FOR INDUSTRY AND SCIENCE for unblocking the flow of innovative solutions in the economy". Acronym INNOPENA ® has a high level of positioning in search engines (e.g. Google) in the context of words like innovation.

The platform provides anonymity, safety, protection of intellectual property rights, business secrets and handles the entire procedure and all processes. Figure 3 shows the sketch of home page of the platform.

Figure 3. Sketch of INNOPENA home page

The platform is to improve information exchange and communication between entities, especially small and medium size, mating the partners of economic sector with R & D (universities, Academy of Sciences, research and developments units) for mutual and faster completion of development and innovative projects in enterprises.

Through **INNOPENA ®** it will become possible for companies to report problems with current business processes and search for ideas for new products, processes or organizational and marketing solutions. This will facilitate horizontal (inside a branch of industry) and / or vertical cooperation (between industries) for the exchange of knowledge, ideas, information about new and innovative ways of feasible solutions (in other firms / branches of industry). This will help to increase the competitiveness of enterprises.

Access to the platform is possible at two levels: a guest (see home page) or a registered user. Guests can learn and hear about the purposes of the platform, the rules and principles of its operation, about nominating the needs of entrepreneurs (called **seekers**) for new solutions and the technologies offered for implementation by **Donors**. This information is provided in a keyword format and in a general fashion (without details). To supplement the common space in the platform there are: cloud tags, information about internships and practices for students, newsletter and e-learning module. After a separate log on to the latter module materials become available to extend the knowledge of innovativeness, competitiveness, entrepreneurship, commercialization of knowledge, etc. The current content of the module (8 subjects) will be expanded as user demand grows.

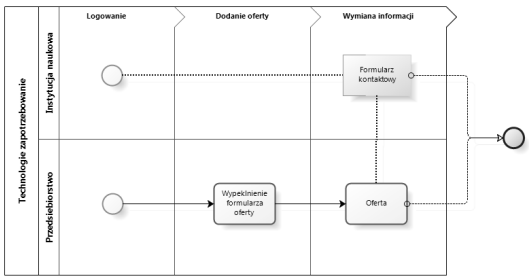
The second level of access to the platform makes full use of its resources after accession (registration) as a business, organization or an individual and determining the purpose (as a client in search of innovation, offering a solution to the donor, or both at once). Data from persons

or companies are checked and then a contract is signed guaranteeing business data protection, commercial data protection, intellectual property rights, etc. Further communication takes place between the donors and seekers and experts. Only the registration allows the platform to carry out its mission, namely:

> ***to provide an active assistance to entrepreneurs in developing cooperation and building a culture of openness and cooperation for innovation; The purpose of this activity is solving your problems with the cooperation and help of others - a method now recognized as one of the most effective for enhancing the competitive advantage of companies seeking partners and associating with those who can offer innovations.***

Customers (K) refer to entrepreneurs seeking quick access to innovative solutions to their problems in the environment and those who encountered problems in their innovative activities – a problem they are unable to solve on their own (see Figure 1). The question is whether the project should be stopped, abandoned, or is there a solution outside? For example, about 15% of the surveyed companies from Greater Poland in 2008-2011 abandoned innovative projects. The costs incurred were usually counted as a loss, and results rarely brought any profit. A spin-off business was not created in any case (Zalewski, in press). Donors (D) - are persons and entities from the science sector (universities, R & D units, institutes), from business environment, other companies and individuals who are able to solve customers' problems quickly by taking part in "open innovation". Donors may know the solution to a similar problem in another area of production, process, organizational or marketing. They possess experience, fresh insights, imagination, knowledge; they love challenges and competition and can quickly find solutions in exchange for gratification, recognition and satisfaction.
The platform provides anonymity, safety, protection of intellectual property rights, business secrets and handles the entire procedure with proper agreements. A map of a part of one of the ongoing processes in which the company reports a need for an innovative solution is shown in Figure 4.

Figure 4. Map of processes in the customer (company) - donor relationship



After registration and logging a customer fills in an offer form briefly describing the problem and the amount of fee that they are willing to pay for a solution. Such an offer is posted on the platform and made available to interested donors. In the meantime the offer goes through an iterative process of validation (feasibility study) with the expert appointed by the platform. After the end of validation the offer awaits for response from the donor, who after registering and logging receives a complete and validated documentation of the problem and may proceed with implementation. The proposed solution may be consulted with an expert and / or customer in the iterative process (the 'negotiations room').
During its initial operation INNOPENA ® platform will mainly aim at companies whose profile of production and innovation is similar to the interests and resources of expertise of the Faculty of Commodity Science of the EU in Poznań and commodity science environments in Kraków (University of Economics), Radom (Technical University of Radom), Olsztyn (University of Warmia and Mazury), Gdynia (Naval Academy) and other affiliated in Commission of Commodity Science at Poznań Branch of Academy of Sciences.
The expertise covers the following areas in particular:
- processing and food technology,
- food packaging materials,
- raw household chemicals and cosmetics,
- plastics,
- market research and development and testing of the marketing concept for
  new products,
- control of production processes,
- assessment of testing and inspection laboratories,
- validation of analytical methods,
- quality and environment management systems,
- safety in the food chain.

Conceptual and implementation works are completed. INNOPENA ® platform is ready to tackle challenges at **www.innopena.pl**

Conclusions

The paper shows the main causes for Poland scoring 22nd in the European ranking for innovation - so called Cumulative Index of Innovation in 2010 and 23rd in 2011 among 27 EU members (EIS 2010). In addition it confirms the existence of barriers that impede innovation and inhibit the activity of almost 84% of enterprises. Without going into intricate and detailed analyses, it can be concluded that companies mostly rely on their own strength and resources in the development of inspiration.

There is a "wall" between enterprises and the sphere of science which often prevents collaboration, sharing of knowledge, ideas, concepts, information about technology markets and the demand for trained personnel.

This paper shows an instrument designated to fill this gap: it is the INNOPENA ® platform. It was built on the open innovation concept and launched at the University of Economics in Poznan. Through **INNOPENA ®** it will become possible for companies to report problems with current business processes and search for ideas for new innovative products, processes or organizational and marketing solutions. This will facilitate horizontal (inside a branch of industry) and/or vertical cooperation (between industries) for the exchange of knowledge, ideas, information about new and innovative ways of feasible solutions (in other firms / branches of industry). This will help to increase the competitiveness of enterprises.

During its initial operation INNOPENA ® platform will mainly aim at companies whose profile of production and innovation is linked to: safety in the food chain, processing and food technology, food packaging materials, raw household chemicals and cosmetics, plastics, market research, development and testing of the marketing concept for new products, statistical control of production processes, assessment of testing and inspection laboratories, validation of analytical methods, quality and environment management systems.

Acknowledgements: The publication was financed by the National Centre for Research and Development (project NR11-0026-10).

References

1. Analiza porównawcza innowacyjności regionów w Polsce w oparciu o metodologię European Innovation Scoreboard, Inst. Techn.. Eksploatacji-Państwowy Instytut Badawczy, Radom, 2008.
2. Badanie kapitału społecznego w Wielkopolsce: diagnoza stanu i perspektywy wzrostu, red. E. Skawińska, Wyd. WSB Poznań 2011.
3. Benchmarking of technological parks in Poland. Report for 2010. PARP, Warszawa 2010.
4. Brodzicki T., Ranking atrakcyjności inwestycyjnej województw Polski, Instytut Rozwoju, Warszawa 2010.
5. Chesbrough H. Open Innovation: The Imperative for Creating and Profiting from Technology, Harvard Business School Press, Boston Mass., 2003.
6. EIS, Innovation Union Scoreboard, 2010 i 2011.
7. Eurostat (http://epp.eurostat.ec.europa.eu)
8. Hausner J. Zielona wyspa samozadowolenia, *Rzeczpospolita* 27.02.2012. str B6.
9. Etzkowitz H., Leydesdorff L. In the Global Knowledge Economy: A Triple Helix of University – Industry – Govermnment Relations, Cassell, London, 1997.
10. Firmy muszą postawić na innowacje, *Rzeczpospolita* 03.11.2011, str B 8-9.
11. Józefiak C., Pobudzanie innowacji w : Raport o innowacyjności gospodarki Polski w 2006 roku, PAN, pod red. T. Baczko.
12. Kurs na południe czy na północ, Rzeczpospolita 20.03.2012, str. B8.
13. Rocznik Statystyczny RP, GUS Warszawa 2011.
14. Talaga Ł, Zalewski R.I., Innovativeness of manufacturing companies in the Greater Poland Voividship in 2011, Towaroznawcze Problemy Jakości 4(29), 2011, 9-21.
15. Talaga Ł., Zalewski R.I. Innowacyjność przedsiębiorstw przetwórstwa przemysłowego w Wielkopolsce w latach 2008-2010 (raport z badań).
16. World Competitive Yearbook 2010, 2011, Institute of Management Development, Lozanna.
17. Zalewski R.I., Rola i miejsce innowacji w burzliwym otoczeniu gospodarczym w: Filozofia TQM w zrównoważonym rozwoju, Wyd. PIB Radom, 2008, 11-27.
18. Zalewski R.I. Aktywność innowacyjna województw w Polsce w latach 2006-08, Przedsiębiorstwo Przyszłości, 4(9), 16-30 (2011).

Refinement of time Petri nets semantics in conflict situations

Adrien Bullich, Hanifa Boucheneb and Olivier H. Roux,

***Abstract*—This paper deals with time Petri nets, where a firing interval is associated with each transition. Three semantics (intermediate, atomic and persistent atomic) are proposed, in the literature for this model, in the context of single-server/multi-server and strong/weak semantics. This paper shows that, in presence of conflicts, these semantics may exhibit some unexpected behaviours and properties. This paper proposes a new semantics more appropriate to deal with conflicts.**

***Index Terms*—Time Petri nets, semantics, conflicts, zone based graph**

I. Introduction

Increasing complexity of systems used nowadays requires rigorous formalisms and tools to automatically verify and control their behaviours. From this perspective, several formalisms such as Petri nets, automata and logics have been developed. Their purpose is to represent, using mathematical concepts, systems in order to be able to verify and control the conformity of their behaviour w.r.t. their expected services.

Automata and Petri nets are designed to model discrete systems. In the context of real time systems, where the behaviour is dependent of time, the used formalisms must integrate explicitly the time factor. Timed automata and time Petri nets appear in order to model hybrids systems, handling discrete systems with continuous variables, i.e. the time. Hybrids systems model both the process and the control system.

Many ways exist to consider time in Petri nets. The time constraints may be expressed in terms of stochastic delays of transitions (stochastic Petri nets), fixed values associated with places or transitions ({P,T}-Timed Petri nets) [13], or intervals labeling places, transitions or arcs ({P,T,A}-Time Petri Nets) [6], [9]–[11], [14].

For {P,T,A}-Time Petri Nets, there are two firing semantics: Weak Time Semantics (WTS) and Strong Time Semantics (STS). For both semantics, each enabled transition has an explicit or implicit firing interval derived from time constraints associated with places, transitions or arcs of the net. A transition cannot be fired outside its firing interval, but in WTS, its firing is not forced when the upper bound of its firing interval is reached. Whereas in STS, it must be fired within its firing interval unless it is disabled. The STS is the most widely used semantics. There are also multiple-server and single-server semantics. The multiple-server semantics allows to handle, at the same time, several time intervals per place (P-TPN), per arc (A-TPN) or per transition (T-TPN), which it is not allowed in the single-server semantics.

In this paper, we consider T-time Petri nets (Merlin's model) [10], called here time Petri nets (TPN in short), in the context of single-server and strong semantics. It seems to be strongly appropriate with communications protocol. In this model, a time interval is associated with each transition. From the semantic point of view, a clock is associated with each transition to measure its enabling time. A transition is firable if its clock has reached its interval and must be fired before overpassing its interval, unless it is disabled by another firing. In [2], three semantics intermediate, atomic and persistent atomic are discussed for time Petri nets. They differ in the way that clocks are handled (memory policies), when a transition is fired. The intermediate semantics resets clocks of all transitions disabled when input tokens of the fired transition are consumed (intermediate marking). The atomic and persistent semantics suppose that the firing of a transition is atomic and do not consider the intermediate marking.

In general, the intermediate semantics is weakly expressive, w.r.t. the weak timed bisimulation, in comparison with the atomic and the atomic persistent ones [2]. But, for time Petri nets with upper-closed intervals[1], the three semantics are equivalent w.r.t. the weak timed bisimulation [2]. From the practical point of view, the atomic and persistent atomic semantics are more appropriate for the specification of observers of systems [2]. The intermediate one seems to be closer to the intuitive interpretation and, for this reason, is widely used. By intuitive interpretation (or semantics) of time Petri nets, we mean that a transition may be fired if it is maintained continuously enabled (using the same tokens) until reaching its firing interval (i.e., its enabling time is inside its firing interval).

In this paper, we first show that, in presence of conflicts, these semantics may exhibit some unexpected behaviours and properties w.r.t. the intuitive semantics. Then, we propose a new semantics more appropriate to deal with conflicts.

This paper is organized as follows. Section II is devoted to the definition of Merlin's model [10] and a short review of the different semantics proposed in the literature for this model. In Section III, we show, by means of examples, that in some conflicting situations, the scenarios obtained w.r.t. the intermediate semantics doesn't respect the constraint of waiting time. In Section IV, we propose a new semantics

A. Bullich and O. H. Roux are with laboratory IRCCyN, 1, rue de la Noë, 44 300 Nantes, France.
E-mail: adrien.bullich@irccyn.ec-nantes.fr, olivier-h.roux@irccyn.ec-nantes.fr

H. Boucheneb is with the Department of Computer Engineering and Software Engineering, École Polytechnique de Montréal, P.O. Box 6079, Station Centre-ville, Montréal, Québec, Canada, H3C 3A7, Québec, Canada.
E-mail: hanifa.boucheneb@polymtl.ca

[1] An upper-closed interval is an interval of the form $[a, b],]a, b], [a, \infty[$ or $]a, \infty[$.

whose idea is to measure the waiting time of each token. Finally, the conclusion is presented in section V.

II. Time Petri nets and their semantics

A. Timed transition system

Usually the semantics of a timed model is defined by means of a timed transition system, where the set of states of the model, its actions as well as its transition relation between states are defined. The actions of a timed model are of two types: discrete actions for events and positive real numbers for time elapsing.

Formally, a timed transition systems is defined by a 4-uplet $< Q, q_0, \Sigma, \rightarrow >$ where Q is a set of states, $q_0 \in Q$ is the initial state, Σ is the set of discrete actions (disjoint from the time domain $\mathbb{R}^+$ of the continuous actions), and $\rightarrow \in Q \times (\Sigma \cup \mathbb{R}^+) \times Q$ is the transition relation. A tuple $(q, a, q') \in \rightarrow$, also denoted $q \xrightarrow{a} q'$, represents the transition from state q to state q' by the discrete or continuous action (time progression) a.

B. Definition of TPN

Petri nets, introduced by Petri in 1962, with their useful abbreviations and extensions are a powerful formalism, which allows precise modeling and analysis of complex systems, using a wide range of methods and tools.

This paper deals with time Petri nets, a simple yet powerful model useful to model and verify real time systems, like communications protocol. This model associates a firing interval with each transition. It allows to model different kinds of time constraints (delays, durations, deadlines, etc.), even if the exact delays or durations of events are not known. Formally, a TPN is defined by a 7-uplet: $< P, T, Pre, Post, \alpha, \beta, M_0 >$ where:

- P is the set of places in the net;
- T is the set of transitions (s.t. $P \cap T = \varnothing$);
- $Pre \in [P \times T \longrightarrow \mathbb{N}]$ is the backward incidence function, indicating, for each transition, the tokens needed for its firing;
- $Post \in [P \times T \longrightarrow \mathbb{N}]$ is the forward incidence function, indicating, for each transition, the tokens produced by its firing (we denote $C = Post - Pre$ the incidence function);
- $\alpha \in [T \rightarrow \mathbb{Q}^+]$ is a function which associates with each transition the lower bound of its firing interval;
- $\beta \in [T \longrightarrow \mathbb{Q}^+ \cup \{\infty\}]$ is a function which associates with each transition the upper bound of its firing interval;
- $M_0 \in [P \longrightarrow \mathbb{N}]$ is the initial distribution of tokens in places, called the initial marking.

For convenience, we denote $t^\bullet = \{p \in P | Post(p,t) > 0\}$ the set of output places of t and ${}^\bullet t = \{p \in P | Pre(p,t) > 0\}$ the set of input places of t. We suppose here that ${}^\bullet t \neq \emptyset$, for every transition of the net.

A time Petri net evolves according to two aspects: the marking and clocks. Thus, we can represent the global state of a time Petri net by a pair (M, ν), where $M \in [P \longrightarrow \mathbb{N}]$ is a marking of the Petri net, and $\nu \in [T \longrightarrow \mathbb{R}^+]$ is a clock valuation over T, which associates with each transition, the value of its clock. Its initial state is $(M_0, \mathbf{0}_T)$, where M_0 is the initial marking and $\mathbf{0}_T$ is the null valuation over T.

A transition t is enabled in M, if there are enough tokens in M for its firing (that means $M \geq Pre(.,t)$). The firing of t takes no time but leads to the marking M' obtained by consuming tokens of $Pre(.,t)$ and producing tokens of $Post(.,t)$: $M' = M - Pre(.,t) + Post(.,t)$.

In time Petri nets, a transition is firable at state (M, ν) iff it is enabled and its clock has reached its associated interval. We denote $firable(M, \nu)$ the set of transitions firable at state (M, ν): $firable(M, \nu) = \{t \in T | M \geq Pre(.,t) \wedge \nu(t) \in [\alpha(t), \beta(t)]\}$.

When a transition t is fired, a new marking is reached, where we can find some newly enabled transitions. We denote $\uparrow enabled(M,t)$ the set of transitions newly enabled in the marking reached from M by firing the transition t. It indicates the clocks that are reseted when t is fired from M.

In time Petri nets, all clocks of transitions evolve uniformly with time. We denote $\nu + d$ the function ν' such that $\forall t \in T, \nu'(t) = \nu(t) + d$. It specifies the evolution of time by d units.

The behaviour of a time Petri net is defined by means of the following timed transition system $< Q, q_0, \Sigma, \rightarrow >$, where $Q = (P \longrightarrow \mathbb{N}) \times (T \longrightarrow \mathbb{R}^+)$ is the set of states of the time Petri net, q_0 is its initial state, $\Sigma = T$, and $\rightarrow$ is composed of continuous and discrete transitions defined as follows:
Let (M, ν) be a state, $d \in \mathbb{R}^+$, $t \in T$, M' a marking and ν' a clock valuation over T.

- Continuous transition:
  $(M, \nu) \xrightarrow{d} (M, \nu')$ if
  $$\begin{cases} \nu' = \nu + d \\ \forall t \in T : t \notin firable(M, \nu + d) \Rightarrow \\ \quad (\forall d' \in [0, d] : t \notin firable(M, \nu + d')) \end{cases}$$
- Discrete transition:
  $(M, \nu) \xrightarrow{t} (M', \nu')$ if
  $$\begin{cases} t \in firable(M, \nu) \\ M' = M - Pre(.,t) + Post(.,t) \\ \forall t' \in T : \nu'(t') = \begin{cases} 0 \text{ if } t' \in \uparrow enabled(M,t), \\ \nu(t') \text{ otherwise.} \end{cases} \end{cases}$$

C. Semantics of TPN in the memory policy

Different semantics can be derived from the transition system given above, depending on the definition of the notion of newly enabled (memory policy) [1], [2]. The definition of this notion has an impact on the behaviour and the properties of the net. In [2], the authors have distinguished three memory policies: intermediate, atomic and persistent atomic semantics.

a) Intermediate semantics: In the intermediate semantics, the firing of a transition consists of two steps: consuming tokens and producing tokens. A distinction is then made between tokens used by a transition and those produced. All transitions not enabled in the marking resulting from the first step (intermediate marking) but enabled, in the marking resulting from the second step, are newly enabled. In other words, let M be a marking, t and t' two transitions. The transition t' is (newly) enabled by firing t from M, if in the intermediate marking of this firing, t' is disabled ($Pre(.,t') \not\leq M - Pre(.,t)$), but is enabled after the firing of t ($Pre(.,t') \leq M + C(.,t)$). Moreover, in the context of single-server semantics, the intermediate semantics resets the clock of the fired transition ($t = t'$). Therefore, the set of the newly enabled transitions by firing t from M is defined by:

$$\uparrow enabled(M,t) = \{t' \in T | Pre(.,t') \leq M + C(.,t) \wedge$$

$$(Pre(.,t') \not\leq M - Pre(.,t) \ \vee \ t = t')\}$$

b) Atomic semantics: In the atomic semantics [2], the firing of a transition is supposed atomic. Unlike the intermediate semantics, the atomic one does not consider the intermediate markings. In this case, all transitions not enabled before firing a transition but enabled after its firing are newly enabled. More precisely, let M be a marking, t and t' two transitions. The transition t' is (newly) enabled by firing t from M, if it belongs to the following set:

$$\uparrow enabled(M,t) = \{t' \in T | Pre(.,t') \leq M + C(.,t) \wedge$$

$$(Pre(.,t') \not\leq M \ \vee \ t = t')\}$$

c) Persistent atomic semantics: In the persistent atomic semantics [2], a transition t' is newly enabled after the firing of t in the marking M if it belongs to the set:

$$\uparrow enabled(M,t) = \{t' \in T | Pre(.,t') \leq M + C(.,t) \wedge$$

$$Pre(.,t') \not\leq M\}$$

The difference between the atomic semantics and the persistent atomic one lies in the particular case of the fired transition. If the fired transition enables again itself, its clock is reseted in the atomic semantics but not reseted in the persistent atomic semantics. It is then considered as a newly enabled transition in the atomic semantics but not newly enabled in the persistent atomic semantics.

D. Age or threshold semantics

Boyer, in [5], considers two kinds of semantics, according to the meaning of clocks. In the first one, the *age semantics*, the firing condition is about the waiting time of tokens. When a transition is labelled $[a; b]$, it means that tokens must wait between a and b time unit to fire. It is the case of the machining, for example. In the second one, the *threshold semantics*, tokens are not distinguished, only matters the load. When a transition is labelled $[a; b]$, it means that the number of tokens must be greater than the weight during a to b time unit to fire the transition. This conceptualisation is adapted to load mechanisms.

Different models with age semantics exist in the literature, like P-time Petri nets and A-time Petri nets. In these two models, the fire of a transition considers the age of tokens and not a number. However, it seems less used in T-time Petri nets. We'll see for example that intermediate, atomic and persistent atomic semantics are threshold semantics.

III. Problem of handling conflicts

The way that conflicts are handled differs from one semantics to the other, leading to different behaviours and properties. So, it is essential for a semantics to be clearly and coherently defined, in order to avoid incoherences, in the manner that similar situations are managed. To be intuitive is a true advantage for a semantics.

The intermediate semantics distinguishes tokens, thanks to the intermediate step, between produced tokens and the others. However, it is important to note that the intermediate semantics is not an age semantics. That's the matter of our first study: what difference we can find in case of conflicts.

Let us explain, by means of examples, such conflicting situations and their impact on the behaviour and properties of the model.

A. Change of behaviour

Consider the net at Fig. 1. According with the waiting time of tokens, the system is expected to behave as follows: the token in place $P1$ goes to the place $P2$ at date 1. The initial token in $P2$ is either consumed by $T2$ or $T3$ at date 1. The token created by $T1$ in $P2$ should be consumed at date 2.

Suppose now that the transition $T1$ is fired before the others, from the initial marking $P1 + P2$. This firing leads to the marking $2P2$, where both transitions $T2$ and $T3$ are enabled and not in conflict. These two transitions were enabled but in conflict in the initial marking (before firing $T1$). Let us examine how the three memory policies handle this situation.

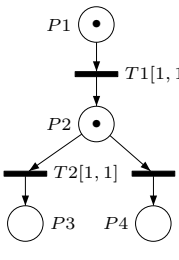


Fig. 1. A TPN with two conflicting transitions

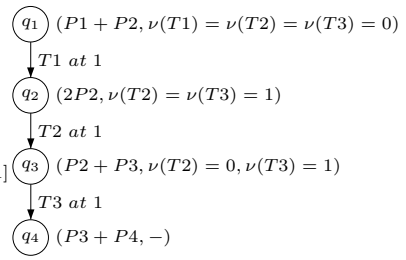


Fig. 2. An unexpected run of the TPN at Fig. 1

In the intermediate semantics, when the transition $T1$ is fired from the state $(P1+P2, \nu(T1) = \nu(T2) = \nu(T3) = 1)$, clocks of both transitions $T2$ and $T3$ are not reseted, since they are enabled before, during and after firing $T1$. With this semantics, the firing of $T1$ leads to the state $(2P2, \nu(T2) = \nu(T3) = 1)$. From this state, the model behaves as if both tokens in $P2$ were created at the same time (at date 0). For instance, if the first token in $P2$ is used by $T2$, then $T3$ is considered as not newly enabled and fired immediately after $T2$ (see Fig. 2). Unlike what is expected, the two tokens reach their destination at the same time. Transition $T3$ is fired even if its token is just created and did not wait the one time unit needed. The constraint of waiting time is then not respected for $T3$. From this point of view, this semantics seems to be incoherent with the context of age semantics.

Both the atomic and persistent atomic semantics accept the same unexpected run.

B. Change of properties

Let us now show that the change in managing conflict, pointed out in the intermediate semantics, may have an impact on the properties of the model. Consider the TPN at Fig. 3. The TPN is bounded w.r.t. the intermediate semantics but is unbounded w.r.t. the intuitive semantics.

In this net, the role of the transition $T3$ should be to empty the place $P4$ and then prevent the system to reach a state where $P5$ is marked. From such a state, the transition $T5$ will be repeatedly fired every one time unit, leading to an infinite number of markings (unbounded net).

In the intermediate semantics, from the initial state $(P1 + P2 + P4, \nu(T1) = \nu(T2) = \nu(T3) = 0)$, the model can fire the transition $T1$ at date 2 to reach the state $(2P2 + P4, \nu(T2) = \nu(T3) = 2)$. From this state, $T2$ and $T3$ are fired successively at dates 3 and 4, leading to the dead marking $P3$ (see Fig.4). Therefore, the enabling time constraint is not respected for $T3$. Indeed, in this scenario, the transition $T3$ uses the token created by $T1$ and then should be fired 4 time units after $T1$ (at date 6), unless it is disabled by firing a conflicting transition. After firing successively transitions $T1$ and $T2$ at dates 2 and 3, both transitions $T3$ and $T4$ are enabled but in conflict. They are both firable at date 6. The firing of $T3$ will disable $T4$ and mark the place $P5$ and then enable $T5$. The model is then unbounded w.r.t. the intuitive semantics.

Note that the run given at Fig. 4 is also valid in the context of the atomic semantics. For the persistent atomic semantics, the accepted run coincides with the one given at Fig. 4 until firing $T2$. In the persistent atomic semantics, as the firing of $T2$, at date 3, enables again $T2$. Its clock is then not reseted and the state reached by $T2$ is $(P2 + P3 + P4, \nu(T2) = \nu(T3) = 3, \nu(T4) = 0)$. From this state, the transition $T2$ is fired again at date 3, which disables $T3$. The reached state is then $(2P3 + P4, -)$.

We have shown that for the same net, the way that conflicting transitions are managed may have an impact on the behaviour and the properties of the model. It is then very important to understand how the different semantics handle the subtle cases of conflicts and to be sure that the behaviour of the model, w.r.t. a given semantics, corresponds exactly to the expected behaviour.

Note that a TPN can be unbounded w.r.t. the intermediate, atomic or persistent atomic semantics but bounded w.r.t. an intuitive semantics about age of tokens. As an example the TPN at Fig. 5 is unbounded w.r.t. the intermediate semantics but bounded w.r.t. the intuitive semantics. Therefore, there is no relationship between properties of the model w.r.t. the intuitive semantics and the others.

IV. A TOKENS SEMANTICS

The enabling time of the transition is different with the age of the youngest tokens used by the transition. To deal with this difference, we need to identify tokens used by each transition and to memorize the age of tokens, as it is evokated in [5]. Doing so, we can make sure that the enabling time of each transition refers to the tokens to be consumed by the transition.

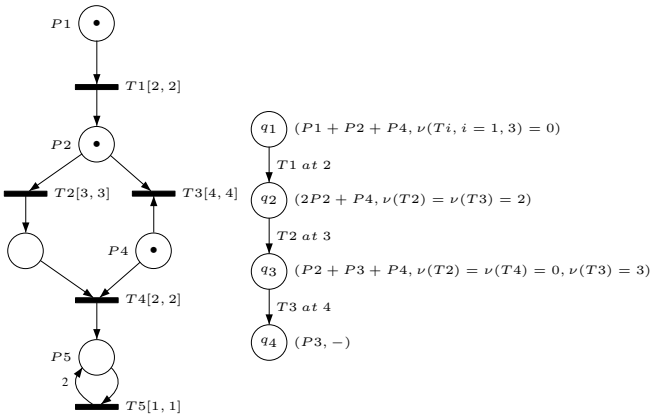


Fig. 3. A bounded TPN w.r.t. the intermediate semantics

Fig. 4. A run of the TPN at Fig. 3 w.r.t. the intermediate semantics

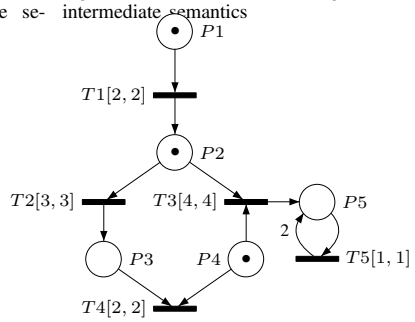


Fig. 5. An unbounded TPN w.r.t. the intermediate semantics

Moreover, in the context of single-server semantics, in the TPN, only one clock is associated with each transition. This clock is used to measure the time elapsed since it was last enabled. The different enabling instances of the same transition are handled sequentially.

To manage well the allocation of tokens to transitions w.r.t. single-server semantics, we associate a clock with each token and a queue of clock values with each place. The tokens of each place p are then handled according to the FIFO (First In First Out) discipline (the first token created in p is the first token consumed from p).

When a token is created in a place p, its clock is set to 0. This value is inserted in the queue of p. All clocks of tokens evolve synchronously with time until they are consumed. An enabled transition is firable if the age of the youngest tokens participating in its enabling has reached the firing interval of the transition. It must be fired without any additional delay, if the age of its youngest tokens has reached the upper bound of its firing interval, unless it is disabled.

A. Formalisation

Formally, we define the TPN state by a pair (M, μ), where M is a marking and μ is a function over P, which associates with each place p a queue of clock values. Each queue is managed FIFO and then ordered from the older to the younger (decreasing order of ages). For example, if a place p has 4 tokens with ages $3, 3, 2$ and 1, its queue is $\mu(p) = [3, 3, 2, 1]$.

The initial state is (M_0, μ_0), where $\forall p \in P, \mu_0(p)[i] = 0$, for $1 \leq i \leq M_0(p)$, if $M_0(p) > 0$, and $\mu_0(p) = []$, otherwise.

As mentioned before, tokens within each place are handled FIFO, an enabled transition will always use the oldest tokens from each place. Thus, if an enabled transition t uses $Pre(p,t)$ tokens from the place p, the age of the youngest token from p used by t is the element $\mu(p)[Pre(p,t)]$ of the queue $\mu(p)$. For example, if $\mu(p) = [3,3,2,1]$ and t needs 3 tokens from p, then it uses 2 tokens with age 3, and 1 token with age 2. The age of its youngest token is exactly its enabling time (i.e., 2).

Let (M,μ) be a state and t a transition enabled in M. We define the enabling time, the firing condition and the set of firable transitions as follows:

- The enabling time of t is the age of the youngest token used by t:
$$\min_{p \in {}^\bullet t} \{\ \mu(p)[Pre(p,t)]\ \}.$$
- The firing condition of t is then:
$$\min_{p \in {}^\bullet t} \{\ \mu(p)[Pre(p,t)]\ \} \ \in\ [\alpha(t), \beta(t)].$$
- The set of firable transitions from a state (M,μ) is:
$$firable(M,\mu) = \{t \in T | M \geq Pre(.,t) \wedge \min_{p \in {}^\bullet t} \{\ \mu(p)[Pre(p,t)]\ \} \in [\alpha(t), \beta(t)]\}$$

Formally, our semantics is defined by the transition system $< Q, (M_0,\mu_0), T, \rightarrow >$ where $Q = ([P \longrightarrow \mathbb{N}]) \times ([P \longrightarrow Queues])$ is the set of states and the transition relation $\rightarrow$ is defined as follows: Let (M,μ) and (M',μ') be two states, t a transition of T and d a nonnegative real number ($d \in \mathbb{R}^+$).

- Continuous transition:
$(M,\mu) \xrightarrow{d} (M',\mu')$ if[2]
$$\begin{cases} M' = M \\ \mu' = \mu + d \\ \forall t \notin firable(M, \mu + d) \Rightarrow \\ \quad \forall d' \in [0;d] t \notin firable(M, \mu + d') \end{cases}$$
- Discrete transition:
$(M,\mu) \xrightarrow{t} (M',\mu')$ if
$$\begin{cases} t \in firable(M,\mu) \\ M' = M - Pre(.,t) + Post(.,t) \\ \mu' \text{ is built this way :} \\ \text{Initialise } \mu' \text{ with } \mu; \\ \forall p \in {}^\bullet t, \text{ pop from } \mu'(p) \text{ the } Pre(p,t) \text{ first elements;} \\ \forall p \in t^\bullet, \text{ push in } \mu'(p) \text{ the element 0, } Post(p,t) \text{ times.} \end{cases}$$

B. Calculus of the state zone graph

The verification of time Petri nets properties is based on abstraction, whose aim is to represent, by removing some irrelevant details, the infinite state space of the model by a finite graph, which preserves properties of interest. There are two well known abstraction techniques used in the literature for the reachability analysis of time Petri nets [4], [8], [12]: the *state class graph* [3] and the *zone based graph* [7]. The basic difference between the two abstractions is the state definition. In the state class graph method, states are defined by a marking and a function, which associates a firing interval with each transition. In the state zone graph, states are defined using clocks as explained in Section II. Both methods can be used in the context of our semantics in a similar way as done for other semantics.

We have implemented a preliminary version of the construction procedure of the zone based graph [4], w.r.t. our semantics. As an example, we report in Fig. 6, the zone based graph obtained for the TPN at Fig. 5.

[2]The operation $\mu + d$ increments with d time units all clock values within queues of μ.

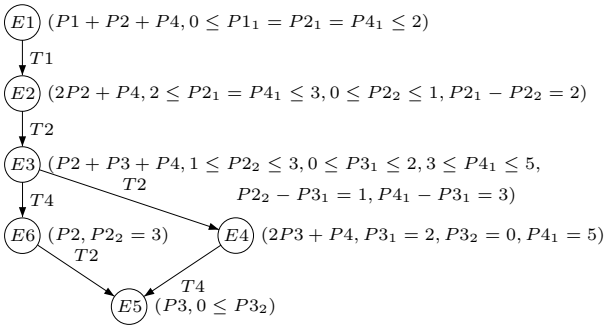


Fig. 6. The state zone w.r.t. our semantics of the TPN at Fig. 5

V. Conclusion

In this paper, we have studied time Petri nets model and its different semantics: the intermediate, the atomic and the persistent atomic one. We have shown that the intermediate semantics is not an age semantics, and wrongly used can involve some unexpected and incoherent behaviours: some tokens may be used by a transition even if, they did not wait the required time. We observ the same phenomenons in the context of the atomic and persistent atomic semantics.

To cope with this problem, we have proposed a semantics based on the age of tokens to deal with conflicts. This semantics associates clocks with tokens and then allows to handle appropriately conflicts, respecting the waiting time of tokens.

We have shown, by means of examples, that a TPN may be bounded w.r.t. the intermediate, atomic or persistent atomic semantics but unbounded w.r.t. our semantics. Reciprocally, a TPN may be unbounded w.r.t. the intermediate, atomic or persistent atomic semantics but bounded w.r.t. our semantics. Therefore, there is no relationship between properties of the TPN w.r.t. our semantics and others.

Note that in non-conflict situation, our semantics can simulate the intermediate semantics by adding a self-loop place to each transition.

Finally, as a perspective, we will investigate the comparison of the expressiveness of our semantics with the others.

References

[1] T. Aura and J. Lilius, *A causal semantics for time Petri nets*, Theoretical Computer Science 243(2), pages 409-447, 2000.

[2] B. Bérard, F. Cassez, S. Haddad, D. Lime, Didier and O. H. Roux, *Comparison of different semantics for time Petri nets*, Third International Symposium on Automated technology for verification and analysis, volume 3707 of LNCS, pages 293-307, 2005.

[3] B. Berthomieu and M. Diaz, *Modeling and verification of time dependent systems using time Petri nets*, IEEE Transactions on Software Engineering, volume 17, number 3, 1991.

[4] H. Boucheneb, G. Gardey, and O. (H.) Roux, *TCTL model checking of time Petri nets*, Journal of Logic and Computation, volume 19(6), pages 1509-1540, December 2009.

[5] M. Boyer and M. Diaz, *Multiple Enabledness of Transitions in Petri Nets with Time*, PNPM01, 2001

[6] M. Boyer and O. H. Roux, *On the compared expressiveness of arc, place and transition time Petri Nets*, journal Fundamenta Informaticae, volume 88, n^o 3, pages 225-249, 2008.

[7] G. Gardey, O. H. Roux and O. F. Roux, *Using zone graph method for computing the state space of a Time Petri Net*, Springer-Verlag, volume 2791 of LNCS, 2004.
[8] R. Hadjidj and H. Boucheneb, *Improving state class constructions for CTL* model checking of Time Petri Nets*, International Journal on Software Tools Technology Transfer (STTT), volume 10, number 2, pages 167-184, 2008.
[9] W. Khansa, J.-P Denat and S. Collart-Dutilleul, *P-Time Petri Nets for manufacturing systems*, International Workshop on Discrete Event Systems, WODES, pages 94-102, 1996.
[10] P. Merlin, *A study of the recoverability of computer system*, PhD Thesis, University of California, 1974.
[11] C. J. Myers, T. G. Rokicki, T. H.-Y. Meng, *POSET timing and its application to the timed cirucits*, in IEEE Transactions on CAD, 18(6), 1999.
[12] W. Penczek and A. Pólrola, *Specification and Model Checking of Temporal Properties in Time Petri Nets and Timed Automata*, In Proc. of ICATPN, Springer-Verlag, volume 3099 of LNCS, pages 37-76, Springer–Verlag, 2004.
[13] C. Ramchandani, *Analysis of asynchronous concurrent systems by timed Petri nets*, PhD Thesis, Massachusetts Institute of Technology, 1974.
[14] B. Walter, *Timed net for modeling and analysing protocols with time*, IFIP Conference on Protocol Specification Testing and Verification, 1983.

Identifying the Evolution Sequence of a Text Document

Katia Mayfield Emanuel S. Grant Eunjin Kim
Department of Computer Science, University of North Dakota
Grand Forks, ND 58201

and

Crystal Alberts
Department of English, University of North Dakota
Grand Forks, ND 58201

ABSTRACT

An issue that has been around in the case of published documents and now with the technology of the Internet is the management of versions of data being created or digitized. The field of Digital Humanities concentrates on the digitization of documents and also supports the creation of "born digital" documents. Considering the texts being digitized and those being created electronically, it can become difficult to determine which published (whether on paper or digital) document is an original, which one is a version of the original, and which one may be a version of a version. In this paper, a sample case of the Emily Dickinson poem *Faith is a fine invention* is analyzed by using the Dempster-Shafer's theory in order to identify the evolution of three versions of the poem.

Keywords: Dempster-Shafer's theory, Digital Humanities, born digital, Text Encoding Initiative, Electronic Literature Directory.

1. INTRODUCTION

One of the main areas of study for scholars in the area of English, under the field of Humanities, is the study of authors and their work. There are several literary works that have different versions created by their own authors or by other agents such as publishers, translators and editors. Depending on the historical time frame of these works, today's scholars may have a difficult time trying to distinguish which one may be the original and which versions were created from the original or from other previous versions. With the new area of Digital Humanities, many documents are now becoming digitized and this creates better chances for scholars to be able to conduct their research by accessing documents electronically, rather than having to travel to locations that house different historical pieces. This electronic access allows the identification of multiple versions of the same text. However, it does not establish the evolutionary path of the document.

Digital Humanities, also sometimes referred to as Digital Computing, is a new field of study where the traditional area of Humanities – English, history, anthropology, literature, etc – work in collaboration with computational sciences [9]. The purpose of this field is to study and implement ways to archive documents that are found on paper or those that are "born digital". "Born digital" data are those that are created by using multimedia systems, basically, that are created on a computer and are specifically made to only be viewable on a computer [9].

According to Matthew Kirschenbaum, within the Humanities field, English departments have been the one's to take advantage of concepts such as Digital Humanities. He gives five reasons to support such claim: text is the most tractable data type to be manipulated by computers, the association between computers and composition, the convergence between editorial theory and method from the 1980s and the means to implement electronic archives and editions of documents, projects around hypertext and other forms of electronic literature, and last is an openness of English departments to be involved in cultural studies where computers and other digital material become centerpieces of analysis [5].

One of the topics that English scholars may focus on is the study of an authors' life. In this area, they may find that an author may have written a piece of work (whether digital or not) and that a few years later the author may have re-written that same piece, creating a different version to it. One well known author that has done this with several of her poems was Elizabeth Barrett Browning. A sample of her poems that can be found to have more than one version are *To Flush, My Dog* (1843, 1844, 1850, 1853, and 1856), *A Child*

Asleep (1840, 1844, 1850, and 1853), *Loved Once* (1844, 1850, and 1853), *To Bettine* (1850, and 1853) [11]. The identification of such different renderings of the same text has been the focus of several studies in Digital Humanities.

Studying how to handle these different versions, Dyreson, Lin and Wang introduced a computational tool that tracks changes to documents, recording modifications applied to the original document or to a previous version [4]. While the tool seems to be very efficient, it requires the versions to be developed under the tool in order to be effective. In another project in this area, Anick and Flynn discuss the implementation of data structures designed to improve the query of database systems where the data content may have been modified, creating different versions of such content [1]. As in the tool proposed by Dyreson, the versioning of the data requires time stamps to be recorded during the modification of the contents [4].

Other studies, covering documents that could have more than one version, did not report such occurrences. For example, Antonacopoulos et. al published a study on the procedures used to transform historical written documents on electronic ones [2]. The study focuses on the recovery of the scanned original text and does not mention the possibility of the existence of more than one version of the same document. Also, Crane, Smith and Wulman describe the implementation of a digital humanities project, involving books, images and maps of London [3]. Such work demonstrates the importance of the digitization of historical documents. However, the authors do not mention the existence of duplicates or second versions of the digitized documents.

This paper focuses on sets of versions of the same text, discussing the use of uncertainty reasoning to identify original documents and the sequence in which versions evolved from that original. A brief introduction to Digital Humanities and a short description of the Dempster-Shafer Theory of Evidence and its association to Belief functions, used in uncertainty reasoning, are presented in the next section. They are followed by the presentation of a poem by Emily Dickinson, which will be used as a case study in this paper. A description on how Belief functions were used to determine the original work and the sequence in which it was modified follows. Finally, a summary of the presented material closes the study.

2. BACKGROUND

According to Kirschenbaum, Digital Humanities is about "scholarship (and a pedagogy) that is publicly visible in ways to which are generally unaccustomed, a scholarship and pedagogy that are bound up with infrastructure in ways that are deeper and more explicit than what scholars are typically accustomed to, a scholarship and pedagogy that are collaborative and depend on networks of people and that live an active 24/7 life online" [5]. Even with this definition of modern digital humanities, there were two challenges that had to be met, those were the establishment of standards and storage.

Today, documents are being created as "born digital" data and old documents are being converted to electronic format for the purposes of archiving. One of the main issues with it all becoming available electronically is establishing a standard that authors and those converting documents should follow. Just like programming with a particular computer language requires a syntax that must be followed for the program to work correctly, with Digital Humanities some authors choose to encode their work with Extensible Markup Language (XML), which has its similarities to HyperText Markup Language (HTML). In particular, when HTML is used, there are standards that must be followed and creators can validate their code through the World Wide Web Consortium [12]. While both of these languages are very similar, XML is used to 'tag' information to make it searchable online while HTML tags provide tools for a web designer to specify how a web page should render. However, XML has no current universal validator like that of W3C for HTML. Considering this being an issue, the Text Encoding Initiative was created to try to establish some sort of standard that those coding in XML should be able to use and "validate" their code [10]. Even though TEI is not a standard used universally, those in the Humanities field are familiar with it and follow these standards when encoding their documents.

Besides trying to establish standards to be followed in this new area, there is also a reasonable concern about storage. There are funded projects that are working on this issue, and there are organizations that have been established to help. The Electronic Literature Directory was created to provide a database for these digital documents to be housed [8]. To submit their documents to the ELD, creators will identify the techniques that were used to create their document, and are required to use XML encoding in their documents where appropriate.

The 24/7 ease of access that Digital Humanities has created for "born digital" documents and those that have been archived, makes it easier not just for scholarly work to be done but also for duplication of

work. Here, it is not being referred to copyright infringements taking place, but by duplication of work considering that there is ease of access, allowing for those who have permissions to work on a document, to be able to collaborate, or work on their own, to create or identify different versions of the document. In such situation, scholars, being able to find different versions of texts that they are researching, may bring up the question of which document is the original, which is a version created from the original, and which are versions created from other version. Scholars may end up with a sense of doubt and uncertainty, which then requires them to conduct additional research to establish the appropriate order of documentation for their research to be accurate. Due to possible lack of information, such order may not be considered exact.

When taking the concept of uncertainty into consideration, in the field of artificial intelligence there are three types of uncertainty that should be considered. First is nonspecificity (or imprecision), which is connected with sized (cardinalities) of relevant sets of alternatives; next is fuzziness (or vagueness), which results from imprecise boundaries of fuzzy sets; and lastly is strife (or discord), which expresses conflicts among the various sets of alternatives [6]. This last type of uncertainty, strife, is the one which best describes the uncertainty being considered among the different versions of documents being examined by scholars.

For the purpose of this paper, the concept of uncertainty is associated to the Dempster-Shafer's theory (DST) which is a mathematical theory of evidence [6]. This theory is based on two dual nonadditive measures: belief measures and plausibility measures [6]. Shafer's framework allows for belief about propositions to be represented as intervals, bounded by two values, *belief* (or *support*) and *plausibility*:

$$belief \leq plausibility$$

The basis for understanding belief and plausibility is to think of having a hypothesis that one must work the basis of their scholarly work on. If we define mass as the proportion of available evidence that supports a claim, the belief measurement is the sum of masses, which are held by all the subsets included in the hypothesis. The belief of a hypothesis will form a lower bound, being its amount that directly supports the hypothesis. The plausibility is considered to be an upper bound of the possibility that the hypothesis is true. By calculating degrees of belief and plausibility on items that describe the sequencing of different versions, a scholar should be able to come to a satisfactory conclusion of the order in which documents were created.

The formal definitions of the Dempster-Shafer's theory establish several mathematical concepts. If one considers X to be universal set: the set representing all possible cases of a problem solution then power set 2^x, known as the power set of X, is the set of all subsets of X, including the empty set. In the Dempster-Schafer theory, each element in the power set can represent the scheme of the state of the system, by representing the states in which the proposition is true. The theory of evidence requires for an expert that can analyze the "scenario" or the existence of statistical data. Based on the expert report or the statistical data, each element that is included in the power set is given a belief mass, which is a value found in the range of [0, 1], and this can be represented by a function:

$$m: 2^x \rightarrow [0, 1]$$

This function is the basic belief assignment and it has a mass for the empty set which is 0, and a second mass for the remaining members of the power set, which all sum up to 1.

$$\sum_{A \in 2^x} m(A) = 1$$

The Dempster-Shafer's theory makes the claim that whichever states belong to the set A but to no specific subset of A. The mass m(A) expresses the proportion of evidence that supports this claim. By assigning masses, the probability interval can be given an upper and lower bound that it falls between. The upper bound is created by the belief measure and the lower bound by the plausibility measure.

$$bel(A) \leq P(A) \leq pl(A)$$

The belief bel(A) for a set A is defined as the sum of all the masses of subsets of the set of interest:

$$\sum_{B|B \subseteq A} m(B)$$

The plausibility pl(A) is the sum of all the masses of the sets B that intersect the set of interest A:

$$\sum_{B|B \cap A \neq \emptyset} m(B),$$

The two measures are related to each other as follows:

$$pl(A) = 1 - bel(\bar{A})$$

To help with analysis, experts can also take into consideration the Evidence Interval. This interval uses the belief and plausibility as minimum and maximum bounds respective. Between these values is where the probability is shown to be true.

$$EI(A) = [bel(A), pl(A)] = [bel(A), 1 - bel(\overline{(A)})]$$

The degree to which A can be disbelieved or refuted can also be calculated and is referred to as Doubt or Dubiety:

$$Dbt(A) = bel(\overline{(A)}) = 1- pl(A)$$

The application of these formulas and theories to a sample case of versioning is discussed in Sections 3 and 4 of this paper.

3. SAMPLE CASE

When documents have more than one version to it, it may become difficult to be able to distinguish which version was created from the original and which versions may have been created from another version. This problem can be considered one of doubt or better yet "uncertainty". According to Klir, "uncertainty can be caused by information being incomplete, imprecise, fragmentary, not fully reliable, vague, contradictory, or deficient in some other way" [7]. Something that can be done is to find a way to be able to measure the amount of uncertainty, focusing next on a new goal which would be to be able to reduce this value of uncertainty.

The poem "Faith is a fine invention" written by Emily Dickinson circa 1860 and originally published in 1891, after her death, is used to demonstrate the concept of versioning.

On a simple online search, different versions of this poem can be extracted from several websites and the different forms it assumes can be easily identified. Three of these different texts, retrieved on November 2011, are:

From http://www.goodreads.com/quotes/show/44777 designated poem 1:

"Faith is a fine invention
When gentlemen can see,
But microscopes are prudent
In an emergency."

From http://www.online-literature.com/dickinson /poems-series-2/32/ designated poem 2:

Faith is a fine invention
For gentlemen who see;
But microscopes are prudent
In an emergency!

From http://www.emilydickinsonmuseum.org/church designated poem 3:

"Faith" is a fine invention
For Gentlemen who see!
But Microscopes are prudent
In an Emergency!

Such poems were target of this study in order to show how the proposed belief functions work in the identification of the versioning process. There are two problems associated to these different versions. The first is to identify the original version, the second is to try to identify which ones were modified to produce the consecutive versions. Therefore, the solution to be considered is that one of the poems is the original in this evolution. Then, assuming that the notation x → y indicates that y is a version derived from x, a relation may be defined showing all possible pairings. However, for a number n of poems, the number of different combination (pairs) of version and original is $n!/(n-2)!$ or simply $n*(n-1)$. In a simple case, for 5 poems, there would be 20 combinations. Then, the superset, which is required in the Dempster-Shafer theory, would have $2^{n(n-1)}$ (in the example $2^{20} \approx$ 1,000,000) elements.

Due to the size of possible combinations to be examined on the Emily Dickinson example, this study was limited to three versions of the poem (designated from now on poem 1, poem 2 and poem 3). In such situation, there are 64 powerset combinations that need to be considered.

In the case of the three poems, six variables were defined to represent the combination pairs:

A = P1→ P2
B = P1→P3
C = P2→P3
D = P2→ P1
E = P3→P1
F = P3→P2

According to the Dempster-Shafer theory, each element of the superset must have a mass value associated to it. In this study, an evidential value was calculated by counting the number of unchanged characters of the poem assumed to be the version (y), based on the poem assumed to be the original (x).

In order to identify the sequence of versions a two-dimensional matrix was created to record the number of unchanged characters between every pair of poems, as seen in Table 1.

Table 1. Unchanged characters between poems

Versions:	Poem 1	Poem 2	Poem 3
Original			
Poem 1	-	83	80
Poem 2	82	-	87
Poem 3	79	87	-

Once the values of the variables A to F were calculated, directly from Table 1, the combinations of those variables had their evidential value computed by adding the individual values. Then, the mass of each variable was established by computing the fraction between the unchanged characters and the total number of evidential values, creating a fraction in the range 0 to 1, with the property of having the summation of all masses equals to 1. Basically, the matrix entries were used to establish the mass (m numbers) to support the beliefs that two versions are connected someway, using the formula

$$m(Z)= \frac{matrix\ count(x,y)}{summation\ value\ of\ matrix\ count},$$

which gives a stronger value to those with smaller changes.

The belief and plausibility of every variable and set could then be calculated using the formulas presented in Section 2. Table 2 shows some of those values.

4. ANALYSIS

In this section, the Dempster-Shafer theory is applied to identify which of the poems is the original version. This is a hypothetical situation where the data was artificially established based on a simplistic assumption that the number of unchanged characters would be less when two poems were related by versioning. This simple assumption allows a demonstration of the effectiveness of the proposed method.

Analyzing the results of the different calculations, the obvious conclusion is that the correct result will come from the variables representing versioning pairs, under the condition that a poem should be the version of another poem and the chosen answer could not contradict itself.

Based on the highest belief value and lowest doubt, it seems that the combinations found in F and C would represent the correct chronological order. However, these two are inverses of each other, and therefore, a selection of one of them must imply the elimination of the other. To further analyze the information and determine a better order, both values are evaluated.

When F is chosen, it represents that poem 4 is a version of poem 5. C at this point is eliminated because it is the inverse of the selected variable F, and A would also be eliminated here because poem 4 cannot have two origins. The next variable with the highest belief is D where poem 3 is a version of poem 4, and the last option to be looked at is E, which shows poem 3 as a version of poem 5. By taking into comparison the values calculated for D and E, it is seen that D has a higher belief, plausibility, evidence interval and lower doubt. Therefore by having a higher belief value, and a lower doubt value, the logical conclusion would be to select D, and therefore the order would be that poem 5 is the original to poem 4 and poem 4 is the original to poem 3.

Next we analyze the information based on variable C, in this case variable F would be eliminated as it is the inverse of C. A would be the next available variable with the highest belief, and therefore D would be eliminated for being its inverse.

Table 2: Partial table of calculated values

Combinations	Unchanged characters (size 93)	changed characters	Mass	Belief	Plausibility	Evidence Interval	Doubt
F	87	6	0.005459	0.005459	0.61314	0.607681	0.38686
C	87	6	0.005459	0.005459	0.61314	0.607681	0.38686
A	83	10	0.005208	0.005208	0.609375	0.603916	0.390625
D	82	11	0.005146	0.005146	0.608434	0.602974	0.391566
B	80	13	0.00502	0.00502	0.606551	0.601092	0.393449
E	79	14	0.004957	0.004957	0.60561	0.600151	0.39439

In this case B would also be eliminated because one poem cannot have two origins, and by selecting B, poem 5 would be evaluated as originating from both, poem 3 and poem 4, which is not possible. In this case, E would be a second possible variable to consider. By comparing A and E, it is seen that A has higher values in all aspects of calculations compared to E. However, there are two values that negatively affect the variable A (evidence interval being higher, and doubt being higher than the values found for E) so it is excluded and variable E is evaluated. In this case it is shown that poem 5 is a version of poem 4 and poem 3 is a version of poem 5.

The conclusions based on these evaluations are contradicting, except for the fact that without prejudice it can be stated that poem 3 is **not** an original poem to either of the other two poems. The difficult part at this point is to determine if it is the first chronological (FD) order or the second (CE) that will result in a feasible statement. It is seen that the values for F and C are exactly the same for all of the calculations and therefore there is no argument that either one of the two could be chosen. However, the comparison must be made according to the variables of D and E, which are the ones that have different values. This comparison has already been shown that D has a higher belief value than E and also has a lower doubt value than E. As a result, it is concluded that the first analysis, with F, is the one that holds true and the proper chronological order of the poems is poem 5 being the original, poem 4 derived from poem 5, and poem 3 derived from poem 4.

Finally, we call the reader's attention to the fact that his research is limited to only one expert and one way of evaluating the poems. A reliable system should depend in more than one set of data and therefore a second, or more, experts, which would allow the analyzer to calculate combined values and be able to obtain a more definite conclusion.

5. SUMMARY

With traditional ways of publishing documents, authors always had the opportunity to decide to re-publish their work with minor changes to them. At the same time, with new technology, those traditional works can now be made available electronically, while authors have the option to create new texts exclusively in electronic format. Frequently, scholars, studying authors and their works, do not have the opportunity to contact the authors to find the chronological order of their texts or the process taken to create new versions of the document. While at times publication dates are associated with the data, it is not a definite conclusion that a later publication is a version of the original, it just means that it became available at a later date. Since scholars do not usually have any other source besides the texts, they depend on techniques that can determine the evolution of a document. This study shows that with the concept of artificial intelligence and the Dempster-Shafer theory, scholarly experts that analyze versions of texts should be able to obtain a reasonable evolution series for documents with different versions.

6. REFERENCES

[1] Anick, P., Flynn, R. "Versioning a Full-text Information Retrieval System," 15th Annual International SIGIR, pp. 98 – 111, Denmark, 1992.

[2] Antonacopoulos , A., Karatzas, D., Krawczyk, H., and Wiszniewski, B. "The Lifecycle of a Digital Historical Document: Structure and Content," DocEng '04, pp. 147-154, 2004.

[3] Crane, G., Smith, D., Wulfman, C. "Building a Hypertextual Digital Library in the Humanities: A Case Study on London." JCDL '01, pp. 426-434, Roanoke, Virginia, 2001.

[4] Dyreson, C., Lin, H., Wang, Y. "Managing Versions of Web Documents in a Transaction-time Web Server," New York, pp. 422-432, May 17-22, 2004.

[5] Kirschenbaum, M. "What is Digital Humanities and What's It Doing in English Departments?," ADE Bulletin, 2010.

[6] Klir, G. "Uncertainty and Information: Foundations of Generalized Information Theory," Wiley – IEEE Press, 2005.

[7] Klir, G., and Yuan, B. "Fuzzy Sets and Fuzzy Logic: Theory and Applications," Prentice Hall, 1995.

[8] National Endowment for the Humanities, Electronic Literature Directory. http://directory. eliterature.org/. June 29, 2011.

[9]Schreibman, S., Siemens, R., Unsworth, J. "A companion to Digital Humanities," Oxford: Blackwell, 2004.

[10] Text Encoding Initiative. http://www.tei-c.org/ index.xml. June 29, 2011.

[11] The University of North Dakota: Elizabeth Barrett Browning Project. http://und.edu/instruct/ sdonaldson/ November 15, 2011.

[12]World Wide Web Consortium (W3C). http://www.w3c.org. November 15, 2011.

Implementing a Dual System of Education to Promote Science, Technology, Engineering and Mathematics

Dr. R. Cherinka, Dr. T. Herring, J. Prezzama and J. Wahnish
The MITRE Corporation
4830 W. Kennedy Blvd., Tampa, FL 33609
Phone: 813-287-9457, Fax: 813-287-9540
rdc@mitre.org

ABSTRACT

There is a growing concern in the United States regarding the declining state of education and college enrollment for degrees in the areas of Science, Technology, Engineering and Math (STEM). As part of a STEM Outreach program, The MITRE Corporation is partnering with educational foundation organizations to encourage school districts to seek out and take advantage of new prospects for students. These include opportunities to learn and apply STEM-related constructs and emerging technologies in a contextually relevant setting. This paper presents a model for integrating classroom and on the job activities to enhance the overall educational experience. A case study is presented highlighting MITRE's STEM program, including: increasing awareness of STEM education and career initiatives, building and maintaining partnerships with Colleges, High School STEM programs and other Community Educational Organizations, and providing internship opportunities.

Keywords: STEM, Engineering, Education, Internship, Apprenticeship, Career Academy

1. INTRODUCTION

In response to a growing concern in the United States regarding the declining state of education and college enrollment for degrees in the areas of Science, Technology, Engineering and Math (STEM), the MITRE Corporation is taking an active role working to encourage school districts to enhance and provide additional opportunities for students to learn and apply STEM-related concepts through a balance of classroom instruction and internship-based hands-on experience. In this role, MITRE has been influential through outreach activities consisting of but not limited to participation in State and Federal Government Summits, providing internships to students, establishing partnerships with College and High School Engineering Magnet programs, teaching at local universities, and judging science and technology events and competitions.

Additionaly, MITRE has initiated partnerships with non-profit organizations such as the Career Technical Education Foundation (CTEF) to help implement a dual system of education that integrates in-class learning at the high school and college level with hands-on internships with Industry to provide an enriching educational experience for students in STEM topics while establishing a pipeline of excellent talent for Industry [1, 12].

This paper presents a case study of MITRE's STEM program, highlighting: paid internships for High School and College students, shadowing programs for High School students, bootcamp training sessions for local students and teachers, summer technology camps and field trips for interns/co-ops, and summer distributed workshops that stimulate interns across several geographic locations by working together to solve realistic sponsor challenges.

2. BACKGROUND

STEM was coined in the 1990s by the National Science Foundation to formally recognize the interlinking and importance of the fields of study in "Science, Technology, Engineering and Mathematics" and to promote overall enhancements in skills, education and standards in these fields as a whole.

STEM Employment and Growth Projections					
	May 2008 Employment	2010 Employment	2018 Employment	Number change: 2008-2018	Percent change: 2008-2018
Life, Physical, and Social Science Occupations	1,110,420	1,255,700	1,469,400	214,000	16.7%
Computer Specialists Occupations	3,464,180	3,717,300	4,529,400	812,200	22.4%
Architecture and Engineering Occupations	2,480,240	2,820,100	3,101,900	282,000	11.6%
Mathematical Occupations	107,150	116,000	139,000	23,000	17.5%
STEM Occupation Employment Total	7,161,990	7,909,100	9,239,700	1,331,200	16.8%
US Total Employment	127,097,160	150,931,700	166,205,600	15,273,900	10.1%
% of Total Employment	5.6%	5.2%	5.6%	n/a	n/a

Figure 1. Projected Market for STEM Professions

As shown in Figure 1, the projected demand for STEM professionals is expected to increase over 16.8% from 2008-2018, adding more than a million new jobs to the workforce [9, 10, 11]. In the May 2011 monthly labor review by the United States Department of Labor, the Bureau of Labor and Statistics listed the top growing needs in the workforce to be STEM occupations, including computer specialists such as developers and systems engineers, network engineers, information security engineers as well as civil and mechanical engineers. This trend is expected to grow.

In order to meet this demand, there must be a qualified workforce that is prepared to fulfill these occupations, thus driving the need for STEM education. However, there are many metrics, reports and examples found in literature today and over the past several years to show concern in the declining state of STEM education in the United States [14, 15]. For example, in the 2006 United States National Academies report to the United States Congress, several alarming points were made related to compromising America's position as a leader in technology, engineering and creating innovations [5]:

- The performance of 15-year-olds in the United States who in 2006 ranked 25th in math achievement and 21st in science achievement.
- The United States has fallen in international rankings from 2nd in 1995 to 15th in 2008.
- The popularity of computer science as a major for incoming college students has fallen more than 60 percent between 2000 and 2004.
- The turning point for interest in science & engineering occurs in Middle School.

To compound this concern further, the Bureau of Labor and Satistics also predicts a trend that due to low salaries for STEM teachers, over 25000 STEM teachers currently in the education system may leave the teaching profession annually [9, 11]. As the United States Department of Education seeks to elevate both the teaching profession and STEM education [4], there is a need for a renaissance in education to include:

- A heightened focus on Science, Technology, Engineering and Math skills;
- Skillfully integrated academic and technical learning paths;
- Increasing America's talent pool by improving K-12 science and mathematics education;
- Strengthening the skills of teachers through additional training in science, math and technology; and
- Enlarging the pipeline of students prepared to enter college and graduate with stem degrees.
- Increased focus on Industry certifications as well as integration with industry to help boost education through understanding real value in the workplace.

Throughout the remainder of this paper, we will discuss a model that attempts to address some of these needs, and how MITRE is responding to help enhance the educational system as a key Industry partner.

3. DUAL SYSTEM EDUCATIONAL MODEL

Quality STEM education programs, such as the program offered by Project Lead The Way (PLTW), seek to engage students in activities, projects, and problem-based learning, which provides hands-on classroom experiences [6]. In such an environment, students create, design, build, discover, collaborate and solve problems while applying what they learn in math and science. They're also exposed to STEM fields through professionals from local industries who supplement the real-world aspect of the curriculum through mentorships and workplace experiences.

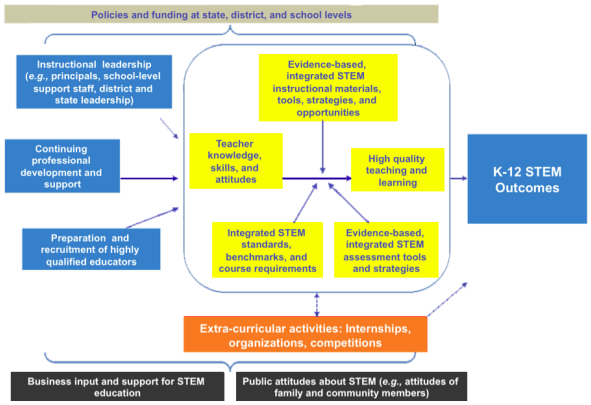


Figure 2. Key variables influencing improvement in STEM education

Figure 2 depicts key variables influencing improvement of STEM teaching and learning in K-12 that have been identified through research [13]. By considering all of the variables and systematically strengthening each component, improved outcomes are more likely to be realized.

A key model to enabling a successful STEM program that attempts to address a balance of each of these variables is referred to as the dual system of education [1, 8]. This

model reflects industry support for educational career academies through focused engagement with schools and through internships and/or apprenticeships. In this model, Academic and Industry partners share responsibility for education and training. The Career Academy assumes the responsibility for teaching the required curriculum content including theory and practical application. Industry provides the career academy financial support and the training necessary to familiarize the trainee with the technological and organizational aspects of the work processes within the company.

Advantages of the Dual System for the Industry partner include:

- Secures the skilled labor needed
- Reduces the costs to train for positions within the company
- Increases motivation and loyalty to the company
- Trainee receives job specific qualifications
- Productive performance of trainee

Advantages of the Dual System for the Student include:

- Recognized Industry Certification
- Increased prospects for employment upon completion
- Theory and practical application of curriculum
- Certain degree of independence through an "earn while you learn" program

The financing of such a model would also come from multiple sources, include Industry partners who would contribute the largest share through donations, mentoring and providing employment opportunities for students. Other financial sources would come from other funding as available from State and Federal Agencies Grants.

This model has been implemented successfully in a number of schools by CTEF, and shows great promise [1]. The potential benefits of this model provide the opportunity for the successful Career Academy Graduate to:

- Earn Industry Certification and/or
- Earn college credit upon successful completion of each course while attending the Career Academy
- Apply to the University of their choice
- Enter gainful employment either with their own training company or another company

The collaborative partnership between Industry and the Educational system represents a model that has been used successfully for years in countries throughout the World, such as the one used throughout the Germany/Austria region [8]. Bringing this model to the United States has shown great promise for the future.

4. MITRE's Role as Industry Partner

From an Industry perspective, science, technology, engineering and mathematics are fundamental components to our professional existence, past and future. STEM initiatives provide youth an opportunity to gain exposure to real world applications of STEM disciplines, helping to see both potential and value.

The goal is to inspire young people to be science and technology leaders. STEM initiatives help to meet this goal by engaging them in exciting programs that build science, engineering and technology skills, inspire innovation and foster well-rounded life capabilities such as self-confidence, communication and leadership.

MITRE is doing its part to help our Nation in addressing STEM concerns. As part of our public interest duty, it is important for MITRE to continue to do this. STEM is fundamental to our business and we need to help groom qualified talent in our core competencies. Our involvement in STEM helps us to do this. By having a vibrant STEM outreach to colleges, k-12 schools and communities, we hope to grow and motivate our future engineers. There are many ways in which MITRE is involved in promoting STEM in our community. Examples include government influence, community and school outreach (e.g. teaching, judging, science fairs, robotics competitions), partnerships with schools (high school and college), and internships. The following section highlights a case study of MITRE's involvement in helping to make the dual system of education a successfule model.

5. CASE STUDY HIGHLIGHTS

Each year, MITRE is actively engaged in STEM outreach across many locations in the United States. At a corporate level, this outreach includes the hiring of over 200 interns each summer from various Universities, colleges and more recently high schools. These interns typically experience a balance of hands-on work experience along with various enrichment activities to promote STEM. Over the past three years in particular, MITRE has partnered with CTEF to become an active sponsor with local High School engineering and technology academies in support of the dual system model. The case study described below highlights the

internship opportunities provided to 19 students across three MITRE site locations during this time period.

The most important part of the internship was working on real world projects alongside MITRE engineers. These MITRE interns gained valuable computer engineering and software development experience through their support to direct funded projects for MITRE's United States Government sponsors as well as for MITRE internal research. A sampling of these projects is described below [2, 3].

Direct Funded Projects

As part of an Agile engineering team, our interns worked to create an infrastructure to collect, store and access socio-cultural information for the Afghanistan/ Pakistan Center of Excellence (AF/PAK COE) at Central Command (CENTCOM). This work entailed network infrastructure integration, software development and social media tool examination.

Several interns were part of a distributed team building gadgets to be used in a composable dashboard for the Partnerships for Americas Collaboration Center. These individulals learned the fundimentals of Java development and widget integration via mashups.

Under a Department of Defense (DoD) sponsor task, a team of two interns created a desktop application, *Moving Target from Video (MTV)* that provided playback, annotation, geo-location and archiving tools for exploiting video streams. This application eventually served as the analyst tool for Full Motion Video Analysts.

In support of the Air Force Enterprise Network, the interns worked on a technology forecasting roadmap, a web application for automatic generation of engineering reports, diagrams, and machine assisted auditing. Working remotely with teams across MITRE, the interns developed Java server pages to query a database and display timeline information.

Learning Ruby on Rails was an opportunity for interns to work on the Semantically Enabled Dynamic Discovery and Delivery (SED3) project. They worked on developing a collection of database tools to make deployment and backup of production data easier to understand and more efficient.

Our interns have supported our customers secure mobile initiative project by demonsrating the flexibility of agile software development techniques in the mobile device development environment. They showcased several on going mobile application development projects as well as the agile process used to achieve them.

In parallel with the stand up of a DoD customer's networking experimemtation lab, several interns were instrumental with the integration, installation and shakedown of OPNET network modeling software. This experience exposed the interns to state of the art network and modeling tools that would never be part of an ordinary engineering curriculum.

Additional Activities

In addition to direct funded projects, the interns participated in other activities to supplement their professional experience and learn more about MITRE and our sponsors.

Bootcamps: Students and incoming interns often lack exposure to many of the key technologies and concepts that play important roles at MITRE. To address this issue, Bootcamps were developed to foster understanding of the core mechanics on these subjects. Bootcamps are short (2-4 hours) classes mixing veteran instruction with hands-on exercise, providing students a fundamental working knowledge of course material. This exposes the students to key terminology, provides a foundation for future learning, and enables immediate participation in related tasks. The end goal is to teach students enough material to hit the ground running in the subject, supporting an effective 'on the job' learning experience. Example bootcamps include topics such as introduction to programming, object-oriented development, mobile development, agile engineering and cyber security.

Fireside Chats: Every Wednesday during lunch local MITRE employees met with interns to discuss their major projects. These "Fireside Chats" provided the intern insight into some of the technical areas that MITRE has focused on recently. Ranging from socio-cultural database management to intelligence integration, these chats were reinforced by visits to our sponsors.

Figure 3. Aerospace Camp

Aerospace Camp: As depicted in Figure 3, interns traveled to Colorado Springs for site visits to MITRE and the United States Air Force (USAF) Academy. This summer field trip provided the interns an introduction to USAF Space Operations; orientation on the USAF Academy and its ongoing research including the Unmanned Airal Survailence (UAS) Program; UAS Research Center and Physics lab; and hands-on participation in live flight UAS experiments.

Distributed Workshops: The Air Operations Center (AOC) workshop was a distributed event between MITRE Bedford, Colorado Springs and Tampa to expose the interns to a real-world Air Operations mission in a simulated environment. The AOC workshop demonstrated distributed AOC with Full Motion Video (FMV) processing, exploitation and dissemination (PED) while collecting and archiving data sets, imagery, and video intelligence products for use in future exercises. The Bedford interns operated an AOC node using common USAF applications for mission planning and ATO production. In responding to various scenario events, they planned out UAS operations for the Colorado Springs node to carry out. The C-Spring interns, with the USAFA cadets, flew the daily missions with Kadet Senior, a commercial RC plane that has been modified to include an autopilot, a GPS system, and a camera. The video and INTs collected in Colorado were sent to the PED node in Tampa, where the intern "analysts" processed and exploited the information looking for clues, patterns and items of interest. Pertinent information was sent back up to the Bedford "AOC" where the interns planned for the next day mission.

Industry Day: MITRE hosted dozens of sophomores from local Florida high schools in an engineering industry day sponsored by CTEF. MITRE introduced the students to cyber security topics, provided an overview of MITRE, and led discussions on what it's like to be an engineer and what we look for when hiring for engineering positions. For fun, the students joined in a marshmallow challenge in which they had to build a tower using marshmallows, spaghetti, string and tape.

Teacher Quest: Teacher Quest is a professional development program for teachers. It works with science, math and technology-based businesses to match teachers with programs to work during the summer and gain industry experience to take back to the classroom. MITRE participated in Teacher Quest by providing summer employment to Computer Science teachers from local high schools. By doing this, teachers can remain current with skills, technologies and application of engineering principals in real world settings. This allows them to return to the class room with refined knowledge and experience to pass along to students.

Robotics: FIRST is an organization focused on inspiring young people to be science and technology leaders, focuses on creating programs to partner education with industry professionals to build science, engineering, and technology skills while inspiring innovation and fostering communication and leadership skills [7]. MITRE actively participates with FIRST by serving as mentors for high school robotics teams in fields such as software, electronics, mechanical and computer-aided design.

Cyber Security Classes: As part of a cyber-focused STEM initiative, MITRE and its interns took part in developing a cyber security course for high school students. The course is aimed at making the students more aware of security issues in the cyber realm. It provides a comprehensive overview of security issues that will provide students with a good foundation for self study or continuing education in the cyber field. In addition, the course will teach concepts that are required for many cyber security certifications.

Capture the Flag Distributed Cyber Event: MITRE hosts an annual Cyber Grand Challenge during the summer internship program. This Grand Challenge, while focused on MITRE interns, is open to particiaption from surrounding colleges and high schools across the US. The objectives of this initiative are to increase awareness about MITRE and its role in information security, facilitate information exchange across institutions, and create a hands-on learning opportunity in security Mission and Information assurance emulating real world situations. To achieve these objectives, the interns and students form teams and participate in a distributed competition that will test the cyber security skills of the participants. The primary focus is on defensive skills, but knowledge of other security related fields will also be tested. MITRE interns will either participate in the high school division or the college division, and the competition is organized in such a way to encourage collaboration and communication between the two divisions.

By combining direct project work with a variety of extra activities designed to provide interns with comprehensive exposure to the engineering profession, each intern is provided multiple opportunities to gain hands-on experience with various engineering disciplines, as well as assess their interest in a possible STEM career. In addition, MITRE is able to assess each student extensively, both from a possible pipeline perspective and for ways to motivate the student further in STEM concepts. The goal is to help each student to be able to return to the classroom and more easily understand the

theory and concepts being taught. When this happens successfully, the full potential of the dual model can be realized.

6. FUTURE WORK

Future STEM work at MITRE will continue to focus on providing educational and internship opportunities for students and teachers in the hopes of enhanceing and motivating STEM interest and education. In particular, MITRE is looking to create a professional cyber security pipeline by partnering with government sponsors, national labs and orgranizations such as CTEF to develop a formal cyber security internship program.

7. CONCLUSIONS

This paper presents the results of implementing a dual system of education through a partnership between MITRE and a number of educational academies.

Through partnerships with organizations such as CTEF that are implementing this model, MITRE has taken an active role as an Industry partner to demonstrate the impact that can be made on students. In the case study example discussed in this paper, of the 19 students who participated during the three year time period across three MITRE locations, 4 students have become MITRE full-time employees, 5 continue as MITRE co-op students in as they complete college, 3 continue as MITRE high school interns, 6 are pursuing other STEM related opportunities and 1 is currently pursuing a non-STEM path. This roughly translates into 60% on track for MITRE's pipeline, and an additional 30% on track for pursuing STEM careers. These results show great promise for this model.

By combining the ingredients of :

- Understanding and embracing the concepts and skills of science, technology, engineering and mathematics
- Passion to make an impact
- Self-confidence to take on new challenges
- Spark to innovate

Young people can accomplish just about anything in superhuman ways, thus creating the kind of breakthroughs and innovations the world desperately needs. Through its STEM initiatives, MITRE is helping to promote future technology and engineering leaders. The dual system of education is a very promising approach to revitalizing STEM education. The partnership with CTEF, a recognized leader of implementing this model in the United States, has proven to be a rewarding experience for both MITRE and the many students that have been part of this experience.

8. REFERENCES

[1] Wahnish, P., "Footprint for Sustainability", Career Technical Education Foundation (CTEF), www.careertechedfoundation.org.

[2] MITRE STEM Initiatives, Technical Report, MITRE Tampa Program Review, The MITRE Corporation, 2011.

[3] MITRE STEM Initiatives, Technical Report, MITRE Tampa Program Review, The MITRE Corporation, 2012.

[4] United States Department of Education, http://www.ed.gov/news/press-releases/obama-administration-seeks-elevate-teaching-profession-duncan-launch-respect-pro, 2012.

[5] 2006 Report to Congress, United States National Academies, http://www.nationalacademies.org.

[6] Project Lead the Way (PLTW), http://www.pltw.org/

[7] FIRST Robotics, http://www.usfirst.org/.

[8] Reform of Vocational Education and Training, http://www.bmbf.de/en/1644.php, 2009.

[9] STEM Employment Statistics, Florida STEM Summit III Presentation, Kelly Services, http://www.careertechedfoundation.org/?page_id=498, 2011.

[10] The STEM Workforce Challenge, US Department of Labor, 04/2007;

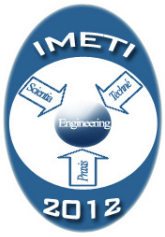
[11] Science, technology, engineering, and mathematics (STEM) occupations: a visual essay, BLS Monthly Labor Review, 2011.

[12] STEMflorida, http://www.stemflorida.net

[13] State STEM Strategic Plan, Florida Center for Research in STEM, Florida State University, 2011.

[14] National Governors Association, Innovation America: A final report. 2010.

[15] National Academy of Sciences, National Academy of Engineering, and Institute of Medicine, "Rising above the gathering storm, revisited: Rapidly approaching category 5", Washington, DC: National Academies Press. p. 18. 2010.



AUTHORS INDEX

(Post-Conference Edition)