RECENT ADVANCES in COMMUNICATIONS

Proceedings of the 19th International Conference on Communications (part of CSCC '15)

Zakynthos Island, Greece
July 16-20, 2015

RECENT ADVANCES in COMMUNICATIONS

Proceedings of the 19th International Conference on Communications (part of CSCC '15)

Zakynthos Island, Greece
July 16-20, 2015

Copyright © 2015, by the editors

All the copyright of the present book belongs to the editors. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the editors.

All papers of the present volume were peer reviewed by no less than two independent reviewers. Acceptance was granted when both reviewers' recommendations were positive.

Series: Recent Advances in Electrical Engineering Series | 50

ISSN: 1790-5117
ISBN: 978-1-61804-318-4

RECENT ADVANCES in COMMUNICATIONS

Proceedings of the 19th International Conference on Communications (part of CSCC '15)

Zakynthos Island, Greece
July 16-20, 2015

Organizing Committee

Editor:
Prof. Valeri Mladenov, Technical University of Sofia, Bulgaria

Associate Editors:
Prof. Dr. Eduardo Mario Dias
Prof. Dorota Jelonek
Assoc. Prof. Miroslav Voznak

Organizing Committee:
Prof. Kleanthis Psarris, The City University of New York, USA (General Chair)
Prof. Pierre Borne, IEEE France Section Chair, IEEE Fellow, Ec Centr de Lille, France (General Chair)
Prof. Panos M. Pardalos, University of Florida, USA (Co-Chair)
Prof. George Vachtsevanos, Georgia Institute of Technology, Atlanta, Georgia, USA (Co-Chair)
Prof. Tadeusz Kaczorek, IEEE Fellow, Warsaw University of Technology, Poland (Co-Chair)
Prof. Nikos Mastorakis, Technical University of Sofia, Bulgaria (Program Chair)
Prof. Branimir Reljin, University of Belgrade, Belgrade, Serbia (International Liaisons)
Prof. Aida Bulucea, University of Craiova, Craiova, Romania (Publicity Chair)
Prof. Valeri Mladenov, Technical University of Sofia, Bulgaria (Publications Chair)
Prof. Imre Rudas, Obuda University, Budapest, Hungary (Tutorials Chair)
Prof. Vladimir Vasek, Tomas Bata University, Zlin, Czech Republic (Special Sessions Chair)
Prof. Anca Croitoru, Al.I. Cuza University, Iasi, Romania (Workshops Chair)

Steering Committee:
Prof. Yuriy S. Shmaliy, IEEE Fellow, Universidad de Guanajuato, Mexico
Prof. Alaa Khamis, IEEE Robotics and Automation Egypt-Chapter Chair, Egypt
Prof. Ioannis Stathopulos, Technical University of Athens, Greece
Prof. Charalambos Arapatsakos, University of Thrace, Greece
Prof. Fragkiskos Topalis, Technical University of Athens, Greece
Prof. Klimis Ntalianis, Technological Educational Institute of Athens, Greece
Prof. Eduardo Mario Dias, University of Sao Paulo, Brazil
Prof. Miroslav Voznak, VSB-Technical University of Ostrava, Czech Republic
Prof. Abdel-Badeeh M. Salem, Ain Shams University, Cairo, Egypt
Prof. Nikolaos Bardis, M.Inst. of Univ. Educ. (ASEI), Hellenic Army Academy, Athens, Greece
Prof. Antoanela Naaji, Vasile Goldis Western University Arad, Romania
Prof. Elena Zamiatina, Perm State University, Perm Krai, Russia
Prof. Pan Agathoklis, University of Victoria, Canada
Prof. George Tsekouras, M.Inst. of Univ. Educ. (ASEI), Hellenic Naval Academy, Athens, Greece
Prof. Claudio Talarico, Gonzaga University, Spokane, USA

International Scientific Committee:
Prof. Lotfi Zadeh (IEEE Fellow, University of Berkeley, USA)
Prof. Leon Chua (IEEE Fellow, University of Berkeley, USA)
Prof. Michio Sugeno (RIKEN Brain Science Institute (RIKEN BSI), Japan)
Prof. Dimitri Bertsekas (IEEE Fellow, MIT, USA)
Prof. Demetri Terzopoulos (IEEE Fellow, ACM Fellow, UCLA, USA)
Prof. Georgios B. Giannakis (IEEE Fellow, University of Minnesota, USA)
Prof. Abraham Bers (IEEE Fellow, MIT, USA)
Prof. Brian Barsky (IEEE Fellow, University of Berkeley, USA)
Prof. Aggelos Katsaggelos (IEEE Fellow, Northwestern University, USA)
Prof. Josef Sifakis (Turing Award 2007, CNRS/Verimag, France)
Prof. Hisashi Kobayashi (Princeton University, USA)
Prof. Kinshuk (Fellow IEEE, Massey Univ. New Zeland),

Prof. Leonid Kazovsky (Stanford University, USA)
Prof. Narsingh Deo (IEEE Fellow, ACM Fellow, University of Central Florida, USA)
Prof. Kamisetty Rao (Fellow IEEE, Univ. of Texas at Arlington,USA)
Prof. Anastassios Venetsanopoulos (Fellow IEEE, University of Toronto, Canada)
Prof. Steven Collicott (Purdue University, West Lafayette, IN, USA)
Prof. Nikolaos Paragios (Ecole Centrale Paris, France)
Prof. Nikolaos G. Bourbakis (IEEE Fellow, Wright State University, USA)
Prof. Stamatios Kartalopoulos (IEEE Fellow, University of Oklahoma, USA)
Prof. Irwin Sandberg (IEEE Fellow, University of Texas at Austin, USA),
Prof. Michael Sebek (IEEE Fellow, Czech Technical University in Prague, Czech Republic)
Prof. Hashem Akbari (University of California, Berkeley, USA)
Prof. Lei Xu (IEEE Fellow, Chinese University of Hong Kong, Hong Kong)
Prof. Paul E. Dimotakis (California Institute of Technology Pasadena, USA)
Prof. Martin Pelikan (UMSL, USA)
Prof. Patrick Wang (MIT, USA)
Prof. Wasfy B Mikhael (IEEE Fellow, University of Central Florida Orlando,USA)
Prof. Sunil Das (IEEE Fellow, University of Ottawa, Canada)
Prof. Nikolaos D. Katopodes (University of Michigan, USA)
Prof. Bimal K. Bose (Life Fellow of IEEE, University of Tennessee, Knoxville, USA)
Prof. Janusz Kacprzyk (IEEE Fellow, Polish Academy of Sciences, Poland)
Prof. Sidney Burrus (IEEE Fellow, Rice University, USA)
Prof. Biswa N. Datta (IEEE Fellow, Northern Illinois University, USA)
Prof. Mihai Putinar (University of California at Santa Barbara, USA)
Prof. Wlodzislaw Duch (Nicolaus Copernicus University, Poland)
Prof. Michael N. Katehakis (Rutgers, The State University of New Jersey, USA)
Prof. Pan Agathoklis (Univ. of Victoria, Canada)
Dr. Subhas C. Misra (Harvard University, USA)
Prof. Martin van den Toorn (Delft University of Technology, The Netherlands)
Prof. Malcolm J. Crocker (Distinguished University Prof., Auburn University,USA)
Prof. Urszula Ledzewicz, Southern Illinois University , USA.
Prof. Dimitri Kazakos, Dean, (Texas Southern University, USA)
Prof. Ronald Yager (Iona College, USA)
Prof. Athanassios Manikas (Imperial College, London, UK)
Prof. Keith L. Clark (Imperial College, London, UK)
Prof. Argyris Varonides (Univ. of Scranton, USA)
Dr. Michelle Luke (Univ. Berkeley, USA)
Prof. Patrice Brault (Univ. Paris-sud, France)
Prof. Jim Cunningham (Imperial College London, UK)
Prof. Philippe Ben-Abdallah (Ecole Polytechnique de l'Universite de Nantes, France)
Prof. Ichiro Hagiwara, (Tokyo Institute of Technology, Japan)
Prof. Akshai Aggarwal (University of Windsor, Canada)
Prof. Ulrich Albrecht (Auburn University, USA)
Prof. Alexey L Sadovski (IEEE Fellow, Texas A&M University, USA)
Prof. Amedeo Andreotti (University of Naples, Italy)
Prof. Ryszard S. Choras (University of Technology and Life Sciences Bydgoszcz, Poland)
Prof. Remi Leandre (Universite de Bourgogne, Dijon, France)
Prof. Moustapha Diaby (University of Connecticut, USA)
Prof. Brian McCartin (New York University, USA)
Prof. Anastasios Lyrintzis (Purdue University, USA)
Prof. Charles Long (Prof. Emeritus University of Wisconsin, USA)
Prof. Marvin Goldstein (NASA Glenn Research Center, USA)
Prof. Ron Goldman (Rice University, USA)
Prof. Ioannis A. Kakadiaris (University of Houston, USA)
Prof. Richard Tapia (Rice University, USA)

Prof. Milivoje M. Kostic (Northern Illinois University, USA)
Prof. Helmut Jaberg (University of Technology Graz, Austria)
Prof. Ardeshir Anjomani (The University of Texas at Arlington, USA)
Prof. Heinz Ulbrich (Technical University Munich, Germany)
Prof. Reinhard Leithner (Technical University Braunschweig, Germany)
Prof. M. Ehsani (Texas A&M University, USA)
Prof. Sesh Commuri (University of Oklahoma, USA)
Prof. Nicolas Galanis (Universite de Sherbrooke, Canada)
Prof. Rui J. P. de Figueiredo (University of California, USA)
Prof. Hiroshi Sakaki (Meisei University, Tokyo, Japan)
Prof. K. D. Klaes, (Head of the EPS Support Science Team in the MET Division at EUMETSAT, France)
Prof. Emira Maljevic (Technical University of Belgrade, Serbia)
Prof. Kazuhiko Tsuda (University of Tsukuba, Tokyo, Japan)
Prof. Nobuoki Mano (Meisei University, Tokyo, Japan)
Prof. Nobuo Nakajima (The University of Electro-Communications, Tokyo, Japan)
Prof. P. Vanderstraeten (Brussels Institute for Environmental Management, Belgium)
Prof. Annaliese Bischoff (University of Massachusetts, Amherst, USA)
Prof. Fumiaki Imado (Shinshu University, Japan)
Prof. Sotirios G. Ziavras (New Jersey Institute of Technology, USA)
Prof. Marc A. Rosen (University of Ontario Institute of Technology, Canada)
Prof. Thomas M. Gatton (National University, San Diego, USA)
Prof. Leonardo Pagnotta (University of Calabria, Italy)
Prof. Yan Wu (Georgia Southern University, USA)
Prof. Daniel N. Riahi (University of Texas-Pan American, USA)
Prof. Alexander Grebennikov (Autonomous University of Puebla, Mexico)
Prof. Bennie F. L. Ward (Baylor University, TX, USA)
Prof. Guennadi A. Kouzaev (Norwegian University of Science and Technology, Norway)
Prof. Geoff Skinner (The University of Newcastle, Australia)
Prof. Hamido Fujita (Iwate Prefectural University(IPU), Japan)
Prof. Francesco Muzi (University of L'Aquila, Italy)
Prof. Claudio Rossi (University of Siena, Italy)
Prof. Sergey B. Leonov (Joint Institute for High Temperature Russian Academy of Science, Russia)
Prof. Lili He (San Jose State University, USA)
Prof. M. Nasseh Tabrizi (East Carolina University, USA)
Prof. Alaa Eldin Fahmy (University Of Calgary, Canada)
Prof. Gh. Pascovici (University of Koeln, Germany)
Prof. Pier Paolo Delsanto (Politecnico of Torino, Italy)
Prof. Radu Munteanu (Rector of the Technical University of Cluj-Napoca, Romania)
Prof. Ioan Dumitrache (Politehnica University of Bucharest, Romania)
Prof. Miquel Salgot (University of Barcelona, Spain)
Prof. Amaury A. Caballero (Florida International University, USA)
Prof. Maria I. Garcia-Planas (Universitat Politecnica de Catalunya, Spain)
Prof. Petar Popivanov (Bulgarian Academy of Sciences, Bulgaria)
Prof. Alexander Gegov (University of Portsmouth, UK)
Prof. Lin Feng (Nanyang Technological University, Singapore)
Prof. Colin Fyfe (University of the West of Scotland, UK)
Prof. Zhaohui Luo (Univ of London, UK)
Prof. Wolfgang Wenzel (Institute for Nanotechnology, Germany)
Prof. Weilian Su (Naval Postgraduate School, USA)
Prof. Phillip G. Bradford (The University of Alabama, USA)
Prof. Hamid Abachi (Monash University, Australia)
Prof. Josef Boercsoek (Universitat Kassel, Germany)
Prof. Eyad H. Abed (University of Maryland, Maryland, USA)
Prof. Andrzej Ordys (Kingston University, UK)

Prof. T Bott (The University of Birmingham, UK)
Prof. T.-W. Lee (Arizona State University, AZ, USA)
Prof. Le Yi Wang (Wayne State University, Detroit, USA)
Prof. Oleksander Markovskyy (National Technical University of Ukraine, Ukraine)
Prof. Suresh P. Sethi (University of Texas at Dallas, USA)
Prof. Hartmut Hillmer(University of Kassel, Germany)
Prof. Bram Van Putten (Wageningen University, The Netherlands)
Prof. Alexander Iomin (Technion - Israel Institute of Technology, Israel)
Prof. Roberto San Jose (Technical University of Madrid, Spain)
Prof. Minvydas Ragulskis (Kaunas University of Technology, Lithuania)
Prof. Arun Kulkarni (The University of Texas at Tyler, USA)
Prof. Joydeep Mitra (New Mexico State University, USA)
Prof. Vincenzo Niola (University of Naples Federico II, Italy)
Prof. S. Y. Chen, (Zhejiang University of Technology, China and University of Hamburg, Germany)
Prof. Duc Nguyen (Old Dominion University, Norfolk, USA)
Prof. Tuan Pham (James Cook University, Townsville, Australia)
Prof. Jiri Klima (Technical Faculty of CZU in Prague, Czech Republic)
Prof. Rossella Cancelliere (University of Torino, Italy)
Prof. Wladyslaw Mielczarski (Technical University of Lodz, Poland)
Prof. Ibrahim Hassan (Concordia University, Montreal, Quebec, Canada)
Prof. Erich Schmidt (Vienna University of Technology, Austria)
Prof. James F. Frenzel (University of Idaho, USA)
Prof. Vilem Srovnal,(Technical University of Ostrava, Czech Republic)
Prof. J. M. Giron-Sierra (Universidad Complutense de Madrid, Spain)
Prof. Rudolf Freund (Vienna University of Technology, Austria)
Prof. Alessandro Genco (University of Palermo, Italy)
Prof. Martin Lopez Morales (Technical University of Monterey, Mexico)
Prof. Ralph W. Oberste-Vorth (Marshall University, USA)
Prof. Photios Anninos, Democritus University of Thrace, Greece

Additional Reviewers

Bazil Taha Ahmed	Universidad Autonoma de Madrid, Spain
James Vance	The University of Virginia's College at Wise, VA, USA
Sorinel Oprisan	College of Charleston, CA, USA
M. Javed Khan	Tuskegee University, AL, USA
Jon Burley	Michigan State University, MI, USA
Xiang Bai	Huazhong University of Science and Technology, China
Hessam Ghasemnejad	Kingston University London, UK
Angel F. Tenorio	Universidad Pablo de Olavide, Spain
Yamagishi Hiromitsu	Ehime University, Japan
Imre Rudas	Obuda University, Budapest, Hungary
Takuya Yamano	Kanagawa University, Japan
Abelha Antonio	Universidade do Minho, Portugal
Andrey Dmitriev	Russian Academy of Sciences, Russia
Valeri Mladenov	Technical University of Sofia, Bulgaria
Francesco Zirilli	Sapienza Universita di Roma, Italy
Ole Christian Boe	Norwegian Military Academy, Norway
Masaji Tanaka	Okayama University of Science, Japan
Jose Flores	The University of South Dakota, SD, USA
Kazuhiko Natori	Toho University, Japan
Matthias Buyle	Artesis Hogeschool Antwerpen, Belgium
Frederic Kuznik	National Institute of Applied Sciences, Lyon, France
Minhui Yan	Shanghai Maritime University, China
Eleazar Jimenez Serrano	Kyushu University, Japan
Konstantin Volkov	Kingston University London, UK
Miguel Carriegos	Universidad de Leon, Spain
Zhong-Jie Han	Tianjin University, China
Francesco Rotondo	Polytechnic of Bari University, Italy
George Barreto	Pontificia Universidad Javeriana, Colombia
Moran Wang	Tsinghua University, China
Alejandro Fuentes-Penna	Universidad Autónoma del Estado de Hidalgo, Mexico
Shinji Osada	Gifu University School of Medicine, Japan
Kei Eguchi	Fukuoka Institute of Technology, Japan
Philippe Dondon	Institut polytechnique de Bordeaux, France
Dmitrijs Serdjuks	Riga Technical University, Latvia
Deolinda Rasteiro	Coimbra Institute of Engineering, Portugal
Stavros Ponis	National Technical University of Athens, Greece
Tetsuya Shimamura	Saitama University, Japan
João Bastos	Instituto Superior de Engenharia do Porto, Portugal
Genqi Xu	Tianjin University, China
Santoso Wibowo	CQ University, Australia
Tetsuya Yoshida	Hokkaido University, Japan
José Carlos Metrôlho	Instituto Politecnico de Castelo Branco, Portugal

Table of Contents

ISBN: 978-1-61804-318-4

ISBN: 978-1-61804-318-4

ISBN: 978-1-61804-318-4

ISBN: 978-1-61804-318-4

ISBN: 978-1-61804-318-4

Plenary Lecture 1

Error Estimation in the Decoupling of Ill-Defined and/or Perturbed Nonlinear Processes

Professor Pierre Borne (IEEE Fellow)
Co-authors Amira Gharbi, Mohamed Benrejeb
Centre de Recherche en Informatique Signal et Automatique de Lille, CRIStAL
Ecole Centrale de Lille
France
E-mail: pierre.borne@ec-lille.fr

Abstract: This lecture deals with the definition of the attractors characterizing the precision of decoupling control laws for a nonlinear process in presence of uncertainties and/or bounded perturbations. This approach is based on the use of aggregation techniques and the definition of a comparison system of the controlled process.

Brief Biography of the Speaker: Pierre BORNE received the Master degree of Physics in 1967 and the Master of Electrical Engineering, the Master of Mechanics and the Master of Applied Mathematics in 1968. The same year he obtained the Diploma of "Ingenieur IDN" (French "Grande Ecole"). He obtained the PhD in Automatic Control of the University of Lille in 1970 and the DSc in physics of the same University in 1976. Dr BORNE is author or co-author of about 200 Publications and book chapters and of about 300 communications in international conferences. He is author of 18 books in Automatic Control, co-author of an english-french, french-english « Systems and Control » dictionary and co-editor of the "Concise Encyclopedia of Modelling and Simulation" published with Pergamon Press. He is Editor of two book series in French and co-editor of a book series in English. He has been invited speaker for 40 plenary lectures or tutorials in International Conferences. He has been supervisor of 76 PhD Thesis and member of the committee for about 300 doctoral thesis . He has participated to the editorial board of 20 International Journals including the IEEE, SMC Transactions, and of the Concise Subject Encyclopedia . Dr BORNE has organized 15 international conferences and symposia, among them the 12th and the 17 th IMACS World Congresses in 1988 and 2005, the IEEE/SMC Conferences of 1993 (Le Touquet – France) and of 2002 (Hammamet - Tunisia) , the CESA IMACS/IEEE-SMC multiconferences of 1996 (Lille – France) , of 1998 (Hammamet – Tunisia) , of 2003 (Lille-France) and of 2006 (Beijing, China) and the 12th IFAC LSS symposium (Lille France, 2010) He was chairman or co-chairman of the IPCs of 34 international conferences (IEEE, IMACS, IFAC) and member of the IPCs of more than 200 international conferences. He was the editor of many volumes and CDROMs of proceedings of conferences. Dr BORNE has participated to the creation and development of two groups of research and two doctoral formations (in Casablanca, Morocco and in Tunis, Tunisia). twenty of his previous PhD students are now full Professors (in France, Morocco, Tunisia, and Poland). In the IEEE/SMC Society Dr BORNE has been AdCom member (1991-1993 ; 1996-1998), Vice President for membership

ISBN: 978-1-61804-318-4

(1992-1993) and Vice President for conferences and meetings (1994-1995, 1998-1999). He has been associate editor of the IEEE Transactions on Systems Man and Cybernetics (1992-2001). Founder of the SMC Technical committee « Mathematical Modelling » he has been president of this committee from 1993 to 1997 and has been president of the « System area » SMC committee from 1997 to 2000. He has been President of the SMC Society in 2000 and 2001, President of the SMC-nomination committee in 2002 and 2003 and President of the SMC-Awards and Fellows committee in 2004 and 2005. He is member of the Advisory Board of the “IEEE Systems Journal” . Dr. Borne received in 1994, 1998 and 2002 Outstanding Awards from the IEEE/SMC Society and has been nominated IEEE Fellow the first of January 1996. He received the Norbert Wiener Award from IEEE/SMC in 1998, the Third Millennium Medal of IEEE in 2000 and the IEEE/SMC Joseph G. Wohl Outstanding Career Award in 2003. He has been vice president of the “IEEE France Section” (2002-2010) and is president of this section since 2011. He has been appointed in 2007 representative of the Division 10 of IEEE for the Region 8 Chapter Coordination sub-committee (2007-2008) He has been member of the IEEE Fellows Committee (2008- 2010) Dr BORNE has been IMACS Vice President (1988-1994). He has been co-chairman of the IMACS Technical Committee on "Robotics and Control Systems" from 1988 to 2005 and in August 1997 he has been nominated Honorary Member of the IMACS Board of Directors. He is since 2008 vice-president of the IFAC technical committee on Large Scale Systems. Dr BORNE is Professor "de Classe Exceptionnelle" at the "Ecole Centrale de Lille" where he has been Head of Research from 1982 to 2005 and Head of the Automatic Control Department from 1982 to 2009. His activities concern automatic control and robust control including implementation of soft computing techniques and applications to large scale and manufacturing systems. He was the principal investigator of many contracts of research with industry and army (for more than three millions €) Dr BORNE is "Commandeur dans l'Ordre des Palmes Acad?miques" since 2007. He obtained in 1994 the french “ Kulman Prize”. Since 1996, he is Fellow of the Russian Academy of Non-Linear Sciences and Permanent Guest Professor of the Tianjin University (China). In July 1997, he has been nominated at the "Tunisian National Order of Merit in Education" by the Republic of Tunisia. In June 1999 he has been nominated « Professor Honoris Causa » of the National Institute of Electronics and Mathematics of Moscow (Russia) and Doctor Honoris Causa of the same Institute in October 1999. In 2006 he has been nominated Doctor Honoris Causa of the University of Waterloo (Canada) and in 2007 Doctor Honoris Causa of the Polytechnic University of Bucharest (Romania). He is “Honorary Member of the Senate” of the AGORA University of Romania since May 2008 He has been Vice President of the SEE (French Society of Electrical and Electronics Engineers) from 2000 to 2006 in charge of the technical committees. He his the director of publication of the SEE electronic Journal e-STA and chair the publication committee of the REE Dr BORNE has been Member of the CNU (French National Council of Universities, in charge of nominations and promotions of French Professors and Associate Professors) 1976-1979, 1992-1999, 2004-2007 He has been Director of the French Group of Research (GDR) of the CNRS in Automatic Control from 2002 to 2005 and of a “plan pluriformations” from 2006 to 2009. Dr BORNE has been member of the Multidisciplinary Assessment Committee of the “Canada Foundation for Innovation” in 2004 and 2009. He has been referee for the nominations of 24 professors in USA and Singapore. He is listed in the "Who is Who in the World" since 1999.

ISBN: 978-1-61804-318-4

Plenary Lecture 2

Applications of Linear Algebra in Signal Processing, Wireless Communications and Bioinformatics

Professor Erchin Serpedin
Department of Electrical and Computer Engineering
Texas A&M University
USA
E-mail: serpedin@ece.tamu.edu

Abstract: In this talk, we will review some of the most important applications of linear algebra in signal processing, wireless communications and bioinformatics, and then outline some of the major open problems which might benefit by the usage of linear algebra concepts and tools.

Brief Biography of the Speaker: Dr. Erchin Serpedin is currently a professor in the Department of Electrical and Computer Engineering at Texas A&M University in College Station. He is the author of 2 research monographs, 1 textbook, 9 book chapters, 105 journal papers and 175 conference papers. Dr. Serpedin serves currently as associate editor for the Physical Communications Journal (Elsevier) and EURASIP Journal on Advances in Signal Processing, and as Editor-in-Chief of the journal EURASIP Journal on Bioinformatics and Systems Biology edited by Springer. He is an IEEE Fellow and his research interests include signal processing, biomedical engineering, bioinformatics, and machine learning.

ISBN: 978-1-61804-318-4

Plenary Lecture 3

Reliability Life Cycle Management for Engineered Systems

Professor George Vachtsevanos
Professor Emeritus
Georgia Institute of Technology
USA
E-mail: george.vachtsevanos@ece.gatech.edu

Abstract: Engineered systems are becoming more complex and by necessity more unreliable resulting in detrimental events for the system itself and its operator. There is evidence to support the contention that industrial and manufacturing processes, transportation and aerospace systems, among many others, are subjected to severe stresses, external and internal, that contribute to increased cost, operator workload, frequent and catastrophic mishaps that require the development and application of new and innovative technologies to improve system reliability, safety, availability and maintainability. These requirements are not true only for strictly engineered systems but are often discussed in business and finance, biological systems and social networks. We introduce in this talk a systematic and verifiable methodology to improve system reliability, reduce operating costs and optimize system design or maintenance practices. The enabling technologies build upon modeling tools to represent critical system functions, a prognostic strategy to predict the long-term behavior of systems under stress, reliability analysis methods exploiting concepts of probabilistic design and an optimization algorithm to arrive at optimum system design for improved reliability. We demonstrate the efficacy of the approach with examples from the engineering domain.

Brief Biography of the Speaker: Dr. George Vachtsevanos is currently serving as Professor Emeritus at the Georgia Institute of Technology. He served as Professor of Electrical and Computer Engineering at the Georgia Institute of Technology from 1984 until September, 2007. Dr Vachtsevanos directs at Georgia Tech the Intelligent Control Systems laboratory where faculty and students began research in diagnostics in 1985 with a series of projects in collaboration with Boeing Aerospace Company funded by NASA and aimed at the development of fuzzy logic based algorithms for fault diagnosis and control of major space station subsystems. His work in Unmanned Aerial Vehicles dates back to 1994 with major projects funded by the U.S. Army and DARPA. He has served as the Co-PI for DARPA's Software Enabled Control program over the past six years and directed the development and flight testing of novel fault-tolerant control algorithms for Unmanned Aerial Vehicles. He has represented Georgia Tech at DARPA's HURT program where multiple UAVs performed surveillance, reconnaissance and tracking missions in an urban environment. Under AFOSR sponsorship, the Impact/Georgia Team is developing a biologically-inspired micro aerial vehicle. His research work has been supported over the years by ONR, NSWC, the MURI Integrated Diagnostic

ISBN: 978-1-61804-318-4

program at Georgia Tech, the U,S. Army's Advanced Diagnostic program, General Dynamics, General Motors Corporation, the Academic Consortium for Aging Aircraft program, the U.S. Air Force Space Command, Bell Helicopter, Fairchild Controls, among others. He has published over 300 technical papers and is the recipient of the 2002-2003 Georgia Tech School of ECE Distinguished Professor Award and the 2003-2004 Georgia Institute of Technology Outstanding Interdisciplinary Activities Award. He is the lead author of a book on Intelligent Fault Diagnosis and Prognosis for Engineering Systems published by Wiley in 2006.

ISBN: 978-1-61804-318-4

Plenary Lecture 4

Augmented Reality: The Emerging Trend in Education

Professor Minjuan Wang
San Diego State University
USA
E-mail: mwang@mail.sdsu.edu

Abstract: Augmented Reality (AR) is the layering of virtual information over the real, 3-D world to produce a blended reality. AR has been considered a significant tool in education for many years. It adds new layers of interactivity, context, and information for learners which can deepen and enrich the learning experience. The combination of real and virtual allows the student to engage in learning about a topic from multiple perspectives and data sources at levels that are not always available in traditional classroom settings and interactions.
As the usage of mobile devices in formal settings continues to rise, so does the opportunity to harness the power of augmented reality (AR) to enhance teaching and learning. Many educators have experimented with AR, but has it proven to improve what students grasp and retain? Is AR just another fun way to engage students, with little transformation of learning? This plenary speaking will introduce augmented reality as an emerging trend in education, provide an overview of its current development, explore examples of curriculum integration, and also suggest approaches for success.

Brief Biography of the Speaker: Dr. Minjuan Wang (Professor of San Diego State University; Distinguished Research Professor of Shanghai International Studies University)
Homepage: http://www.tinyurl.com/minjuan
Minjuan is Professor of Learning, Design, and Technology at San Diego State University (California, USA), and distinguished professor of Shanghai International Studies University (Shanghai, China). She was recently selected as the “Oriental Scholar” by the Municipal Educational Committee of Shanghai). In addition, she and her American colleagues obtained a four-year 1.3 million grant to study environment protection (including the Golden monkeys) in Fanjingshan, Guizhou province.
Minjuan’s work has been highly interdisciplinary, covering the field of education, technology, computer science, geography, and communication. In her 14 years at SDSU, she teaches Designing and Developing Learning for the Global Audience, Mobile Learning Development, Technologies for Course Delivery, and Methods of Inquiry. Her research specialties focus on online learning, mobile learning, Cloud Learning, and intelligent learning (as part of the Intelligent Camps initiative launched by British Telecom). Minjuan is the Editor-in-Chief of a newly established journal-- EAI Transactions on Future Intelligent Educational Environments. She also serves on the editorial boards for four indexed journals: Open Education Research,

ISBN: 978-1-61804-318-4

International Journal on E-Learning (IJEL), the Open Education Journal, and Journal of Information Technology Application in Education.

As a winner of several research awards, Minjuan is recognized as one of the high impact authors in blended and mobile learning. She has more than 80 peer-reviewed articles published in indexed journals, such as Educational Technology Research and Development, IEEE Transactions on Education, and British Journal of Educational Technology. She was a keynote and invited speaker to 11 international conferences. In addition, she is also an accomplished creative writer and an amateur flamenco dancer. Her recent Novel--Walking in this Beautiful World—has inspired many young people around the world.

ISBN: 978-1-61804-318-4

Plenary Lecture 5

Application of Multivariate Empirical Mode Decomposition in EEG Signals for Subject Independent Affective States Classification

Prof. Konstantinos N. Plataniotis
Department of Electrical and Computer Engineering
University of Toronto
CANADA
E-mail: kostas@ece.utoronto.ca

Abstract: Physiological signals, EEG in particular, are inherently noisy and non-linear in nature which are challenging to work with using conventional linear signal processing methods. In this paper, we are adopting a new signal processing method, Multivariate Empirical Mode Decomposition, as a preprocessing method to reconstruct EEG signals according to its instantaneous frequencies. To test its effectiveness, we applied this signal reconstruction technique to analyze EEG signals for a 2-dimensional affect states classification application. To evaluated the proposed EEG signal processing system, a three-class classification experiment were carried out on the "Emobrain" dataset from eNTERFACE'06 with K-nearest neighbors (KNN) and Linear Discriminate Analysis (LDA) as classifiers. A leave-one-subject out cross validation process were used and an averaged correct classification rate of 90.77% were achieved. Another main contribution of this paper was inspired by the growth of non-medical grade EEG headsets and its potential in advanced human-computer interface design. However, to reduce cost and invasiveness, consumer grade EEG headsets have far less number of electrodes. In this paper, we used emotion recognition as a case study, and applied Genetic Algorithm to systematically select the critical channels (or sensor locations) for this application. The results of this study will shed lights on the sensor configuration challenges faced by most consumer-grade EEG headset design projects.

Brief Biography of the Speaker: Konstantinos N. (Kostas) Plataniotis received his B. Eng. degree in Computer Engineering from University of Patras, Greece and his M.S. and Ph.D. degrees in Electrical Engineering from Florida Institute of Technology Melbourne, Florida. He was with the Computer Science Department at Ryerson University, Ontario, Canada from July 1997 to June 1999. Dr. Plataniotis is currently a Professor with The Edward S. Rogers Sr. Department of Electrical and Computer Engineering at the University of Toronto in Toronto, Ontario, Canada, where he directs the Multimedia Laboratory. He is a founding member and the inaugural Director – Research of the Identity, Privacy and Security Institute, IPSI, (www.ipsi.utoronto.ca). Kostas was the Director (January 2010- June 2012) of the Knowledge Media Design Institute, KMDI, (www.kmdi.utoronto.ca) at the University of Toronto.
Dr. Plataniotis was the Guest Editor for the March 2005 IEEE Signal Processing Magazine special issue on "Surveillance Networks and Services", and the Guest Editor for the EURASIP Applied

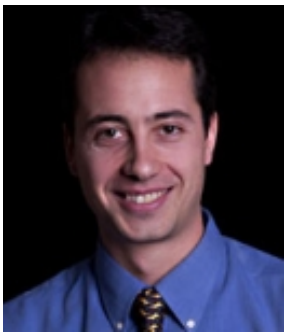
Signal Processing Journal's special issue on "Advanced Signal Processing & Pattern Recognition Methods for Biometrics". He is a member of the IEEE Periodicals Review and Advisory Committee (2011-2013); he has served as a member of the 2008 IEEE Educational Activities Board; he chaired of the IEEE EAB Continuing Professional Education Committee, and he served as the 2008 representative of the Computational Intelligence Society to the IEEE Biometrics Council. Dr. Plataniotis chaired the 2009 Examination Committee for the IEEE Certified Biometrics Professional (CBP) Program (www.ieeebiometricscertification.org) and he served on the Nominations Committee for the IEEE Council on Biometrics. He was a member of the Steering Committee for the IEEE Transaction on Mobile Computing, an Associate Editor for the IEEE Signal Processing Letters as well as the IEEE Transactions on Neural Networks and Adaptive Systems and he has served as the Editor-in-Chief for the IEEE Signal Processing Letters from January 1, 2009 to December 31, 2011. Dr. Plataniotis chaired the IEEE Toronto Signal Processing and Applications Toronto Chapter from 2000 to 2002, he was the 2004-05 Chair of the IEEE Toronto Section and a member of the 2006 as well as 2007 IEEE Admissions & Advancement Committees. He served as the Technical Program Committee Co-Chair for the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2013) and he is the Vice President – Membership for the IEEE Signal Processing Society (2014-2016).

Dr. Plataniotis is a Fellow of IEEE, Fellow of the Engineering Institute of Canada, a registered professional engineer in the province of Ontario, and a member of the Technical Chamber of Greece.

The recipient of numerous grants and research contracts as the principal investigator, he speaks internationally and writes extensively in his field and he has been a consultant to a number of companies. He has served as lecturer in 12 short courses to industry and continuing education programs; he is a contributor to seventeen books, the co-author of "Color Image Processing and Applications", Springer Verlag, 2000, (ISBN-3-540-66953-1) and "WLAN Positioning Systems: Principles & applications in Location-based Services", Cambridge University Press, 2012 (ISBN 978-0-521-9185-2), "Multi-linear Subspace Learning: Reduction of multi-dimensional data}, CRC Press, 2013, (ISBN: 978-14398557243). He is the co-editor of "Color Imaging: Methods and Applications", CRC Press, September 2006, (ISBN 084939774X) and the Guest Editor of the IEEE/Wiley Press volume on "Biometrics: Theory, Methods and Applications" published in October 2009 (ISBN: 9780470247822). Dr. Plataniotis has published more than 400 papers in refereed journals and conference proceedings. In 2005 he became the recipient of the IEEE Canada Engineering Educator Award for "contributions to engineering education and inspirational guidance of graduate students". Dr. Plataniotis is the joint recipient of the "2006 IEEE Trans. on Neural Networks Outstanding Paper Award" for the published in 2003 "Face recognition using kernel direct discriminant analysis algorithms", IEEE Trans. on Neural Networks, Vol. 14, No 1, 2003.

Plenary Lecture 6

State of the Art and Recent Progress in Uncertainty Quantification for Electronic Systems (i.e. Variation-Aware or Stochastic Simulation)



Professor Luca Daniel
Electrical Engin. & Computer Science
Massachusetts Institute of Technology (MIT)
Cambridge, MA, USA
E-mail: luca@mit.edu

Abstract: On-chip and off chip fabrication process variations have become a major concern in today's electronic systems design since they can significantly degrade systems' performance. Existing commercial circuit and MEMS simulators mostly rely on the well known Monte Carlo algorithm in order to predict and quantify such performance degradation. However during the last decade a large variety of more sophisticated and efficient alternative approaches have been proposed to accelerate such critical task. This talk will first review the state of the art of most modern uncertainty quantification techniques including intrusive and sampling-based ones. It will be shown in particular how parameterized model order reduction, and low-rank tensor based representations can be used to accelerate most uncertainty quantification tools and to handle the curse of dimensionality. Examples will be presented including amplifiers, mixers, voltage controlled oscillators with tunable micro-electro-mechanical capacitors and phase locked loops.

Brief Biography of the Speaker: Luca Daniel is an Associate Professor in the Electrical Engineering and Computer Science Department of the Massachusetts Institute of Technology (MIT). Prof. Daniel received the Ph.D. degree in Electrical Engineering from the University of California, Berkeley, in 2003. In 1998, he was with HP Research Labs, Palo Alto. In 2001, he was with Cadence Berkeley Labs.

Dr. Daniel research interests include development of integral equation solvers for very large complex systems, stochastic field solvers for large number of uncertainties, and automatic generation of parameterized stable compact models for linear and nonlinear dynamical systems. Applications of interest include simulation, modeling and optimization for mixed-signal/RF/mm-wave circuits, power electronics, MEMs, nanotechnologies, materials, MRI, and the human cardiovascular system.

Prof. Daniel has received the 1999 IEEE Trans. on Power Electronics best paper award; the 2003 best PhD thesis awards from both the Electrical Engineering and the Applied Math departments at UC Berkeley; the 2003 ACM Outstanding Ph.D. Dissertation Award in Electronic Design Automation; 5 best paper awards in international conferences, 8 additional nominations for best paper award; the 2009 IBM Corporation Faculty Award; and the 2010 IEEE Early Career Award in Electronic Design Automation.

ISBN: 978-1-61804-318-4

ISBN: 978-1-61804-318-4

Co-design: An Assistive Technology Acceptance Approach

Bryan R.M. Manning and Stephen Benton

Abstract**—**Demographic change in the percentage of the elderly has the potential to destabilise health and social care service provision, unless they can be persuaded to accept Assistive Technology as a means to help them to remain independent. Unfortunately the design approach used so far has alienated the vast majority of them, as it has failed to recognise the need for a whole systems approach that can adapt to their evolving needs in the context of ageing as it affects them.

This paper describes a co-design process in which they are equal partners in an inter-disciplinary development team that accepts and manages incremental change as a means of optimising its output to satisfy the need for easy use and inherent usefulness of its products and services. It highlights the crucial need to recognise that understanding and accommodating the psychology of the ageing and the aged is the critical failure avoidance factor.

It outlines the psychological and behavioural profiling approach used in the context of using clothing as a combined communications and physiological monitoring platform acceptable to the elderly. After which it projects this approach forward to investigate ways of identifying early stage indications of latent dementia.

Keywords**—** Co-design, Behavioural Profiling, Smart Wearables, Whole-systems Design.

I. Introduction

THE developed and developing nations are almost all beset by the problem of rapid growth in the percentage of the elderly in their populations [1,2] to the extent that their health and social care services will potentially no longer be able to cope. So an urgent search is on for the best way of helping the ageing to maintain their independence, autonomy and quality of life against a background of increasing personal physical, cognitive and social frailty, at a time when fewer professional resources are likely to be available [3].

Although assistive technology is seen as the primary solution to these looming shortages, it is also seriously handicapped by the lack of effective cross-platform inter-operability needed to respond to the complex problems that the ageing process presents [4]. Then as if the technical problems of device plug-and-play extensibility needed to keep pace with each individual's hugely variable and evolving needs is not difficult enough – there is the issue of a significant "digital disconnection" between the elderly and technology that works against its successful adoption.

Unfortunately this "digital divide" stems from on-going cultural change and the inevitable life-experience gap that exists between generations. Moreover this is not just centred on the human-computer interface concerns, but is actually affected by deep-seated psychological factors that slowly build up throughout the life-course of each person.

These factors are inevitably person-specific and frequently hidden from others, who are not of their generation. As a result they do not readily surface in clinical or care profession driven case reviews or surveys, but can usually be drawn out in one-to-one sessions carefully led by someone of their own age group. The resulting behavioural profiling across an appropriately-sized peer-group cohort is then a key element in subsequent co-design product and system development.

II. Elderly Attitudes

Sadly changes in demographics and lifestyles over recent decades have meant that families and friends have moved further apart, with a resultant loosening of bonds coupled with a radical alteration in the means and level of communication between them. As these old certainties of life slip away, the elderly then face increasing isolation and loss of frequent direct face-to-face social contact.

As the pace of life has accelerated [5], this disconnection has also disrupted the once implicit levels of trust in professions and service providers that used to exist [6], and was once so freely given. At the same time confidence that their reasonable expectations of good quality results have begun to ebb away, leaving them evermore cautious and suspicious lest their trust may be betrayed [7].

This is not helped by age-differential effects, where the good intentions of the younger generations can easily end up quietly ditched or ignored by the elderly for a variety of hidden causes or misperceptions across this inevitable gap. Whilst the younger generation are directly exposed to latest innovations and developments, the elderly are likely to be less so – yet not necessarily sufficiently unaware or disinterested not to have a valid opinion on them.

What sets them apart is experience. The elderly will have gathered a wealth of this together throughout their lengthy transit to old age – yet this remains uncollected and unused. Then up-coming generations are still slowly collecting theirs, but are more closely exposed to innovation and its

Bryan R.M. Manning is Senior Associate at the Business Psychology Centre, Department of Psychology, Faculty of Science and Technology, University of Westminster, London, England (e-mail: bryan.manning@btinternet.com)

Stephen Benton is Director of the Business Psychology Centre, Department of Psychology, Faculty of Science and Technology, University of Westminster, London, England (email: bentons@westminster.ac.uk)

ISBN: 978-1-61804-318-4

opportunities. The effect is that their perspectives, and therefore their judgements are based on radically different frames of reference and resultant attitudes.

Strategic initiatives derived from the basic concept of a universal digital economy generally do not take account of such issues and thus effectively design-in inherent critical failure factors. In the UK the idea that interaction between the state and citizens should be progressively switched to digital [8], has resulted in "digital exclusion" for many, not least the elderly [9].

This has not been helped by a number of ill-founded assumptions, not least where the elderly are concerned that: they will have been provided with, or can afford, an appropriate telecommunications link and equipment; the various types of services available can really suit their actual and prospective needs; they have the desire and the ability to use it effectively [10]; and reliable and effective support is available at all times to resolve all problems, perceived or actual. However even if these basic criteria are met, the critical success factor will depend on whether such systems are designed so that they can be configured and adjusted to satisfy each older person's evolving requirements.

A. Acceptability Criteria Factors

Optimising the alignment between design criteria and end-user requirements ideally needs to be based on clear insights into how the user will judge the efficacy of the final design. This particularly so, where the target audience are elderly, and where the age-differential can have such a dominant effect [11,12].

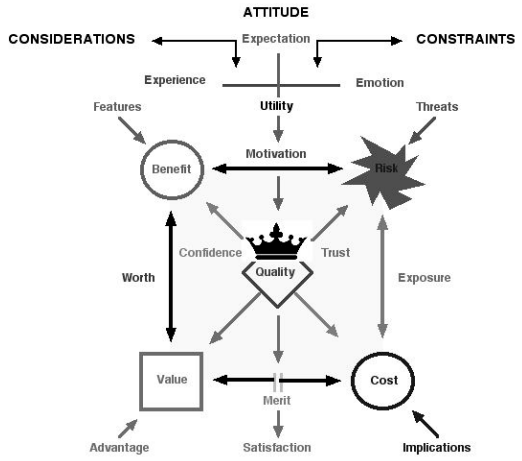


Fig.1 User Perceptions

The first step is to "get inside" the generic thought processes that are involved in any such transaction. These broadly come down to: what considerations are driving the user's needs and motivation; what constraints are likely to close down a decision to proceed; and what are the critical success factor criteria underpinning the satisfaction of expectations, [Fig.1].

The first aspect of the decision process begins with identifying in broad terms what features are considered necessary to deliver the key benefits that the user has in mind. These benefits are then weighed against the perceived advantages that each will deliver in terms of practical and financial value over any other possible alternatives to establish net worth of the options available.

Next, constraints are considered by assessing possible risks in terms of likely threats and associated hazards that could arise during the complete process and might influence a decision to proceed. The implications of any of these risks arising are then judged against any perceived image and financial impacts that could be incurred to give a view of the level of exposure that might occur.

In most cases the motivation, to proceed or not, is basically made by striking a reasonable balance between perceived benefits and risk, supported by a judgement on relative merits of value set against cost. However, much will depend on the attitude of the individual concerned and their perception of the overall utility of the offering in terms of its usefulness and ease of use [13,14]. Whilst heavily influenced by past experience, emotional aspects may well intervene to inject less rational bias into the decision process.

However this may not be the end of the matter. Unfortunately there can often be downstream problems where the recipient is dissatisfied with the results. This stems from the expectation that the outcome will satisfy the decision criteria by appropriately delivering a good standard of quality and resultant utility – thus justifying the degree of confidence and trust placed in the provider at the outset.

The underlying causes usually arise out of the inherent discrepancies in perspectives and perceptions and between provider and recipient, and their failure to communicate effectively as the project proceeds. Once again the age-differential is likely to be the source, since the standpoints of each party will be rooted in different eras, terminology and technologies [15].

Resolving such issues – or better still obviating them before they become an unbridgeable gap – comes down to recognising the likely pitfalls well in advance and negotiating ways of circumventing them. Achieving this depends on establishing an effective on-going dialogue between the parties at the outset – based on mutual best interests – that unearths hidden problems and attitudes that go with them, and manages variations and adaptions accordingly.

III. Implications of Ageing

Whilst it should be self-evident that designing products and services should take great care to understand the views and circumstances of their target audience, time-to-market and cost-to-margin pressures result in outsourcing this key aspect to marketeers. Unfortunately they are often constrained to using survey questionnaires to maximise coverage, rather than collect in-depth knowledge of motivational issues from within the potential client base.

ISBN: 978-1-61804-318-4

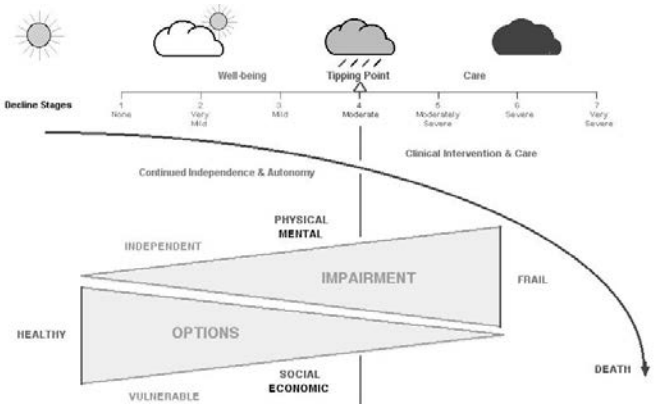


Fig.2 Decline Stages and Trajectory

As a result their input is often confined to collecting aggregate views on design concepts, as opposed to identifying what the public really wants, and why. Moreover these commissions tend to be "one-shot" exercises, possibly with a late stage validation of the design, rather than as an iterative and integral part of the design process.

Where the audience is elderly, bridging this latent age-differential gap is essential. At the very minimum design groups need to understand the trajectory and its stage-by stage impact on the changing situation that the elderly are faced with (Fig.2). Initially decline is relatively slow until a tipping point is reached, after which the process begins to accelerate toward its appointed end.

This trajectory passes through seven distinct stages of increasing severity, identified on the basis of a simple scale originally devised to categorise severity of psychotic conditions [16]. In this case it is generalised to cover the complex mix of physiological, cognitive and socio-economic conditions in the transition from a good overall state of health through to severe frailty.

In essence this journey is characterised by increasing impairment accompanied by ever more restricted lifestyle options. The ultimate challenge is therefore how to delay the decline, and also to mitigate its effects by delaying the point at which significant demand on care services becomes a necessity.

Whilst the former depends on helping the ageing to improve their overall well-being, and the latter requires assistance to continue an independent mode of living, the answer lies in joint application of technology and good design, in the context of effective communication between all concerned. Since good design is the key to success in this complex domain, it needs to be iterative, inclusive, and inter-disciplinary.

A. Psychology of Well-Being

Although the maintenance of good health in all respects is obviously a sensible objective, it is amazing how humanity in general only tends "to pay lip service" to the idea. So it is not too surprising that, with the constraints of ageing, the elderly tend to "opt out" en mass, or "drop out" later, despite its potential to radically extend their continued independence and autonomy [17].

Reversing this trend presents a massive "hill to climb", as it presents a change in embedded lifestyle that they are not keen to embark upon as it is seen an added "clinical burden". One way around this is to place it into a "whole life" setting that they control – albeit with optional support (Fig.3).

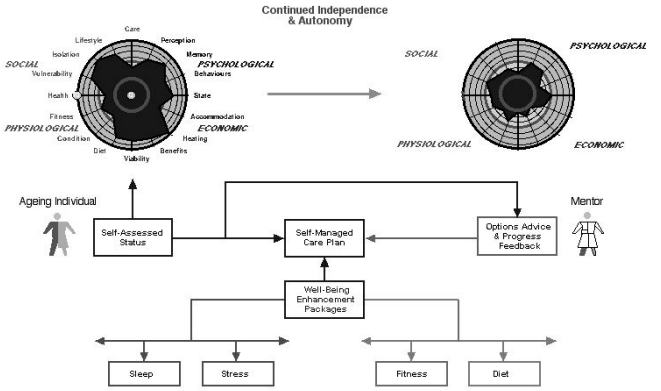


Fig.3 Technology enabled Well-being

This self-managed approach utilises aspects of game theory [18], where ageing individuals can regularly self-assess their perceived condition across a wide range of characteristics of physiological, psychological, social and economic topic areas, at regular intervals. The scores for each selected category use a common sliding 0 – 10 scale [where 10 is bad and 0 is good].

The results are then mapped onto a "radar" plot, on radial arms from the central zero point, and then linked and in-filled to create an overview "patch" [19]. The "game" is then to shrink the "patch" by applying selected packages focused on improving well-being in specific target areas. At its most simplistic this can be done using a pencil and chart and a phone link to a mentor, otherwise this can be done using a suitably designed "app" providing well-being enhancement packages.

B. Characteristics of Ageing

Gain is an almost unrecognised general feature in the life of the up-coming generations, e.g. knowledge; experience; home and family; social circle, wealth, etc. Whilst this can inevitably be set back, on occasion, by troubles of varying degrees of severity and resulting losses, even these can produce gains in personal resilience.

Loss is a dominant feature of ageing that is not just confined to personal health and increased impairment, but impinges on most aspects of life as personal "horizons" slowly implode. This reinforces innate adversity to change and attitudes to risk, both of which make it difficult to persuade the elderly to consider or accept help in one form or other.

It is particularly unfortunate that key elements of these losses are that of trust and confidence in others to recognise, respect and accept the massive contribution that those who are now elderly have made to provide the platform that enables the up-coming generations to thrive. Moreover they tend to become particularly disenchanted by implications, that they

ISBN: 978-1-61804-318-4

are an undeserving drag on society and who are somehow of no account with nothing to offer.

Unless this trust is renewed, and they are enlisted as pro-active partners in the development of ways to enable them to extend their period of independence and autonomy within their communities, a *tsunami* of ever-mounting rising demand will undermine care services provision.

IV. Generic co-design

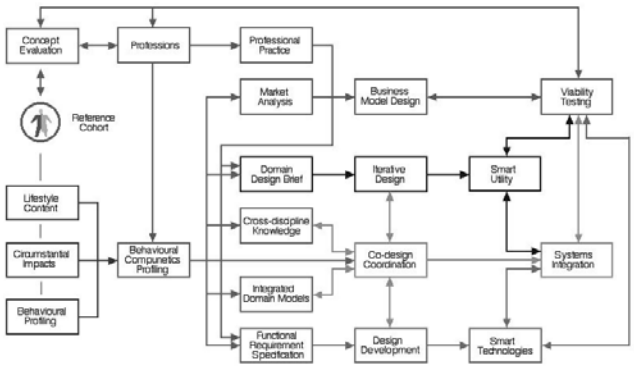


Fig.4 A Generic Co-design Process Structure

A potential answer to this looming problem is contained with the generic concept of co-design (Fig.4), which centres on an approach that incorporates a reference cohort as an integral part of the design and development of systems focused on complex domain issues [20,21]. In what may seem to some as over-complicating a predominantly technical issue by the introduction of multi-disciplinary and outside expertise into an iterative process cycle, there are already more than enough examples of IT disasters to prove the point [22].

The concept on which most major IT enabled projects are based usually emerges out of an amorphous set of diverse inputs from a disparate group of interested parties, professions, pressure groups and political interests. This is then interpreted and "worked up" into a function requirements specification by a technical team, after which it is submitted and then authorised by the appropriate funding organisation – who may, or may not be advised of an independent consultancy practice.

It is interesting to note that for most service oriented projects there is minimal, if any, input from a truly representative group of users with in-depth knowledge of the domain – or indeed often a similar professional grouping.

As a counter to this notable error, a reference cohort of appropriate size, experiential diversity and backgrounds has to be recruited. The role of this group is to act as a source of lifestyle expertise spanning issues and impacts, as well as acting as coordinators for additional local in-depth research surveys [23].

The group itself needs to be facilitated by one or more psychology researchers, and must be of an appropriate statistical size and range to yield valid data from a suitably diverse set of locations. From an organisational perspective, there are two additional sub-groups, the large one of which acts as a quality assurance team validating the design as it evolves, whilst the second smaller one is directly involved as part of the design team.

With the assistance of the psychologists, the design team members work with the relevant service professionals to validate/correct aspects of the system concept as the behavioural profiling progresses. The output of this process then feeds into the main design and business process streams.

The design stream is divided into two sub-streams whose respective roles are to provide the "smart" enabling technology that delivers service utility content to the end-user community which comprises both from the professional and reference cohort domains. The whole process is coordinated by a close-knit co-design management team that iteratively adapts and optimises the overall programme to ensure that the utility content is properly suited to the proven needs of its end-users.

The whole package output is subject to a rolling sequence of incremental systems integration and viability testing, that includes stage-by-stage cross-checking and improving the business model, that itself is regularly validated against current market conditions.

This generic approach is a synthesis of design management techniques applied across a ranges of service sector domains and is the basis of the approach used in Case History 1 and planned for Case History 2.

V. Case Study 1 – "Smart Wearables" in a "Smart Infrastructure"

This New Dynamics of Ageing project was primarily focused on the development of a "proof of concept" multi-layered clothing-based "communications platform" incorporating a variety of physiological monitoring devices that could be linked back to remote services. It also applied the generic co-design methodology (Fig.4) to two seemingly incompatible design approaches to overcome the problem of creating a garment design capable of incorporating a systems infrastructure that was acceptable and attractive to the elderly.

The basic dichotomy that this presented was how to bridge the gap between a formalised and highly structured sequential technology design methodology, and an interactive arts and crafts approach, in which the design evolved iteratively as part of a dialogue between tailor and client.

This was resolved by creating a management team comprising a lead member of each of the three design segments together coordinated by the project director with support of an external strategic consultant. The role of the reference cohort lead proved to be crucial in that he ensured that the team recognised the need to accept iterative adjustment to the design brief as the way to respond to emerging style and functional requirements resulting from the "front-end" research into overall design acceptability criteria within the elderly.

ISBN: 978-1-61804-318-4

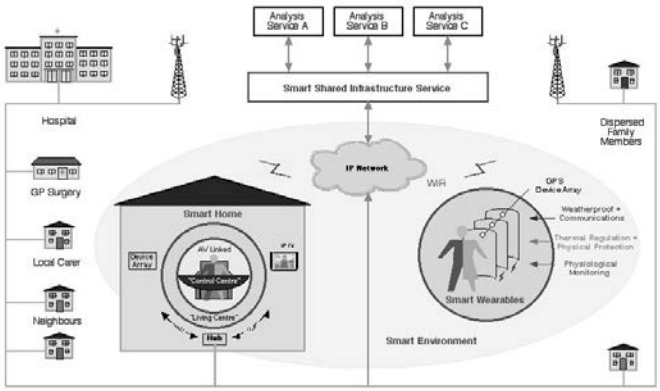


Fig.5 An Assistive "Smart Services Infrastructure"

On the whole the technology group found it more difficult to adapt their way of working to accept the need to deviate from a purely functional perspective to accommodate the idiosynchratic views of the elderly. This was overcome by dividing the technical design into design modules for delivery and subsequent integration into the garment structure.

Whilst the garment design broadly followed the usual adaptive bespoke tailoring approach, it had to modify its approach to accommodate the device positioning requirements and links that needed to be integrated into the fabrics. The other significant design research involved a surface body scanning procedure for number of volunteers to establish body shapes and sizes of the elderly, since these change radically over the life-course. This short procedure was able to rapidly produce 3D avatars for use as virtual tailors dummies making it much easier for garment designers to style and produce cutting patterns directly.

The direct involvement of these volunteers not only ensured that the overall design was not only optimised to deliver attractive and desirable clothing for this long neglected group but also ensured the engagement of the whole group through the complete process through to the final stage of trial usage within the community.

A. "Smart Wearables" in a "Smart Services Infrastructure"

The longer-term aim of the Smart Wearables project was envisaged to be set within a wider strategic design concept (Fig.5), as it would open up safer access to the wider community landscape [26].

The basic of the Smart Wearables approach centred on the creation of a three layer clothing design that comprised of: a close fitting inner layer enabling a range of physiological monitoring devices to be incorporated within it; an intermediate layer using various leading-edge materials to provide both thermal regulation and physical protection; and weather-proof outer layer with an in-built 4G mobile-phone system that also acted as a communications hub for clothing embedded monitoring devices [26].

This system could either function autonomously, or linked back as necessary to the wearer's Smart Home hub to activate or control other device arrays there or pass data onward.

VI. Case Study 2 – Dementia: Pre-Diagnosis Possibilities

This embryo project is aimed at shedding light on early indications of latent dementia or other mental health issues prior to the emergence of recognisable signs and symptoms. Whilst this project is in its early conceptual design stage, it is planned to apply the same generic design management approach as outlined previously to deal with the wide range of complex problems that cognitive abnormalities create for all concerned.

Any project focusing on this extremely sensitive and highly charged issue, with a multiplicity of stakeholders involved, who all have widely differing and strongly held views, is challenging to say the very least. Given this situation, it is imperative that the design process is fully inclusive with all interested parties involved in contributing to the debate – which demands considerable care and tact in achieving an acceptable balance and engagement of all throughout.

In this situation the "front-end" reference cohort will not only be larger, but also will need to be divided into sub-groups to get a set valid behavioural profiles from each of those directly involved spanning: the elderly; their families; and their carers [28]. Paradoxically because of the multiple professions involved in mental health service provision a similar but smaller cohort will be needed to bring together the disparate perspectives and expertise of this grouping.

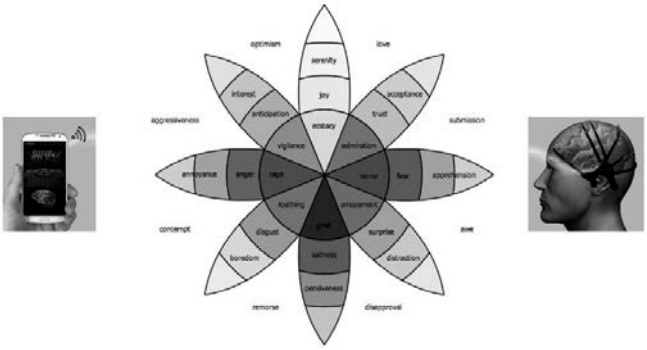


Fig.6 Mood Tracking and Profiling

By contrast mobile technology and its integration with the smart wearables research fits neatly into the approach used in the first of these two case studies. The other likely difference will be need for in-depth analysis of the results, both before and in subsequent use of mood and emotional tracking and profiling processes. This will need to be handled as an "off-line" activity, probably needing longer term monitoring in conjunction with the reference groups.

A. Mood and Emotional Tracking

The aim of this projected project is to determine the degree to which changes in mood may be an early indicator of susceptibility to dementia [29]. In this case the design approach is predominantly focused on the "fronted-end" research as outlined in the previous Case Study above, but with much less emphasis on product or service design.

The intention is to use a variety of mood monitoring techniques as currently used by self-help groups interested in improving personal well-being [30], and to combine this with

behavioural profiling and brain activity research data in the hope that this may lead to the creation of an early stage dementia detection technique. However the key to gathering data depends once again on a combination of mobile-phone app-based brain activity monitoring [31], which will only be acceptable to the elderly if the "headset" is hidden within a suitable hat or cap. In due time it is hoped this will be combined with home-based monoamine measurement [32] to provide a viable assessment procedure.

VII. Conclusion

The pressure for the development of effective assistive technologies is intense and according to the data, set to further intensify as these global demographic trends driven by economic development become entrenched. This is already having an impact on social and health care provision as increasing longevity and socio-medically enhanced survivability radically alters demand levels imposed by an ageing population. However beside the considerable added civic value these achievements have delivered, they contain both a threat and an opportunity.

Whilst medical developments offer unprecedented ageing trajectories for all, the impact of the sheer weight of demand from the growing numbers of the elderly combined with current economic recovery pressures threatens the viability and sustainability of care provision. This can only worsen as the size of the gap between required and available physical resources widens and costs escalate.

Although the "opportunities" presented by assistive technology to close the gap by enabling the elderly to retain their independence with its support and that of family, friends and others – it crucially depends on its acceptance by the bulk of the elderly population. Unfortunately this is singularly lacking - due to significant failures by technologists and care professionals to understand the psychological impacts of ageing that set the elderly against accepting the undoubted benefits that the technology could bring.

This, together with the pressure for an all-encompassing technological solution that can be rolled out to bridge the gap, merely creates an unnecessarily high level of "risk" of continued rejection. Moreover it is particularly likely to be the case if pre-conceived "user aspirations" are driving the design and development process on a "one-size-fits-all" basis – especially in a situation in which actual system requirements are essentially idiosyncratic and evolve with each individual case.

In view of the hugely complex range of circumstances, attitudes and issues within the ageing population, the best way forward is to incorporate behavioural profile prototyping into a user-centred co-design process - if the current situation is to be turned around. Ideally technology needs to be seen to offer the opportunity to achieve a new level of autonomy, quality of life and activity, whilst being attractive and easy to use - as with enhanced communications offered by tablets.

If this opportunity is not taken, it is likely to once again bear witness to yet another scenario where the failure of a good idea has been "snatched from the jaws of success", by systems where the lowest common denominator of "independent living" scores low on its capacity to enhance "quality of living" and even meet citizen-users limited expectations.

References

[1] National Research Council**,** Preparing for an Aging World: The Case for Cross-National Research. Washington, D.C.: National Academy Press, 2001.

[2] Commission of the European Communities**,** The social situation in the EU, 2004. http://europa.eu.int/comm/employment_en

[3] Commission of the European Communities**,** e-Health ‿ making healthcare better for European citizens: An action plan for a European e-Health Area, 2004.

[4] Commission of the European Communities (2004). European Interoperability Framework for pan-European eGovernment Services http://ec.europa.eu/idabc/servlets/Docd552.pdf?id=19529

[5] R.V. Levine, A Norenzayan. *The Pace of Life in 31 Countries*. Journal of Cross-Cultural Psychology, Vol. 30 No. 2, March 1999, 178-205

[6] S. Jauhar. *Why Doctors Are Sick of Their Profession* The Saturday Essay, The Wall Street Journal Aug. 29, 2014 http://www.wsj.com/articles/the-u-s-s-ailing-medical-system-a-doctors-perspective-1409325361

[7] G.Hosking, review of *Why We Need a History of Trust*, (review no. 287a) http://www.history.ac.uk/reviews/review/287a

[8] Cabinet Office Policy Paper: *Government Digital Strategy: 2013* https://www.gov.uk/government/publications/government-digital-strategy/government-digital-strategy

[9] Age UK: *Digital Inclusion Evidence Report 2013* http://www.ageuk.org.uk/Documents/EN-GB/For-professionals/Research/Age%20UK%20Digital%20Inclusion%20Evidence%20Review%202013.pdf?dtrk=true

[10] V. Venkatesh, and H. Bala,"TAM 3: Advancing the Technology Acceptance Model with a Focus on Interventions," Manuscript in-preparation. http://www.vvenkatesh.com/it/organizations/Theoretical_Models.asp#Con=structdefs

[11] B.R.M.Manning, S.Benton. *Qualitative and qualitative methods applied to active ageing – Section 6.11 Future trends.* Textile-led Design for the Active Population. Eds.J.McCann and D.Bryson. Woodhead Publishing 2015, 86-89. ISBN 978-0-85709-538-1

[12] S.Gasson. *Human-centered vs. User-centred approaches to Information systems design.* Journal of Information Technology Theory and Application (JITTA), 5:2, 2003, 29-46

[13] F.D. Davis Perceived usefulness, perceived ease of use, and user acceptance of information technology. Management Information Systems Q, 1989, Sep, 13(3):319-339.

[14] J. Kim, H.A. Park *Development of a Health Information Technology Acceptance Model Using Consumers' Health Behavior Intention*. J Medical Internet Res 2012;14(5):e133

[15] N.J.Adler *Communicating across Cultural Barriers* https://global.duke.edu/sites/default/files/images/NancyAdlerCrossCultComm.pdf

[16] Alzheimer's Association Stages of Alzheimer's Disease Updated October 2003 http://www.bu.edu/alzresearch/files/pdf/StagesofADAlzAssoc3.pdf

[17] S.T.George *The Effect of Therapeutic Alliance on Client Dropout: Hierarchical Modeling of Client Feedback*. Proquest, Umi Dissertation Publishing (2011) ISBN-10: 1243428708

[18] Stanford Encyclopedia of Philosophy. *Game Theory* http://plato.stanford.edu/entries/game-theory/

[19] Institute for Research and Innovation in Social Services (IRISS). *How to make a Radar chart/spider chart.* http://make /sites/default/files/resources/radar_chart_1.pdf

[20] B.R.M.Manning, S.Benton. *Qualitative and qualitative methods applied to active ageing – Section 6.10 Qualitative research aspects of co-design.* Textile-led Design for the Active Population. Eds.J.McCann and D.Bryson. Woodhead Publishing 2015, 84-86. ISBN 978-0-85709-538-1

[21] Manning B.R.M., McCann J., Benton S., Bougourd J. Active Ageing: Independence through Technology Assisted Optimisation. Medical and Care Compunetics 5, June 2008. ISBN 978-1-58603-868-7

[22] UK Parliament – *Public administration Committee Report: Section 2 The Public Sector's Record*, July 2011. http://www.publications.parliament.uk/pa/cm201012/cmselect/cmpubadm/715/71505.htm

[23] T.Williamson, S.Hinder *Public involvement in garment design research.* Textile-led Design for the Active Population Eds.J.McCann and D.Bryson. Woodhead Publishing 2015, 245-255. ISBN 978-0-85709-538-1

[24] J.McCann. *Co-design principles and practice: working with the active ageing.* . Textile-led Design for the Active Population Eds.J.McCann and D.Bryson. Woodhead Publishing 2015, 215-243. ISBN 978-0-85709-538-1

[25] J.Bougourd *Ageing Populations: 3D scanning for apparel size and shape.* Textile-led Design for the Active Population Eds.J.McCann and D.Bryson. Woodhead Publishing 2015, 139-166. ISBN 978-0-85709-538-1

[26] B.Manning, L.Kun. *Information Highway to Home and Back: A Smart Systems Review*. Eds.K.Yogesan, L.Bos, P.Brett and M.C Gibbons. Springer 2009, 5-32. ISBN 978-3-642-01386-7

[27] B.R. Manning,J. McCann, S. Benton, J. Bougourd *Active ageing: independence through technology assisted health optimisation.* Studies in Health Technology and Informatics, IOS Press. 2008; 137:257-62. PMID:18560086

[28] S. Benton, B. Altemeyer, and B. Manning. Behavioural Prototyping: Making Interactive and Intelligent Systems Meaningful for the User. In Daradoumis, T., Demetriades, S., and Xhafa, F. (Eds.) (2012): Intelligent Adaptation and Personalization Techniques in Computer-Supported Collaborative Learning. Studies in Computational Intelligence. Springer Verlag. ISBN 978-3-642-28586-8, 2012.

[29] R.Plutchik *The Nature of Emotions.* American Scientist Vol 89 344-350 http://www.emotionalcompetency.com/papers/plutchiknatureofemotions%202001.pdf

[30] A.Carmichael, R.Barooah *Get your mood on.* 2012 QS Worldwide [Chapters 1-6] http://quantifiedself.com/2012/12/get-your-mood-on-part-1/

[31] R.Trenholm. *Samsung prototypes brainwave-reading wearable stroke detector*. January 22, 2015 http://www.cnet.com/news/samsung-prototypes-brainwave-reading-wearable-stroke-detector/

[32] H. Lövheim *A new three-dimensional model for emotions and monoamine neurotransmitters*. Med Hypotheses (2011),

Implementation and Real-time Verification of an Automatic Modulation Classification Algorithm for Cognitive Vehicle-to-X Communication

Sebastian Sichelschmidt and Dieter Brückmann
Faculty of Electrical, Information and Media Engineering
University of Wuppertal, Germany
(sichelschmidt, brueckm)@uni-wuppertal.de

***Abstract*—**Lack of bandwidth has been and more so will be a major problem in mobile communication. Cognitive Radio offers a solution by utilizing unused licensed spectrum. Automatic Modulation Classification (AMC) is a part of this process. It allows assumptions not only about the channel state, but also about a detected signal's origin. This paper presents a comprehensive AMC classifier, which is capable of detecting a wide range of differently modulated signals even under tough channel conditions. To achieve this, an extensive simulation model has been developed, whose results are validated with real-time tests. These results lead to the conclusion that this kind of AMC is well suited for Vehicle-to-X environments, which will supposably also be affected by the problem of low bandwidth.

***Keywords*—**Automatic Modulation Classification, Car-to-X Communication, Cognitive Radio, Vehicle-to-X Communication

I. Introduction

DUE to worldwide increasing numbers of internet-ready mobile devices and rising amount of data transferred by applications, demands on soft- and hardware to meet these challenges are growing. Four of these causes are:

1) The number of mobile phone accounts has been growing steadily since the introduction of cellular communication. In the last seven years alone, the number of contracts has doubled [1], [2]. This trend will be continued, since more and more machines will be connected to the Internet of Things (IoT). 43% more devices will be getting online this way during the next years [3].
2) Broadband deployment in rural areas nowadays is mainly carried out via mobile networks. Since 2008 the number of households connected via radio link is higher than that of optical fiber connections and is still rising [1].
3) Currently a shift in consumer behavior can be observed from using a computer to using a smartphone for accessing the internet [4]. In many developing countries, smartphones are by far the primary way to connect to the internet.
4) Growing spread of smart phones and highly developed hardware lead to more complex applications and games with an increasing demand on bandwidth. The amount of transmitted application data has been increased from second quarter 2013 to second quarter 2014 by 60% [2].

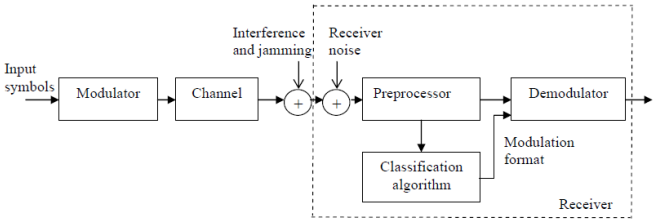


Fig. 1: AMC system block diagram [12]

These four causes led to a growth of transmitted data by a factor of eight in the years 2009 to 2013 [5]. This trend is predicted to continue [3] and will cause a major shortage in provided bandwidth [6].

To address this issue, different countermeasures have been discussed and developed:

1) *Enhancement of spectral efficiency*, for example by implementing new coding and modulation techniques or utilizing intelligent antenna arrays.
2) *Capitalization of new frequency bands*, for example by developing transceivers [7] and antennas [8] which can transmit and receive millimeter waves in urban areas.
3) *Sharing spectrum*, for example by using Cognitive Radio technologies to access external licensed spectrum as a secondary user [9], [10].

In this paper, Cognitive Radio serves as the enabling technology to overcome the described bandwidth issue. One important step of the so called Cognition Cycle [10] is the

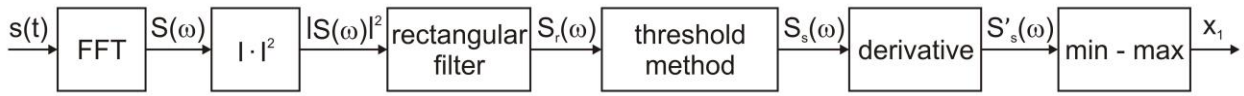


Fig. 2: Bandwidth estimation algorithm

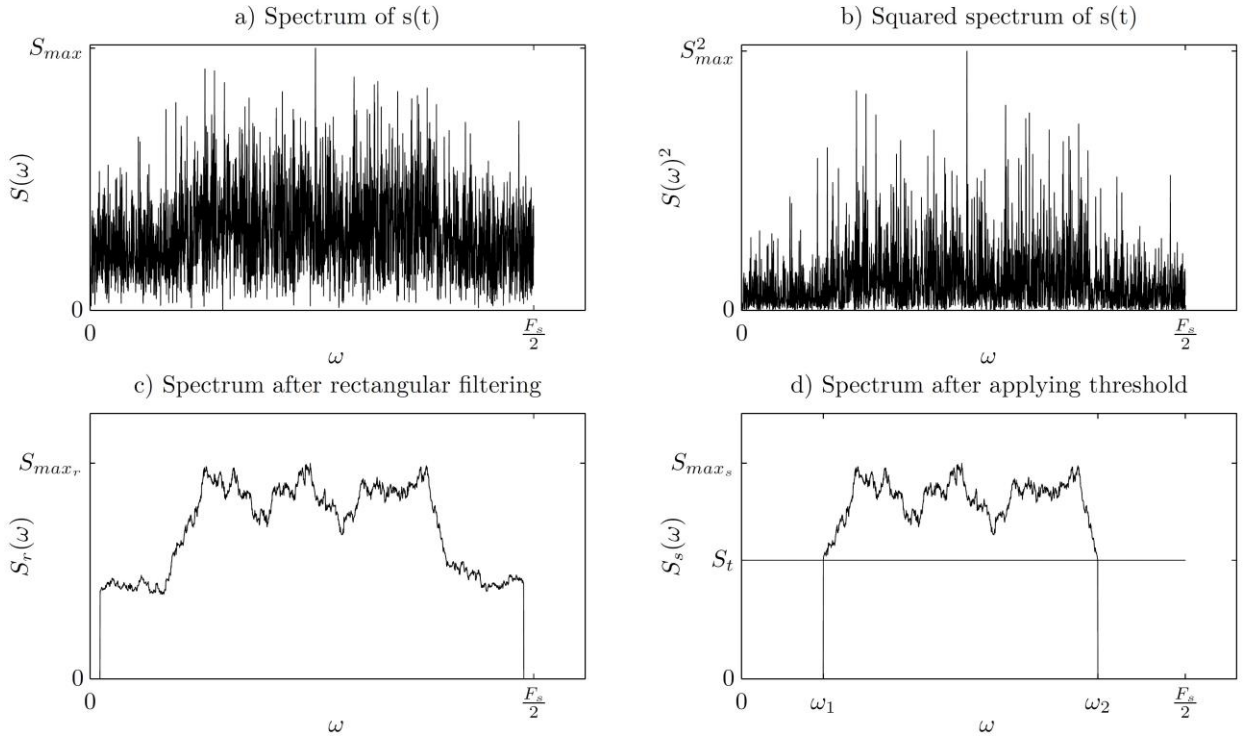


Fig. 3: Bandwidth estimation of an OFDM signal with AWGN (SNR: 5dB)

analysis of the examined RF scene. This can be done by using Automatic Modulation Classification (AMC) to identify communication standards by examining details of a signal's underlying modulation technique [11], [12]. Fig. 1 shows the basic principle with the inherent classification algorithm at its core. In the following a robust classifier is presented, followed by the introduction of an dedicated simulation system and real-time tests.

II. Classification Features

After segmentation feature extraction is applied. This step is most crucial, since the quality of the classification process is determined by the quality of the features. The feature vector $\vec{x}$ is composed of four features x_1 to x_4. Parts of these features are derived from experiments which can be found in literature and therefore are referenced accordingly. These established algorithms are improved by new enhancements and the innovative combination of these features.

x_1 - Bandwidth Estimation

Different bandwidths are used in different wireless transmission standards. LTE for instance uses up to six different sets of bandwidth from 1.25 to 20 MHz. This can be used to classify a signal, since transmission bands are surrounded by smaller guard bands which allow discrimination from one another.

Simple energy detectors do not work well with low SNR [13] and therefore have been enhanced, for example by smoothing algorithms [14]. The x_1 algorithm is based on this kind of smoothing, as can be seen in Fig. 2. In order to suppress noise, the obtained spectrum $S(\omega)$ is squared and filtered by the rectangular window

$$S_r(\omega) = \frac{1}{D} \sum_{k=\omega-\frac{D}{2}}^{\omega+\frac{D}{2}+1} |S(k)|^2. \quad (1)$$

The window width D has been evaluated and set to 100 Hz, which turned out to be a good tradeoff between speed and accuracy. As can be seen in Fig. 3c, the bandwidth can be determined much more clearly from $S_r(\omega)$, than from $S(\omega)$ in Fig. 3a. By applying a threshold method, signal and noise are separated, as is shown in Fig. 3d. To make the algorithm more robust, the actual bandwidth is finally extracted from the derivative $S'_s(\omega)$ of the threshold-spectrum.

x_2 – Guard Interval Estimation

The majority of modern communication standards uses a so called guard interval to reduce errors due to inter-symbol-

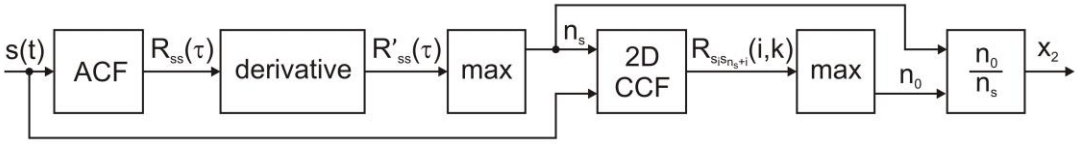


Fig. 4: Guard interval estimation algorithm

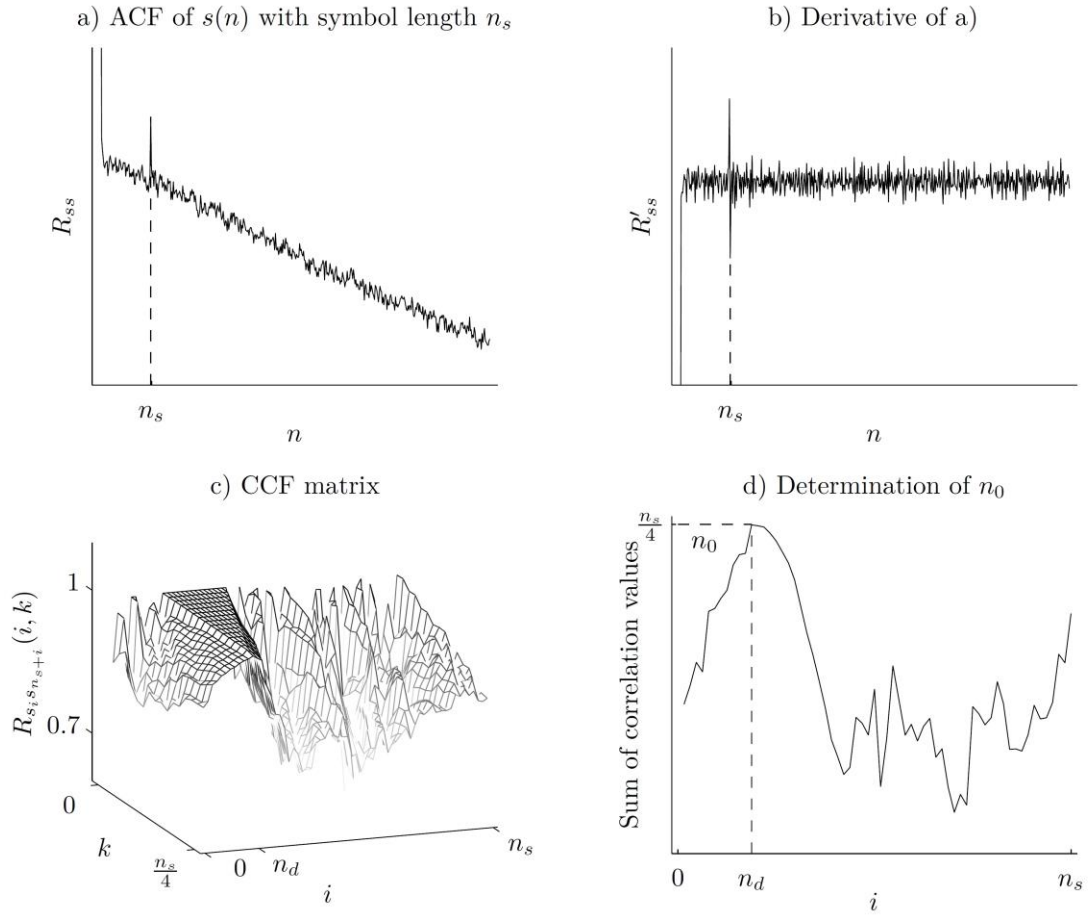


Fig. 5: Intermediate results of guard interval estimation of an IEEE 802.11a OFDM signal

interference, which occur because of multipath fading. The ratio between guard interval and frame length is defined in the standards and therefore holds information about the sender, which can be extracted by an AMC system [15].

Signals, including a guard interval can be considered as wide sense cyclostationary processes. This means that the inherent periodicity, which may not be derivable from the signal itself, can be seen in its autocorrelation function (ACF). For a time-discrete signal $s(n)$ of finite length N, the ACF can be written as

$$R_{ss}(m) = \sum_{n=1}^{N-m} s(n) \cdot s(n+m). \quad (2)$$

Periodicity with n_0 along with finite signal length leads to a decrease of amplitude by $\frac{kn_0}{2}$. (2) can therefore be rewritten as

$$R_{ss}(m) = R_{ss}(m + kn_0) + \frac{kn_0}{2}. \quad (3)$$

From (3), n_0 can be extracted easily. This is still the case, if only a small part of the signal is periodic, as in signals with guard intervals. Since the guard interval usually is put in front of the partly replicated symbol, n_0 cannot be derived from the ACF directly. What can be extracted though, is the symbol length n_s, as can be seen in Fig. 5a.

In the following step n_0 is derived from n_s by applying a two-dimensional crosscorrelation function (CCF) over n_s and the maximum possible guard interval length $\frac{n_s}{4}$ as can be seen in Fig. 5c. Line by line summation of these CCF values reveals not only the position of the guard interval, but also its actual length n_0 (Fig. 4d) which is returned to the classifier as the ratio between guard interval and symbol length.

ISBN: 978-1-61804-318-4

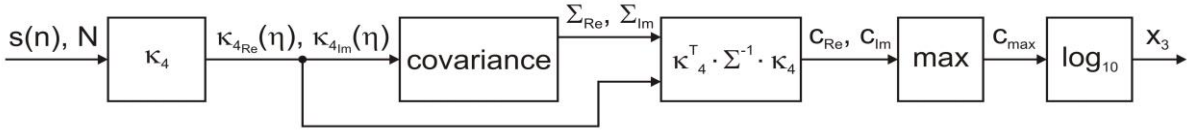


Fig. 6: Algorithm for single-/multi-carrier signal discrimination

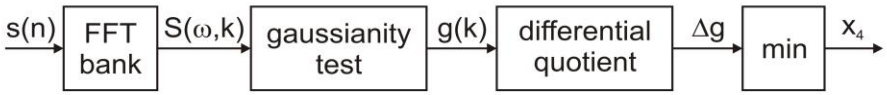


Fig. 7: Algorithm for single-/multi-carrier signal discrimination

x_3 – Single-/ Multi-Carrier Discrimination

With the third feature, separation between single carrier and OFDM modulated multi carrier signals can be achieved. The amplitudes of an OFDM signal are approximately normally distributed over time. Therefore so called "Gaussianity tests" can be applied to numeralize the similarity to normal distribution. In time domain, instead of mean or variance, higher order cumulants are used in these tests, since cumulants of orders greater than two become zero for normally distributed signals [16].

A respective block diagram of the algorithm presented in this paper is shown in Fig. 6. It is based on fourth order cumulants κ_4 as has been shown to be viable in [17]. Based on the Giannakis–Tsatsanis tests [16] and function parameters from [17] the cumulants of signal $s(n)$ with length N can be calculated as

$$\kappa_4(0,\eta,\eta) = \frac{1}{N}\sum_{n=0}^{N-\eta-1} -s^4(n) - s^2(n)\cdot s^2(n+\eta) \qquad (4)$$

with $0 \le \eta \le \mathrm{N}^{0.4} - 1$ for real and imaginary part separately.

These vectors are decorrelated by covariance matrices via Leonov-Shiryaev formula [18] and the logarithmized maximum is returned to the classifier. Some example results are listed in table I, where the discrepancy between single- and multi-carrier signals becomes obvious. Since different standards of the same kind result in slightly different values for x_3, the exact results are used for classification, instead of only differentiating between single- and multi-carrier signals.

x_4 – Estimation of Number of Subcarriers

Since OFDM modulation plays a significant role in modern transmission standards, the fourth feature exclusively addresses the number of subcarriers in OFDM modulated signals. More exactly, the number of IFFT points can be examined, which again - with knowledge of the other classification features - leads to the number of subcarriers.

The number of IFFT points can be extracted by a FFT filter bank [19]. The resulting matrix is examined by the same Gaussianity test, that has been used in the extraction of feature x_3. Since values of the FFT bank output signal are correlated, the FFT with the correct number of points results in a local maximum of the Gaussianity values.

Table I: x_3 values for different modulation techniques and standards

	s(n)	x_3 in dB
multi-carrier	LTE	1.57
	802.11a	0.55
	DVB-T	-1.16
single-carrier	16QAM	5.47
	8PSK	6.62
	GSM	12.63
	Bluetooth	12.06

For robustness reasons, the minimum of the differential quotient is returned to the classifier.

III. Classifier

Best performance, out of several tested classification algorithms, showed the *k*-Nearest Neighbors algorithm (KNN) [20]. A feature vector $\vec{x}$ is assigned to a class, depending on the *k* nearest neighbors' classes. Additional parameters are the *distance metric*, on which base the nearest neighbors are chosen, and the *decision rule*, which states what class is to be selected in the case of a tie.

The simulation results in Chapter IV will document the fact, that for this feature constellation, a setting of *k*=3 with *Manhattan* distance metric and *nearest* neighbor decision rule lead to optimum results. Further optimization like boosting doesn't lead to better results.

IV. Simulation

Previous AMC systems concentrated on specific modulation schemes [12]. The algorithm proposed in this paper has been developed to distinguish between a large number of different transmission standards and also to differentiate between small parameter disparities (i.e. bandwidth, guard interval ratio, number of subcarriers) within

Table II: Classes and Subclasses for classifier training and validation

class	transmission standard	modulation type	band-width	number of subcarriers	guard interv. ratio
1	Bluetooth	BPSK	1 MHz	-	-
2	GSM	GMSK	200 kHz	-	-
3	UMTS/HSPA	W-CDMA	5 MHz	3	-
4	LTE	OFDM/ 16-QAM	1.25 MHz	76	1/4
5		OFDM/ 16-QAM	2.5 MHz	151	1/4
6		OFDM/ 16-QAM	2.5 MHz	151	9/128
7	IEEE 802.11a	OFDM/ 16-QAM	20 MHz	52	1/4
8		OFDM/ 64-QAM	20 MHz	52	1/4
9	white Gaussian noise				

these standards. This way, a huge variety of potential primary and secondary users can be identified. A subset of classes that represent both, a big diversity of transmission standards and also a differentiation in smaller details is listed in Table II. 100 signal samples of each class are generated by a sophisticated signal generator, which has been developed for testing the AMC system presented in this paper.

In a second step, the signals are overlain by noise and interference by using a channel simulator, which has also been developed for this work. A receiver signal $s(t)$ with an optional direct and *N* indirect signal paths can be written as

$$s(t) = \sum_{i=0}^{N} \left(a_i(t)\, cos\left(\omega_S t + \omega_{d_i} t + \phi_i\right)\right) + n(t), \qquad (5)$$

with amplitude a_i, signal frequency ω_S, Doppler frequency shift ω_{d_i}, phase shift ϕ_i and additive noise n.

Previous AMC systems were mostly tested with Additive White Gaussian Noise (AWGN) only [12]. To get information about applicability in various channel conditions the simulations for this paper have been carried out for three channel types with variable parameters (see table III):

1) *AWGN channel – rural areas:* To simulate a channel with a small number of different data paths and a small delay deviation, White Gaussian Noise with adjustable SNR level is added.
2) *Rician channel – suburban areas:* Multiple signal paths can be simulated by setting the different amplitudes via Rice distribution and its *K*-factor. Additionally, Doppler shift can be added and set via the antennas' relative speed v_{rel}.
3) *Rayleigh channel – urban areas:* This channel simulates the case of no direct but many indirect signal paths with a predefined amplitude distribution and optional Doppler shift.

Table III: Channel condition parameters

channel model	parameter	min	step	max
AWGN	SNR in dB	5	5	40
Rician	*K* in dB	1	1	10
	v_{rel} in km/h	0	40	440
Rayleigh	v_{rel} in km/h	0	40	440

The 900 test signals combined with 140 different channel settings as listed in table III result in 126000 different signal samples. Two third are used for training and one third for testing the classifier.

The classification results are shown in Fig. 8. Several observations and conclusions can be drawn from these graphs:

- The higher the signal length, the better the correct classification rate. This difference decreases with higher *k*.
- The *nearest* decision rule leads to slightly better results than the *random* rule, making it a valid choice
- *Manhattan* distance metric should be favored over *Euclidean* distance
- Higher *k* results in better classification results with saturation for *k* greater than three.

Configuring the classifier to *k*=3, *Manhattan* distance metric and *nearest* decision rule results with the given feature space and demanding channel conditions in a correct classification rate of 95.52 %.

V. Real-Time Tests

To validate the simulation results, a real-time environment test bed has been designed. Two USRP N210 Software Defined Radios (SDR) with RFX2400 daughterboards (see Fig. 9a) are set up to transmit the generated signals via 2.4 GHz ISM band on the one side and determine the transmission standard on the other side. Even with no line of sight and high interference the correct classification rate reached 100 %.

A second SDR test setup has been arranged in two cars, as can be seen in Fig. 9b. Due to very low transmission power (-6 dBm), proof of concept could only be given for a range up to 20 meters. From 20 to 50 meters, the classification became partly erroneous and over 50 meters at -93 dBm signal strength no classification could be performed anymore.

VI. Conclusion

Simulations of the proposed AMC system showed the capability of classifying a huge variety of transmission standards under diverse channel conditions. Real-time tests generally proofed the simulation results.

Although the real-time vehicle test could not be executed over greater distances, the AMC system might be suited for implementation in Car2X communication. Low transmission

ISBN: 978-1-61804-318-4

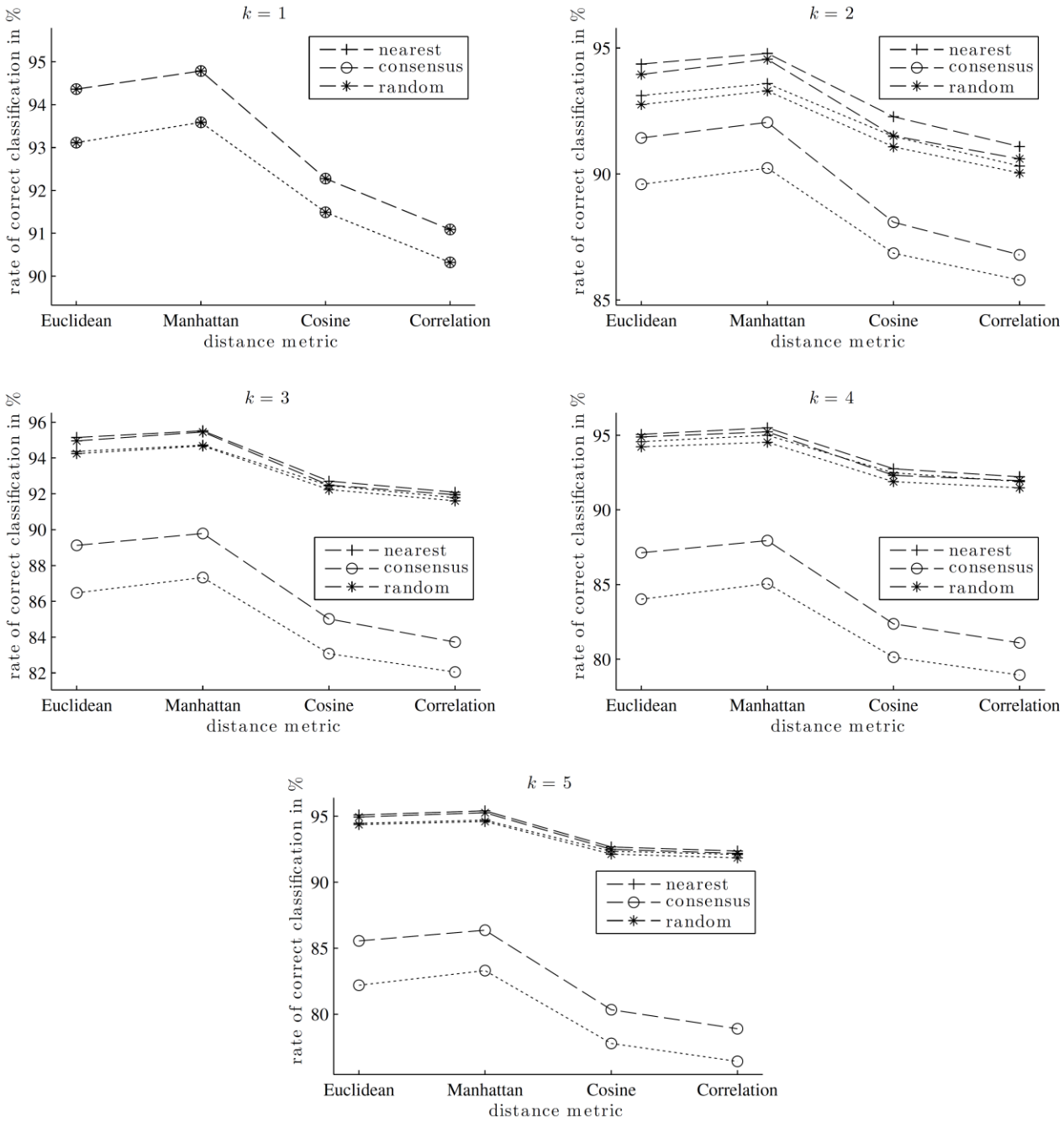
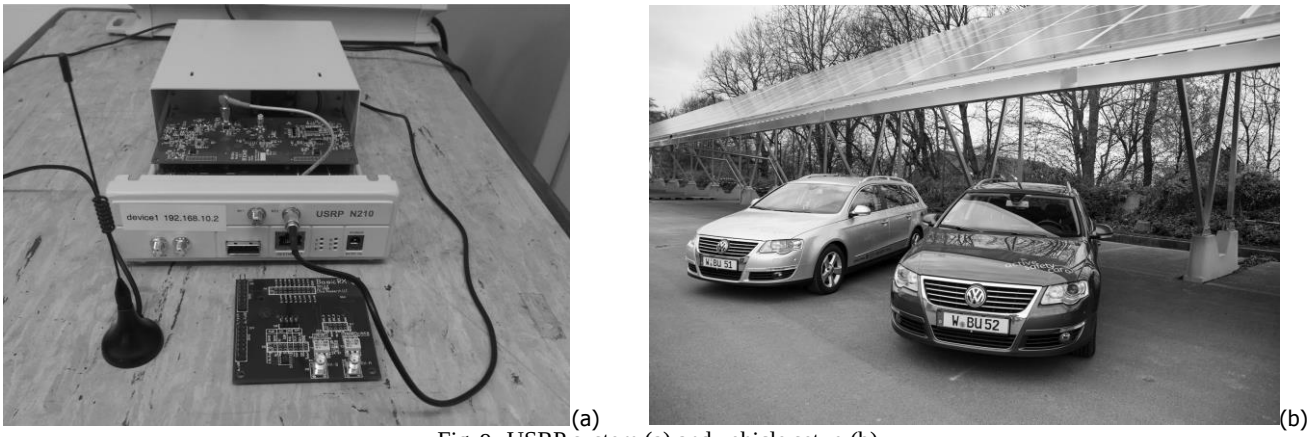


Fig. 8: Simulation results for different signal lengths (doted line: 1000 bit, dashed line: 2000 bit), decision rules (see legend), distance metrics and neighbors *k*



(a) (b)

Fig. 9: USRP system (a) and vehicle setup (b)

ISBN: 978-1-61804-318-4

rates in IEEE 802.11p, network overload due to heavy traffic [21] and high data rates (caused i.e. by transmitted video data [22]) may lead to a shortage of bandwidth in Car2X networks. CR with inbuilt AMC could be a solution to this problem. The great number of in-car antenna systems as well as meeting the high level of power consumption, requested by SDRs can help to make the implementation into vehicles relatively cheap and easy.

REFERENCES

[1] ITU World Telecommunication / ICT Indicators database, "Key 2005-2014 ICT Data For The World," 2014.

[2] Ericsson, "Ericsson Mobility Report," no. August, 2014.

[3] Cisco, "Cisco Visual Networking Index : Global Mobile Data Traffic Forecast Update , 2013 – 2018," 2014.

[4] TNS Infratest Germany, "The Connected Consumer," tech. rep., 2014.

[5] Bundesnetzagentur, "Jahresbericht 2013 Starke Netze im Fokus. Verbraucherschutz im Blick.," [German], 2014.

[6] International Telecommunication Union, "World Radiocommunication Conference," in Final Acts WRC-12, (Geneva, Switzerland), 2012.

[7] S. K. Reynolds, B. A. Floyd, U. R. Pfeiffer, S. Member, T. Beukema, J. Grzyb, C. Haymes, B. Gaucher, and S. Bicmos, "A Silicon 60-GHz Receiver and Transmitter Chipset for Broadband Communications," vol. 41, no. 12, pp. 2820–2831, 2006.

[8] W. Hong, K. Baek, Y. Lee, and Y. Kim, "Design and analysis of a lowprofile 28 GHz beam steering antenna solution for Future 5G cellular applications," in Microwave Symposium (IMS), pp. 14–17, 2014.

[9] J. Mitola III, Cognitive Radio An Integrated Agent Architecture for Software Defined Radio Dissertation. Dissertation, Royal Institute of Technology, 2000.

[10] S. Haykin, "Cognitive Radio: Brain-Empowered," vol. 23, no. 2, pp. 201–220, 2005.

[11] F. Liedtke, "Computer simulation of an automatic classification procedure for digitally modulated communication signals with unknown parameters," Signal Processing, vol. 6, no. 4, pp. 311–323, 1984.

[12] O. A. Dobre, A. Abdi, Y. Bar-ness, and W. Su, "A Survey of Automatic Modulation Classification Techniques : Classical Approaches and New Trends," no. April 2005, pp. 18–19, 2005.

[13] H. Tang, "Some physical layer issues of wide-band cognitive radio systems," in IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, DySPAN, pp. 151–159, 2005.

[14] T. Yucek and H. Arslan, "Spectrum characterization for opportunistic cognitive radio systems," in IEEE Military Communications Conference, MILCOM, 2006.

[15] B. Ramkumar, "Automatic modulation classification for cognitive radios using cyclic feature detection," Circuits and Systems Magazine, IEEE, pp. 27–45, 2009.

[16] G. Giannakis and M. Tsatsanis, "Time-domain tests for Gaussianity and time-reversibility," IEEE Transactions on Signal Processing, vol. 42, no. 12, pp. 3460–3472, 1994.

[17] W. Akmouche, "Detection of multicarrier modulations using 4th-order cumulants," in IEEE Military Communications Conference, vol. 1, pp. 432–436, IEEE, 1999.

[18] D. Brillinger, Time series: data analysis and theory. Holden Day, 1974.

[19] H. Li, Y. Bar-Ness, and A. Abdi, "OFDM modulation classification and parameters extraction," in 1st International Conference on Cognitive Radio Oriented Wireless Networks and Communications, 2006.

[20] T. Cover and P. Hart, "Nearest neighbor pattern classification," IEEE Transactions on Information Theory, vol. I, 1967.

[21] P. Szczurek, B. Xu, O. Wolfson, J. Lin, and N. Rishe, "Learning the relevance of parking information in VANETs," ACM international workshop on VehiculAr InterNETworking - VANET, p. 81, 2010.

[22] K.-h. Lee, J.-n. Hwang, G. Okapal, and J. Pitton, "Driving Recorder Based On-Road Pedestrian Tracking Using Visual SLAM and Constrained Multiple-Kernel," in International IEEE Conference on Intelligent Transportation Systems, 2014.

A Minimax Approach for Robust Estimation of Clock Offset in Wireless Sensor Networks

Xu Wang, Erchin Serpedin, and Khalid Qaraqe

***Abstract*—Recently, wireless sensor networks (WSNs) have received great attention due to their outstanding monitoring and management potential in industrial, medical and environmental applications. WSNs require all the nodes to run on a common time scale in order to perform tasks such as localization, tracking, data fusion, and sleeping/waking up nodes. Operation of WSNs highly depends on the synchronization among the nodes clocks. Assuming a two-way message exchange mechanism, the maximum likelihood estimators (MLEs) for the clock offset between the nodes of a WSN were derived for exponential and Gaussian random network delays. However, the derived MLEs are not robust in the presence of non-Gaussian and non-exponential delays. In this paper, a minimax approach is considered to derive two robust estimators to cope with the possible variations in the distribution of Gaussian and exponential random delays, respectively. Based on the simulation results, it is shown that the proposed robust estimators outperform the corresponding MLEs in terms of robustness.**

***Index Terms*—Clock synchronization, wireless sensor networks (WSNs), minimax, robust estimators.**

I. Introduction

WIRELESS sensor networks (WSNs) present many applications such as in health care and pollution monitoring, monitoring of industrial processes and data fusion. Most of these applications require the clocks of the nodes to be synchronized [1], [2], [3], [4]. Hence, developing efficient clock synchronization protocols is drawing a great deal of interest. In literature, most of the widely-used clock synchronization protocols, such as the Network Time Protocol (NTP) [5] and the Timing Synch Protocol for Sensor Networks (TPSN) [6], employ a two-way message exchange mechanism. In this mechanism, the estimation of the clock offset between two nodes plays a critical role to adjust and synchronize the nodes in one common reference time.

Maximum Likelihood Estimator (MLE) is known to be a good estimator since in general it achieves asymptotically the Cramér-Rao lower bound (CRLB) as the number of observations goes to infinity. Assuming a known fixed delay and a symmetric exponential distributed random delay with a known mean, Abdel-Ghaffar [7] claimed that the MLE of the clock offset does not exist. However, Jeske [8] clarified the result stated in [7] and derived the MLE by assuming an unknown fixed delay and without knowledge of the mean of the symmetric exponential random delay. The minimum variance unbiased estimator (MVUE) for the clock offset was proposed in [9] in the presence of both symmetric and asymmetric exponential delay models. In addition, it is shown in [9] that MVUE coincides with MLE [8] under the symmetric exponential model. Noh et al. [10] derived the MLE of the clock offset by assuming a symmetric normally distributed random delay model. For the study with other random delay distributions, such as Weibull and Gamma, the reader is referred to [11], [12] for more details.

In practice, it is difficult to determine which random delay model should be selected in a given WSN. Also, the distribution of the random delay may change as the time evolves. In this case, the random delay does not follow a single distribution. Therefore, in this paper, it is shown that the MLEs in [8] and [10] for the exponential and Gaussian distributions are not robust to the time-varying nature of the random delay. Moreover, using the minimax technique, we propose two robust estimators based on the exponential and Gaussian models, respectively.

The rest of the paper is structured as follows. Section II describes the system model that will be used throughout the paper. In Section III, the state-of-the-art MLEs under Gaussian and exponential scenarios are briefly mentioned and the corresponding robust estimators are proposed using the minimax approach. In Section IV, the simulation results are shown. Section V concludes the paper.

II. System Model

A two-way message exchange mechanism in a WSN between node A and node B is depicted in Fig. 1 [9]. During message exchange round i, the synchronization starts at node A and the time is measured as T_i^a by the clock of node A. Node A sends a synchronization message containing the time measurement T_i^a to node B. Then, node B records the reception time T_i^b and sends an acknowledgement message containing the time measurements T_i^b and T_i^c back to node A at T_i^c. The acknowledgement message is received by node A at T_i^d and the message exchange round i ends. This exchange process is repeated for N times where N denotes the required number of observations.

Based on the system model shown in Fig. 1, the synchronization problem can be expressed as

$$T_i^b = T_i^a + d + \theta + P_i,$$
$$T_i^d = T_i^c + d - \theta + Q_i,$$

where T_i^a, T_i^d and T_i^b, T_i^c are time measurements recorded at node A and B, respectively, d denotes the fixed portion of the delays and θ stands for the clock offset. Moreover, P_i and Q_i represent the uplink and downlink random delays and they are

X. Wang, E. Serpedin and K. Qaraqe are with the Department of Electrical and Computer Engineering, Texas A&M University, College Station (e-mail:serpedin@ece.tamu.edu).

ISBN: 978-1-61804-318-4

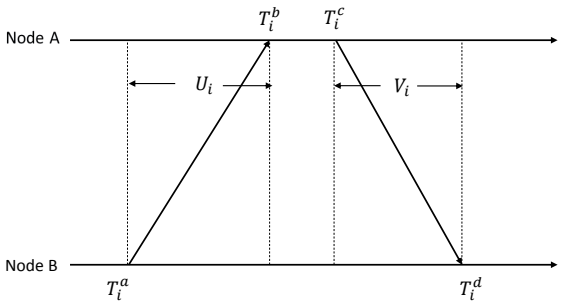


Fig. 1. Two-way message exchange mechanism

assumed to be independent and identically distributed random variables. Furthermore, the above equations can be simplified as

$$U_i = d + \theta + P_i, \quad (1)$$
$$V_i = d - \theta + Q_i, \quad (2)$$

where U_i and V_i are observations defined as $U_i = T_i^b - T_i^a$ and $V_i = T_i^d - T_i^c$.

III. Main Results

Under the assumption of an exponential delay model, i.e., $P, Q \sim \exp(\lambda)$, the MLE for the clock offset (EML) was reported in [8]:

$$\hat{\theta}_{EML} = \frac{U_{(1)} - V_{(1)}}{2}, \quad (3)$$

where $U_{(1)}$ and $V_{(1)}$ denote the minimum value of the observations $\{U_i\}_{i=1}^N$ and $\{V_i\}_{i=1}^N$, respectively. On the other hand, under the assumption of a Gaussian delay model, i.e., $P, Q \sim \mathcal{N}(\mu, \sigma)$, the MLE (GML) was expressed in [10] as:

$$\hat{\theta}_{GML} = \frac{\sum_{i=1}^{N}(U_i - V_i)}{2N} = \frac{\overline{U} - \overline{V}}{2}, \quad (4)$$

where $\overline{U}$ and $\overline{V}$ stand for the average values of observations $\{U_i\}_{i=1}^N$ and $\{V_i\}_{i=1}^N$, respectively. However, the simulation results in [10] show that the performance of these MLEs highly depends on the type of random delay models used. Therefore, these estimators are not expected to be robust to possible variations or changes in the random portion of the delays. In accordance with the uncertainties of the random delay, we formulate a simplified (sub-optimal) estimation problem by subtracting (2) from (1) and then dividing the result by 2:

$$\frac{U_i - V_i}{2} = \theta + \frac{P_i - Q_i}{2}, \quad (5)$$

which resumes to

$$X_i = \theta + Z_i, \quad (6)$$

where $X_i = \frac{U_i - V_i}{2}$ and $Z_i = \frac{P_i - Q_i}{2}$. In this way, we get rid of some nuisance parameters, i.e., the fixed portion of the delays d, and formulate a location estimation problem [13].

Huber's M-estimators [14] are employed herein paper to estimate the location parameter θ in (6). Specifically, the M-estimators admit the general form

$$\hat{\theta}_M = \arg\min_{\theta} \sum_{i=1}^{N} \Psi(X_i - \theta), \quad (7)$$

where Ψ is a function that determines the estimator. Under the assumption that Ψ is a convex and symmetric function, the asymptotic variance of the estimator $\hat{\theta}_M$ is given by

$$V(\psi, f) = \frac{\int \psi^2 f}{\left(\int \psi' f\right)^2}, \quad (8)$$

where $\psi = \Psi'$ is defined as the influence curve and f denotes the density function of Z. To increase the robustness of the resulting estimator, Huber applied the minimax formulation

$$\min_{\psi} \max_{f \in \mathcal{F}} V(\psi, f), \quad (9)$$

where f belongs to a certain compact distribution set $\mathcal{F}$. The solution to (9) is given by $\psi^* = -(f^*)'/f^*$, where f^* represents the least-favorable distribution that minimizes the Fisher information $I(f) = \int (f'/f)^2 f dx$ over the set $\mathcal{F}$:

$$f^* = \arg\min_{f \in \mathcal{F}} I(f). \quad (10)$$

Thus, the pair (ψ^*, f^*) satisfies the minimax property

$$V(\psi^*, f) \le V(\psi^*, f^*) \le V(\psi, f^*), \quad (11)$$

for all $f \in \mathcal{F}$ and ψ. The left-hand side (LHS) in (11) implies that if the estimator based on ψ^* is designed, the estimator will outperform $V(\psi^*, f^*)$ irrespective of f if $f \in \mathcal{F}$. This property ensures the robustness of the designed estimator among a certain class of $f \in \mathcal{F}$.

A. Robust Estimator based on the Exponential Model

In this section, the M-estimator under symmetric exponential random delays (EM) is derived based on the aforementioned minimax technique. Specifically, we assume that the random portion of delays P and Q mainly follow from an exponential distribution $\exp(\lambda)$. It turns out that $Z = (P - Q)/2$ mainly follows from a Laplace distribution with location parameter 0 and scale parameter $\frac{1}{2\lambda}$. As discussed earlier, the distribution of the random delay may be time-varying; this is the reason why we use "mainly" here to describe the distributions. More accurately, we assume that the distribution of Z belongs to a class

$$\mathcal{F}_{\mathcal{L}} = \{f : f = (1 - \epsilon)\mathcal{L}(0, 1/2\lambda) + \epsilon h\}, \quad (12)$$

where $\mathcal{L}(\cdot)$ stands for a Laplace distribution, $0 \le \epsilon < 1$ is a contamination parameter and h is an arbitrary distribution. Intuitively, it means that the distribution of Z follows from a Laplace distribution $1 - \epsilon$ of the time and some other distribution h ϵ of the time.

In order to derive the EM, equation (10) is first employed to calculate the least-favorable distribution that belongs to the set in (12). The influence curve ψ^* is then given by $\psi^* =$

$-(f^*)'/f^*$. In this case, it is shown in Appendix A that the influence curve ψ^* admits the form

$$\psi^*_{\mathcal{L}}(x) = \begin{cases} k, & x > 0, \\ -k, & x < 0, \end{cases} \tag{13}$$

where k, λ and ϵ share the relationship

$$\frac{k}{2\lambda} = 1 - \epsilon.$$

Since $\psi = \Psi'$, equation (7) implies that

$$\sum_{i=1}^{N} \psi^*_{\mathcal{L}}(X_i - \hat{\theta}_{EM}) = 0.$$

However, it is reported in [14] that M-estimates of location have to be supplemented by a simultaneous estimate of a scale parameter S due to the time-invariant property of the M-estimates. Thus we are actually facing a two-parameter problem

$$\sum_{i=1}^{N} \psi^*_{\mathcal{L}}\left(\frac{X_i - \hat{\theta}_{EM}}{\hat{S}}\right) = 0. \tag{14}$$

Since the joint estimation of the location and the scale parameter is time-consuming and computationally complex [15], the methods based on a pre-set value of the scale parameter S are usually adopted to estimate the location parameter θ. By setting the initial value of $\hat{\theta}_{EM}$ as

$$\hat{\theta}_{EM}(0) = \text{med}\{X_i\},$$

and pre-setting

$$\hat{S} = \text{med}\{|X_i - \hat{\theta}_{EM}(0)|\},$$

an iterative method named as the modified residuals [14] is employed herein paper to derive the M-estimate $\hat{\theta}_{EM}$. Specifically, the iteration follows the equation

$$\hat{\theta}_{EM}(m+1) = \hat{\theta}_{EM}(m) + \frac{\hat{S}}{N}\sum_{i=1}^{M} \psi\left(\frac{X_i - \hat{\theta}_{EM}(m)}{\hat{S}}\right). \tag{15}$$

The proof of convergence can be found in [14] and it is also shown that the converged result will always lead to the solution of (14), where $\hat{S}$ is pre-set as $\text{med}\{|X_i - \hat{\theta}_{EM}(0)|\}$.

B. Robust Estimator based on the Gaussian Model

In this section, the M-estimator under symmetric Gaussian random delays (GM) is proposed. Assuming the random portion of delays P and Q mainly follow from an Gaussian distribution $\mathcal{N} \sim (\mu, \sigma)$, it implies that $Z = (P-Q)/2$ mainly follows from a Gaussian distribution with mean 0 and variance $\sigma^2/2$. Similarly, it is assumed that the distribution Z belongs to a class

$$\mathcal{F}_{\mathcal{N}} = \{f : f = (1-\epsilon)\mathcal{N}(0, \sqrt{\sigma^2/2}) + \epsilon h\}. \tag{16}$$

Based on the derivations in Appendix A, the influence curve is given by

$$\psi^*_{\mathcal{N}}(x) = \begin{cases} 2x/\sigma^2, & |x| < k\sigma^2/2, \\ k, & x \geq k\sigma^2/2, \\ -k, & x \leq -k\sigma^2/2, \end{cases} \tag{17}$$

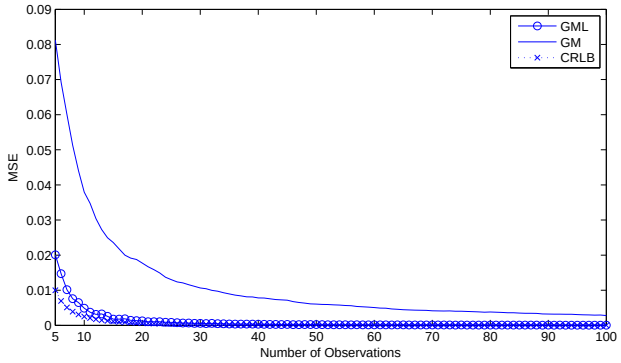


Fig. 2. MSEs of clock offset estimators for exponential random delays ($\lambda = 1$).

where k, σ and ϵ share the relationship

$$\int_{-k\sigma^2/2}^{k\sigma^2/2} g(x)dx + \frac{g(-k\sigma^2/2) + g(k\sigma^2/2)}{k} = \frac{1}{1-\epsilon}.$$

As described in Section III-A, the M-estimate $\hat{\theta}_{GM}$ can be obtained from a two-parameter problem

$$\sum_{i=1}^{N} \psi^*_{\mathcal{N}}\left(\frac{X_i - \hat{\theta}_{GM}}{\hat{S}}\right) = 0. \tag{18}$$

Replacing $\psi^*_{\mathcal{L}}$ with $\psi^*_{\mathcal{N}}$ and following the same iterative steps mentioned in Section III-A (the modified residuals method) leads to the solution of $\hat{\theta}_{GM}$.

IV. Simulation Results

In this section, computer simulation results are presented to illustrate the performance of EM and GM with respect to EML, GML and CRLB in terms of estimation accuracy of the clock offset. Herein paper, the estimation accuracy is measured as the mean square error (MSE). The contamination parameter ϵ is selected as 0.1 throughout the section. In other words, with $\epsilon = 0.1$, EM and GM are designed by assuming z belongs to the distribution sets (12) and (16), respectively.

In Figs. 2, 3 and 4, an EM based on (12) and (13) is simulated. Fig. 2 shows the MSE of EM and EML when the random delay distributions P and Q are symmetric exponential pdfs. In addition, the CRLB is also provided as a reference. It can be seen that the performance of EM is comparable to EML especially when the number of observations becomes large. Fig. 3 depicts the MSE of EM and EML when the random delay distributions are exponential and contaminated with 20% Gaussian. It is observed that EML is very unstable and performs poorly. Additionally, EM outperforms EML in terms of stability and accuracy. A similar trend can be observed in Fig. 4 in which the random delays are assumed to be exponential and contaminated with 40% Gaussian.

In Figs. 5, 6 and 7, a GM based on (16) and (17) is illustrated. In reference to CRLB, Fig. 5 illustrates the MSE of GM and GML when the random delay distributions P and Q are symmetric Gaussian pdfs. The simulation results show that the performance of GM is comparable to that

ISBN: 978-1-61804-318-4

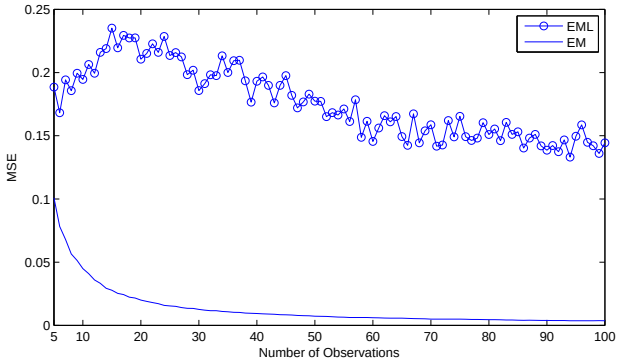


Fig. 3. MSEs of clock offset estimators for exponential random delays contaminated with 20% Gaussian random delays ($\lambda = 1, \mu = 0, \sigma = 1$).

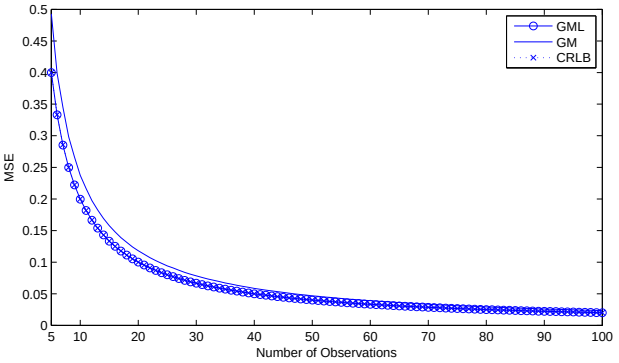


Fig. 5. MSEs of clock offset estimators for Gaussian random delays ($\mu = 0, \sigma = 2$).

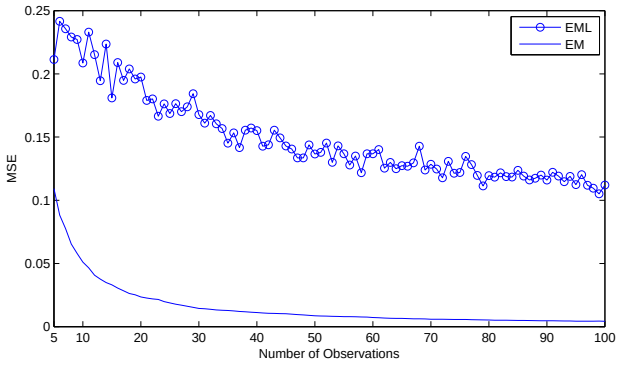


Fig. 4. MSEs of clock offset estimators for exponential random delays contaminated with 40% Gaussian random delays ($\lambda = 1, \mu = 0, \sigma = 1$).

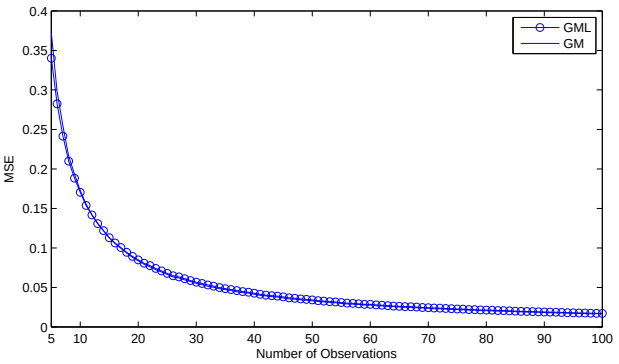


Fig. 6. MSEs of clock offset estimators for Gaussian random delays contaminated with 20% exponential random delays ($\mu = 0, \sigma = 2, \lambda = 1$).

of GML even for a small number of observations. Fig. 6 depicts the MSE of GM and GML when the random delay distributions are Gaussian contaminated with 20% exponential. It is observed that their performances are almost identical and the MSEs nearly overlap. However, when random delays are modeled as Gaussian contaminated with 40% exponential, it is shown in Fig. 7 that GM is more robust and outperforms GML in terms of the MSE. Numerous other simulations, such as the comparison of GM and GML assuming a Gaussian contaminated with Weibull and Gamma, are not shown herein due to the space limitation. However, these simulations all share a similar trend to the ones depicted in Figs. 5, 6 and 7.

V. Conclusions

The paper provides a minimax approach to design robust estimators (EM and GM) for estimating the clock offset under the assumption of an exponential and Gaussian random delay in a two-way message exchange mechanism, respectively. The simulation results illustrate that the proposed estimators are comparable to the state-of-the-art maximum likelihood estimators when the random portion of delays are pure exponential or Gaussian. In addition, in the presence of some contaminated random delays, EM and GM outperform the MLEs in terms of MSE. Therefore, the designed estimators are more robust when the random delays are time-varying.

Appendix A
Derivation of the Influence Curves

Let $\mathcal{F}$ be the set of all density functions coming from distribution g through ϵ contamination

$$\mathcal{F} = \{f : f = (1-\epsilon)g + \epsilon h\}.$$

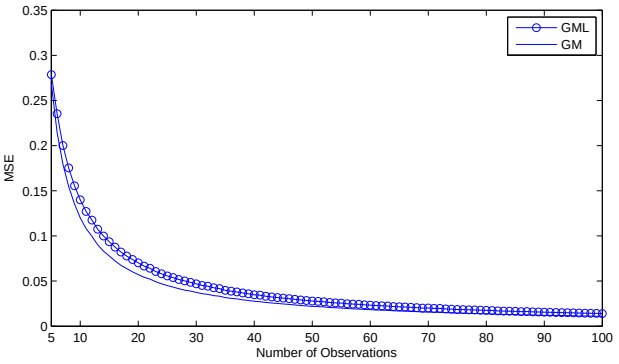


Fig. 7. MSEs of clock offset estimators for Gaussian random delays contaminated with 40% exponential random delays ($\mu = 0, \sigma = 2, \lambda = 1$).

ISBN: 978-1-61804-318-4

Let $x_0 < x_1$ be the endpoints of the interval $|g'/g| \leq k$, where k is related to ϵ by:

$$\int_{x_0}^{x_1} g(x)dx + \frac{g(x_0)+g(x_1)}{k} = \frac{1}{1-\epsilon}. \tag{19}$$

The least-favorable distribution [14] that minimizes the Fisher information $I(f) = \int (f'/f)^2 f dx$ is expressed as :

$$f^*(x) = \begin{cases} (1-\epsilon)g(x_0)e^{k(x-x_0)}, & x \leq x_0, \\ (1-\epsilon)g(x), & x < x_0 < x_1, \\ (1-\epsilon)g(x_1)e^{-k(x-x_1)}, & x \geq x_1. \end{cases} \tag{20}$$

Thus, the influence curve, given by $\psi^* = -(f^*)'/f^*$, admits the form

$$\psi^*(x) = \begin{cases} -k, & x \leq x_0, \\ -g'(x)/g(x), & x < x_0 < x_1, \\ k, & x \geq x_1. \end{cases} \tag{21}$$

In the case of a contaminated Laplace distribution in (12) with $g(x) = \mathcal{L}(0,b)(b>0)$ where $b = 1/(2\lambda)$, i.e.,

$$g(x) = \frac{1}{2b}\exp(-\frac{|x|}{b}).$$

If follows that $|g'(x)/g(x)| = 1/b$ for $x \neq 0$. It is first assumed that $1/b \leq k$, and we can obtain that $x_0 = -\infty$ and $x_1 = \infty$. Plugging these values back into (19), the LHS equals

$$\int_{-\infty}^{\infty} g(x)dx + \frac{g(-\infty)+g(\infty)}{k} = 1,$$

which contradicts the fact that ϵ is any real number in the range $[0,1)$. Therefore, it can be concluded that $1/b > k$ and $x_0 = 0^-,\ x_1 = 0^+$. Based on (20) and (21), it follows that

$$f^*_{\mathcal{L}}(x) = \begin{cases} \frac{1-\epsilon}{2b}e^{kx}, & x > 0, \\ \frac{1-\epsilon}{2b}e^{-kx}, & x < 0, \end{cases} \tag{22}$$

and

$$\psi^*_{\mathcal{L}}(x) = \begin{cases} k, & x > 0, \\ -k, & x < 0, \end{cases} \tag{23}$$

where k, b and ϵ are related as follows:

$$bk = 1 - \epsilon. \tag{24}$$

On the other hand, in the case of a contaminated Gaussian distribution in (16) with $g(x) = \mathcal{N}(0,\sigma_0)$ where $\sigma_0 = \sqrt{\sigma^2/2}$, i.e.,

$$g(x) = \frac{1}{\sqrt{2\pi\sigma_0^2}}\exp -\frac{x^2}{2\sigma_0^2},$$

following the same steps above, it turns out that the least-favorable distribution and the influence curve [16] for a contaminated Gaussian set (16) are given by

$$f^*_{\mathcal{N}}(x) = \begin{cases} \frac{1-\epsilon}{\sqrt{2\pi\sigma_0^2}}e^{-\frac{x^2}{2\sigma_0^2}}, & |x| < k\sigma_0^2, \\ \frac{1-\epsilon}{\sqrt{2\pi\sigma_0^2}}e^{-kx+k^2\sigma_0^2/2}, & x \geq k\sigma_0^2, \\ \frac{1-\epsilon}{\sqrt{2\pi\sigma_0^2}}e^{kx+k^2\sigma_0^2/2}, & x \leq -k\sigma_0^2, \end{cases} \tag{25}$$

and

$$\psi^*_{\mathcal{N}}(x) = \begin{cases} x/\sigma_0^2, & |x| < k\sigma_0^2, \\ k, & x \geq k\sigma_0^2, \\ -k, & x \leq -k\sigma_0^2, \end{cases}$$

where k and ϵ are related by

$$\int_{-k\sigma_0^2}^{k\sigma_0^2} g(x)dx + \frac{g(-k\sigma_0^2)+g(k\sigma_0^2)}{k} = \frac{1}{1-\epsilon}.$$

ACKNOWLEDGMENT

This paper was made possible by NPRP grant NPRP 4-1293-2-513 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

REFERENCES

[1] I. Sari, E. Serpedin, K. L. Noh, Q. Chaudhari, and B. Suter, "On the joint synchronization of clock offset and skew in RBS-protocol," *IEEE Trans. Commn.*, vol. 56, no. 5, pp. 700–703, 2008.

[2] K. L. Noh and E. Serpedin, "Pairwise broadcast clock synchronization for wireless sensor networks," in *IEEE International Symposium on World of Wireless, Mobile and Multimedia Networks*, Espoo, Finland, Jun. 2007, pp. 1–6.

[3] J. S. Kim, J. Lee, E. Serpedin, and K. Qaraqe, "A robust estimation scheme for clock phase offsets in wireless sensor networks in the presence of non-Gaussian random delays," *Signal Processing*, vol. 89, no. 6, pp. 1155–1161, 2009.

[4] Q. M. Chaudhari, E. Serpedin, and K. Qaraqe, "Some improved and generalized estimation schemes for clock synchronization of listening nodes in wireless sensor networks," *IEEE Trans. Commn.*, vol. 58, no. 1, pp. 63–67, 2010.

[5] D. Mills, "Internet time synchronization: The network time protocol," *IEEE Trans. Commn.*, vol. 39, no. 10, pp. 1482–1493, 1991.

[6] S. Ganeriwal, R. Kumar, and M. B. Srivastava, "Timing synch protocol for sensor networks," in *Proc. 1st International Conference on Embedded Network Sensor Systems*, Los Angeles, CA, USA, Nov. 2003.

[7] H. S. Abdel-Ghaffar, "Analysis of synchronization algorithms with time-out control over networks with exponentially symmetric delays," *IEEE Trans. Commn.*, vol. 50, no. 10, pp. 1652–1661, 2002.

[8] D. R. Jeske, "On maximum-likelihood estimation of clock offset," *IEEE Trans. Commn.*, vol. 53, no. 1, pp. 53–54, 2005.

[9] Q. M. Chaudhari, E. Serpedin, and K. Qaraqe, "On minimum variance unbiased estimation of clock offset in a two-way message exchange mechanism," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2893–2904, 2010.

[10] K. L. Noh, Q. M. Chaudhari, E. Serpedin, and B. W. Suter, "Novel clock phase offset and skew estimation using two-way timing message exchagnes for wireless sensor networks," *IEEE Trans. Commn.*, vol. 55, no. 4, pp. 766–777, 2007.

[11] A. Ahmad, A. Noor, E. Serpedin, H. Nounou, and M. Nounou, "On clock offset estimation in wireless sensor networks with weibull distributed network delays," in *Proc. 20th International Conference on Pattern Recognition*, Istanbul, Turkey, Aug. 2010, pp. 2322–2325.

[12] A. Papoulis, *Random Variable and Stochastic Processes*. New York, NY, USA: McGraw-Hill, 1991.

[13] S. A. Kassam and V. Poor, "Robust techniques for signal processing: A survey," *Proceedings of the IEEE*, vol. 73, no. 3, pp. 433–481, 1985.

[14] P. J. Huber and E. M. Ronchetti, *Robust Statistics*. Hoboken, NJ, USA: Wiley, 2009.

[15] F. R. Hampel, *Robust statistics: The approach based on influence functions*. Hoboken, NJ, USA: Wiley, 1986.

[16] K. Kim and G. Shevlyakov, "Why Guassianity?" *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 102–113, 2008.

Extended Watchdog Mechanism for Byzantine Failure Resilient Ad-Hoc Networks

Norihiro Sota and Hiroaki Higaki
Tokyo Denki University, Japan

***Abstract*—Wireless multihop networks consist of numbers of wireless nodes. Hence, introduction of failure detection and recovery is mandatory. Until now, various failure detection and recovery methods such as route switch and multiple routee detection have been proposed based on an assumption with stop failure model. However, the assumption that failed wireless nodes never transmit any messages is too restrict the area where the proposed methods can be applied. In order to solve this problem, we propose a novel failure detection and notification method that suppots not only stop failure but also Byzantine failure. That is, it is possible for failed wireless nodes to transmit malicious messages not according to the data message transmission and the failure detection and notification protocols unconsciously due to failure or even intentionally. Here, the design of failure detection and notification protocols is critical. In this paper, Byzantine failures in an intermediate node are detected by its multiple neighbor wireless nodes cooperatively since the neighbor wireless nodes are also vulnerable and might transmit erroneous failure notifications. From the performance viewpoint, no additional control messages are required to be transmitted while no failure wireless node is detected, i.e., in usual data message transmissions.**

***Index Terms*—Ad-Hoc Networks, Fault-Tolerant Wireless Networks, Byzantine Failure, Cooperative Watchdog, Protocol, Ad-Hoc Routing.**

I. Introduction

In mobile wireless ad-hoc networks (MANETs) and wireless sensor networks, data messages are transmitted according to wireless multihop transmissions where each intermediate wireless nodes along the wireless multihop transmission route forwards them from the source wireless node to the destination one. Usually, the wireless transmission range of each wireless node is limited and the wireless nodes are assumed to be distributed densely enough for all the wireless nodes to be possible to communicate with some neighbor wireless nodes directly and to communicate with almost all the other wireless nodes by the wireless multihop communication. This is because, all the observation area is required to be covered by at least one sensor node and the sensor data messages are required to be transmitted to one of the sink nodes in sensor networks and enough high connectivity by wireless multihop transmissions is required in usual mobile wireless ad-hoc networks.

Such wireless multihop networks consist of numbers of wireless nodes. Hence, it is impossible to operate such wireless multihop networks continuously without failure detection, notificatio and recovery mechanisms. That is, higher resilient wireless multihop networks are required. Until now, various techniques for fault-tolerant distributed systems such as distributed failure detection, notificatio and recovery algorithms and systems have been proposed [3], [10]. For wireless multihop networks, only a naive watchdog method and its slight extensions have been proposed. Here, almost only the stop failure model in which failed wireless nodes become silent and never transmit any data and control messages is supported. Even though some methods support the Byzantine failure model, desirable behavior such as only erroneous data messages are transmitted is assumed. As discussed in this paper, erroneous and/or malicious data message transmissions deviated from the application protocols and erroneous and/or malicious failure detection and notificatio transmissions are required to be supported. This paper proposes a novel cooperative watchdog method and designs a data message transmission protocol with an extension of the Byzantine failure detection and notificatio and a routing protocol for detection of watchdoggable wireless multihop transmission routes based on floodin based ad-hoc routing protocols such as AODV [7].

II. Related Works

Suppose a wireless multihop transmission route $\mathcal{R} := ||N_0(= N^s) \dots N_n(= N^d)\rangle\rangle$ from a source wireless node N^s to a destination one N^d in a wireless multihop network such as a mobile wireless ad-hoc network and a wireless sensor network. If one of the intermediate wireless nodes N_f $(0 < f < n)$ is detected to be failed by one of its neighbor wireless nodes, a failure notificatio message is transmitted to the source node N^s and another wireless transmission route $\mathcal{R}'$ without N_f is searched and detected. Then, data messages are transmitted through not $\mathcal{R}$ but $\mathcal{R}'$. Until now, some failure detection, notificatio and recovery by re-routing have been proposed [4], [9]. In addition, for avoidance of high communication and time overhead for search of a detour wireless multihop transmission route, various multiple route detection protocol have also been proposed where multiple wireless multihop transmission routes are detected in a routing protocol and the routes are switched each time a failed intermediate wireless node is detected along an available wireless multihop transmission route [1], [5], [8]. These papers only discuss the

ISBN: 978-1-61804-318-4

methods to switch wireless multihop transmission routes after detection of failure of one of the intermediate wireless nodes. The discussion of failure detection and notificatio is almost out of range.

There are some various failure model for wireless nodes [10]. Almost all of them assume that wireless nodes fail according to the following stop failure model where the failed wireless nodes become silent and the stop failure is detected by at least one of the other wireless nodes by using periodically transmitted "Hello" or "I'm alive" messages.

[Stop Failure Model]

A failed wireless node stops. It becomes silent, i.e., it never transmits and receives any data and control messages. □

A stop failure usually detected by using a timer [2]. A neighbor wireless node Q of another wireless node P sets its timer. If Q does not receive a message from P before the expiration of the timer, Q detects failure of P. In wireless multihop data message transmissions along $\mathcal{R}$, in cases that there are no failed wireless nodes in $\mathcal{R}$, within a certain interval after the time when an intermediate wireless node N_{i-1} forwards a data message m to its next-hop wireless node N_i, N_i forwards m to its next-hop wireless node N_{i+1}. As shown in Figure 1, under an assumption of the disk model wireless signal transmissions, N_{i-1} is surely within the wireless transmission range of N_i and m transmitted from N_i to N_{i+1} is surely overheard by N_{i-1}. Hence, if N_{i-1} does not overhear m forwarded by N_i to N_{i+1} during a certain interval after N_{i-1} forwards m to N_i, N_{i-1} detects that N_i is failed.

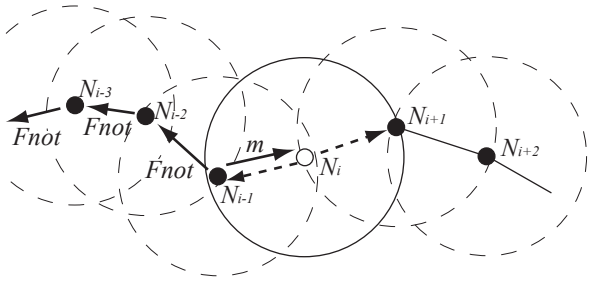


Figure 1: Stop Failure Detection in Wireless Multihop Networks.

III. Proposal

A. Problems

As discussed in the previous section, the stop failure model is supported in various fault-tolerant methods for wireless multihop networks. The Byzantine failure model is more general than the stop failure model and it is much difficul to support [10].

[Byzantine Failure Model]

Failed wireless nodes do not always become silent. They might transmit and receive data and control messages. In addition, the transmission of the messages are not always according to application protocols. The failed wireless nodes might transmit erroneous and/or malicious data and control messages. □

Different from the stop failed wireless nodes, the Byzantine failed intermediate wireless nodes in a wireless multihop transmission route might transmits different data messages from those they have received to their next-hop wireless nodes and might transmits data messages to their next-hop wireless nodes even though they have not yet receive any messages from their previous-hop wireless nodes. For such problems, some watchdog methods by the previous-hop nodes have been proposed [6]. If the wireless transmissions are based on the disk model, the transmitted data message from an intermediate node N_i to its next-hop wireless node N_{i+1} is overheard by its previous-hop node N_{i-1}. As shown in Figure 2, if N_i transmits a different data message m' to N_{i+1} from m that N_i has received from N_{i-1}, N_{i-1} detects the failure of N_i by receipt of m' different from m. That is, the Byzantine failure in N_i is detected by N_{i-1} by the comparison of data messages received and transmitted by N_i.

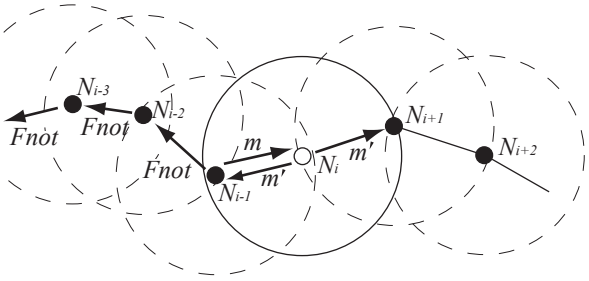


Figure 2: Byzantine Failure Detection in Wireless Multihop Networks.

However, if two successive intermediate wireless nodes N_i and N_{i+1} simultaneously fail, it is impossible for N_{i-1} to detect the failure especially in N_{i+1} in Figure 3. Though N_i correctly forwards a data message m received from N_{i-1} to N_{i+1}, a failed intermediate wireless node N_{i+1} transmits a different data message m' from m to its next-hop wireless node N_{i+2}. Since N_i overhears m' from N_{i+1}, it can detect the failure in N_{i+1} due to the comparison of m and m'. However, if N_i also fails, N_i does not transmits any failure notificatio control messages to its neighbor wireless nodes and no failure recovery such as rerouting without failed wireless nodes is initiated. Generally, $n-$simultaneous failure is define as follows [3]:

[$n-$simultaneous Failure]

The number of failed wireless nodes are at most n at any instance. Failed wireless nodes are never recovered by themselves and removed from the wireless network system by a certain maintenance procedure. □

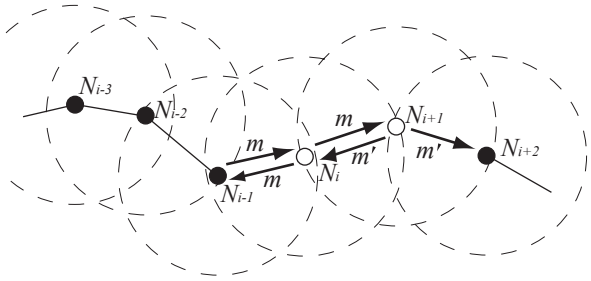


Figure 3: Simultaneous Byzantine Failures in Wireless Multihop Networks.

On detect the failure of an intermediate wireless node N_i, a wireless multihop transmission of a failure notificatio message $Fnot$ to the source node N^s is initiated by N_{i-1}. On receipt of the $Fnot$, N^s searches a wireless multihop transmission route $\mathcal{R}'$ to the destination wireless node N^d without the failed intermediate wireless node N_i. Until now, the failure detection is assumed to be correctly done in any intermediate wireless node. However, the failed intermediate wireless node N_{i-1} might erroneously detect a failure of its neighbor wireless node especially its next-hop intermediate wireless node N_i and initiate the transmission of the failure notificatio control message by transmission of a failure notificatio message $Fnot$ of N_{i+1} to its previous-hop wireless node N_{i-2} even though N_i does not fail as shown in Figure 4. Since it is impossible for N_{i-2} to fin the $Fnot$ is transmitted by N_{i-1} erroneously, N_{i-2} and the other intermediate wireless nodes forwards the message to their previous-hop wireless nodes along $\mathcal{R}$. Here, the source node is notifie for requirement of re-routing due to failure not in N_{i-1} but in N_i. Hence, newly detected wireless multihop transmission route surely excludes not N_{i-1} but N_i, which is a serious problem to be solved.

The failure notificatio control message $Fnot$ of N_i transmitted by N_{i-1} is also received by N_i. Hence, it can detect the erroneous or malicious transmission of $Fnot$. In order to notify the failure of N_{i-1} to N^s, an additional wireless transmission route from N_i to N^s without N_{i-1} is required. In addition, since N^s receives two different failure notificatio messages from N_i and N_{i-1}, N^s is required to select one of them for recovery.

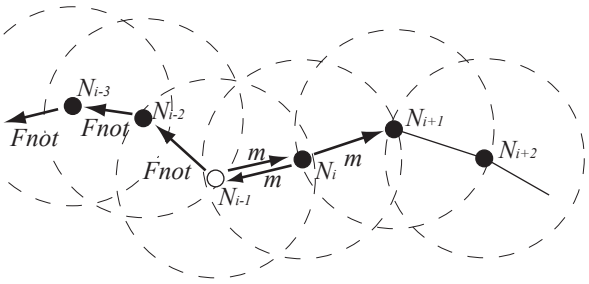


Figure 4: Erroneous Failure Detection of Byzantine Failure.

B. Neighbor Watchdog Wireless Nodes

In order to solve the problem discussed in the previous subsection, that is, under the 1-simultaneous Byzantine failure assumption, one of the intermediate wireless nodes along a wireless multihop transmission route might erroneously or maliciously transmit a failure notificatio control message, this paper proposes a cooperative watchdog method with the help of a neighbor wireless node O_i of N_{i-1} and N_i as shown in Figure 5. Here, a neighbor watchdog wireless node O_i is within the wireless transmission ranges of both N_{i-1} and N_i. Hence, O_i overhears the data messages transmitted both from N_{i-1} to N_i and from N_i to N_{i+1}. Hence, same as N_{i-1}, O_i also detects the failure of N_i by comparison of data messages transmitted from N_{i-1} to N_i and from N_i to N_{i+1}. Therefore, even if N_{i-1} erroneously or maliciously transmits a failure notificatio message $Fnot$ of N_i to N_{i-2}, O_i detects that the $Fnot$ message while N_i correctly works.

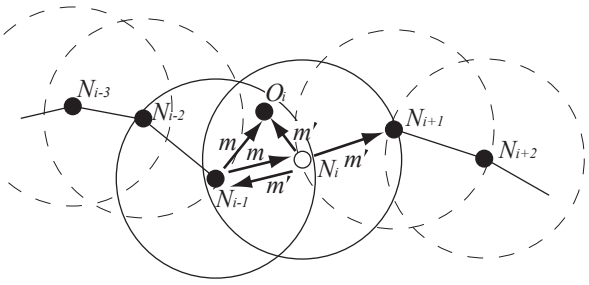


Figure 5: Cooperative Watchdog Neighbor Wireless Nodes.

In cases that O_i detects the erroneous transmission of the $\mathcal{F}not$ message, O_i should prevent the wireless multihop transmission of $Fnot$ of N_i to N^s and initiate the wireless multihop transmission of $Fnot$ of N_{i-1} since O_i has detected

the failure of N_{i-1}. Hence, a control message $Fnot$ for notificatio of failure of N_{i-1} is transmitted from O_i to N^s through N_{i-2}. However, N_{i-2} is not always a neighbor wireless node of O_i and the $Fnot$ message is required to be transmitted not through the failed intermediate wireless node N_{i-1}. In order to realize the later discussed lower overhead route detection based only on the neighbor node information in each wireless node, O_i and N_{i-2} are required to be 1-hop neighbor or 2-hop neighbor through an intermediator wireless node I_i as shown in Figure 6. The role of I_i is only forwarding the $Fnot$ message from O_i to N_{i-2}.

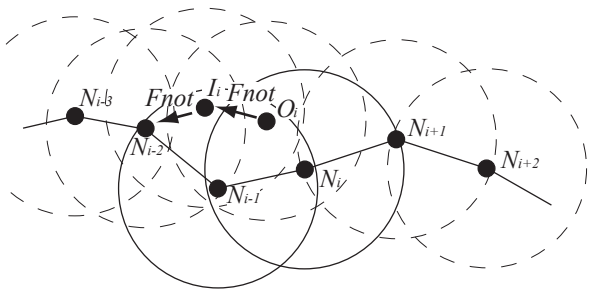


Figure 6: Intermediator Wireless Nodes for Notification

Now, we discuss the procedure in wireless nodes N_{i-1}, N_i, O_i and I_i for detection and notificatio of the 1-simultaneous Byzantine failure of one of these nodes to N_{i-2}. In the following discussion, the $Fnot$ message from O_i is transmitted to N_{i-2} through I_i; however, almost the same procedure is possible to be applied without the intermediator node I_i.

First, in the cases free from the Byzantine failures of all the intermediate, the neighbor watchdog and the intermediator wireless nodes, a data message m is transmitted through the wireless transmission route $\mathcal{R}$ according to the forward of m by the intermediate wireless nodes N_i as shown in Figure 7. There are no additional control message is required to be transmitted.

In cases that the intermediate wireless node N_i fails according to the Byzantine failure model, the data message m forwarded from N_{i-1} to N_i is not transmitted from N_i to N_{i+1}, a different data message m' from m is transmitted from N_i to N_{i+1} or a data message m'' is transmitted from N_i to N_{i+1} even though no data message is transmitted from N_{i-1} to N_i. Anyway, as shown in Figure 8, both N_{i-1} and the neighbor watchdog wireless node O_i detect the difference of data messages transmitted through the wireless links from N_{i-1} to N_i and from N_i to N_{i+1}. At this time, the same failure notificatio control messages $Fnot$ for the failure of N_i are transmitted from N_{i-1} to N_{i-2} and from O_i to N_{i-2} through I_i. Thus, N_{i-2} receives these two $Fnot$ messages.

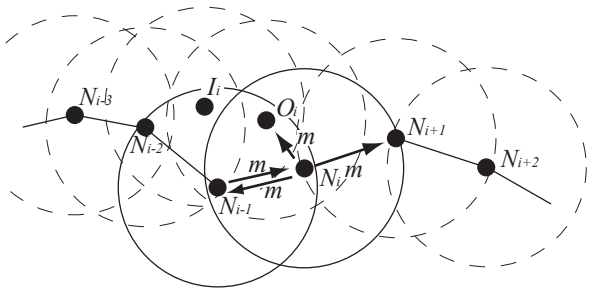


Figure 7: Data Message Transmissions with No Node Failure.

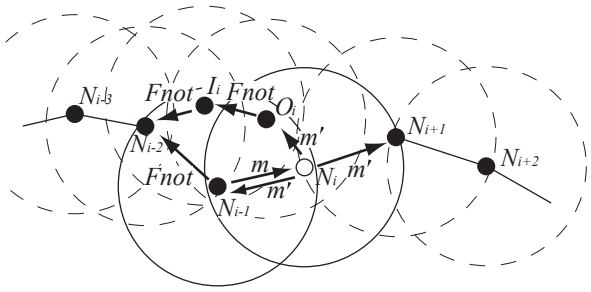


Figure 8: Detection of Failure in N_i.

In cases that N_{i-1} transmits a failure notificatio message $Fnot$ for N_i to N_{i-2} though N_i works correctly, N_{i-1} fails according to the Byzantine failure model as shown in Figure 9. Due to the 1-simultaneous Byzantine failure assumption, N_i does not fail. O_i detects that N_{i-1} transmits the $Fnot$ message for N_i to N_{i-2} though N_i does not fail by overhearing the transmitted data and control messages. Thus, O_i transmits a failure notificatio message $Fnot$ for N_{i-1} to N_{i-2} through I_i.

Same as the previous cases, even though N_i does not fail and works correctly, O_i erroneously detects the failure of N_i and notifie it to N_{i-2} through I_i as shown in Figure 10. Due to the 1-simultaneous Byzantine failure assumption, N_{i-1} does not fail. N_{i-1} detects that O_i transmits a failure notificatio control message $Fnot$ for N_i though N_i does not fail by overhearing the transmitted data and control messages. Then, N_{i-1} transmits a failure notificatio message $Fnot$ for O_i to N_{i-2}. Thus, N_{i-2} receives two different failure notificatio messages $Fnot$ for N_i from O_i and for O_i from N_{i-1}.

Finally, in cases that N_i does not fail and one of O_i and N_{i-1} fails according to the Byzantine failure model and transmits a failure notificatio control message $Fnot$ for the

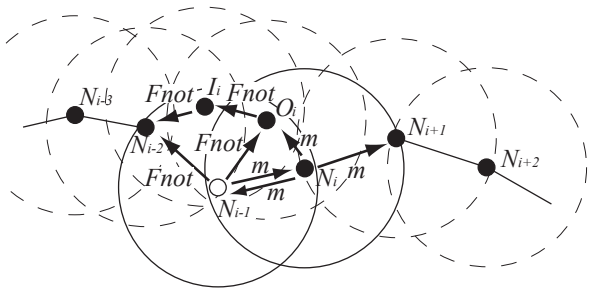


Figure 9: Detection of Failure in N_{i-1}.

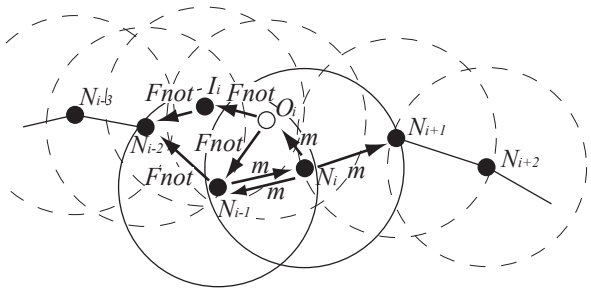


Figure 10: Detection of Failure in Watchdog Neighbor Wireless Nodes.

other to N_{i-2} as shown in Figure 11. Here, the correct wireless node detects the erroneous or malicious transmission of the failure notificatio control message $Fnot$ from the failed one. Thus, it transmits another failure notificatio control message $Fnot$ to N_{i-2}. Hence, N_{i-2} receives two different $Fnot$ messages for N_{i-1} and O_i.

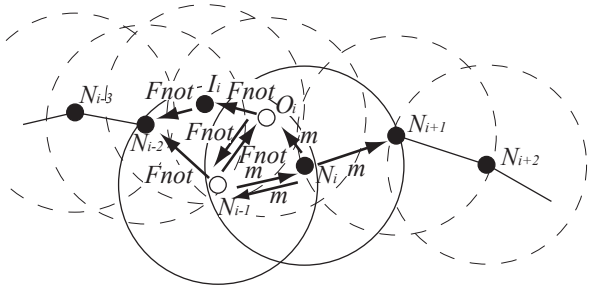


Figure 11: Failure Detection and Notificatio in O_i or N_{i-1}.

The following Table 1 summarizes the above discussion. If one of the wireless nodes N_{i-1}, N_i and O_i fails, two failure notificatio control message $Fnot$ from O_i and N_{i-1} are transmitted to N_{i-2}. Thus, when N_{i-2} receives one $Fnot$ message for one of the wireless nodes N_{i-1}, N_i and O_i from I_i or N_{i-1}, it waits for receiving another $Fnot$ message. Then, N_{i-2} determines the really failed wireless node in accordance with Table 1 and transmits a composite failure notificatio control message to N_{i-3}, which is transmitted to N^s along $\mathcal{R}$ for re-routing for the removal of the failed wireless node.

Table I: Failure Node Determination in N_{i-2}.

Failure Node in $Fnot$ from N_{i-1}	Failure Node in $Fnot$ from O_i	Failure Node
N_i	N_i	N_i
N_i	N_{i-1}	N_{i-1}
O_i	N_i	O_i
O_i	N_{i-1}	N_{i-1} or O_i

Usually, a failure of an intermediate wireless node N_j is detected by its neighbor watchdog wireless node O_j and/or its previous-hop wireless node N_{j-1} and a transmission of a failure notificatio control message $Fnot$ is initiated. Based on the 1-simultaneous Byzantine failure assumption, all the intermediate wireless node between N_{j-2} and N^s are surely correct. So that, these intermediate wireless nodes safely forward the failure notificatio control message to their previous-hop nodes. However, since the Byzantine failure model is assumed, a transmission of a failure notificatio message for an intermediate node N_j might be initiated by another intermediate wireless node N_i $(i < j - 1)$ erroneously or maliciously. As a result, an intermediate wireless nodes in a wireless multihop transmission route $\mathcal{R}$ might forward an erroneous or malicious failure notificatio control message which increases the communication overhead in the wireless multihop network.

The unique chance to detect the erroneous or malicious failure notificatio control message is when the message is initiated. If the $Fnot$ message for N_j is initiated by N_i, N_i transits a $Fnot$ message for N_j though it has not received the message from N_{i+1}, all of which is observed by the neighbor watchdog wireless node O_{i+1}. Hence, it is possible for O_{i+1} to transmit the $Fnot$ message for N_i to N_{i-1} and to induce the confirmatio procedure in N_{i-1}. However, if this confirmatio procedure is introduced in each intermediate wireless node for transmission of $Fnot$ message hop-by-hop, longer transmission delay is required for $Fnot$ transmission since transmitted $Fnot$ message and the additional $Fnot$ message from O_{i+1} are required to be synchronized at N_{i-1} for confirmation The transmission delay overhead for the failure notificatio control message is too high for realization of fault-tolerant wireless multihop networks. Thus, in our protocol, for confirmatio of the failure notificatio message, digital signature of the initial wireless node of the failure

ISBN: 978-1-61804-318-4

notificatio control message is attached to the $Fnot$ control message.

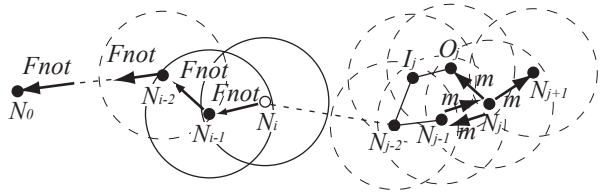


Figure 12: Erroneous or Malicious Failure Notificatio to Source Node.

C. Watchdoggable Wireless Multihop Transmission Route

As discussed in the previous subsection, for realizing 1-simultaneous Byzantine failure detection in wireless multihop transmissions, all the wireless communication links $|N_iN_{i+1}\rangle$ in a wireless multihop transmission route $\mathcal{R} = ||N_0 \ldots N_n\rangle\rangle$ should be $watchdoggable$. The condition for a watchdoggable wireless communication link is as follows:

[Watchdoggable Wireless Communication Links]

A wireless communication link $|N_iN_{i+1}\rangle$ is watchdoggable if and only if there is a neighbor watchdog wireless node O_{i+1} satisfying the following conditions (Figure 13):

- O_{i+1} is a neighbor wireless node of N_{i+1}.
- O_{i+1} is a neighbor wireless node of N_{i-1} or there is a intermediator wireless node I_i neighboring to N_{i-1}, N_i and O_{i+1}. □

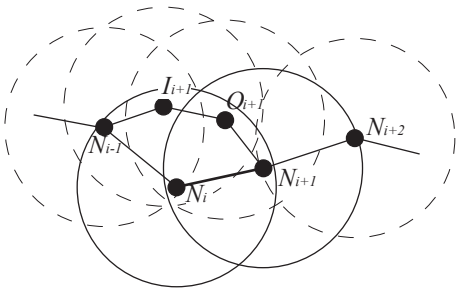


Figure 13: Watchdoggable Wireless Links.

For determination whether a wireless communication link $|N_iN_{i+1}\rangle$ is a watchdoggable one or not, the neighbor relation with N_{i-1} is required. Hence, in order to determine the possible next-hop wireless nodes satisfying the watchdoggable wireless communication links, each node requires the neighbor relation of two hop neighbor wireless nodes. Thus, each wireless node achieves its location information by using GPS and advertise the location information to its 2-hop neighbor nodes.

The detailed proposed protocol would be discussed in our future research papers.

IV. Evaluation

By using the data message transmission protocol with the Byzantine failure detection and notification fault-tolerant wireless multihop transmissions of data messages are provided. In order to apply the proposed failure detection and notification the wireless multihop transmission route is required to consist of only watchdoggable wireless communication links. Such a route is able to be detected by a flooding-base routing protocol such as AODV. Here, the protocol has two phases; a floodin phase for a route request control message $Rreq$ transmissions and a unicast phase for a route reply control message $Rrep$ along a detected wireless multihop transmission route $\mathcal{R}$. There are no additional control message transmissions and no additional synchronization overhead for data message transmissions without failure of intermediate wireless nodes.

However, in order to detect the watchdoggable wireless multihop transmission route based on the floodin of an $Rreq$ control message as discussed in the previous section, each candidate of an intermediate node is required to keep the two-hop neighbor relation as discussed in subsection 3.3. That is, each wireless node broadcasts its location information to all its 2-hop neighbor nodes by using TTL centric broadcasts independently of the transmission requests. For data message transmissions, no additional data and control messages are required to be transmitted. Additional control message transmissions are only required to detect and notify the failure of N_{i-1}, N_i and O_i to N^s. These $Fnot$ control messages are transmitted to N_{i-2} and synchronized there which requires communication and synchronization overhead.

In the proposed method, a wireless multihop transmission route is required to consist of only watchdoggable wireless communication links. Hence, a part of wireless communication links are not included in the wireless multihop transmission routes and the available wireless communication links ratio is expected to depend on the density of wireless nodes. Thus, we evaluate the effect on the route detection ratio by the restriction on the wireless communication links in the proposed method in simulation experiments. Figure 14 shows the simulation settings. N^d is a destination wireless node and N_i^ss are a source wireless node or intermediate ones. Additionally 1,000–20,000 wireless nodes are randomly distributed in the 600m×600m simulation area whose wireless transmission ranges are 10m.

ISBN: 978-1-61804-318-4

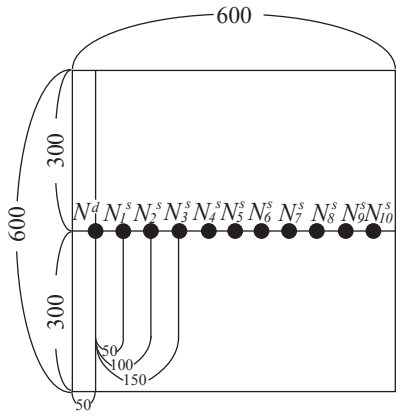


Figure 14: Simulation Setting.

Figure 15 shows the simulation results. The x-axis represents the numbers of wireless nodes, y-axis represents the distance from the source wireless node to the destination one, and z-axis represents the successful route detection ratio. For comparison, the route detection ratio in AODV is also evaluated. In both method, the route detection ratio monotonically increases according to the number of wireless nodes and is almost independent of the distance from the source wireless node to the destination one. In highly dense and sparse distribution of wireless nodes environment, the route detection ratio is almost constant. In the middle range, the route detection ratio steeply changed. In AODV, the threshold of high route detection ratio is 8,000 and the threshold of low route detection ratio is 6,000. On the other hand, in the proposed method, the threshold of high route detection ration is 11,000 and the threshold of low route detection is 6,000. Thus, in the range 8,000-11,000, the proposed method reduces the route detection ratio, which is almost only the disadvantage of the proposed method. The detection, notificatio and recovery of the Byzantine failed wireless nodes are critical technique for achieving the fault-tolerant wireless multihop networks and the merits of the proposed method surpass the disadvantage for reliable wireless multihop transmission requirements.

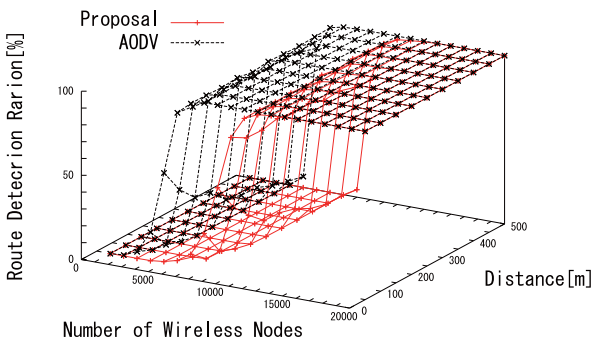


Figure 15: Route Detection Ratio (Simulation Results).

V. Concluding Remarks

This paper has proposed a novel communication protocols, i.e., for wireless transmission route detection and for data message transmissions in wireless multihop networks with failure detection, notificatio and recovery. Though almost all the conventional methods only support the stop failure, the proposed method supports the Byzantine failure where failed wireless nodes does not become silent and continues to communicate with the others out of their application protocols, i.e., erroneous and malicious data messages are transmitted independently of the application protocols. In addition, various erroneous and malicious control messages are also transmitted. This makes difficul to realize the failure detection and notification The proposed method introduces the cooperative watchdog method where two successive intermediate wireless nodes and an additional neighbor watchdog wireless node cooperate. In the proposed protocol, no additional control message transmissions are needed and the failed wireless node is correctly removed. In addition, the simulation experiments show that the proposed method has a little disadvantage on the successful route detection ratio. However, in the usual density of wireless node to assure the wireless multihop connectivity, almost no reduction in route detection ratio is expected.

References

[1] Adibi, S. and Erfani, S., “A Multipath Routing Survey for Mobile Ad-Hoc Networks,” Proceedings of the 2nd Annual IEEE Consumer Communications and Networking Conference, pp. 984–988 (2005).

[2] Cho, Y., Qu, G. and Wu, Y., “Insider Threats against Trust Mechanism with Watchdog and Defending Approaches in Wireless Sensor Networks,” Proceesings of the IEEE Symposium on Security and Privacy Workshop, pp. 134–141 (2012).

[3] Fokkink, W., “Distributed Algorithms: An Intuitive Approach,” *The MIT Press* (2013).

[4] Harada, Y., Wang Hui, Fukushima, Y., Yokohira, T., "A reroute method to recover fast from network failure", Information and Communication Technology Convergence , pp. 903 – 908 (2014).

[5] Kaur, R., Mahajan, R. and Singh, A., “A Survey on Multipath Routing Protocols for MANETs,” International Journal of Emerging Treands and Technology in Computer Science, vol. 2, no. 2, pp. 42–45 (2013).

[6] Pandit, V., Jung, H. and Agrawal,D.P ., “Inherent Security Benefit of Analog Network Coding for the Detection of Byzantine Attacks in Multi-Hop Wireless Networks,” ,IEEE 8th International Conference on Mobile Adhoc and Sensor Systems , pp. 697–702 (2011).

[7] Perkins, C., Belding-Royer, E. and Das, S., “Ad Hoc On-Demand Distance Vector (AODV) Routing,” RFC 3561 (2003).

[8] Perlyasamy, P. and Kathikeyan, E., “Survey of Current Multipath Routing Protocols for Mobile Ad Hoc Networks,” International Journal of Computer Network and Information Security, vol. 5, no. 12, pp. 68–79 (2013).

[9] Po-Kai Tseng and We-Ho Chung ., “Local Rerouting and Channel Recovery for Robust Multi-Hop Congnitive Radio Networks ” IEEE International Conference on Comunications, pp. 2895–2899 (2013).

[10] Raynal, M., “Distributed Algorithms for Message-Passing Systems,” *Springer* (2013).

ISBN: 978-1-61804-318-4

Analysis of the polarization on the bidirectional channel characteristics in an outdoor-to-indoor office scenario

I. Vin, D. P. Gaillot, P. Laly, J. M. Molina-Garcia-Pardo, M. Lienard and P. Degauque

***Abstract*—**This paper presents experimental results on the influence of the polarization on the outdoor-to-indoor channel characteristics and for a frequency range around 1.3 GHz. Virtual antenna arrays have been used both at the transmitting and the receiving sites. Path loss and delay spread are deduced from measurements carried out for different positions of the receiver, from indoor light to deep indoor. The angles of departure/arrival of the rays are obtained by applying a high resolution algorithm and values of the angular spread are given. Lastly, the maximum capacity of MIMO channels is presented both for co-and cross-polarization.

***Keywords*—**Indoor propagation, indoor penetration, polarization diversity, polarization diversity. MIMO

I. Introduction

With the ever-growing development of high data rate mobile communication, the characterization of signal penetration into buildings is of prime importance. Numerous measurement campaigns have thus been carried out, the objective of most of them being to study the additional propagation loss for estimation of indoor coverage. A classification of the indoor environment was proposed in [1] by defining different categories as: indoor light, indoor and deep indoor, according to the position of the room inside the building referred to the face of the building illuminated by the base station. Measured path loss at 780 MHz [2] and 1800 MHz [3] was compared to values predicted by the COST 231 model. This model is based on empirical formulas expressed, among other parameters, in terms of the number of internal walls and the through-wall propagation loss. The additional attenuation at 1.8 GHz in comparison to the results obtained at 900 MHz was outlined in [4], the authors also emphasizing the influence of the position of the building referred to the base station (BS), i.e. if the building is in the Line Of Sight (LOS) or not (NLOS) of the BS. A statistical approach made in the 0.8-8 GHz band in numerous office buildings and multistory car parks was proposed in [5].

This work was partly supported as a CISIT project.

I. Vin, D. P. Gaillot, P. Laly, M. Lienard and P. Degauque are with the University of Lille, IEMN Laboratory, TELICE group, Villeneuve d'Ascq, 59655 France, phone: 33 3 20 43 48 49; email: kyoko.vin@ed.univ-lille1.fr; {davy.gaillot, pierre.laly, martine.lienard, pierre.degauque@univ-lille1.fr}

J. M. Molina-Garcia-Pardo is with the Technical University of Cartagena, Information Technologies and Communication Dept., 30202 Spain (email: josemaria.molina@upct.es).

Nevertheless, the inaccuracy in a path loss prediction method based on the distance between the receiver and the penetration point such as a window, was discussed in [6]. To get a better insight on the influence of the walls, penetration loss and reflection coefficient were measured at 5.8 GHz for different types of wall as dry wall, wood, and for various polarizations [7]. The difference in attenuation due to materials used either in old or in new constructions is studied in [8] in the 800 MHz-18 GHz frequency band. Path loss is of course one of the main characteristics to determine the coverage in an indoor environment. Nevertheless, recent wireless systems include Multiple-Input Multiple-Output (MIMO) transmission schemes. In this case, directional channel characteristics as the Angle of Arrival (AoA), Angle of Departure (AoD), correlation between array elements, strongly influence the performances of the link.

For a vertical polarization VV, i.e. used both at the transmitter and at the receiver, the power of the multipath components (MPC) and the values of AoA/AoD were statistically studied in [9]. The role of the polarization and the value of the Cross Polarization Discrimination (XPD) was briefly presented in [10] but for a purely indoor or outdoor scenario.

The objective of this work is thus to study the influence of the polarization of the incident wave not only on the penetration loss but also on the delay spread, AoA and XPD in different rooms inside the building to cover scenarios from indoor light to deep indoor. The center frequency is 1.3 GHz.

After describing in Section II, the geometrical configuration of the building and the measurement setup, the distribution of path loss and XPD are studied in Section III depending on the polarization of the transmitter and on the position of Rx inside the building. Similarly, the distribution of the AoA in the various rooms is presented in Section IV, MIMO capacity being studied in Section V.

II. Measurement Setup and Antenna Configuration

A. Measurement Setup

The transmitting (Tx) and and receiving (Rx) antenna arrays were situated at the first floor of 2 buildings, 50 m apart. The Tx antenna was put at a window facing the other building in which the Rx array was placed. Successive scenarios were considered as shown in Fig. 1, the Rx array being nearly in the

ISBN: 978-1-61804-318-4

center of each room or in a corridor (Positions P6 and P7). P4 and P5 correspond to an "indoor light", the windows of the room being illuminated by the incident wave.

However the windows for the case P5 are partly shadowed by the leaves of a tree. Lastly, P1, P2 and P3 are related to deep indoor scenarios. In this office building, all rooms are entirely furnished.

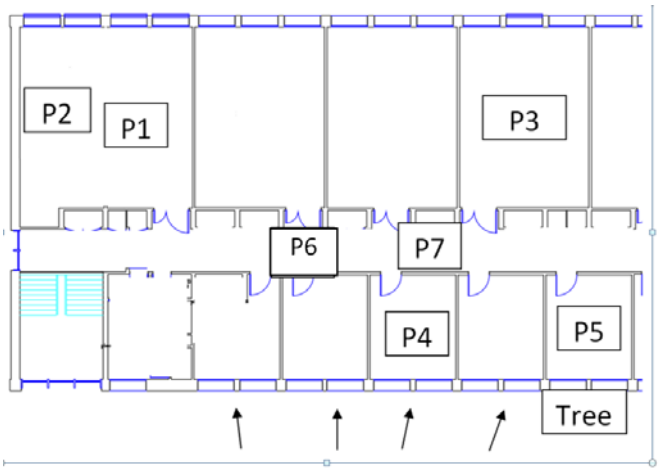


Fig. 1. Geometrical configuration of the measurement campaign. Successive positions of the Rx array are labeled P1 to P7.

The Tx antenna is a virtual 4-element uniform linear array (ULA), the element being a dual-polarized patch antenna whose center frequency is 1.3 GHz. Its orientation is such that the main lobe of its radiation pattern is directed towards the Rx building, the direction of the transmitting rays being given by the arrows in Fig. 1. At the receiving side, a 3x3 virtual uniform rectangular array (URA) is used, each element being also a dual-polarized patch antenna. The distance between the successive positions of the antenna is 10 cm, i.e. 0.43 λ, both for the ULA and the URA.

For each position, P1 to P7, the complex channel transfer functions H(f) between each element of the Tx and Rx arrays were measured with a vector network analyzer (VNA) on 512 equally spaced frequency points in a 22 MHz band with a center frequency of 1.3 GHz. The Rx antenna is directly connected to one port of the VNA using a low attenuation coaxial cable, 4 m long, a 30 dB low-noise amplifier being inserted or not, depending on the received power. Using a coaxial cable to connect the Tx antenna to the other port of the VNA leading to prohibitive attenuation, the signal of the Tx port of the VNA is converted to an optical signal sent through fiber optics, converted back to radio frequency and amplified. The phase stability of the fiber optics link has been checked and the calibration of the VNA takes amplifiers, cables, and optic coupler into account.

B. Radiation Pattern of the Antennas

One of the objective of the paper being to compare the channel characteristics for horizontal and vertical polarizations, it will be interesting to base this comparison on an Rx antenna presenting an omnidirectional radiation pattern, at least in the horizontal plane. Indeed, one can expect that the most powerful reflected rays will propagate inside the building with a small angle of elevation. To achieve this goal, the Rx patch antenna was rotated around its vertical axis, by steps of 90°. Summing the complex value of the field successively received for the 4 positions of the Rx antenna leads to a nearly omnidirectional pattern in the horizontal plane, as shown in Fig. 2, curve (a). This curve was obtained for an HH polarization, the first and the second letter referring to the polarization of the Tx and Rx antenna, respectively. As a comparison, the pattern of a single patch antenna is given by curve (b). A similar curve was obtained for VV.

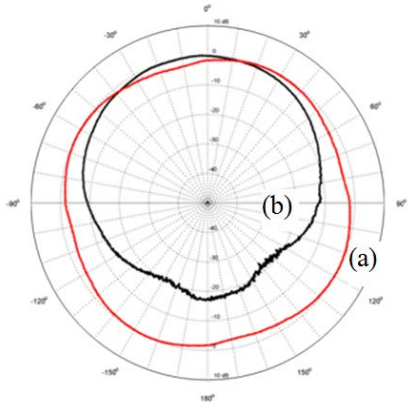


Fig. 2. Radiation pattern in the horizontal plane for HH polarization; (a) by summing the field radiated by the patch antenna in 4 orthogonal positions; (b) for a single patch antenna.

Lastly, to increase the signal-to-noise ratio (SNR), 10 successive measurements are averaged. The approach is thus rather long but, during the experiments, nobody was present in the building and the channel can be considered as stationary.

III. Path Loss and Cross Polarization Discrimination Factor from Indoor Light to Deep Indoor

A. Path loss

Let us first consider the variation of the signal amplitude when the receiver moves from P4 to P1 or P2. At P4, i.e. in indoor light conditions, the mean value of the received signal is maximum for an HH polarization (both Tx and Rx are horizontally polarized). Since we want to outline the additional attenuation from indoor light to deep indoor, this mean value is chosen as a reference (0 dB). Curves in Fig. 3 show the cumulative distribution function (cdf) of the relative loss for 4 possible combinations of the polarization.

This cdf was deduced from measurements made for the 9 positions of the Rx antenna (URA) at point P4, the 4 positions of Tx (ULA) and the 512 frequency points. For this scenario, waves remain polarized and HH is slightly better than VV.

Curves in Fig. 4 have been plotted for a deep indoor scenario (Position P1), the reference being always the mean value of the received signal HH in P4 (indoor light). The minimum loss is still obtained for co-polarized antennas, HH or VV. However, in this case, the waves are strongly depolarized; a cross polarization scenario (HV) presenting an attenuation less than 3 dB compared to HH.

ISBN: 978-1-61804-318-4

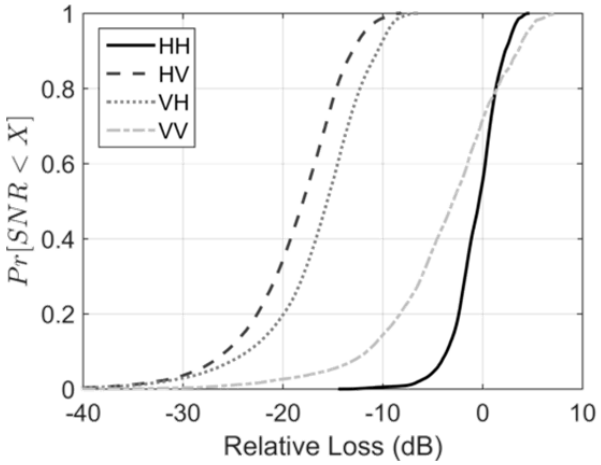


Fig. 3. Indoor light: cumulative distribution function of the additional loss referred to the mean value of the received signal for HH polarization.

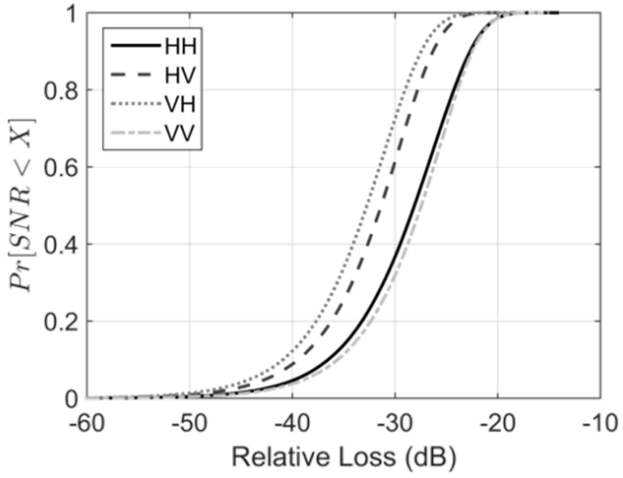


Fig. 4. Same as in Fig. 3 but for the deep indoor scenario P2.

Table I summarizes results obtained in the different scenarios. By choosing, as previously, the reference of 0 dB for "light indoor" and HH polarization, the mean additional path loss is given for co-polar and cross-polar configurations for: i) light indoor P4, ii) light indoor but in presence of a dense vegetation masking the line of sight ("Light obstructed" P5), iii) in the corridor (P6 and P7) and lastly iv) in deep indoor (P1, P2 and P3).

The averaging made over various positions inside the different rooms confirms the less important additional losses for co-polarization even in deep indoor. Furthermore, if we compare results for light indoor, obstructed or not by a tree with dense foliage, it appears that the presence of such a tree near a window increases the attenuation of 6 to 10 dB.

B. Delay Spread

Curves in Fig. 5 give the power delay profile (pdp) for a deep indoor scenario and for different polarizations. The shapes of the pdp for other scenario are similar and one can expect that the rms delay spread will have the same order of magnitude.

Table I. Mean additional path loss (in dB) for various receiving scenarios: Light indoor ("Light"); LOS obstructed ("Light. obst."), corridor and deep indoor.

	Light	Light obst.	Corridor	Deep
HH	0	-9	-12	-25
HV	-17	-18	-23	-30
VV	-1	-11	-12	-27
VH	-14	-19	-21	-30

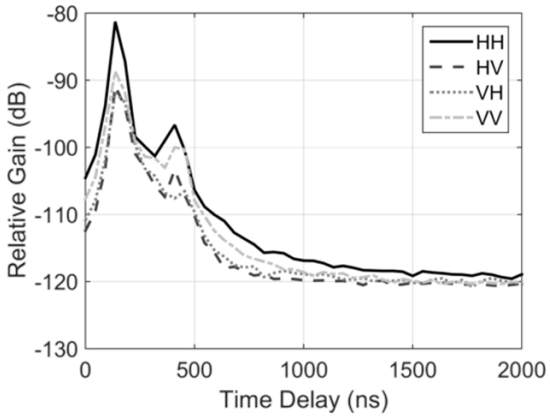


Fig. 5. Power delay profile for the deep indoor scenario P1.

This appears in Table II summarizing results obtained in the different scenarios, considering a threshold of -20 dB.

Table II. Mean rms delay spread (in ns) for the different scenarios

	Light	Light obst.	Corri.	Deep
HH	39	31	36	70
HV	39	36	41	108
VV	35	39	35	80
VH	32	37	34	100

As expected, the delay spread increases when the receiver moves from indoor light to deep indoor, varying from about 35-40 ns to 70-100 ns. Furthermore, the difference between co-and cross-polarization configurations is not important, except in deep indoor.

IV. Channel Characteristics Deduced from High Resolution Algorithm

Directional channel characteristics are deduced from the previous results by applying the RiMAX High Resolution Algorithm (HRA). It allows getting a joint estimation of the relative time of arrival (delay), the angle of arrival (AoA) and the angle of departure (AoD). It must be emphasized that the performance of an HRA is highly sensitive to the frequency band, to the number of frequency points in this band, and to the number of spatial samples corresponding to the number of array elements which is rather small in our application.

The presence of correlated paths (e.g. clusters) in the channel also strongly affects parameter estimation. In our case, the multipath components (MPC) were extracted from RiMAX by considering 351 frequency points in the 22 MHz

band, the 3x3 URA and the 4-ULA used for the Rx and Tx array, respectively. An example of the transmitting and receiving rays deduced from RiMAX is given in Fig. 6 where Rx is situated in position P2, i.e. in deep indoor for HH polarization. In this case, 15 rays having an attenuation less than 20 dB, referred to the most powerful ray, have been obtained. At Rx, a wide spread of the AoA is observed, this spread being much smaller at the Tx site, the antenna directly illuminating the Rx building.

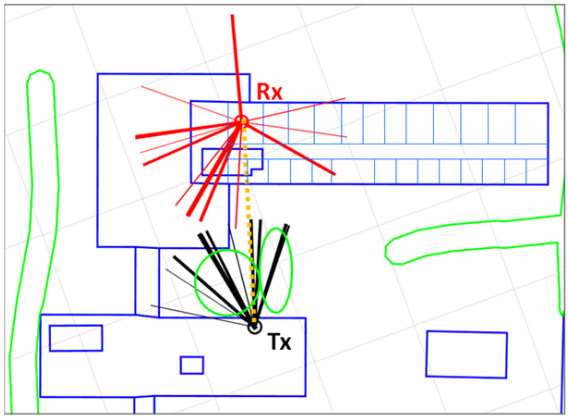


Fig. 6. Ray-tracing of the estimated paths for P2. The path thickness is proportional to the power.

Table III gives the rms angular spread of the AoA/AoD for the different positions of Rx in the building and, in each case, for the best polarization, i.e. presenting the lowest path loss.

Table III. Mean rms angular spread for the different scenarios

	Light	Light obst.	Corridor	Deep
AoA	12°	64°	40°	83°
AoD	12°	23°	20°	40°

V. MIMO CAPACITY WITH UNIFORM LINEAR ARRAY

For a MIMO transmission based on ULA, it is interesting to know if the orientation of the ULA inside the building has a great influence or not on the channel capacity. We will thus consider a (4,3) MIMO system as shown in Fig. 7.

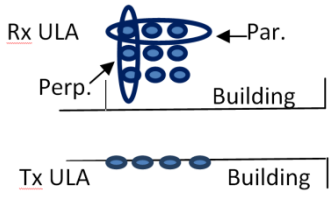


Fig. 7. Geometrical configuration of the simulated MIMO system.

The alignment of the Rx ULA is either parallel or perpendicular to the Tx ULA, and noted in Fig. 7 "Par.' and "Perp.", respectively. The channel capacity was calculated by assuming an SNR of 35 dB in room P4, for HH polarization and takes the additional losses into the rooms into account.

Let us recall that this position was chosen as a reference point in Section III, Table I. Results detailed in Table IV are related to the parallel configuration, but nearly the same results were obtained for the Rx ULA in the "perpendicular" orientation. The orientation of the ULA inside the room is thus not at all a critical parameter.

Table IV. Capacity (bits/s/Hz) for the different scenarios. "Parallel configuration".

	Light	Light obst.	Corri.	Deep
HH	33	27	24	16
HV	25	23	18	10
VV	37	31	27	13
VH	26	22	19	10

VI. CONCLUSION

We have shown that the cross-polar discrimination factor varies from 13-15 dB for indoor light to only few dBs for deep indoor. The rms delay spread, varies from 30 to 100 ns and is slightly higher for a cross-polar configuration, mainly for deep indoor. The angular spread at the receiver reaches 80° for deep indoor. Lastly, for a MIMO link using a linear array, the best results in terms of capacity were obtained for a co-polar configuration. Furthermore, the capacity is not strongly dependent on the orientation of the receiving array inside the building.

REFERENCES

[1] L. Ferreira, M. Kuipers, C. Rodrigues and L. M. Correia, "Characterisation of signal penetration into buildings for GSM and UMTS," in *Proc. Int. Symp. Wireless and Commun. Systems*," 2006, pp. 63-67.

[2] E. Suikkanen, A. Tölli and M. Latva-aho, "Characterization of propagation in an outdoor to indoor scenario at 780 MHz," in *Proc. 21st Int. Symp. on PIMRC*, 2010, pp. 70-74

[3] R. Visbrot, A. Kozinsky, A. Freedman, A. Reichman and T. Blaunstein, "Measurement campaign to determine and validate outdoor to indoor propagation models for GSM signals in various environments," in *Proc. IEEE Int. Conf. on COMCAS*, 2011, pp. 1-5

[4] D. M. Rose and T. Kurner, "Outdoor to indoor propagation – Accurate measuring and modeling of indoor environment at 900 and 1800 MHz," in *Proc. European Conf. Antennas and Propag. (EUCAP)*, 2012, pp. 1440-1444.

[5] H. Okamoto, K. Kitao and S. Ichitsubo, "Outdoor-to-indoor propagation loss prediction in 800 MHz to 8 GHz band for an urban area," *IEEE Trans. Vehicular Techno.*, vol. 3, pp. 1059-1067, March 2009.

[6] Y. Hirota, H. Izumikawa and C. Ono, "Outdoor-to-indoor radio propagation characteristics with 800 MHz band in an urban environment," in *Proc. IEEE Antennas Propagation/URSI Symp.*, 2014, pp. 697-698.

[7] Y. Azar, H. Zhao and M. E. Knox, "Polarization diversity measurements at 5.8 GHz for penetration loss and reflectivity of common building materials in an indoor environment," in *Proc. 3rd Int. Conf. Future Generation Commun. Techno,"* pp. 50-54, Aug. 2014

[8] I. Rodriguez, H. C. Nguyen, N. Jorgensen, T. Sorensen and P. Mogensen, "Radio propagation into modern buildings: Attenuation measurements in the range from 800 MHz to 18 GHz," in *Proc. 80th IEEE Vehicular Techno. Conf.*, 2014, pp. 1-5

[9] S. Wyne, A. F. Molisch, P. Almers, G. Eriksson, J. Kardeal and F. Tufvesson, "Outdoor to indoor MIMO measurements and analysis at 5.2 GHz," *IEEE Trans. Vehicular Techno.,"* vol. 57, pp. 1374-1386, May 2008

[10] E. M. Vitucci, F. Mani, C. Oestges and V. Degli-Esposti, "Analysis and modeling of the polarization characteristics of diffuse scattering in indoor and outdoor radio propagation", in *Proc. of Int. Conf. on applied EM and Commun.*, 2013, pp. 1-5

ISBN: 978-1-61804-318-4

A Minimax Approach in Training Sequence Design for Carrier Frequency Offset Estimation in Frequency-Selective Channels

Xu Wang, Erchin Serpedin, and Khalid Qaraqe

***Abstract*—In this paper, the problem of carrier frequency offset estimation for frequency-selective channels in a data-aided context is considered. The channel is assumed to be contaminated with an unknown noise and the Cramér-Rao bound (CRB) is used as an estimation accuracy criterion. The minimax approach consists in minimizing the worst-case (maximum) CRB of the frequency offset. Under certain assumptions, it is shown that the worst-case CRB of the frequency offset is achieved when the noise is modeled as circularly symmetric complex-valued Gaussian. In addition, it is illustrated that a white training sequence minimizes the worst-case CRB, or equivalently, achieves the minimax optimality in this case.**

***Index Terms*—Frequency offset estimation, frequency-selective channels, circularly symmetric complex-valued Gaussian noise, Cramér-Rao bound, training sequence.**

I. Introduction

CARRIER frequency offset is one of the most common impairments found in wireless communications systems. Caused by a local oscillator drift or a Doppler effect, CFO may result in a distortion of the transmitted signal [1]. Hence, the frequency offset needs to be accurately estimated in order to retrieve the transmitted signal. Many current wireless communications systems employ a training sequence to mitigate the effect of the frequency-selective channels and estimate the frequency offset [2]. References [3], [4], [5] developed an optimal training sequence that minimizes the variance of the channel parameter estimates based on a specific estimation approach. Alternatively, instead of considering a specific estimation method, the CRB was considered as an optimization criterion for training sequence selection. Specifically, a training sequence design for frequency offset estimation in frequency selective channels was reported in [2] using a minimax approach that minimizes the worst-case CRB of the frequency offset. Stoica et al [6] extended their results in [2] by developing an optimal training sequence for both frequency offset and channel estimation. The optimal training problem was also discussed in [1], [7] by averaging the CRB over Gaussian and Ricean channels, respectively.

X. Wang, E. Serpedin and K. Qaraqe are with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, 77843 USA (e-mail:serpedin@ece.tamu.edu).

In this paper, we re-investigate the results in [2] and [6] by assuming an unknown complex-valued noise distribution. More precisely, under the assumption that the noise has a fixed covariance matrix, it is shown herein paper that the circularly symmetric complex-valued Gaussian is the least-favorable distribution that achieves the worst-case CRB and the white training sequence is minimax optimal.

The rest of the paper is structured as follows. In Section II, the system model that will be used throughout the paper is described. In Section III, the training sequence selection problem is formulated and solved as a minimax optimization of the CRB of frequency offset. Finally, Section IV concludes the paper.

II. System Model

Consider a linear signal through a frequency-selective channel with a larger coherence bandwidth compared to the signal bandwidth. The channel system can be modeled as

$$y(n) = e^{i2\pi f_0 n} \sum_{l=0}^{L-1} h(l)t(n-l) + \nu(n), \tag{1}$$

where $\mathbf{y} = [y(0), y(1), \cdots, y(N-1)]^T$ denotes the received signal, f_0 represents the frequency offset, $\mathbf{h} = [h(0), h(1), \cdots, h(L-1)]^T$ stands for the channel impulse response, $\mathbf{t} = [t(-L+1), t(-L+2), \cdots, t(N-1)]^T$ denotes the training sequence, and $\boldsymbol{\nu} = [\nu(0), \nu(1), \cdots, \nu(N-1)]^T$ is the complex-valued noise. In a more compact form, the system model can also be expressed as

$$\mathbf{y} = \boldsymbol{\Gamma}(\omega_0)\mathbf{T}\mathbf{h} + \boldsymbol{\nu}, \tag{2}$$

where $\omega_0 = 2\pi f_0$ is the angular frequency offset,

$$\boldsymbol{\Gamma}(\omega_0) = \text{diag}\left(1, e^{i\omega_0}, \cdots, e^{i(N-1)\omega_0}\right),$$

$$\mathbf{T}(k,l) = t(k-l),\ k = 1, \cdots, N,\ l = 1, \cdots, L.$$

Specifically, the complex-valued noise $\boldsymbol{\nu}$ admits the from

$$\begin{aligned}\boldsymbol{\nu} =& \mathbf{a} + i\mathbf{b} \\ =& [a_0 + ib_0, a_1 + ib_1, \cdots, a_{N-1} + ib_{N-1}]^T,\end{aligned}$$

where

$$\mathbf{a} = [a_0, a_1, \cdots, a_{N-1}]^T$$

ISBN: 978-1-61804-318-4

and

$$\mathbf{b} = [b_0, b_1, \cdots, b_{N-1}]^T$$

stand for the real and imaginary part of the noise, respectively.

III. Minimax Optimization

The CRB of the frequency offset ω_0, denoted as $\text{CRB}(\omega_0)$, may depend on the selection of training sequence $\mathbf{T}$ and the noise $\boldsymbol{\nu}$. Additionally, the channel impulse response $\mathbf{h}$ is usually unknown in practice. Herein paper, we consider the following minimax problem:

$$\min_{\mathbf{T}} \max_{||\mathbf{h}||=\rho, f_{\boldsymbol{\nu}} \in \mathcal{F}} \text{CRB}(\omega_0). \tag{3}$$

The constraint $||\mathbf{h}|| = \rho > 0$ is used to prevent the trivial solution $\mathbf{h} = \mathbf{0}$. Since ρ can be arbitrarily close to 0, this constraint actually does not place any restriction to the problem [6]. The constraint $f_{\boldsymbol{\nu}} \in \mathcal{F}$ denotes that $f_{\boldsymbol{\nu}}$ belongs to a certain probability set $\mathcal{F}$, where $f_{\boldsymbol{\nu}}$ stands for the probability density function (PDF) of the noise $\boldsymbol{\nu}$. More precisely, for $\boldsymbol{\nu} = \mathbf{a} + i\mathbf{b}$, it is assumed that each component of $\mathbf{a}$ and $\mathbf{b}$ is identically distributed and independent of each other with a zero mean and fixed variance $\sigma^2/2$. Additionally, the PDF of each component has a zero value at the endpoints $-\infty$ and ∞.

The reason for using the minimax formulation is that the optimal training sequence obtained from the minimax problem does not depend on $\mathbf{h}$ and $\boldsymbol{\nu}$. More importantly, the optimal training sequence is robust to the possible variation of the noise and channel since it minimizes the worst-case CRB.

We begin with maximizing $\text{CRB}(\omega_0)$ with respect to $f_{\boldsymbol{\nu}}$ for given $\mathbf{h}$ and $\mathbf{T}$. Let $\boldsymbol{\theta} = [\omega_0, \mathbf{h}_R, \mathbf{h}_I]^T$ represent the unknown parameter vector to be estimated, where $\mathbf{h}_R$ and h_I stand for the real and imaginary part of $\mathbf{h}$, respectively. It is observed that directly calculating $\text{CRB}(\omega_0)$ for unknown noise consists of deriving the inverse of the Fisher information matrix (FIM) of $\boldsymbol{\theta}$ and finding the upper left entry corresponding to ω_0. This process is time-consuming and computationally costly. Alternatively, we are seeking for the minimum FIM in the sense of order of positive semi-definite matrices. In this way, if we denote $\mathbf{F}^*$ to be the minimum FIM and $\mathbf{F}$ to be any FIM of $\boldsymbol{\theta}$, it yields that

$$\begin{aligned}
&\mathbf{F}^* \preceq \mathbf{F} \\
\Rightarrow &\mathbf{C}^* \succeq \mathbf{C} \\
\Rightarrow &[1, 0, \cdots, 0]\mathbf{C}^* \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \geq [1, 0, \cdots, 0]\mathbf{C} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\
\Rightarrow &\text{CRB}^*(\omega_0) \geq \text{CRB}(\omega_0),
\end{aligned} \tag{4}$$

where $\preceq$ and $\succeq$ stand for the inequalities in the sense of ordering positive semi-definite matrices, $\mathbf{C}^*$ and $\mathbf{C}$ are corresponding CRB matrices. Therefore, it can be concluded that minimizing the CRB of the frequency offset $\text{CRB}(\omega_0)$ is equivalent of minimizing the FIM. The latter, compared to the former, is relatively easier to analyze and tackle as shown in the rest of the section.

Let $\mathbf{x}_\theta = \boldsymbol{\Gamma}(\omega_0)\mathbf{T}\mathbf{h}$, the channel in (2) can be expressed as

$$\mathbf{y} = \mathbf{x}_\theta + \boldsymbol{\nu}. \tag{5}$$

Based on (5), the score function is defined as

$$\begin{aligned}
s(\boldsymbol{\theta}) &= \frac{\partial}{\partial \boldsymbol{\theta}} \log f_{\mathbf{y}|\boldsymbol{\theta}}(\mathbf{y}|\boldsymbol{\theta}) \\
&= \frac{\partial}{\partial \boldsymbol{\theta}} \log f_{\mathbf{y}|\mathbf{x}_\theta}(\mathbf{y}|\mathbf{x}_\theta) \\
&= \frac{\partial}{\partial \boldsymbol{\theta}} \log f_{\boldsymbol{\nu}}(\mathbf{y} - \mathbf{x}_\theta) \\
&\overset{(a)}{=} -\left(\frac{\partial \mathbf{x}_\theta}{\partial \boldsymbol{\theta}}\right)^T \frac{\partial \log f_{\boldsymbol{\nu}}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}} - \left(\frac{\partial \mathbf{x}_\theta^*}{\partial \boldsymbol{\theta}}\right)^T \frac{\partial \log f_{\boldsymbol{\nu}}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}^*},
\end{aligned} \tag{6}$$

where "$*$" stands for the complex conjugate and equation (a) is due to the chain rule for Wirtinger (complex) derivatives [8]. The FIM is then represented as

$$\begin{aligned}
\mathbf{F} =& \text{E}(s(\boldsymbol{\theta})s^T(\boldsymbol{\theta})) \\
=& \text{E}\Bigg[\left(\frac{\partial \mathbf{x}_\theta}{\partial \boldsymbol{\theta}}\right)^T \left(\frac{\partial \log f_{\boldsymbol{\nu}}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}}\right) \left(\frac{\partial \log f_{\boldsymbol{\nu}}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}}\right)^T \left(\frac{\partial \mathbf{x}_\theta}{\partial \boldsymbol{\theta}}\right) \\
&+ \left(\frac{\partial \mathbf{x}_\theta}{\partial \boldsymbol{\theta}}\right)^T \left(\frac{\partial \log f_{\boldsymbol{\nu}}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}}\right) \left(\frac{\partial \log f_{\boldsymbol{\nu}}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}^*}\right)^T \left(\frac{\partial \mathbf{x}_\theta^*}{\partial \boldsymbol{\theta}}\right) \\
&+ \left(\frac{\partial \mathbf{x}_\theta^*}{\partial \boldsymbol{\theta}}\right)^T \left(\frac{\partial \log f_{\boldsymbol{\nu}}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}^*}\right) \left(\frac{\partial \log f_{\boldsymbol{\nu}}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}}\right)^T \left(\frac{\partial \mathbf{x}_\theta}{\partial \boldsymbol{\theta}}\right) \\
&+ \left(\frac{\partial \mathbf{x}_\theta^*}{\partial \boldsymbol{\theta}}\right)^T \left(\frac{\partial \log f_{\boldsymbol{\nu}}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}^*}\right) \left(\frac{\partial \log f_{\boldsymbol{\nu}}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}^*}\right)^T \left(\frac{\partial \mathbf{x}_\theta^*}{\partial \boldsymbol{\theta}}\right) \Bigg]
\end{aligned} \tag{7}$$

Since the components of vectors $\mathbf{a}$ and $\mathbf{b}$ are independent and identically distributed (i.i.d) and the complex PDF $f_{\boldsymbol{\nu}}(\boldsymbol{\nu})$ is the joint PDF of its real and imaginary parts, $f_{\boldsymbol{\nu}}(\boldsymbol{\nu})$ can be expressed as

$$f_{\boldsymbol{\nu}}(\boldsymbol{\nu}) = f_{\mathbf{a}}(\mathbf{a}) f_{\mathbf{b}}(\mathbf{b}). \tag{8}$$

Based on the definition of Wirtinger derivatives [8], it follows that

$$\begin{aligned}
\frac{\partial \log f_{\boldsymbol{\nu}}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}} &= \frac{1}{2}\left[\frac{\partial \log f_{\boldsymbol{\nu}}(\boldsymbol{\nu})}{\partial \mathbf{a}} - i\frac{\partial \log f_{\boldsymbol{\nu}}(\boldsymbol{\nu})}{\partial \mathbf{b}}\right] \\
&= \frac{1}{2}\left[\frac{\partial \log f_{\mathbf{a}}(\mathbf{a})}{\partial \mathbf{a}} - i\frac{\partial \log f_{\mathbf{b}}(\mathbf{b})}{\partial \mathbf{b}}\right]
\end{aligned} \tag{9}$$

$$\begin{aligned}
\frac{\partial \log f_{\boldsymbol{\nu}}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}^*} &= \frac{1}{2}\left[\frac{\partial \log f_{\boldsymbol{\nu}}(\boldsymbol{\nu})}{\partial \mathbf{a}} + i\frac{\partial \log f_{\boldsymbol{\nu}}(\boldsymbol{\nu})}{\partial \mathbf{b}}\right] \\
&= \frac{1}{2}\left[\frac{\partial \log f_{\mathbf{a}}(\mathbf{a})}{\partial \mathbf{a}} + i\frac{\partial \log f_{\mathbf{b}}(\mathbf{b})}{\partial \mathbf{b}}\right]
\end{aligned} \tag{10}$$

ISBN: 978-1-61804-318-4

Plugging (9) and (10) into (7) leads to

$$\begin{aligned}\mathbf{F} =&\frac{1}{4}\Bigg[-i\left(\frac{\partial \mathbf{x}_\theta}{\partial \boldsymbol{\theta}}\right)^T \left(E_{\mathbf{a}}E_{\mathbf{b}}^T + E_{\mathbf{b}}E_{\mathbf{a}}^T\right)\left(\frac{\partial \mathbf{x}_\theta}{\partial \boldsymbol{\theta}}\right)\\ &+\left(\frac{\partial \mathbf{x}_\theta}{\partial \boldsymbol{\theta}}\right)^T (\mathbf{F}(\mathbf{a})+\mathbf{F}(\mathbf{b}))\left(\frac{\partial \mathbf{x}_\theta^*}{\partial \boldsymbol{\theta}}\right)\\ &+\left(\frac{\partial \mathbf{x}_\theta^*}{\partial \boldsymbol{\theta}}\right)^T (\mathbf{F}(\mathbf{a})+\mathbf{F}(\mathbf{b}))\left(\frac{\partial \mathbf{x}_\theta}{\partial \boldsymbol{\theta}}\right)\\ &+i\left(\frac{\partial \mathbf{x}_\theta^*}{\partial \boldsymbol{\theta}}\right)^T \left(E_{\mathbf{a}}E_{\mathbf{b}}^T + E_{\mathbf{b}}E_{\mathbf{a}}^T\right)\left(\frac{\partial \mathbf{x}_\theta^*}{\partial \boldsymbol{\theta}}\right)\Bigg],\end{aligned} \tag{11}$$

where

$$E_{\mathbf{a}} = \mathrm{E}\left(\frac{\partial \log f_{\mathbf{a}}(\mathbf{a})}{\partial \mathbf{a}}\right), \tag{12}$$

$$E_{\mathbf{b}} = \mathrm{E}\left(\frac{\partial \log f_{\mathbf{b}}(\mathbf{b})}{\partial \mathbf{b}}\right), \tag{13}$$

$$\mathbf{F}(\mathbf{a}) = \mathrm{E}\left[\left(\frac{\partial \log f_{\mathbf{a}}(\mathbf{a})}{\partial \mathbf{a}}\right)\left(\frac{\partial \log f_{\mathbf{a}}(\mathbf{a})}{\partial \mathbf{a}}\right)^T\right],$$

$$\mathbf{F}(\mathbf{b}) = \mathrm{E}\left[\left(\frac{\partial \log f_{\mathbf{b}}(\mathbf{b})}{\partial \mathbf{b}}\right)\left(\frac{\partial \log f_{\mathbf{b}}(\mathbf{b})}{\partial \mathbf{b}}\right)^T\right].$$

It is observed that $\mathbf{F}(\mathbf{a}) = \mathbf{F}(\mathbf{b})$ represent the Fisher information matrices of the real and imaginary part of the complex-valued noise, where $\mathbf{a}$ and $\mathbf{b}$ are identical multivariate variables with zero mean and covariance matrix $\sigma^2\mathbf{I}/2$. Furthermore, it is shown in Appendix A that $\mathrm{E}_{\mathbf{a}} = \mathrm{E}_{\mathbf{b}} = \mathbf{0}$.

Thus, (11) can be simplified as

$$\begin{aligned}\mathbf{F} =&\frac{1}{2}\left[\left(\frac{\partial \mathbf{x}_\theta^*}{\partial \boldsymbol{\theta}}+\frac{\partial \mathbf{x}_\theta}{\partial \boldsymbol{\theta}}\right)^T \mathbf{F}(\mathbf{a})\left(\frac{\partial \mathbf{x}_\theta^*}{\partial \boldsymbol{\theta}}+\frac{\partial \mathbf{x}_\theta}{\partial \boldsymbol{\theta}}\right)\right]\\ =&2\mathbf{P}^T\mathbf{F}(\mathbf{a})\mathbf{P},\end{aligned} \tag{14}$$

where $\mathbf{P} = \mathrm{Re}\left(\frac{\partial \mathbf{x}_\theta^*}{\partial \boldsymbol{\theta}}\right) = \mathrm{Re}\left(\frac{\partial \mathbf{x}_\theta}{\partial \boldsymbol{\theta}}\right)$ denotes the real part of the matrices $\frac{\partial \mathbf{x}_\theta^*}{\partial \boldsymbol{\theta}}$ and $\frac{\partial \mathbf{x}_\theta}{\partial \boldsymbol{\theta}}$. In this way, a problem in complex domain is simplified to a real domain problem. Furthermore, finding the minimum $\mathbf{F}$ amounts to minimizing $\mathbf{F}(\mathbf{a})$ since $\mathbf{F}^*(\mathbf{a}) \preceq \mathbf{F}(\mathbf{a})$ implies $\mathbf{P}^T\mathbf{F}^*(\mathbf{a})\mathbf{P} \preceq \mathbf{P}^T\mathbf{F}(\mathbf{a})\mathbf{P}$ for any matrix $\mathbf{P}$ [9].

Since $\mathbf{a}$ and $\mathbf{b}$ are identical multivariate variables with zero mean and covariance matrix $\sigma^2\mathbf{I}/2$, the following lemma [10] states that the $\mathbf{F}(\mathbf{a})$ is lower bounded by the $\mathbf{F}(\mathbf{a}_G)$ of a normally distributed random vector $\mathbf{a}_G$.

Lemma 1. *For a random vector* $\mathbf{a}$ *and a Gaussian random vector* $\mathbf{a}_G$ *whose covariance matrices are identical, the following inequality holds*

$$\mathbf{F}(\mathbf{a}_G) \preceq \mathbf{F}(\mathbf{a}). \tag{15}$$

The lemma can be proved by numerous methods, and the reader is referred to [11], [9], [12] for more details.

Therefore, when the real and imaginary part of the complex-valued noise follow a Gaussian distribution with zero mean and covariance matrix $\sigma^2\mathbf{I}/2$, or equivalently, when the complex-valued noise follows a circularly symmetric complex Gaussian distribution with zero mean and covariance matrix $\sigma^2\mathbf{I}$, the FIM of $\boldsymbol{\theta}$ is minimized.

Towards this end, the FIM $\mathbf{F}$ resumes to

$$\mathbf{F} = \frac{2}{\sigma^2}\mathrm{Re}\left(\frac{\partial \mathbf{x}_\theta}{\partial \boldsymbol{\theta}}^H\frac{\partial \mathbf{x}_\theta}{\partial \boldsymbol{\theta}^T}\right), \tag{16}$$

and the CRB for ω_0 is found by taking the inverse of (16), which is given by [2]

$$\mathrm{CRB}(\omega_0) = \frac{\sigma^2}{2}\left[\mathbf{h}^H\mathbf{T}^H\mathbf{D}\Pi\mathbf{D}\mathbf{T}\mathbf{h}\right]^{-1}, \tag{17}$$

where $\Pi = \mathbf{I} - \mathbf{T}(\mathbf{T}^H\mathbf{T})^{-1}\mathbf{T}^H$. Since searching for the optimal training sequence for (17) is hardly feasible, [2] provided a closed-form asymptotic CRB as an approximation of the CRB in (17) as follows:

$$\mathrm{asCRB}(\omega_0) = \frac{1}{N^3}\frac{6\sigma^2}{\mathbf{h}^H\mathbf{R}\mathbf{h}},$$

where $\mathbf{R}$ is an $L \times L$ covariance matrix for the training sequence.

Thus, the minimax optimization problem (3) resumes to

$$\min_{\mathbf{T}} \max_{||\mathbf{h}||=\rho} \mathrm{asCRB}(\omega_0). \tag{18}$$

Following the proof in [2] yields that

$$\begin{aligned}\max_{||\mathbf{h}||=\rho} \mathrm{asCRB}(\omega_0) =& \max_{||\mathbf{h}||=\rho} \frac{1}{N^3}\frac{6\sigma^2}{\mathbf{h}^H\mathbf{R}\mathbf{h}}\\ =&\frac{6\sigma^2}{N^3}\left[\min_{||\mathbf{h}||=\rho} \mathbf{h}^H\mathbf{R}\mathbf{h}\right]^{-1}\\ =&\frac{6\sigma^2}{N^3}\frac{1}{\rho^2\lambda_{\min}(\mathbf{R})},\end{aligned} \tag{19}$$

where $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue of the matrix between brackets. In this way, the minimax problem amounts to finding the training sequence that minimizes $\frac{1}{\lambda_{\min}(\mathbf{R})}$, or equivalently, maximizes $\lambda_{\min}(\mathbf{R})$. If a transmitted power constraint is added for $\mathbf{R}$, i.e., $\mathrm{Tr}(\mathbf{R}) \le \alpha$, where $\mathrm{Tr}(\cdot)$ stands for the trace of the matrix. The training sequence can be obtained from the following problem

$$\begin{aligned}&\max\ \lambda_{\min}(\mathbf{R})\\ &\text{s.t. } \mathrm{Tr}(\mathbf{R}) \le \alpha,\end{aligned}$$

and the solution is given by [13]

$$\mathbf{R}^* = \frac{\alpha}{L}\mathbf{I}, \tag{20}$$

which corresponds to a white training sequence.

IV. Conclusions

The training sequence design problem for frequency offset estimation in frequency-selective channels is considered in this paper. The results in [2], [6] are re-established assuming an unknown complex-valued noise distribution with a fixed covariance matrix. A minimax optimization problem is formulated and it is illustrated that the circularly symmetric complex-valued Gaussian maximizes the CRB of the frequency offset.

ISBN: 978-1-61804-318-4

Moreover, it is shown herein paper that the white training sequence minimizes the worst CRB of the frequency offset and thus it is optimal for the minimax problem.

APPENDIX A
DERIVATIONS OF $\mathrm{E}_{\mathbf{a}}$ AND $\mathrm{E}_{\mathbf{b}}$

Based on (12) and (13), it follows that

$$\begin{aligned}
&\frac{\partial \log f_{\mathbf{a}}(\mathbf{a})}{\partial \mathbf{a}} \\
&= \left[\frac{\partial \log f_{\mathbf{a}}(\mathbf{a})}{\partial a_0}, \cdots, \frac{\partial \log f_{\mathbf{a}}(\mathbf{a})}{\partial a_{N-1}}\right]^T \\
&= \left[\frac{\partial \log f_{a_0}(a_0)}{\partial a_0}, \cdots, \frac{\partial \log f_{a_{N-1}}(a_{N-1})}{\partial a_{N-1}}\right]^T \\
&= \left[\frac{\partial f_{a_0}(a_0)}{\partial a_0}\frac{1}{f_{a_0}(a_0)}, \cdots, \frac{\partial f_{a_{N-1}}(a_{N-1})}{\partial a_{N-1}}\frac{1}{f_{a_{N-1}}(a_{N-1})}\right]^T,
\end{aligned} \tag{21}$$

where the second equality is due to the fact that

$$f_{\mathbf{a}}(\mathbf{a}) = \prod_{j=0}^{N-1} f_{a_j}(a_j).$$

The expectation of any element in vector (21) can be expressed as

$$\begin{aligned}
&\mathrm{E}\left[\frac{\partial f_{a_j}(a_j)}{\partial a_j}\frac{1}{f_{a_j}(a_j)}\right] \\
=&\int_{-\infty}^{\infty} f_{a_j}(a_j)\frac{\partial f_{a_j}(a_j)}{\partial a_j}\frac{1}{f_{a_j}(a_j)}da_j \\
=&f_{a_i}(a_i)|_{-\infty}^{\infty} \\
=&0,
\end{aligned}$$

where $j = 0, 1, \cdots, N-1$ and the last equality follows from the assumption that the PDF of each i.i.d component has a zero value at the endpoints $-\infty$ and ∞.

Thus, it is concluded that

$$\mathrm{E}_{\mathbf{a}} = \mathrm{E}\left[\frac{\partial \log f_{\mathbf{a}}(\mathbf{a})}{\partial \mathbf{a}}\right] = \mathbf{0}.$$

Following the aforementioned steps will lead to the same result for $\mathrm{E}_{\mathbf{b}}$.

ACKNOWLEDGMENT

This paper was made possible by NPRP grant NPRP 4-1293-2-513 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

REFERENCES

[1] P. Ciblat, P. Bianchi, and M. Ghogho, "Optimal training for frequency offset estimation in correlated-Rice frequency-selective channel," in *9th workshop on Signal Processing Advances in Wireless Communications*, Recife, Brazil, Jul. 2008, pp. 6–10.

[2] O. Besson and P. Stoica, "Training sequence selection for frequency offset estimation in frequency selective channels," *Digital Signal Processing*, vol. 13, pp. 106–127, 2003.

[3] S. Crozier, D. Falconer, and S. Mahmoud, "Least sum of squared errors (LSSE) channel estimation," *Inst. Elect. Eng. Proc. F*, vol. 138, no. 4, pp. 371–378, 1991.

[4] C. Tellambura, M. Parker, Y. J. Guo, S. Sheperd, and S. Barton, "Optimal sequences for channel estimation using discrete Fourier transform techniques," *IEEE Trans. Commun.*, vol. 47, no. 2, pp. 230–238, 1999.

[5] W. Chen and U. Mitra, "Training sequence optimization: Comparisons and an alternative criterion," *IEEE Trans. Commun.*, vol. 48, no. 12, pp. 1987–1991, 2000.

[6] P. Stoica and O. Besson, "Training sequence design for frequency offset and frequency-seletive channel estimation," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1910–1917, 2003.

[7] Y. D. Kim, J. K. Lim, C. Suh, and Y. H. Lee, "Designing training sequences for carrier frequency estimation in frequency-selective channels," *IEEE Trans. Veh. Technol.*, vol. 55, no. 1, pp. 151–157, 2006.

[8] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data*. Cambridge, UK: Cambridge, 2010.

[9] R. Rioul, "Information theoretic proofs of entropy power inequalities," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 33–55, 2011.

[10] S. Park, E. Serpedin, and K. Qaraqe, "Gaussian assumption: The least favorable but the most useful," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 183–186, 2013.

[11] P. Stoica and P. Babu, "The Gaussian data assumption leads to the largest Cramér-Rao bound," *IEEE Signal Process. Mag.*, vol. 23, no. 3, pp. 132–133, 2011.

[12] J. F. Bercher and C. Vignat, "On minimum Fisher information distributions with restricted support and fixed variance," *Information Sciences*, vol. 179, pp. 3832–3842, 2009.

[13] T. Soderstrom and P. Stoicce, *System Identification*. London, UK: Prentice Hall International, 1989.

YouTube's DASH implementation analysis

Javier Añorga, Saioa Arrizabalaga, Beatriz Sedano, Maykel Alonso-Arce, and Jaizki Mendizabal

***Abstract*—**As long as *YouTube* is one of the most used services of the World Wide Web, it consumes an enormous quantity of bandwidth along the network. In order to maintain a high efficiency in this bandwidth management requirement, *YouTube* needs to adopt (and renew) highly efficient video streaming techniques. These changes produce a deprecated literature about *YouTube* service traffic characterization. This work reports an analysis of the recent DASH adaptive video streaming technique adopted by *YouTube*. In addition, this article includes the state of art about the literature with regard to the *YouTube* traffic characterization and analyses the DASH implementation of *YouTube*, reporting the relationship between download bandwidth consumption and video quality obtained and its performance.

***Keywords*—**YouTube, DASH, streaming, bandwidth, video quality.

I. Introduction

YouTube is one of the most popular services on internet, being the third most visited web site in the world [1], and its traffic has a great impact over mobile and fixed networks. With this great popularity and bandwidth intensive demand, *YouTube* presents a challenge for Internet Service Providers (ISPs) in order to offer a good quality for the consumed download services by clients. Therefore, the analysis and characterization pattern of this service traffic is important.

YouTube's traffic has been studied and documented, such as in [2] and [3]. *YouTube*'s videos are transported by HTTP over TCP. Consequently, YouTube has not to cope with lost or reordered packets, and the only quality degradation which may be caused by transmission, is a stalling of the video. However, using the Dynamic Adaptive Streaming over HTTP (DASH) technique, *YouTube* is able to switch the video quality based on the link capabilities. The main outcome of this feature is that if on *YouTube*'s player quality parameter is set on "*auto*", *YouTube* can adapt the bitrate of the video based on the client's available bandwidth.

The main objective of this work is to find the relationship among the *YouTube*'s downloaded video quality level, the consumed *YouTube*'s video bandwidth and the available download bandwidth from the perspective of an access point to the Internet. This work uses a Home Gateway (HG) as the access point to the Internet with a *YouTube* client inside the Local Area Network (LAN). In addition, the time that *YouTube*'s DASH implementation needs to adapt to available bandwidth fluctuations is characterized.

This paper is organized as follows. Section II describes the state of art about DASH and the adaptive streaming. Section III focuses on the related work about *YouTube* characterization, describing some key points of *YouTube*'s behavior. The test-bed deployment for measurements is depicted in Section IV. Results obtained from the measurements are shown and analyzed in Section V. Finally, conclusions and future lines are exposed in Section VI.

II. DASH And adaptive streaming

In the past, streaming services were offered over UDP (User Datagram Protocol) transport protocol; however, nowadays, with the increasing bandwidth connection at households and the popularity of World Wide Web, the media content can be efficiently delivered now in larger segments using HTTP (HyperText Transfer Protocol). It is motivated due to two main reasons. Firstly, HTTP is more firewall friendly because most of the firewalls are configured to allow HTTP outgoing connections, and, secondly, with HTTP streaming, the client manages the streaming without having to maintain a session state on the server.

However, basic progressive HTTP based streaming is not suitable for environments which may have a considerable high bandwidth fluctuations. The video stream has to adapt to the varying bandwidth capabilities in order to deliver the user a continuous video stream without any stalls at the best possible quality for the moment. This is achieved by adaptive streaming over HTTP.

There are adaptive HTTP-Streaming based proprietary systems like *Smooth Streaming* (from *Microsoft*), *HTTP Dynamic Streaming* (from *Adobe*) or HTTP Live Streaming (from *Apple*). Each solution reports its advantages and drawbacks depending on the circumstances, as it is detailed in [4]-[6]. DASH [7]-[8], or *MPEG-DASH*, is an emerging ISO/IEC MPEG standard that is an extension of the classic HTTP streaming. Through DASH technology, the video

J. Añorga is with Ceit and Tecnun, Parque Tecnológico de San Sebastián, Paseo Mikeletegi, Nº 48, 20009, Donostia - San Sebastián (phone: +34 943 212800 ext. 2983; fax: +34 943 213076; e-mail: jabenito@ceit.es).

S. Arrizabalaga is with Ceit and Tecnun, Parque Tecnológico de San Sebastián, Paseo Mikeletegi, Nº 48, 20009, Donostia - San Sebastián (e-mail: sarrizabalaga@ceit.es).

B. Sedano is with Ceit and Tecnun, Parque Tecnológico de San Sebastián, Paseo Mikeletegi, Nº 48, 20009, Donostia - San Sebastián (e-mail: bsedano@ceit.es).

M. Alonso-Arce is with Ceit and Tecnun, Parque Tecnológico de San Sebastián, Paseo Mikeletegi, Nº 48, 20009, Donostia - San Sebastián (e-mail: maarce@ceit.es).

J. Mendizabal is with Ceit and Tecnun, Parque Tecnológico de San Sebastián, Paseo Mikeletegi, Nº 48, 20009, Donostia - San Sebastián (e-mail: jmendizabal@ceit.es).

ISBN: 978-1-61804-318-4

quality level can be delivered according to the current network conditions. Several representations of a video clip are generated based on quality/bit rate level and each representation is divided into fragments (usually from 2 to 10 seconds of length) [9]. In order to avoid rebuffering due to buffer starvation, the video player usually chooses a quality level that has a lower bit rate than the measured available bandwidth. In this sense, the video download rate can be higher than the video playback rate (or at the least the same if a full video buffer is present). Moreover, the use of DASH also results in bandwidth consumption saving as it is reported in [10].

Because of the benefits that HTTP-Streaming based technology implies and due to the fact that *DASH* is a company-independent standard, and it allows saving bandwidth resources, nowadays, *YouTube* and other popular services, such as *Netflix*, have implemented DASH as the preferred streaming technology rather than FLV (Flash Video) streaming.

III. Related Work

There are several studies that have characterized the *YouTube* traffic along the past years. However, until the date when this work was written, there are no so much studies about the *YouTube* behavior with DASH streaming standard [10]. For example, in [3], the *YouTube* service from the viewpoint of traffic generation in the server's application layer is characterized focused on FLV based video clips. An analysis of how content distribution in *YouTube* is realized is done in [11], conducting a measurement study of *YouTube* traffic in a large university campus network.

The article [12], published in 2011, studies the *YouTube* streaming characteristics and operation, and it reports for mobile devices: "In fact, the mobile devices cannot buffer the entire video so the player progressively requests portions according to the evolution of the playback". This reported issue in 2011 presents the same properties as the DASH operation detailed in previous section.

In [3] also two phases of *YouTube* streaming is reported. Frist occurs an initial phase where there is a significant burst of data. This phase is called "burst phase". After this initial burst, the receiving download data rate at *YouTube*'s player is considerably reduced. This second phase is called "throttling phase". In addition, from the analysis of FLV *YouTube* videos, [13] describes that during the initial burst phase, the amount of data sent by the *YouTube* server is related to the transmission rate during the throttling phase.

This work is focused on the analysis of DASH behaviour of *YouTube* video streaming to desktop devices, more precisely using *Google Chrome* as web browser software. This article tries to fill the gap in literature (due to the constant changes in streaming techniques) about *YouTube*'s DASH implementation characterization providing a data report about *YouTube* download bandwidth consumption versus the *YouTube* quality level obtained. In addition, this work reports how *YouTube*'s DASH implementation behaves in terms of time response to available bandwidth fluctuations.

IV. Deployed Test-bed

In order to carry out this the *YouTube*'s *DASH* analysis, the architecture shown in Fig. 1 is used. A PC is used with *Windows 7* as running operative system and *Google Chrome* as web browser. Client machine is connected to the Internet through the HG. *Client – YouTube's server* download bandwidth measurements are done in the HG. An Additional *Java* application is running on the client machine. This *Java* application is called *Commander* and commands *Google Chrome* to open a *YouTube*'s testing web page served from the HG.

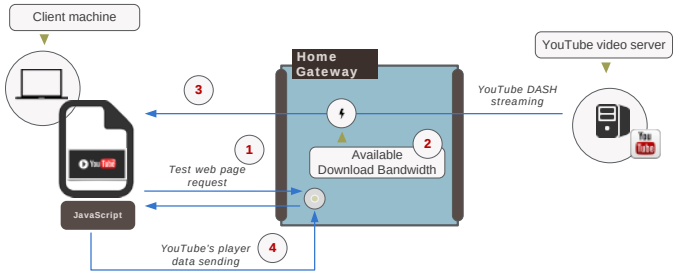


Fig. 1 Test-bed architecture and test steps.

The measurements trend to be as non- invasive as possible. The following steps are done:

1) *Commander* on client machine calls the testing web page from HG with 3 possible arguments:
   - *YouTube*'s video id: the *YouTube*'s video id to invoke at *YouTube*'s video player (which video reproduce from *YouTube*).
   - *YouTube's* video quality: fix the quality of the video. Possible values are: *tiny* (240p), *medium* (360p), *large* (480p), *hd720* (720p), *hd1080* (1080p) and *auto*.
   - HG allowed download bandwidth: the allowed download bandwidth from *Client – YouTube's server* download connection at HG.
2) The requested available download bandwidth is applied at HG.
3) The served web page contains an embedded *YouTube*'s player configured to automatically start to play the *YouTube*'s video id requested at configured quality level.
4) The variables of this *YouTube* embedded player are extracted by *JavaScript* code and sent back to the HG in order to centralize all the measurements.

Once the test finishes a post process of the monitored data is done with *Excel* or *Matlab* software.

V. Testing and Results

This section reports the results obtained from the realization of 4 different tests. The first test (A), details the *YouTube* DASH video streaming phases with no bandwidth limitations at the HG. The next three tests (*B, C, D*) study the relationship between the average download bandwidth consumption of the requested video and the *YouTube* quality level obtained. These

tests are replicated over the same 199 videos, and all of these videos are requested from *YouTube Data Api* [14] with parameters *videoEmbeddable* (only videos that can be embedded in a web page) and *videoSyndicated* (only videos that can be played outside of *youtube.com*) set on *true*. Test *B* tries first to obtain the average download bandwidth consumption of a *YouTube*'s video for a fixed *YouTube*'s quality level. Then, the obtained average bandwidth is used to set an allowed download rate and it is checked whether the corresponding quality level is obtained. Test *C* reports a ratio out of the 199 test videos of which *YouTube* streaming video quality level is obtained restricting the available download bandwidth at the HG.

Finally, test *D* depicts the detailed DASH adaptation pattern for available bandwidth changes, where the time the player needs to notice the change has been measured. In addition, the time delay between the quality request and the video quality change is also quantified.

A. *ZeroTest*

This test reports the *YouTube* DASH implementation behavior with no bandwidth restrictions at HG. The video quality level is set on *auto* and *YouTube* player decides to request *hd1080* because of the high available bandwidth that sees on the link. This test is shown in Fig. 2 illustrating the download video rate and the buffer state.

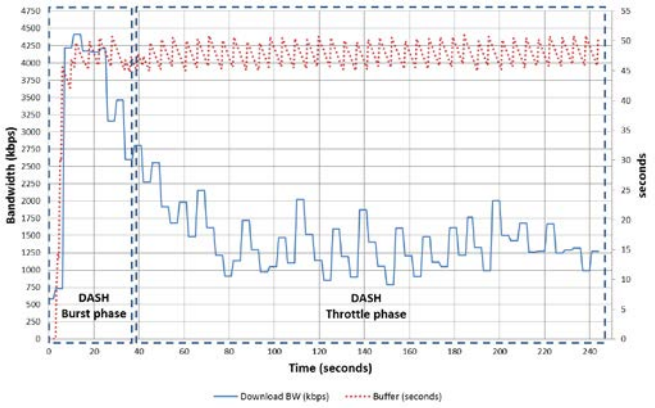


Fig. 2 *ZeroTest* result graph. From approximately second 0 to 35 the first DASH burst phase occurs. After this period, it follows the throttle phase.

From second 0 to approximately second 35 occurs the DASH burst phase and the download rate of video reaches 4400 kbps as maximum value. At this phase the *YouTube*'s player buffer is being filled through a high request of video fragments. When the buffer is full, in this case storing about 45 seconds of video playback, a second throttle phase follows. At this stage the amount of received data is significantly reduced (about an average of 1300kbps), maintaining the download rate (and the player buffer state) according to the video playback rate.

B. *avgBW@fixedYTquality(4min)*

This test is intended to report the average download bandwidth consumption of *YouTube*'s videos given a fixed *YouTube*'s requested quality. The test is done over the reproduction of the first four minutes of each one of the 199 videos. For five different quality levels (*small*, *medium*, *large*, *hd720* and *hd1080*) the average bandwidth consumption per video is shown in Fig. 3. It can be appreciated that values show an increasing average bandwidth consumption when requested quality level is increased also.

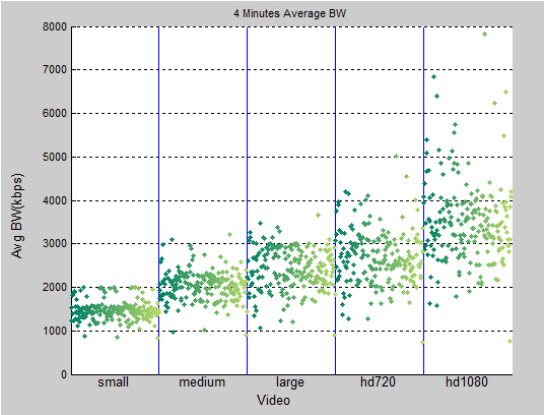


Fig. 3 Average bandwidth consumption (kbps) in first 4 minutes of *YouTube* video playback.

Table I summarizes the measures obtained and depicted in Fig. 3. The average bandwidth consumption per *YouTube*'s quality level of the 199 videos and the standard deviation per *YouTube*'s quality level measurements are shown. It is relevant to remark that the dispersion of the data measured increases with the quality level.

Table I Average bandwidth (kbps) and std. deviation per *YouTube*'s quality level (199 videos). 4 minutes of each video playback.

Quality	**Avg. BW (kbps)**	**Avg. BW std. dev.**
small (240p)	1485.11	225.04
medium (360p)	2061.22	348.46
large (480p)	2446.99	485.66
hd720 (720p)	2737.27	632.22
hd1080 (1080p)	3541.30	951.19

Then, the maximum available download bandwidth is fixed in the HG with the average bandwidth acquired from Table I. The result obtained is that most of the *YouTube*'s videos take a superior quality level than it is expected from Table I, especially for lower intended qualities.

As a second approach, the average bandwidth consumption of the throttle phase has been calculated for the same set of videos. Fig. 4 depicts the results obtained discarding the data of the first two minutes of the previous test. As it is expected, the average bandwidth consumption of throttling phase per quality level also increases when the video quality level increases. Since the throttle phase maintains the player buffer

ISBN: 978-1-61804-318-4

with an approximately constant buffer length, the download bandwidth consumption depends on video playback bit rate.

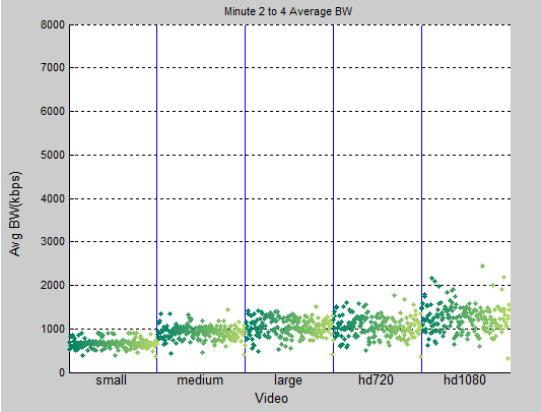


Fig. 4 Average bandwidth consumption (kbps) in minute 2 to 4 of *YouTube* video playback (throttle phase).

Table II summarizes the measurements obtained and depicted in Fig. 4. The average bandwidth consumption per *YouTube*'s quality level of the 199 videos and the standard deviation per *YouTube*'s quality level measurements are shown. Again the dispersion of the average bandwidth consumed per video increases with the quality level, although this dispersion is sensibly lower. The average bandwidth consumption at throttle phase is nearly reduced by a percentage of 60%.

Table II Average bandwidth (kbps) and std. deviation per *YouTube*'s quality level (199 videos). 2 minutes of each video playback.

Quality	**Avg. BW (kbps)**	**Avg. BW std. dev.**	**% of BW reduction**
small (240p)	668.38	103.17	54.99
medium (360p)	922.04	157.14	55.27
large (480p)	1052.77	211.05	56.98
hd720 (720p)	1065.22	240.72	61.08
hd1080 (1080p)	1252.65	318.61	64.63

Again, the same proof is done. If the maximum available download bandwidth is fixed with the average bandwidth obtained in Table II the result obtained this time is that most of the *YouTube*'s videos trend to take an inferior quality level than it is expected from Table II, this time especially if the high definition qualities (*hd720* and *hd1080*) are tested.

C. *YTQuality@fixedAvailBW(4min)*

This test is intended to report the quality level that the *YouTube*'s DASH implementation finally choses when a fixed available downstream bandwidth is applied at HG and *auto* quality if configured. As the previous case, the test is done over the first four minutes of video playback.

Fig. 5 it shows the percentage of the tested videos that take a determined *YouTube*'s quality level with a fixed available download bandwidth (*tiny* quality stands for 144p resolution). The figure shows, once more, the increasing bandwidth consumption with the quality level. Whereas the *small* and *medium* qualities present high percentage of videos concentered around a determined value of available bandwidth, *hd720* presents the most entropy in finding a corresponding available bandwidth value.

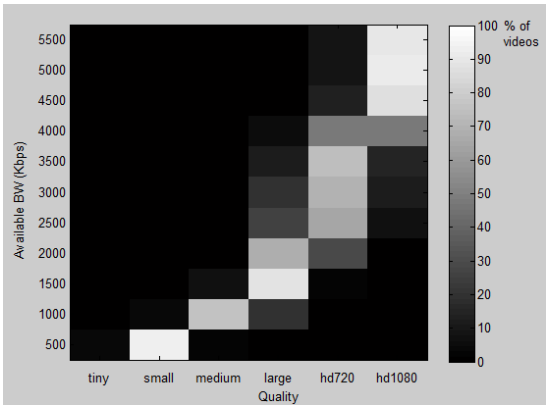


Fig. 5 Percentage of *YouTube*'s tested videos that finally take a determined quality (Xaxis) with a fixed available download bandwidth (Yaxis).

In Fig. 6 it is shown the percentage of *YouTube*'s tested videos that take a determined quality level or superior given a fixed allowed download bandwidth. For example, for the case of *hd720*, it is needed to set an available download bandwidth of 4000kbps to estimate with around 80% of probability that the video quality level is going to be set as *hd720* when the *auto* quality is selected.

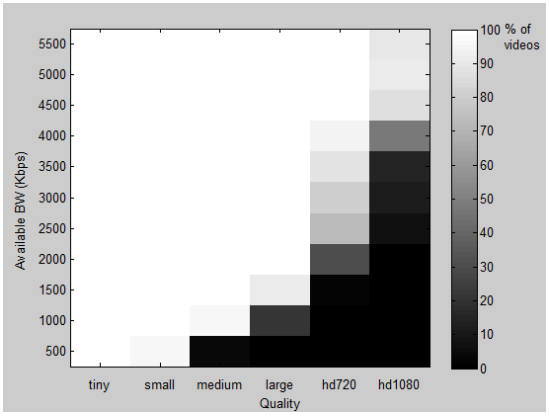


Fig. 6 Percentage of *YouTube*'s tested videos that finally take a determined quality or superior (Xaxis) with a fixed available download bandwidth (Yaxis).

Fig. 6 can be used to obtain an estimation of the probability of watch *YouTube*'s videos at a determined quality level or superior with an allowed maximum download rate. It provides valuable information for bandwidth management in the HG settings.

ISBN: 978-1-61804-318-4

Table III Summary of results.

Allowed BW change (seconds)	Quality requested (seconds)	Player quality request (seconds)	Video quality change (seconds)	Delay for quality request (seconds)	Delay quality change (seconds)	Player buffer state (seconds)
0	*small* (240p)	0	Not applicable	Not applicable	Not applicable	Not applicable
35,83	*medium* (360p)	58,05	104,46	22,22	46,42	45,23
120,83	*large* (480p)	137,70	244,47	16,86	106,78	105,58
262,82	*hd720* (720p)	275,82	464,41	13,00	188,59	187,84
493,83	*hd1080* (1080p)	532,77	619,51	38,94	86,74	85,97

D. YTQuality@BWScale

This last test is intended to report the time that *YouTube* player lasts to show a requested player quality operating in *auto* mode. A monotonically increasing scale of maximum available download bandwidth steps at HG is used for this test, which have been selected based on the knowledge acquired in the previous tests. This test is applied over a selected *YouTube*'s video as an example, and it is depicted in Fig. 7. The figure shows the download bandwidth consumption of the *YouTube*'s player in kbps, the available download bandwidth set in the HG in kbps, *YouTube*'s player buffer in seconds and the marks representing the instant of *YouTube* quality request (triangle marks) and effective quality change in the played video (diamond marks).

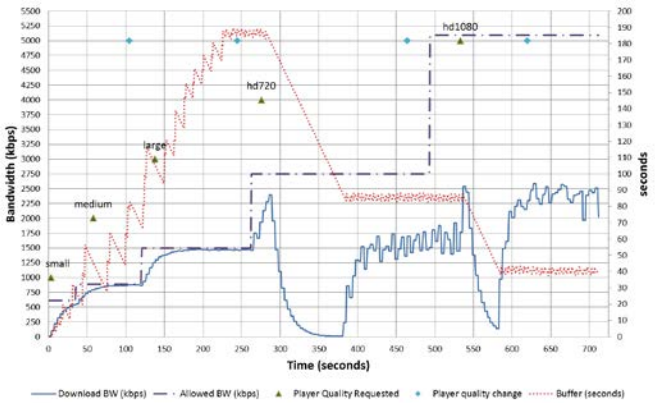


Fig. 7 Monotonically increasing scale of maximum available download bandwidth test.

Results obtained in Fig. 7 are summarized in Table III. The table shows that the time delay for the player to detect network bandwidth fluctuations goes in the range from 13 to 40 seconds. Then, the player requests the new quality but this is not immediately shown to the user. In fact, the time delay for the quality change measured in the test and shown in the table (column *delay quality change*) is approximately the same as the amount of seconds of video that embedded player's video stores. Due to this fact, when the request is made with a high amount of video stored in buffer the expected quality is shown with a relative high delay of time (approximately 3 minutes for *hd720* quality request at the depicted test).

The use of a large buffer involves a lack of response in showing the quality requested with regard to a decision of change the quality requested by the player due to an available bandwidth fluctuation. However, a large buffer prevent video stalling and rebuffering events, and it also allows the user to instantly seek and play any part of the video that is inside the buffer range without the stalling of the rebuffering event. This presents a trade-off analysis in order to obtain an optimum quality of experience by the end user.

VI. Conclusion

This article has exposed the state of the art about DASH standard and adaptive video streaming and also has made a review of the related work on *YouTube*'s traffic analysis. From this review can be concluded that changes in technology, and more precisely in streaming techniques, occurs frequently, deprecating the analysis and reports of a determined service made by researches along the recent years. Due to this constantly changing world, this work tries to fill the gap about the characterization of the recent use of DASH implementation as the preferred streaming technique by *YouTube*.

This article has reported the results obtained from the analysis of the *YouTube*'s DASH implementation download traffic patterns. The relationship between the average download rate of *YouTube*'s video streaming and the quality that *YouTube*'s player operating in *auto* mode finally decides to request from server. In addition, the *YouTube*'s DASH quality adaptation performance with regard to a bandwidth fluctuation test is also reported. From this last analysis, it is concluded that the use of a large video buffer involves a lack of response in showing the quality requested with regard to available bandwidth fluctuations. However, a large video buffer prevent video stalling and rebuffering events and also allows the end user to instantly seek and play any part of the video that is inside the buffer range without the stalling of the rebuffering event.

Future lines of this work involve to expand the bandwidth fluctuation test analyzing the buffer filling and varying some network conditions, such as the latency to the server. Also a comparison analysis between *YouTube*'s video streaming to PCs and mobile devices could be done.

References

[1] Alexa Corporation. *The top 500 sites on the web*. Available: http://www.alexa.com/topsites. Accessed 2015.

[2] A. Rao, A. Legout, Y. Lim, D. Towsley, C. Barakat and W. Dabbous, "Network characteristics of video streaming traffic," *in Proceedings of the Seventh COnference on emerging Networking EXperiments and Technologies*, pp. 25, 2011.
[3] P. Ameigeiras, J. J. Ramos-Munoz, J. Navarro-Ortiz and J. M. Lopez-Soler, "Analysis and modelling of YouTube traffic," *Transactions on Emerging Telecommunications Technologies*, vol. 23, pp. 360-377, 2012.
[4] C. Müller, S. Lederer and C. Timmerer, "An evaluation of dynamic adaptive streaming over http in vehicular environments," *in Proceedings of the 4th Workshop on Mobile Video*, pp. 37-42, 2012.
[5] S. Akhshabi, A. C. Begen and C. Dovrolis, "An experimental evaluation of rate-adaptation algorithms in adaptive streaming over http," *in Proceedings of the second annual ACM conference on Multimedia systems*, pp. 157-168, 2011.
[6] S. Akhshabi, S. Narayanaswamy, A. C. Begen and C. Dovrolis, "An experimental evaluation of rate-adaptive video players over HTTP," *Signal Process Image Commun*, vol. 27, pp. 271-287, 4, 2012.
[7] C. Timmerer and C. Müller, "HTTP streaming of MPEG media," *Streaming Day*, 2010.
[8] I. Sodagar, "The mpeg-dash standard for multimedia streaming over the internet," *IEEE Multimedia*, pp. 62-67, 2011.
[9] S. Lederer, C. Müller and C. Timmerer, "Dynamic adaptive streaming over http dataset," *in Proceedings of the 3rd Multimedia Systems Conference*, pp. 89-94, 2012.
[10] D. K. Krishnappa, D. Bhat and M. Zink, "DASHing YouTube: An analysis of using DASH in YouTube video service," *in Local Computer Networks (LCN), 2013 IEEE 38th Conference on*, pp. 407-415, 2013.
[11] M. Zink, K. Suh, Y. Gu and J. Kurose, "Characteristics of YouTube network traffic at a campus network–measurements, models, and implications," *Computer Networks*, vol. 53, pp. 501-514, 2009.
[12] A. Finamore, M. Mellia, M. M. Munafò, R. Torres and S. G. Rao, "Youtube everywhere: Impact of device and infrastructure synergies on user experience," *in Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pp. 345-360, 2011.
[13] S. Alcock and R. Nelson, "Application flow control in YouTube video streams," *ACM SIGCOMM Computer Communication Review*, vol. 41, pp. 24-30, 2011.
[14] YouTube. *YouTube Data API*. Available: https://developers.google.com/youtube/v3/. Accessed 2015.

Javier Añorga was born in Logroño, Spain, in 1987. He received his MSc degree in Telecommunications Engineering from Tecnun (School of Engineering at San Sebastián), University of Navarra, Spain, in 2011.

He joined the CEIT Research Centre in San Sebastián in 2011, and he is currently working on his PhD in CEIT's Electronics and Communications Department. His professional research activity is in the field of communication protocols, QoS, QoE, Residential Gateways, Embedded Systems and Information Technology. Currently he is also lecturer at the Engineering School of the University of Navarra (Tecnun) in San Sebastián.

Saioa Arrizabalaga was born in Azkoitia in 1979. She received her MS degree in Telecommunication Engineering from the Faculty of Engineering in Bilbao (UPV-EHU) in 2003 and obtained her PhD degree from the University of Navarra in 2009.

She has been involved in several projects regarding remote monitoring using embedded systems, Internet access sharing, Residential Gateways or QoS management in Multi-Dwelling Units. She has published a book and several articles in international journals and conferences. Currently she is also lecturer of the Computer Architecture subject at the Engineering School of the University of Navarra (Tecnun) in San Sebastián.

Beatriz Sedano was born in Eibar in 1978. She received her degree in Telecommunication Technical Engineering in the Engineering School from Santander (University of Cantabria) in 2000 and in Telecommunication Engineering in the same university in 2003. Afterwards she received her PhD. In Engineering (about ultra-wideband microwave oscillators) in 2009 in the Engineering School of the University of Navarra (TECNUN).

She has been involved in research projects related to monitoring using embedded systems: RF domotic devices for intelligent housings and a system based on GSM-R for railway communications. Currently she works in a research project related to a wireless communication system for bioengineering, and she is lecturer of the Microwave subject and its laboratory in the Engineering School of the University of Navarra in San Sebastian. She has published several articles in international journals and conferences.

Maykel Alonso-Arce was born in Vitoria-Gasteiz (Spain) in 1985. He received his Technical Telecommunications Engineering degree, majoring in Communications Systems, in 2007 and his Telecommunications Engineering degree in 2009 from the Faculty of Engineering Mondragon University (MGEP), Spain.

During his studies he worked for Fagor Electronics porting uC-Linux to a prototype board as an intern of the Faculty of Engineering - Mondragon University (MGEP). He also studied one semester at the Aalborg University (AAU), Denmark, through an ERASMUS scholarship. In 2009 he joined the Electronics & Communication Department at CEIT, and started his PhD studies, through the Iñaki Goenaga (FCT-IG) Technology Centers Foundation grant. At present, his research interest field is focused on the "Design of embedded communications systems inside human bodies".

Jaizki Mendizabal is a lecturer at Tecnun, the Technological Campus of University of Navarra, San Sebastián, Spain, and a researcher in the Electronics and Communications Department at CEIT. He was born in Zarautz and received his MSc and PhD degrees in Electrical Engineering from Tecnun (University of Navarra, San Sebastian, Spain) in 2000 and 2006 respectively.

He joined Fraunhofer Institut fr Integrierte Schaltungen, Erlangen (Germany) from 2000 to 2002 and SANYO Electric Ltd, in Gifu (Japan) from 2005 to 2006 as RF-IC designer. He obtained his PhD in the field of monolithic RF design for GNSS systems. He currently works in CEIT where his research interests include RFICs and analogue safety systems for the railway industry. He has participated in more than 8 research projects, has directed 2 doctoral theses, is author or co-author of some 21 scientific and technical publications in national and international journals and conferences and is the author of the book GPS and Galileo Dual RF Front-end receiver and Design, Fabrication, Test published by McGraw-Hill.

ISBN: 978-1-61804-318-4

Comparison of Evolutionary Optimization Algorithms for FM-TV Broadcasting Antenna Array Null Filling

Emmanouil Tziris, Pavlos I. Lazaridis, Bruce Mehrdadi, Violeta Holmes, Ian A. Glover, Zaharias D. Zaharis, Aristotelis Bizopoulos, and, John P. Cosmas.

***Abstract* — Broadcasting antenna array null filling is a very challenging problem for antenna design optimization. This paper compares five antenna design optimization algorithms (Differential Evolution, Particle Swarm, Taguchi, Invasive Weed, Adaptive Invasive Weed) as solutions to the antenna array null filling problem. The algorithms compared are evolutionary algorithms which use mechanisms inspired by biological evolution, such as reproduction, mutation, recombination, and selection. The focus of the comparison is given to the algorithm with the best results, nevertheless, it becomes obvious that the algorithm which produces the best fitness (Invasive Weed Optimization) requires very substantial computational resources due to its random search nature.**

***Keywords*— antenna array, null filling, evolutionary optimization algorithms, Particle Swarm, PSO, Differential evolution, Invasive Weed Optimization, IWO.**

I. INTRODUCTION

Research on antennas has become very challenging, especially in the area of broadcasting [1-2]. A lot of techniques have been proposed for the design of base station antenna arrays in order to satisfy requirements, which are essential for broadcasting applications [3-4]. These requirements usually considered by a broadcasting antenna array are given below: (a) due to the large distance between the transmitting base station and the service area, the antenna array needs to produce a very narrow main lobe which, in conjunction with the need for reduction of the spatial spread of radiated power, results in the requirement of maximum gain. (b) Provided that the broadcasting base station is usually located at higher places relative to the service area, the main lobe is required to be tilted from the horizontal plane. Due to the large distance from the service area, the tilting angle is usually small (between 2 and 4 degrees).

Emmanouil Tziris, and John Cosmas are with the School of Engineering and Design, Brunel University, London UB8 3PH, UK.
Pavlos I. Lazaridis, Bruce Mehrdadi, Violeta Holmes, and Ian A. Glover are with the Department of Engineering and Technology, University of Huddersfield, HD1 3DH, UK.
Zaharias D. Zaharis is with Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece. Aristotelis Bizopoulos is with the Department of Electronics, Alexander Technological Educational Institute of Thessaloniki, GR-57400 Thessaloniki, Greece.

(c), in order to have satisfactory reception of transmitted signal inside an angular sector under the main lobe, the directional gain is not permitted to fall below a certain value in relation to the maximum gain value, which results in filling of radiation pattern nulls inside the above-mentioned angular sector. The level of null filling depends on the service type (e.g., FM radio, TV DVB-T) and the value of signal-to-noise ratio (SNR), (iv) In order to reduce the power reflection along the feeding lines and thus increase the efficiency of the whole feeding network, the impedance matching condition is required for every array element, which means that the standing wave ratio (SWR) of every element must be close to unity.

It is obvious that the design of such an antenna array is a multi-objective problem, since the above requirements must be simultaneously satisfied. Therefore, an optimization method is necessary to solve this kind of problem, [5-7]. Such an efficient method recently proposed is the Invasive Weed Optimization (IWO) method [5-12]. The IWO is an evolutionary method inspired from the invasive nature of weeds. Due to its fast convergence and performance, the IWO has been chosen to solve many problems in the area of electromagnetics. The optimization methods under study have been applied to optimize linear arrays according to the above-specified requirements. In all the cases studied here, a uniform-amplitude excitation distribution is considered to be applied on the array elements, since excitations of equal amplitudes are easily implemented in practice. In the two studied cases, linear arrays of 8 and 16 isotropic sources, respectively, are optimised for maximum gain, main lobe tilting and null filling, while the impedance matching condition is not required due to the use of isotropic sources. The radiation characteristics of each array need to be calculated for every evaluation of the fitness function, which is going to be minimized by the optimization methods. The optimization results exhibit the relative effectiveness of the proposed methods. More specifically, the IWO method has initially been proposed by Mehrabian and Lucas [5]. The IWO algorithm simulates the colonizing behavior of weeds in nature. Initially, a population of weeds is dispersed at random positions inside an N-dimensional search space, where N is the number of parameters to be optimized by the IWO algorithm for the given problem. These positions are produced by a uniform random number generator. The optimization algorithm is an iterative process and consists of three basic steps repeatedly applied on each iteration.

ISBN: 978-1-61804-318-4

In artificial intelligence, an evolutionary algorithm (EA) is a generic population-based metaheuristic optimization algorithm. An EA uses mechanisms inspired by biological evolution, such as reproduction, mutation, recombination, and selection. Candidate solutions to the optimization problem play the role of individuals in a population, and the fitness function determines the environment within which the solutions "live". Evolution of the population then takes place after the repeated application of the above operators. Artificial evolution (AE) describes a process involving individual evolutionary algorithms. EAs are individual components that participate in an EA.

Antenna arrays play an important role in detecting and processing signals arriving from different directions. The goal in antenna array geometry synthesis is to determine the physical layout of the array that produces a radiation pattern that is closest to the desired pattern. The shape of the desired pattern can vary widely depending on the application.

Before starting to use an EA, setting up the problem is required, which means making sure than an EA is the optimal solution to the problem. Secondly, the parameters that need optimization must be decided. The parameter which needs to be maximized is the fitness of the population and it is used to generate the next population after being evaluated.

Some basic optimization concepts for electromagnetic applications will be evaluated for this project and these are the following: 1. Differential Evolution (DE), 2. Particle Swarm Optimization (PSO), 3. Invasive Weed Optimization (IWO), 4. Taguchi's Optimization Method, 5. Adaptive IWO (ADIWO).

The main steps of an EA are explained and shown on the flowchart below for a better understanding.

1. Initialization of Population: Initially a random population size is generated. Size differs depending on the problem, so that the entire range of possible solutions is allowed.

2. Evaluation of Fitness: Each individual of the population has a fitness value which is evaluated to decide which individuals have the best fitness.

3. Selection of Population with the Best Fitness: After the fitness evaluation the individuals with the best fitness values are chosen and are used for the next population.

4. Termination: Steps 2 and 3 are repeated until the best fitness is found and the process is terminated.

II. Evolutionary Algorithms

A. *Differential Evolution*

The general problem that an optimization algorithm is concerned with, is to determine the vector variable x so as to optimise:

$$f(x); x = \{x_1, x_2, ..., x_D\} \quad (1)$$

Where, D is the dimensionality of the function. The variable domains are defined by their lower and upper bounds:

$$x_{j,low}, x_{j,upp}; j \in \{1, ..., D\}.$$

The population of the original DE algorithm contains NP D-dimensional vectors:

$$x_{i,G} = \{x_{i,1,G}, x_{i,2,G}, ..., x_{i,D,G}\}, i = 1, 2, ..., NP \quad (2)$$

Where, G is the generation
During one generation for each vector, DE employs mutation and crossover operations to produce a trial vector:

$$u_{i,G} = \{u_{i,1,G}, u_{i,2,G}, ..., u_{i,D,G}\}, i = 1, 2, ..., NP \quad (3)$$

Then, a selection operation is used to choose vectors for the next generation (G+1). The initial population is selected uniform randomly between the lower $(x_{j,low})$ and upper $(x_{j,upp})$ bounds defined for each variable x_j. These bounds are specified by the user according to the nature of the problem. After initialization, DE performs several vector transforms (with the above mentioned operations), in a process called evolution.

B. *Particle Swarm*

In PSO terminology, [13-14], every individual in the swarm is called “particle” or “agent”. The number S of the particles that compose the swarm is called “population size”. A population size between 10 and 50 is optimal for many problems. All the particles act in the same way like bees do, they move in the search space and update their velocity according to the best positions already found by themselves and by their neighbors, trying to find an even better position. Each particle is treated as point in an N-dimensional space. The position of the i-th particle $(i = 1, ...S)$ is represented as $x_i = (x_{i1}, x_{i2}, ..., x_{iN})$, where $x_{in} (n = 1, ..., N)$ are the position coordinates. Each coordinate x_{in} may be limited in the respective (n-th) dimension between an upper boundary U_n and a lower boundary L_n, so that $L_n \le x_n \le U_n (n = 1, ..., N)$. The difference $R_n = U_n - L_n$ between the two boundaries is called “dynamic range” of the n-th dimension. The performance of each particle is measured according to a predefined mathematical function F called “fitness function”, which is related to the problem to be solved. The value of the fitness function depends on the position coordinates, i.e, $F = F(x_i) = F(x_{i1}, x_{i2}, ..., x_{iN})$. Actually, the particle position is considered to be improved as the value of the fitness function is increased/or decreased (maximization or minimization problem). The best previous position (best position) of the i-th particle is recorded and represented as $p_i = (p_{i1}, p_{i2}, ..., p_{iN})$.
The change of x_i is:

$$\Delta x_i = u_i \Delta \tau \quad (4)$$

Δt is the time interval, $\upsilon_i = (v_{i1}, v_{i2}, \ldots, v_{iN})$ is the velocity of the i-th particle, and $\upsilon_{in} (n = 1, \ldots, N)$ are the velocity coordinates.

Calculation of velocity:
Considering that $\Delta t = 1$, the position change becomes $\Delta x_i = \upsilon_i$. Thus, the new position of the i-th particle after a time step is given by:

$$x_i(t+1) = x_i(t) + \upsilon_i(t+1) \quad (5)$$

Particle swarms have been studied in two types of neighborhood, called "gbest" and "lbest". In the gbest neighborhood, every particle is attracted to the best position found by any particle of the swarm which is called "gbest position".
In the lbest neighborhood, each (i-th) individual is affected by the best performance of its Ki immediate neighbors which is called "lbest position". The equation of velocity for gbest model is:

$$u_i(t+1) = w*u_i(t) + c_1 rand(t)* \left[p_i(t) - x_i(t)\right] + c_2 rand(t)*\left[g(t) - x_i(t)\right] \quad (6)$$

Where, w = inertia weight (0.0 - 0.1), c1 and c2 are cognitive coefficient, and social coefficient respectively, and rand(t) is a function that generates random numbers from a uniform distribution between 0.0 and 1.0. The equation of velocity for lbest model is:

$$u_i(t+1) = w*u_i(t) + c_1 rand(t)* \left[p_i(t) - x_i(t)\right] + c_2 rand(t)*\left[l_i(t) - x_i(t)\right] \quad (7)$$

C. *Taguchi*

The development of Taguchi's method is based on orthogonal arrays (OAs) that have a profound background in statistics. Orthogonal arrays were introduced in the 1940s and have been widely used in designing experiments. They provide an efficient and systematic way to determine control parameters so that the optimal result can be found with only a few experimental runs. This section briefly reviews the fundamental concepts of OAs, such as their definition, important properties, and constructions. The procedure of Taguchi algorithm consists of five stages. These stages are the following:

1. Problem Initialization: The optimization procedure starts with the problem initialization, which includes the selection of a proper OA and the design of a suitable fitness function. The selection of an OA (E, P, L, t) mainly depends on the number of optimization parameters. Where E is the number of Experiments, P is the number of Parameters, L is the number of Levels, and t is the strength.

2. Input Parameters Designation: The input parameters need to be selected to conduct the experiments. When the OA is used, the corresponding numerical values for the levels of each input parameter should be determined. For each i_th iteration and each p_th parameter, the level difference ⟦LD⟧_pi is calculated by the following formula:

$$LD_{pi} = rr^{i-1} LD_{p1},\ p = 1,\ldots,P \quad (8)$$

Where, $LD_{p1} = \dfrac{\left(max_p - min_p A\right)}{\left(L+1\right)},\ p = 1,\ldots,P$ (9)

is the initial level difference and rr is the reduced rate. Also, max_p and min_p are respectively the upper and the lower bound of the p_th parameter.

3. Experiments Conduction and Response Table Building: The fitness function fit_{ei} for each experiment (e) can be calculated and the fitness value is converted to the signal-to-noise (S/N) ratio (η) in Taguchi's method using the following formula:

$$\eta = -20\log\log\left(Fitness\right) \quad (10)$$

A small fitness value results in a large S/N ratio. After conducting all experiments in the first iteration, the fitness values and corresponding S/N ratios are obtained and listed. The average fitness values in dB are then extracted for each parameter and each level to build the response table by applying the expression:

$$\bar{\eta}_{lpi} = \left(L/E\right) \sum_{e, OA(e,p)=l} \eta_{ei},\ p = 1,\ldots,P\ \&\ l = 1,\ldots,L \quad (11)$$

4. Optimal Level Values Identification: Finding the largest S/N ratio in each column of response table can identify the optimal level for that parameter. When the optimal levels are identified, a confirmation experiment is performed using the combination of the optimal levels identified in the response table. This confirmation test is not repetitious because the OA-based experiment is a fractional factorial experiment. The fitness value obtained from the optimal combination is regarded as the fitness value of the current iteration.

5. Optimization Range Reduction: If the results of the current iteration do not meet the termination criteria, which are discussed in the following subsection, the process is repeated in the next iteration, otherwise, the procedure is terminated.

D. *Invasive Weed & Adaptive Invasive Weed*

The Invasive Weed Optimization (IWO) is an optimization algorithm that is also proposed for Electromagnetic applications. The IWO is a numerical optimization algorithm inspired from weed colonization and it was first introduced by Mehrabian and Lucas in 2006, [5]. This optimizer can in certain instances outperform other algorithms like the particle swarm optimization (PSO) and is able to handle new electromagnetic optimization problems. The colonization behavior of weeds follows the steps bellow:

1. First, there is a set of variables that are in need of optimizing. Once these variables are selected the minimum and maximum values for these variables are set.
2. Once the variables are set, the seeds are randomly positioned in an N-dimensional problem space. Each seed position is considered to be a solution. These positions will contain a value for each variable previously set. That means N values for N variables.

3. Subsequently, each seed will grow into a plant. The fitness function returns a fitness value that represents how good the solution will be for each individual seed. Once each seed is assigned a fitness value, it is called a plant.
4. In order for a plant to produce new seeds, and how many seeds, it must meet certain fitness values. Based on the fitness value rank every plant has, it produces a number of seeds between a minimum and maximum possible number. The closer to the set variables a plant is, the more seeds it is allowed to produce.
5. The seeds created in the previous step are spread over the search space. Every new seed is distributed using random numbers for the values of its location but with the numbers whose average value equal to the parent plants location as well as varying standard deviations. The standard deviation (SD) at the present time step can be expressed by:

$$\sigma = \frac{(I_{MAX} - I)^n}{(I_{MAX})^n}(\sigma_{in} - \sigma_{fi}) + \sigma_{fi} \quad (12)$$

Where, I is the number of iterations and I_{MAX} the maximum number of iterations. σ_{in} and σ_{fi} are defined as the initial and final standard deviations respectively and n is the nonlinear modulation index.
6. Once all seeds have found a position over the search area they become plants and take fitness values and rank along with their parents. In order to keep the maximum number of plants in the colony, plants that are not fit are discarded.
7. The plants that survive produce in turn new seeds and the process is repeated until the maximum number of iterations is reached or the desired fitness achieved.

In the Adaptive IWO (ADIWO), the standard deviation σ of the dispersion of the seeds produced by a weed is a linear function of the fitness value f of this weed. Considering that the goal is the minimization of the fitness function, σ can be estimated according to the following expression:

$$\sigma = \frac{\sigma_{MAX} - \sigma_{min}}{f_{MAX} - f_{min}} f + \frac{\sigma_{min} f_{MAX} - \sigma_{MAX} f_{min}}{f_{MAX} - f_{min}} \quad (13)$$

Where, σ_{MAX} and σ_{min} are the standard deviation limits defined in the same way as in the original IWO algorithm, while f_{MAX} and f_{min} represent respectively the maximum and minimum fitness values at a certain iteration. The ADIWO algorithm has the same structure as the original IWO algorithm. The only difference lies in the calculation of σ which is performed by using (13). It is easy to realize that the best weed ($f = f_{min}$) disperses its seeds with the minimum σ ($\sigma = \sigma_{min}$), while the worst weed ($f = f_{MAX}$) disperses its seeds with the maximum σ ($\sigma = \sigma_{MAX}$). Therefore, the weeds have different behavior depending on their fitness values. As the fitness value gets closer to f_{min}, the exploration ability of the weed is reduced and thus the weed can only fine-tune its near-optimal position. On the contrary, as f gets closer to f_{MAX}, the exploration ability of the weed increases and thus the weed is capable of exploring the search space to find better positions. In this way, the exploration ability of the weed colony is maintained until the end of the optimization process. Moreover, the adaptive seed dispersion makes the ADIWO converge faster than the original IWO although it is less accurate.

III. Results

The evolutionary optimisation algorithms were applied to two cases of linear array optimisation. A uniform-amplitude excitation distribution is considered in every case. The two cases considered concern a theoretical aspect of linear array design and therefore the arrays are considered to be composed respectively of 8 (case 1) and 16 (case 2) isotropic sources. In these cases, the optimization is performed for maximum array gain G_p, $\Delta\theta_{des} = 2^o$ (downward main lobe tilting), and $g_{des} = -20\text{dB}$ (null-filling) inside a sector from 90° to 120°, which are achieved by minimizing the fitness function. Since G_p is required to be maximised without reaching any desired value, two reference values of directional gain are calculated in order to be used for comparison with G_p. These values are: (i) the maximum directional gain G_{bp} of a broadside linear array (i.e., array without main lobe tilting, $\Delta\theta_{des} = 0^o$) composed respectively of 8 (for case 1) and 16 (for case 2) isotropic sources with equal inter-element distances d and equal excitation phases, and without the requirement for null-filling, and (ii) the maximum directional gain G_{tp} of a linear array composed respectively of 8 (for case 1) and 16 (for case 2) isotropic sources with equal inter-element distances d and equal excitation phase differences between adjacent sources given by the expression

$$\Delta\phi = \frac{2\pi}{\lambda} d \sin(\Delta\theta_{des}) \quad (14)$$

where $\Delta\theta_{des} = 2^o$, and finally without the requirement for null-filling. In all the cases, the IWO algorithm is applied with $ns_{\min} = 0$, $ns_{\max} = 5$, $\sigma_{\min} = 0$, $\sigma_{\max} = 0.5$ and $\mu = 2.5$. In cases 1, where N=8, 14 parameters need to be optimised. A population of 82 weeds is used. Also, the algorithm terminates after 5,000 iterations. In cases 2, where N=16, 30 parameters must be optimized. The IWO algorithm again is using a population of 82 weeds. Due to the large number of optimisation parameters in case 2 (30 parameters), 10,000 iterations are used to complete the execution of the algorithm.

All of the optimization algorithms were applied for two different scenarios. One scenario is an antenna array with eight elements and another is with sixteen elements. The chosen total number of iterations of each case was selected so that the algorithms will be able to pick the best possible final population for each case. Each case was run 20 times for every algorithm, which is enough for an average fitness evaluation of every algorithm, except for the Taguchi algorithm which automatically selects the total number of iterations.

ISBN: 978-1-61804-318-4

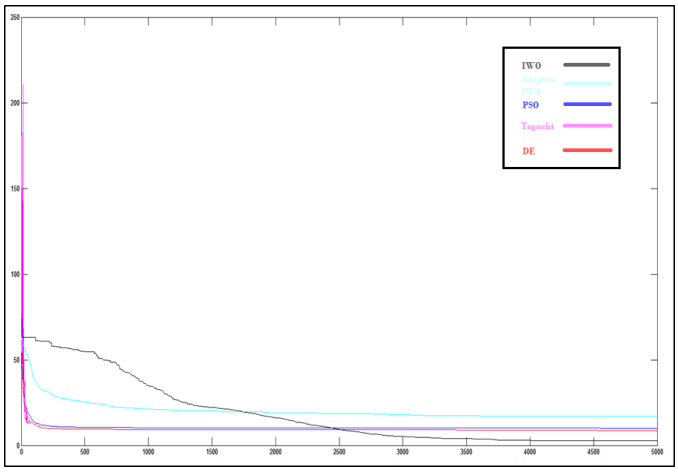


Fig. 1. Convergence (Fitness/Iterations) diagram of all the algorithms for the antenna array with 8 elements (Differential Evolution, PSO, Taguchi, IWO, Adaptive IWO).

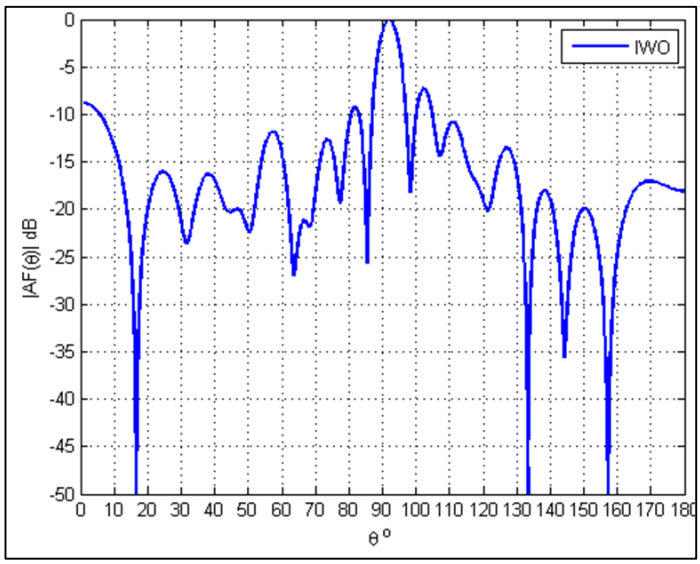


Fig. 2. Radiation Pattern of IWO optimized antenna array with 8 elements.

The target of the simulations was to maximize gain of the derived antenna (optimization variables are: dipole element distances, positions and phases) and the gain not to drop below -20dB from the peak value between the 92º and 120º azimuth angle. The fitness values per iteration for both the antenna array with eight and sixteen elements of all the algorithms are shown and a final comparison can be obtained concerning the behavior of each algorithm. The graphs depict the average convergence of the algorithms in 20 executions. In both scenarios all of the algorithms produced a radiation pattern which satisfies an antenna design with broadcasting capabilities for UHF-VHF frequencies (relative gain is higher than -20dB between 92º and 120º, no deep null). The important observation is that the best fitness is produced by the IWO algorithm. Although, the rest of the algorithms produce initial populations with better fitness values, IWO optimizes the fitness value per iteration at a slower rate compared to the rest of the algorithms, thus it needs a more computation time. These facts indicate the possibility of upgrades with a possible combination of algorithms.

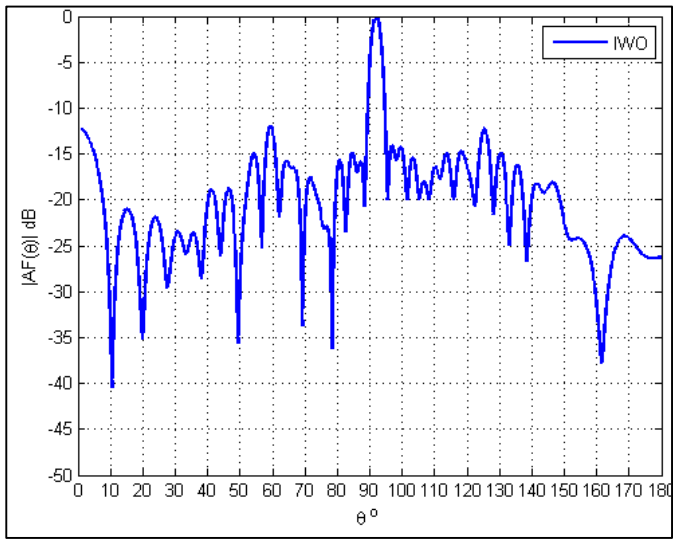


Fig. 3. Radiation Pattern of IWO optimized antenna array with 16 elements.

IV. Conclusions

Several evolutionary optimization algorithms are used in the design of an optimized broadcasting antenna array with null-filling. It is established that IWO produces the best results since it gives the lowest fitness value in comparison with the other examined algorithms. Another very important factor is the time of completion needed for every algorithm, and it is seen that improved and accelerated versions of the algorithms are required.

Acknowledgment

Parts of this work were performed within the NATO SfP-984409 project “Optimization and Rational Use of Wireless Communications Bands” (ORCA). The authors would like to thank everyone involved.

References

[1] I. Kosalay, “Estimation of RF electromagnetic levels around TV broadcast antennas using fuzzy logic,” IEEE Trans. Broadcast., vol. 56, no. 1, pp. 36–43, Mar. 2010.

[2] P. Mousavi, M. Fakharzadeh, and S. Safavi-Naeini, “1K element antenna system for mobile direct broadcasting satellite reception,” IEEE Trans. Broadcast., vol. 56, no. 3, pp. 340–349, Sept. 2010.

[3] F. J. Ares-Pena, J. A. Rodriguez-Gonzalez, E. Villanueva-Lopez, and S. R. Rengarajan, “Genetic algorithms in the design and optimization of antenna array patterns,” IEEE Trans. Antennas Propagat., vol. 47, no. 3, pp. 506–510, Mar. 1999.

[4] W. Shen and W. X. Zhang, “Pattern synthesis of non-symmetric tapered slotline antenna,” Electronics Letters, vol. 42, no. 8, pp. 443–444, Apr. 2006.

[5] A. R. Mehrabian and C. Lucas, "A novel numerical optimization algorithm inspired from weed colonization," Ecological Informatics, vol. 1, pp. 355–366, 2006.

[6] S. Karimkashi, A. A. Kishk, and D. Kajfez, "Antenna array optimization using dipole models for MIMO applications," IEEE Trans. Antennas Propagat., vol. 59, no. 8, pp. 3112–3116, Aug. 2011.

[7] N. Nemri, A. Smida, R. Ghayoula, H. Trabelsi, and A. Gharsallah, "Phase-only array beam control using a Taguchi optimization method," in Proc. 11th Mediterranean Microwave Symposium (MMS), Sept. 2011, pp. 97–100.

[8] Z. Zaharis, P. Lazaridis, J. Cosmas, C. Skeberis and T. Xenos, 'Synthesis of a Near-Optimal High-Gain Antenna Array With Main Lobe Tilting and Null Filling Using Taguchi Initialized Invasive Weed Optimization', *IEEE Trans. on Broadcast.*, vol. 60, no. 1, pp. 120-127, 2014.

[9] P. I. Lazaridis, Z. D. Zaharis, C. Skeberis, T. Xenos, E. Tziris, and P. Gallion, 'Optimal design of UHF TV band log-periodic antenna using invasive weed optimization', *Wireless VITAE conference*, Aalborg, Denmark, May 2014.

[10] Z. D. Zaharis, C. Skeberis, T. D. Xenos, P. I. Lazaridis, and D. I. Stratakis, 'IWO-based synthesis of log-periodic dipole array', *TEMU International conference*, Crete, Greece, July 2014.

[11] Z. Zaharis, C. Skeberis, P. Lazaridis, and T. Xenos, 'Optimal wideband LPDA design for efficient multimedia content delivery over emerging mobile computing systems', *IEEE Systems*, vol. PP, no. 99, Jan. 2015.

[12] Z. Zaharis, C. Skeberis, T. Xenos, P. Lazaridis and J. Cosmas, 'Design of a Novel Antenna Array Beamformer Using Neural Networks Trained by Modified Adaptive Dispersion Invasive Weed Optimization Based Data', *IEEE Trans. on Broadcast.*, vol. 59, no. 3, pp. 455-460, 2013.

[13] Z. Zaharis, D. Kampitaki, A. Papastergiou, A. Hatzigaidas, P. Lazaridis, 'Optimal design of a linear antenna array under the restriction of uniform excitation distribution using a particle swarm based optimization method',*WSEAS Trans. on Communications*, vol. 6, No.1, pp. 52-59, Jan. 2007.

[14] Z. D. Zaharis, D. G. Kampitaki, P. I. Lazaridis, A. I. Papastergiou, and P. B. Gallion, 'On the design of multifrequency dividers suitable for GSM/DCS/PCS/UMTS applications by using a particle swarm optimization-base technique,' Microw. Opt. Technol. Lett., vol. 49, no. 9, pp. 2138-2144, Sept. 2007.

Network Connection Fault Injection in Virtual Laboratory

Javier Añorga, Leonardo Valdivia, Gonzalo Solas, Saioa Arrizabalaga and Jaizki Mendizabal

***Abstract*—**As safety critical systems introduce new safety functionalities that need to be tested, fault injection techniques must be provided. This work focuses on the design and implementation of a network saboteur which injects faults at network communication level between devices working into a virtual laboratory. The combination of a virtual laboratory and the network saboteurs offers a portable and scalable testing environment reducing the money and time costs. The inclusion of the described saboteur into the EATS (ETCS Advanced Testing and Smart train positioning system) project laboratory is also detailed as a case study of the exposed work.

***Keywords*—**saboteur, safety critical, fault injection, virtual laboratory, network hacking.

I. Introduction

Critical applications can be divided into two sections depending on their impact. When the failure impacts on human beings, they are called safety-critical and they are called mission-critical if the failure only impacts finances [1]:

- Transportation, aerospace and nuclear plants are good examples of safety-critical applications.
- The Internet, bank information and Web servers are examples of mission-critical application.

At present, most of the modern systems depend on computers to operate, regardless if they are safety or mission critical. Therefore a failure in these computers can cause the loss of human lives.

Any critical system must meet two attributes: safety and availability. A safe system must behave correctly in all operating and environmental conditions. The availability ensures continuous operation of the system and it is also correlated with the capacity to restore after a failure [2]. Railways are one example of a safety-critical application, since a failure has a relevant impact on human beings. In Europe, the European Train Control System (ETCS) ensures safety in railways. As the International Union on Railways [3] defines, "ETCS is a signaling, control and train protection system designed to replace the many incompatible safety systems currently used by European railways, especially on high-speed lines".

Computers and embedded systems employed in railway applications often incorporate redundancy to tolerate faults that would otherwise cause a system failure. A fault tolerant computer system's dependability must be validated to ensure that its redundancy has been correctly implemented and the system will provide the desired level of reliable service. Fault injection (the deliberate insertion of faults) into an operational system to determine its response offers an effective solution to this problem.

In railway systems, the devices conforming the on-board equipment usually operate using a networked communication. Some fault injection can be done at this network communication level by applying fault injection on the packets transmitted in the network and observe the response of the system. If a network access point exists that allows inserting the intended faults, applying a network level fault injection seems to be an interesting and useful technique due to the fact that there is no need to apply hardware of software modifications into the original on-board devices. In addition, the inclusion of these techniques into virtual testing laboratories can increment the testing efficiency, reducing the time and cost.

This paper is organized as follows. Section II briefly reports the state of the art about safety specifications requirements, virtual laboratories and fault injection techniques. Section III details the design and implementation of the deployed network saboteur. Section IV depicts the ETCS fault injection and the ETCS Advanced Laboratory as a case study for the detailed network communication fault injection. Finally, conclusions are shown in Section V.

II. State of the Art

A. Safety

In the developing of embedded systems that perform safety-

This work was supported by the European Community's Framework Program FP7/2007-2013 in the frame of EATS project under the grant agreement nr. 31419.

Javier Añorga is with Ceit, Parque Tecnológico de San Sebastián, Paseo Mikeletegi, Nº 48, 20009, Donostia - San Sebastián (phone: +34 943 212800 ext. 2983; fax: +34 943 213076; e-mail: jabenito@ceit.es).

Leonardo Valdivia is with Ceit, Parque Tecnológico de San Sebastián, Paseo Mikeletegi, Nº 48, 20009, Donostia - San Sebastián (e-mail: lvaldivia@ceit.es).

Gonzalo Solas is with Ceit, Parque Tecnológico de San Sebastián, Paseo Mikeletegi, Nº 48, 20009, Donostia - San Sebastián (e-mail: gsolas@ceit.es).

Saioa Arrizabalaga is with Ceit, Parque Tecnológico de San Sebastián, Paseo Mikeletegi, Nº 48, 20009, Donostia - San Sebastián (e-mail: sarrizabalaga@ceit.es).

Jaizki Mendizabal is with Ceit, Parque Tecnológico de San Sebastián, Paseo Mikeletegi, Nº 48, 20009, Donostia - San Sebastián (e-mail: jmendizabal@ceit.es).

related functions, it is mandatory to adhere to applicable standards. These kinds of standards define the Safety Integrity Level (SIL) for safety-related functions depending on the maximum Tolerable Hazard Rate (THR) assigned to these functions.

The standards that govern the development of safety software for railway signaling systems are IEC61508 [5] and EN50129 [6]. These norms specify design and testing rules in order to attain a particular safety specification, which guarantees that the system continues to fulfill its safety requirements in case of random hardware failure. This means that the following must be considered:

- Individual failure effects.
- Independence between components.
- Detection of individual failures.
- Reaction after detection.
- Multiple failure effects.
- Defense against systematic failures.

When a system has programmable devices, it is not possible to guarantee that the system is not going to have dangerous failures. To attain the safety level, it is necessary for some of the parts that carry out the safety functions to be independent.

EN50129 [6] states that for SIL3 and SIL4 dynamic tests and analyses shall be done. The different methods for dynamic testing and analysis are listed in Table I, where HR is highly recommended and R is recommended.

Table I Dynamic Testing Analysis

Technique	SIL1	SIL2	SIL3	SIL4
Test case execution from extreme value analysis	HR	HR	HR	HR
Test case execution from error intuition	R	R	HR	HR
Test case execution from error introduction	R	R	R	R
Performance modelling	R	R	HR	HR
Equivalence classes and input partition test	R	R	HR	HR
Structure-based test	R	R	HR	HR

B. Virtual Laboratory

A virtual laboratory is defined as a testing laboratory where physical equipment devices are replaced by computational models of them [4].

Virtual laboratories lead to costs reduction in terms of space and money, as long only the computational model replaces the expensive hardware equipment, offering portability and scalability advantages.

C. Fault Injection Techniques

The functionality of most systems is usually evaluated with black box tests: an action or event is commanded through the accessible inputs of the system and the response is checked in the outputs. This type of test facilitates independence between the tests and the design, as the inputs and outputs are defined in the system requirements in the early stages of design.

Nevertheless, safety-related systems make use of fault tolerance techniques, which introduce additional functionality in the system. Due to this functionality cannot be excited by any accessible inputs and it is only exercised in the presence of internal failures, fault injection testing techniques are required.

The past 20 years of investigation in fault injection techniques can be classified into three areas: simulation, hardware and software based [5]-[6].

a) Simulation based

This type of fault injection consists of evaluating the performance of the system in the presence of faults using simulation tools [7].

b) Hardware-implemented fault injection (HWIFI)

HWIFI uses additional hardware to introduce faults. Normally, HWFI is used to test final products to test their behavior and time response against external failures.

c) Software-implemented fault injection (SWIFI)

The objective of this technique is to introduce faults in software to observe the behavior of the system. The main advantage of SWIFI is that no hardware is required to perform the test.

SWIFI approaches can be categorized on the basis of when the faults are injected: during compile-time or during runtime.

- Compile-time: The faults are introduced into the program before being loaded in the system, simulating possible faults in the programming phase.
- Runtime: This type of SWIFI does not modify the source code. Different methods are used to change the memory information or introduce faults in the communication network at runtime. *Network communication level fault injection* is a runtime technique concerned with the corruption (or message manipulation), loss or reordering of network packets at the network interface. The faults injected are based on injecting corrupt bytes or entire packets into the original raw network traffic.

III. Network Saboteur Design and Implementation

For network fault injection a new device is added to the test site, intercepting the communications with the System Under Test (SUT). The SUT can be one device or a combination of devices which response to deliberated injected faults is desired to be evaluated. This new added device is denominated as saboteur, and it will inject faults to the SUT. As Fig. 1 depicts, the saboteur acts as a “Man In The Middle” (MITM) for the connections to/from SUT, hijacking (taking control) the intended network traffic flow.

ISBN: 978-1-61804-318-4

In this type of test sites, there is a device sending the test events to the SUT to be tested. These test events are IP (Internet Protocol) packets which contain information, such as *event_time* (time of event) and *event_type* (time of event). A connection flow is dedicated for each type of events. A connection flow is determined in this work for a destination IP address and a destination TCP (Transmission Control Protocol) or UDP (User Datagram Protocol) port. When the event sending device tries to send an event to the SUT, the saboteur hijacks the original redirecting the connection flow to itself and establishes two connections:

- event sending device – saboteur: this connection is established between the device that intends to send the event to the SUT and the saboteur. The saboteur forges the SUT respond.
- saboteur – SUT: this connection is established between the saboteur and the SUT. The saboteur sends corrupted messages to the SUT.

Fig. 1 also illustrates the network connection hijacking. Once both connections are established, the saboteur reads the packet corresponding to an event and checks from its database if there are any faults to be injected for this event or at this simulation time. If there are any, a fault is injected to the SUT. The SUT response is then copied as the response sent back to the event sending device.

The use of this two side connection technique is motivated due to the necessity of injecting faults that are related to packet suppression or packet reordering. When TCP is being used as transport protocol, if a packet is dropped or reordered over the original *device – SUT* connection, the TCP control mechanism would try to resend the dropped packet. This fault injection is not intended to test the TCP connection but to test faults events sent to SUT. In this section TCP is used, but the saboteur architecture is the equivalent for the UDP case.

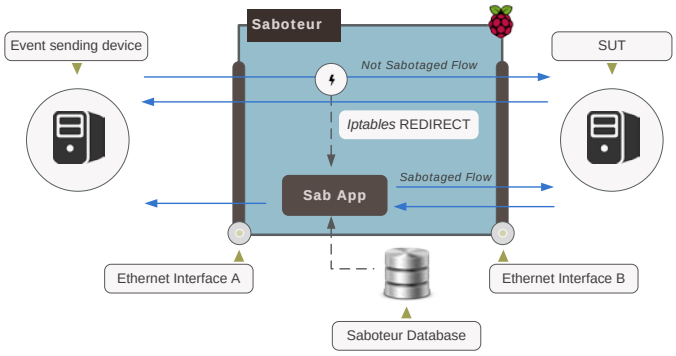


Fig. 1 Saboteur actions. Packets that match destination IP and TCP destination port are redirected to Sab App.

Next points briefly describes the core functions used in order to build the fault injection packets, the saboteur implementation and the performance impact that the saboteurs have in the communication.

A. *Saboteur core functions*

The saboteur device makes use of four core functions in order to build the fault injection packets:

a) *Event_suppression:*

This function drops a message from the original flow. Simply this event is not transmitted to the SUT.

b) Event_creation:

This function creates a new event. These new events are sent to the SUT at the end of each simulation.

c) Event_random_bytes:

This function replaces a range of contiguous bytes into the original event by random bytes.

d) Event_flip:

This function flips the *event_time* of two consecutive events. This means, interchanging the *event_time* value of two consecutive events.

B. *Implementation of the Saboteur*

The hardware used to deploy the saboteur is a Raspberry Pi 1 model B (*CPU*: *ARM11* @700MHz; *SDRAM*: 512MiB). The Raspberry Pi only has one built-in Ethernet interface, so, an USB – Ethernet adapter is added in order to get two Ethernet interfaces. The operative system working on the saboteur is *Raspbian*, which is based on the ARM hard-float (*armhf*) *Debian 7 'Wheezy'* architecture and compiled for the more limited ARMv6 instruction set of the Raspberry Pi. *Iptables* software has been used for packet management rules at saboteur [8]. *Iptables* is defined as the user-space command line program used to configure the *Linux* 2.4.x and later packet filtering ruleset.

An application written in *C* programming language (*Sab App*) is deployed and listening for incoming packets in saboteur device. *Sab App* is listening on a TCP not used by test events transmission. The saboteur device also disposes a database stored in the same device. This database is implemented with *MySQL*, and it stores the information about *iptables* rules to be set before the starting of the simulation and which faults to apply. For each fault list in database a *fault time window* (time window for a fault is intended to be injected) is used to decide which testing events are sabotaged.

For a network fault injection on a desired event transmission, *iptables* rules are configured to redirect the incoming packets that match the SUT IP as destination IP and event destination TCP port to *Sab App* (Fig. 1). As Fig. 2 depicts, when *Sab App* detects a *SYN* packet, it establishes a new connection to the SUT on the corresponding event TCP destination port (Sabotaged Flow). Once two side connection has been established, *Sab App* checks the saboteur database for each incoming event on this connection and decides if a fault injection must be done, and which core function must be used and with which arguments depending on the

ISBN: 978-1-61804-318-4

corresponding hazard. Once the fault is prepared, the corresponding packets are sent (injected) to the SUT from *Sab App*. If a fault injection must not be done for this event, simply *Sab App* replicates the event to the SUT. The response of each event from *Sab App* to the event sending device is done just replicating the SUT response received at *Sab App*. When the event sending device decides to close the connection, *Sab App* also closes its SUT connection (if there are events to be created, *Sab App* injects them just before closing the SUT connection). The packets that don't match the *iptables* redirect rule are directly sent and responded to/from the SUT (Not Sabotaged Flow).

Notice that a *Sab App* running on a saboteur is attached to an event traffic flow (defined by the IP destination address and TCP destination port); it is possible to run more than one instance of the *Sab App* attached to different event traffic flows in the same saboteur device.

Fig. 2 summarizes the saboteur fault injection workflow:

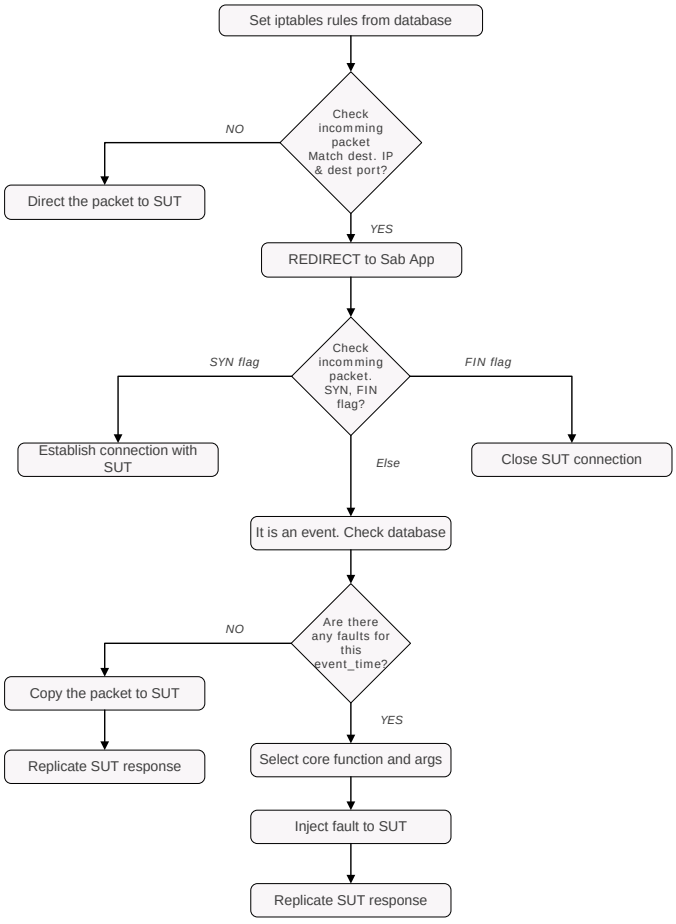


Fig. 2 Saboteur workflow.

C. Performance Analysis of the Saboteur

On a virtual laboratory based on batch simulation the additional delay that the saboteur may introduce does not affect the simulation result, only the simulation time. To quantify that the delay introduced by this process is negligible, the following test on a *Raspberry Pi* acting as a saboteur has been performed:

- A TCP connection is used between network devices.
- Events are sent through *event sending device – saboteur* connection and they are redirected to *Sab App* at saboteur device.
- *Sab App* makes queries to the saboteur database and decides if random bytes are introduced to the event packet.
- The event is sent to the SUT through the *saboteur – SUT* connection.
- The SUT reply coming from the *saboteur – SUT* connection is copied into the *event sending device – saboteur* connection.

The delay introduced by the saboteur (doing the described operations) is checked running an instance of *tcpdump Linux* application on both saboteur network interfaces. A total of 23 events are sent form the event sending device to the SUT and 10 out of these 23 events are configured to be corrupted. The average time delay introduced by the network saboteur in packets being transmitted from the event sending device to the SUT is 110.5 milliseconds. There is a negligible delay difference between packets that are finally corrupted and packets that are not, due to the most of the time spent on saboteur actions is consumed making database queries (for corrupted packets and not corrupted packets the same number of database queries are done). The average time delay introduced by the network saboteur in reply packets being transmitted from the SUT to the event sending device is 3.7 milliseconds. Reply packets are simply copied from *saboteur – SUT* connection *to saboteur – event sending device* connection and database queries are not done.

This test proofs that, especially in case of non-real time based virtual laboratory simulation, the delay introduced by network saboteurs can be considered as negligible.

IV. ETCS Fault Injection and ETCS Advanced Laboratory

Railways, just like any other safety-critical application, require a control system to ensure safety, and ETCS provides such a system for Europe. ETCS sets the norms that ensure the desired safety level, and for railways SIL4 is required. This section describes the ETCS on-board interfaces and, in particular, the Balise Transmission Module (BTM) interface; specifications that describe the reference testing architecture and fault lists are described, and the chosen testing site (which is presented as a case study of the implementation of the proposed saboteur) is briefly depicted.

A. ETCS on-board interfaces and Faults Lists

The ETCS reference architecture is shown in Fig. 3. It indicates the principal elements of ETCS on-board equipment and the reference document for each one:

1) Train Interface Unit (TIU): This is the means by which

ETCS controls the train's on-board equipment [9].

2) Driver Machine Interface (DMI): DMI is how the ETCS on-board equipment communicates with the driver, consisting of a screen placed in the driver's desk [10].

3) Specific Transmission Module (STM): This module allows reading the track signals. When the STM is used the DMI shows all the information received (for example, the brakes system) [11].

4) Balise Transmission Module (BTM): A balise is a transponder placed in the rails that sends information to the train. The module that reads that information is the balise transmission module. When the BTM reads a telegram sends the information to the European Vital Computer (EVC) [13].

5) Loop Transmission Module (LTM): This module is similar to the BTM but instead of receiving telegrams from balises the LTM reads messages from Euroloop [14].

6) Juridical Recorder Unit (JRU): It is a system that records information about the train, being possible to associate each event at a specific time [15].

7) Odometer (ODO): Odometer is the module responsible for informing about the speed and traveled distance of the train. This information is sent to the EVC to calculate the acceleration and running direction.

8) Euroradio: Euroradio transmits the position measured to the control center via GSM-R [16].

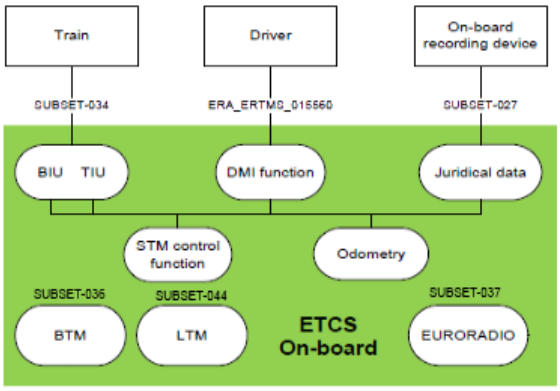


Fig. 3 ETCS reference architecture [12].

The standard interface for BTM and LTM is an air-gap, as shown in Fig. 4. However, the BTM-EVC interface is not standardized so each manufacturer uses a custom interface, mainly Ethernet (RJ-45).

A dangerous failure for the ETCS on-board equipment is defined as: *failure to provide on-board supervision and protection according to the information provided to the ETCS on-board from external entities*, and only failures that lead to the ETCS Core Hazard (exceeding the safe speed or distance information provided to ETCS) need to be considered.

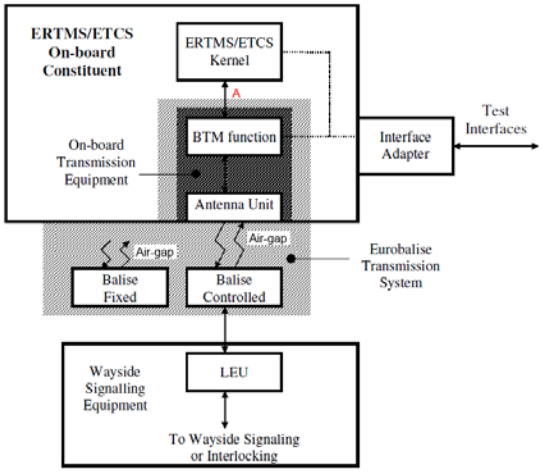


Fig. 4 Eurobalise Transmission System [11].

As a starting point, SUBSET-091 [17] has been taken into account. This specification identifies a list of failure events inside ETCS that might cause the ETCS Core Hazard to occur, either alone or in combination with other failures.

However, it does not specify the testing laboratory to be used to include this failure events. SUBSET-094 [18] specifies the reference architecture for any testing laboratory designed to test ERTMS on-board equipment. Thus, a testing site has been chosen in order to test and validate the saboteur presented in this research work. That testing site is the laboratory described in [19], which has been designed based on the reference architecture defined in [18].

B. ETCS Advanced Laboratory

EATS (ETCS Advanced Testing and Smart train positioning system) project [20] is making progress beyond the state of the art providing a model of the complete on-board ERTMS system behavior to eliminate interpretation differences. EATS project defines a virtual testing laboratory called ETCS Advanced Laboratory, using the Golden Reference Model (GRM) as reference to build and test it [19]. The GRM consists of a computational model of generic on-board equipment, designed according to the specifications in SUBSET-026.

ETCS Advanced Laboratory is composed of several building blocks and tools that have been implemented integrated and tested [19].

The communication between each of the involved sub-systems is based on Ethernet and performed by using IP addresses and TCP ports conforming connection flows as described in Section III. The application layer of each module generate the corresponding so called "EATS Events", which are serialized by using the EATS Protocol (a protocol defined and developed for the ETCS Advanced Laboratory); next, they are then encapsulated on TCP packets and sent to the destination via the TCP connection previously established.

EATS event packets structure is composed of a header section and a data section. The header section (18 bytes) consists of four fields (Table II). The data section contains

ISBN: 978-1-61804-318-4

byte fields corresponding to the applicative data of the event and it has a variable length structure.

Table II EATS event packet structure.

Bytes	Name	Description
0	*event_type*	One byte defining the type of EATS event.
1-8	*sim_time*	Eight bytes defining the time relative to simulation time.
9-10	*length*	Two bytes defining the length of the of the data section.
11-18	*event_time*	Eight bytes defining the time relative to event time as figures in journey plan. This value is used to order the EATS journey events in time.

In this testing laboratory the saboteur has been successfully integrated to test the system response to faults injected over the BTM events flow, as shown in the Fig. 5. In this case the network simulation is controlled and previously configured by the Laboratory Controller device, for example, saboteur database is filled with the information introduced at Laboratory Controller device.

In the saboteur database different faults can be configured based on the fault list identified in the SUBSET-091. Table III shows the BTM-related hazard definition and associated saboteur core functions. For this specific BTM-related hazard definition a new core function have to be defined *ad-hoc* so as to implement the *erroneous localization of a balise group*. This function has been titled as *Event_shift_balise_group* and it takes a balise group (bunch of balise event packets that correspond to the same balise group) and replaces the *event_time* value of each balise of a balise group by a random value (each balise of the balise group will contain the same randomized value).

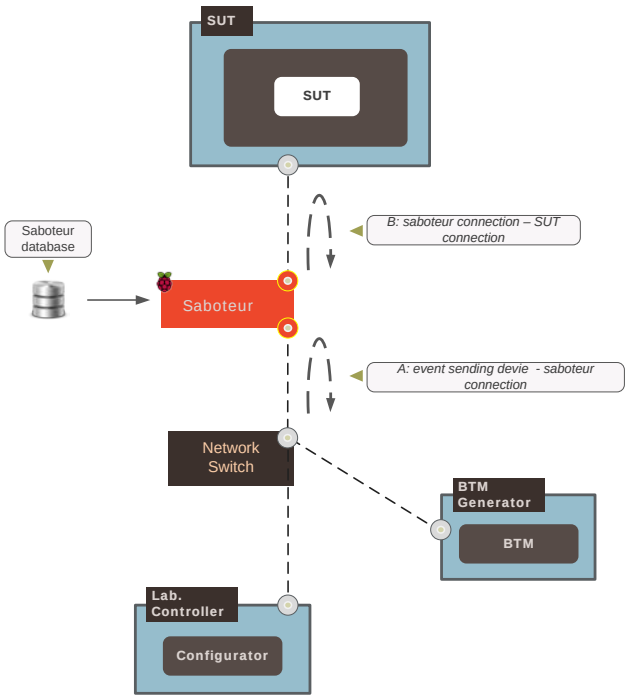


Fig. 5 Saboteur deployment into the virtual laboratory. The saboteur hijacks the *event sending device – SUT* connection and establishes two connections: *event sending device* (point A in Figure) – *saboteur*, and *saboteur – SUT* (point B in Figure).

V. Conclusion

This work has presented the integration of network communication fault injection into the EATS project, more specifically inside the ETCS Advanced Laboratory. The design and implementation of the network saboteurs has been detailed and, in order to contextualize this work, it has been previously reported a brief description of EATS project objectives, and the devices and workflow taking part of the ETCS Advanced Laboratory also has been exposed.

It can be concluded that, if the network point to include this saboteur devices is accessible (this means that network interfaces can be reconnected to other third-party hardware), this fault injection technique results to be non-invasive to the existing hardware and software under test, as there is no need for software or hardware modifications.

Table III ETCS Advanced Laboratory hazard definitions and associated saboteur core functions.

Description	Core Function
A balise group is not detected, due to failure at the onboard BTM function.	*Event_suppression*
Transmission to the on-board kernel of an erroneous telegram, interpretable as correct, due to failure at the onboard BTM function.	*Event_random_bytes*
Erroneous localization of a balise group, with reception of valid telegrams, due to failure within the on-board BTM function (erroneous threshold or excessive Tele-powering).	*Event_shift_balise_group*
The order of reported balises, with reception of valid telegrams, is erroneous due to failure within the on-board BTM function (erroneous threshold or excessive Tele-powering).	*Event_flip*
Erroneous reporting of a balise group in a different track, with reception of valid telegrams, due to failure within the on-board BTM function (erroneous threshold or excessive Tele-powering).	*Event_creation*

ISBN: 978-1-61804-318-4

As long as the implementation of the detailed network saboteur is done by the deployment of low-end market devices into the system under test, there is no need for complex and expensive equipment, resulting in cost saving.

In addition, this technique is a fast method for fault injection because it is based on simple actions on the network communication into a virtual laboratory, in contrast with more complex time-costing fault injection models. The combination of virtual laboratory testing architectures and the use of the network communication fault injection offers the testing environment scalability and portability features.

Besides, this work shows that network hacking techniques can also be used for testing safety-critical systems and conform a worthy techniques and tools to perform fault injections and test safety functionalities of the systems (not only for testing security functionalities).

Future lines of this work involve adding extra network hacking techniques to the fault injection. Some MITM techniques could be added, such as ARP poisoning techniques. These techniques would allow simply adding the saboteur to the star network architecture instead of physically deploy the saboteur in the middle of the connection link. Also this work can be extended designing and implementing network saboteur actions in case of secure network connections.

REFERENCES

[1] F. Flammini, Dependability Assurance of Real-Time Embedded Control Systems. Nova Science Publisher's, 2010.

[2] W. R. Dunn, Practical Design of Safety-Critical Computer Systems. Reliability Press, 2002.

[3] International Union Of Railways. UIC - International Union Of Railways. Available: http://www.uic.org.

[4] P. Polly, N. Marcus, D. Maguire, Z. Belinson and G. M. Velan, "Evaluation of an adaptive virtual laboratory environment using Western Blotting for diagnosis of disease," BMC Medical Education, vol. 14, pp. 222, 2014.

[5] J. Mendizabal, "Methodology & Tools for the Design & Verification of SIL4 SW Based on MDD," 2012.

[6] M. Hsueh, T. K. Tsai and R. K. Iyer, "Fault injection techniques and tools," Computer, vol. 30, pp. 75-82, 1997.

[7] J. C. Baraza, J. Gracia, D. Gil and P. J. Gil, "A prototype of a VHDL-based fault injection tool: description and application," J. Syst. Archit., vol. 47, pp. 847-867, 2002.

[8] L. Xuan and P. Wu, "The Optimization and Implementation of Iptables Rules Set on Linux," in Information Science and Control Engineering (ICISCE), 2015 2nd International Conference on, pp. 988-991, 2015.

[9] UNISIG, "Train interface FIS," Tech. Rep. SUBSET-034, 2012.

[10] European Railway Agency, "ETCS driver machine interface," 2012.

[11] UNISIG, "Specific transmission module FFFIS," Tech. Rep. SUBSET-035, 2012.

[12] UNISIG, "ERTMS/ETCS system requirements specification," Tech. Rep. SUBSET-026, 2012.

[13] UNISIG, "FFFIS for eurobalise," Tech. Rep. SUBSET-036, 2012.

[14] UNISIG, "FFFIS for euroloop," Tech. Rep. SUBSET-044, 2008.

[15] UNISIG, "FIS juridical recording," Tech. Rep. SUBSET-027, 2012.

[16] UNISIG, "Euroradio FIS," Tech. Rep. SUBSET-037, 2005.

[17] UNISIG, "Safety requirements for the technical interoperability of ETCS in levels 1 & 2," Tech. Rep. SUBSET-091, 2012.

[18] UNISIG, "Functional requirements for on-board reference test facility," Tech. Rep. SUBSET-094, 2014.

[19] G. Solas, L. J. Valdivia, J. Añorga, A. Podhorski, J. Mendizabal, S. Pinte and L. Marcos, "Virtual Laboratory for on-board ETCS equipment," in IEEE 18th International Conference on Intelligent Trasnportation Systems, Las Palmas de Gran Canaria (Spain), 2015.

[20] EATS. EATS FP7 Project. Available: http://www.eats-eu.org/.

Javier Añorga was born in Logroño, Spain, in 1987. He received his MSc degree in Telecommunications Engineering from Tecnun (School of Engineering at San Sebastián), University of Navarra, Spain, in 2011.

He joined the CEIT Research Centre in San Sebastián in 2011, and he is currently working on his PhD in CEIT's Electronics and Communications Department. His professional research activity is in the field of communication protocols, QoS, QoE, Residential Gateways, Embedded Systems and Information Technology. Currently he is also lecturer at the Engineering School of the University of Navarra (Tecnun) in San Sebastián.

Leonardo Valdivia was born in Guadalajara, Mexico, in 1988. He received a B.S. in Mechatronic Engineering from the Panamericana University in 2013 and an M.S. degree in Telecommunications Engineering from the University Of Navarra School Of Engineering (TECNUN) in 2014.

After spending three years working on software for the automotive sector, in 2013 he joined the Embedded Systems and Networks Area at CEIT. At present his research interests are focused on the design of a device that injects faults in train modules to ensure safety.

Gonzalo Solas received his MSc degree in Telecommunications Engineering from ESIDE, the Engineering Faculty of University of Deusto, in 2005. During the last year of his degree, he became a Cisco Certified Network Associate (CCNA).

He joined the Embedded Systems Group (GEMESYS), within the Electronics and Communications department, at CEIT in the fall of 2005. Currently, he is a PhD student in this group, researching mobility mechanisms for the integration of wireless sensor networks and infrastructure networks. His research interests are focused on wireless sensor networks, mobility mechanisms and wireless communication protocols. He has taken part in the elaboration of several EU funded projects and has been working on two European research projects: SUMO (Service Ubiquity in Mobile and Wireless Realm) and PERFORM (A sophisticated multiparametric system for the continuous effective assessment and monitoring of motor status in Parkinson's disease and other neurodegenerative diseases). He is author and co-author of several research papers for international conferences.

Saioa Arrizabalaga was born in Azkoitia in 1979. She received her MS degree in Telecommunication Engineering from the Faculty of Engineering in Bilbao (UPV-EHU) in 2003 and obtained her PhD degree from the University of Navarra in 2009.

She has been involved in several projects regarding remote monitoring using embedded systems, Internet access sharing, Residential Gateways or QoS management in Multi-Dwelling Units. She has published a book and several articles in international journals and conferences. Currently she is also lecturer of the Computer Architecture subject at the Engineering School of the University of Navarra (Tecnun) in San Sebastián.

Jaizki Mendizabal is a lecturer at Tecnun, the Technological Campus of University of Navarra, San Sebastián, Spain, and a researcher in the Electronics and Communications Department at CEIT. He was born in Zarautz and received his MSc and PhD degrees in Electrical Engineering from Tecnun (University of Navarra, San Sebastian, Spain) in 2000 and 2006 respectively.

He joined Fraunhofer Institut fr Integrierte Schaltungen, Erlangen (Germany) from 2000 to 2002 and SANYO Electric Ltd, in Gifu (Japan) from 2005 to 2006 as RF-IC designer. He obtained his PhD in the field of

monolithic RF design for GNSS systems. He currently works in CEIT where his research interests include RFICs and analogue safety systems for the railway industry. He has participated in more than 8 research projects, has directed 2 doctoral theses, is author or co-author of some 21 scientific and technical publications in national and international journals and conferences and is the author of the book GPS and Galileo Dual RF Front-end receiver and Design, Fabrication, Test published by McGraw-Hill.

ISBN: 978-1-61804-318-4

Adaptive Modulation and Coding for Unmanned Aerial Vehicle (UAV) Radio Channel

Amirhossein Fereidountabar,Gian Carlo Cardarilli, Rocco Fazzolari,Luca Di Nunzio

***Abstract*—**In wireless radio channels, a signal from the transmitter may arrive at the receiver antenna through several different paths. The transmitted electromagnetic wave may be reflected, diffracted, and scattered by surrounding buildings and the objects in the way of radio communications, or by troposphere and ionosphere in the case of long-distance radio communications. As a result, the signal picked up by the receiver antenna is a composite signal consisting of these multipath signals. Sometimes a line-of-sight (LOS) signal may exist. The multipath signals arrive at the receiver at slightly different delays and have different amplitudes. The different delays translate to different phases. Transmission of the signals can done by different modulations related to the application also coding for correcting errors during the transmission, applying some coding techniques are common. The goal of this paper is designing an adaptive radio link for Unmanned Aerial Vehicles (UAVs) when transmitter uses PSK family modulation with different Space Time Block Code (STBC) somehow the transmitted signal has the best quality for detection in the receiver.

***Keywords*—**Coding, Modulation, PSK, STBC

I.INTRODUCTION

Variations in the property of the propagation medium, such as the occurrence of rain or snow, also can cause fading. However, this type of fading is long-term fading, which we will not consider here. Multipath also causes inter symbol interference for digital signals.

For vehicular radio channels, there is also the Doppler frequency shift. Doppler shift causes carrier frequency drift and signal bandwidth spread. All these matters cause degradation in performance of modulation schemes in comparison with that in AWGN channels. In this paper we study performances of modulation schemes in fading channels. After that we first study flat-fading-channel performances of M-PSK, modulation scheme. Now an introduction to fading is described.

Slow fading: In a slow fading channel, the channel impulse response changes at a much slower rate than the symbol rate. The channel coherence time is much greater than the symbol duration, or equivalently, the Doppler spreading is much smaller than the signal bandwidth.

Fast fading: If the channel impulse response changes rapidly within a signal symbol duration, the channel is classified as a fast fading channel, otherwise it is classified as a slow fading channel. The fast change of the channel impulse response is caused by the motion, or equivalently, the Doppler spreading. Quantitatively when the channel coherence time is smaller than the symbol duration, or equivalent, the Doppler spreading is greater than the signal bandwidth, a signal undergoes fast fading.

Flat fading: Flat fading is also called Frequency nonselective fading. If a wireless channel has a constant gain and linear phase response over a bandwidth which is greater than the signal bandwidth, then the signal will undergo flat or frequency nonselective fading [1]-[2].

This type of fading is historically the most common fading model used in the literature. In flat fading, the multipath structure is such that the spectral characteristics of the transmitted signal is preserved at the receiver. However, the strength of the signal changes with time, due to the variation of the gain of the channel caused by multipath [3]-[4]-[5].

Frequency selective fading: If the channel has a constant gain and a linear phase response over a bandwidth which is smaller than the signal bandwidth, then the signal undergoes frequency selective fading. This is caused by such a multipath structure that the received signal contains multiple versions of the transmitted signal with different attenuations and time delays. Thus the received signal is unclear. Viewed in the frequency domain, some frequency components have greater gains than others. Frequency selective fading channels are much more difficult to model than flat fading channels. Each multipath signal must be modeled and the channel is considered as a linear filter. Models are usually developed based on wideband measurement.

Most common coding technique for error correction in flat fading channels is Alamouti Coding [5].This paper purpose is showing that in some situations(depend to SNR) Orthogonal and Quasi Orthogonal Space Time Block Codes(OSTBC and QOSTBC) have better performance[6]-[11].

In telecommunications technical differentiation relates to a method for improving the reliable transmission of a signal using two or more communication channels with different characteristics. The segregation plays an important role in combating interference thus avoiding errors. The strong fluctuation of signal strength in adverse environments can reach 20-30 dB and has even lead to the interruption of communication when is compared to received signal levels fall too low. The diversity technique is based on the fact that individual channels are characterized by different levels of interference. Multiple copy of the same signal can be transmitted from the transmitter and then be taken and attached to the receiver. Alternatively error detection code (forward error corrector) can be added so that different parts of the message to be transferred to different channels. It is important to ensure that different copies of the original signal are independent, i.e. are affected differently by the channel. The advantage of this concept is easily understood if you consider the simple case of having two versions of the signal arriving at the receiver. This idea, although very simple in understanding has been highly effective. For this reason, many different approaches have been developed differentiation. Indicatively [4]-[7]:

- spatial diversity
- frequency diversity
- time diversity

ISBN: 978-1-61804-318-4

- polarization diversity
- multiuser diversity

The spatial diversity, also called Antenna Diversity is a simple, efficient and widely used technique applied to reduce the negative effects of multipath fading environments from many scatterers. The diversification of space is to use multiple antennas transmitting and / or receiving stations, which are located some distance from each other that the different versions of the signal arriving at each of the receive antennas to be subject to different fading. The distance between the antennas must be such as to ensure that the different versions of the signal are uncorrelated, i.e. affected by uncorrelated manner of their arrival from the channel. Typically, this distance should be sometimes greater than the signal wavelength [6].Originally developed diversity reception techniques using multiple antennas at the receiver at distances sufficient to obtain uncorrelated signals. Systems operating with a transmitting antenna and multiple receive antennas as mentioned previously called SIMO (Single Input-Multiple Output) systems. The main disadvantage of diversity reception is that it makes the receivers more complex and more expensive. For this reason, making the diversity mainly applied to the base stations to improve the performance of the systems. The receiving stations serving hundreds or thousands of terminals, and so it is economical to add equipment to base stations to achieve diversity. Another reason why making diversity was not extended to the terminals is the lack of space [8].With these data, the technical diversity transmission emerges as an interesting alternative. The technical diversity emission developed more recently and consists in having multiple antennas transmitting at distances sufficient to transmit signals to undergo uncorrelated fading on the channel. The diversification of transmission has the advantage that by simply adding some transmitting antennas at the base station ensures diversity gain for all users. Furthermore, it has been shown that the same antenna can be used for differentiation of transmission at the downlink, i.e. the communication base station to the terminals, and for diversity reception in the uplink, namely the communication terminal to the base station. With the differentiation time the same data is transmitted multiple times resulting errors resulting diffuse in time [9]-[10]. Finally, techniques have been developed space diversity transmitter-receiver, using multiple transmitting antennas and multiple receiving antennas simultaneously. These are the so-called MIMO (Multiple Input - Multiple Output) systems have the advantage of providing even greater diversity gain using the appropriate mechanism making. Various techniques have been proposed for these transmission systems, the technical space-time coding and spatial multiplexing techniques are essential [5]-[11]. A typical example of the latter case is the diversification of polarization (polarization diversity). It is known that some of the characteristics transmitted in wireless communications are different for waves with horizontal and vertical polarization waves. Multiple reflections between the transmitter and the receiver lead to a change in polarization of radio waves, while conveying some of the energy of the transmitted signal in orthogonal polarized wave. Because of this characteristic of multichannel radio vertical / horizontal polarized transmitted signals are also horizontal / vertical components. A very important parameter that describes the polarization diversity system is the correlation coefficient between the obtained spectra of the signals. Since the diversity bias requires use of a dual-polarized antenna only the final state necessarily leads to certain correlation signals. But studies show that the systems of multiple antennas can achieve a significant diversity gain [9]-[12]-[13].There are various diversity reception techniques used in these systems and will be presented below.

A. Method of maximal ratio (Maximum Ratio Combining - MRC)

We consider a system which takes M copies of the transmitted signal through M different routes. Assuming that r_m is the m-th received signal which is determined as follows: $r_m = a_m * s + n_m$ where n_m is the sample of AWGN. A Maximum Likelihood (ML) decoder combines the M signals are transmitted in order to find the signal that is most likely to have been transmitted. Consider a phase detection where the receiver knows the channel gain a_m .Once the noise samples are independent Gaussian random variables, the received signal is also independent Gaussian random variable for a given channel gain and transmitted signal. For this reason, the conditional probability density function of the received signal is [4]-[7]-[14]:

$$f(r_1,r_2,....,r_m \mid s,a_1,a_2,...,a_M) = \frac{1}{(\pi N_0)^{\frac{M}{2}}} \exp\{\frac{-\sum_{m=1}^{M} \mid r_m - sa_m \mid^2}{N_0}\} \quad (1)$$

where $N_0 / 2$ is the square of the standard deviation of the real and imaginary part of the complex variable Gaussian noise. To maximize this show the receiver must find the most suitable transmitted signal which minimizes the average $\sum_{m=1}^{M} \mid r_m - sa_m \mid^2$ .We note that no diversity, M = 1, the function that minimizes the above condition is $\mid r_1 - s * a_1 \mid^2$ or $\mid r - s * a \mid^2$. This

is equivalent to calculate the closest among all the possible transmitted signals. For a constellation with equal energy symbols we have, for example PSK, resulting:

$$\begin{aligned} \hat{s} &= \arg\min\left(\sum_{m=1}^{M} \mid r_m - s a_m \mid^2\right) \\ &= \arg\min\left(\left|-s\sum_{m=1}^{M} a_m r_m^{*} - \sum_{m=1}^{M} a_m^{*} r_m\right|\right) \\ &= \arg\min\left(\left|\sum_{m=1}^{M} a_m^{*} r_m - s\right|^2\right) \end{aligned} \quad (2)$$

For this reason, the ML decoding is similar to the system with no differentiation if instead of the quantity ra^* use an average of the received signals $\sum_{m=1}^{M} a_m^{*} r_m$ .Summarizing the MRC using a filter, which is the optimal receiver for each of the received signals and

using the quantity $\omega_m = a_m^*$ combines the outputs of the filters. This process is known as MRC, is effective but complicated as it requires information of all aspects of fading channels.

B. Select of better signal (Selection combining)

The receiver selects the best received signal for demodulation and sensed, in accordance with certain criteria. These criteria relate to the total received signal power, the relationship is:

$$r = \mathrm{C}\left(r_1, r_2, ..., r_m\right) \tag{3}$$

Where C represents the selection of the signal. In practice, however, these figures are difficult to control because its control implies the existence of a mechanism of assessing these parameters in each antenna. A variation of this method is the existence of a switch, which connects one of the antennas to the receiving system. Where the received signal falls below a certain threshold, the switch selects another antenna for continuing the reception.

C. Cumulative Shooting with Weight Coefficients (Gain combining)

With this method, the signal is used by the receiver is derived as a linear combination of the received signals.

$$r = \sum_{m=1}^{M} a_i \, r_i \tag{4}$$

D. Method of Equal Weight Coefficients (Equal Gain Combining)

In the method of equal weight coefficients or simple aggregate making (equal gain combining) the coefficients are chosen so that the signals from the antennas are in phase and added. Although it is less suitable, this method with in-phase detection is often an attractive solution as it does not require an estimate of the amplitude and therefore gives results less complex than the optimum MRC. However, this method limited in practice when we refer to M-PSK signals. Indeed, for signals with unequal energy symbols such as M-QAM necessarily to estimate the width of the channel and therefore in such configurations must be used MRC for best performance [11]-[13].

II. SYSTEM ARCHITECTURES

Depending on the number of antennas that are on the show but also in making a data transmission system, the system is characterized as a system of SISO, SIMO, MISO and MIMO.SISO systems are less complex than a MIMO in these systems there is a transmitting antenna and one receiving antenna. This makes it easier to predict the behavior of systems as the parameters to be taken into account is less than in MIMO in which interactions are numerous and cannot be determined without detailed studies. Therefore the next section will be described in the simplest case MIMO system with Tx = Rx = 1.

A. Space Time Coding (STBC)

The space-time coding is widely used technique in wireless communications for transmitting different copies of information from multiple antennas and the use of different versions of the same information to the receiver in such a way as to improve the system performance. The fact that the transmitted signal propagates in fading environments and thermal noise has the effect of altering the original information and any copies of this information, they arrive at the receiver, to be more accurate than others. The abundance of such signal components arriving at the receiver enables to exploit one or more copies of the original signal for accurate decoding of the signal. The space-time coding basically combines (combining) all copies of the original signal, obtained in the most appropriate manner in order to recover from them the best possible information. The space-time coding is usually denoted by a symbol table. Each series represents a time (timeslot), in which the transmitted symbol, and each column the number of transmitting antennas which send symbols for time [1, T]. The block of symbols is the set of symbols that are transmitted from all the antennas the period T. Each modulated symbol s_{ij} denotes the symbol sent at time i from antenna j. For example, the element of the second row and third column of the matrix, s_{23} ,is the symbol transmitted from the third antenna to the second time duration of the block[12].

$$\begin{array}{c} \textit{transmit}\ \text{antennas} \\ \textit{time}\ \text{slots} \end{array} \begin{pmatrix} s_{11} & \cdots & s_{1n_T} \\ \vdots & \ddots & \vdots \\ s_{T1} & \cdots & s_{Tn_T} \end{pmatrix} \tag{5}$$

The symbol transmission rate depends on the specific space-time coding. We consider that a configuration is used with data and transmitted constellation different symbols during a block. This means that during a block entering the encoder $K\,2^M$ bits, which are assigned into symbols. Rate of the code, i.e. the average number of symbols transmitted in the duration of a block is defined as transmission. Consider a MIMO transmission system transmitting antennas and receiving antennas. The transmitter antennas simultaneously transmitting symbols S_{ij} from the matrix below (transmission matrix), where each column corresponds to the transmitted symbols from all the transmitting antennas [7]-[9].

$$s = \begin{pmatrix} s_{11} & \cdots & s_{1n} \\ \vdots & \ddots & \vdots \\ s_{m1} & \cdots & s_{mn} \end{pmatrix} \textit{for}\ \mathrm{i} = 1,2,\ldots \mathrm{m} \quad \textit{and}\ \mathrm{j} = 1,2,\ldots n \tag{6}$$

With assumption that the symbols in the matrix are independent and are selected from a constellation in data transmission, depending on the configuration selected.

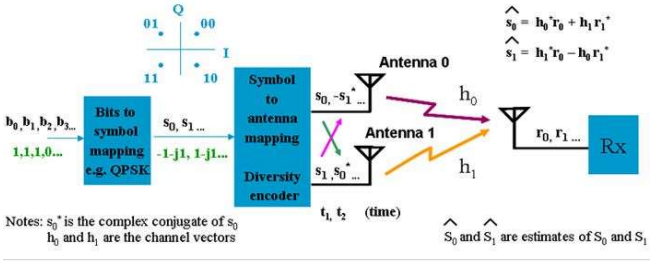


Fig.1MIMO transmission

These symbols are passed through a multipath fading channel, which is quasi-stationary i.e. varied, but slowly enough to be regarded constant during at least T moments required for transmission of all the columns of the matrix unit. The following table gives the factors for each signal multiplied table transmission, when crossing the channel.

$$H = \begin{pmatrix} h_{11} & \cdots & h_{1N} \\ \vdots & \ddots & \vdots \\ h_{M1} & \cdots & h_{Mn} \end{pmatrix} \quad (7)$$

The point is the intermittency factor (fading coefficient) between the transmitting antenna and the receiving antenna is given by the relationship: $h_{ij} = |h_{ij}| e^{j\varphi}$ where $|h_{ij}|$ and φ the amplitude and phase of the complex gain of the channel, respectively.

The data in the matrix h_{ij} should are independent. This independence is ensured by placing the antenna at sufficient distance between them.

Statistical models for fading channels mentioned above can be applied to profits h_{ij} the MIMO channel. For example, in a channel with uncorrelated Rayleigh fading each element of a channel will be an independent and identical distributed (independent and identical distributed - iid) complex Gaussian variable, while the width of $|h_{ij}|$ distribution will follow Rayleigh. The communication system is called open-loop (open-loop systems), when the receiver has full information of the channel (Channel State Information(CSI), while the transmitter has no information on his condition. The receiver that knows the coefficients h_{ij} panel of the channel at any time and may use them in decoding and demodulation and sensed. By contrast, in closed loop systems (closed-loop systems), the receiver sends back some information at the transmitter for channel through a feedback channel (feedback channel). This information is used by the transmitter to improve system performance. By doing this, of course, increases the complexity of the telecommunications system. The systems studied in this work are open loop[3]-[11]. Based on the above, the equation describing the transmission in MIMO system are:

$$Y = H * S + N \quad (8)$$

Where the matrix includes the baseband complex signal received by the receiving antennas in time:

$$Y = \begin{pmatrix} y_{11} & \cdots & y_{1N} \\ \vdots & \ddots & \vdots \\ y_{M1} & \cdots & y_{Mn} \end{pmatrix} \quad (9)$$

The maximum value that can get the transmission rate of the code is the unit (full rate). Generally the higher the transmission rate, the smaller the gain diversity, so chosen depending on the application, the appropriate code. The only code that achieves maximum diversity gain with simultaneous rate equal to the unit belongs to the class of orthogonal codes and presented extensively then.

The space-time codes are divided into two major categories:

- (Orthogonal Space Time Block Codes - OSTBCs)
- (Quasi-Orthogonal Space Time Block Codes - QOSTBCs)

In this section, OSTBCs is described.

B. Orthogonal Space-Time Codes

A rectangular space-time code is a linear code that has the following property:

$$S * S^H = \sum_{n=1}^{N} |s_n|^2 * I \quad (10)$$

Where the identity matrix and the index denote the Hermitian complex inverse. The basic property of OSTBCs is that any two columns of the matrix between the transceiver is rectangular. This means that the sequences of signals transmitted from any two antennas are orthogonal. The orthogonality property of the columns is one that gives the great advantage of orthogonal space-time codes, which is the ability of simple linear decoding at the receiver with the maximum likelihood criterion (Maximum Likelihood criterion-ML). Thus, each symbol is decoded separately at the receiver using only linear processes. To achieve linear decoding, the receiver is necessary to have full knowledge of the channel, which remains constant for the duration of a block. Another advantage is that the OSTBCs achieve maximum diversity gain, which for transmitting antennas and receiving antennas in Rayleigh fading environments has proven to be equal to $N * M$ .However, OSTBCs cannot get maximum diversity gain and maximum transmission rate together with the sole exception of the code of Alamouti (Alamouti code).

With entries $\pm s_{ij}$. Real OSTBCs that provide maximum diversity gain, maximum code rate and ML decoding are simple for n = 2, 4 and 8 antennas. Generalized real OSTBCs: The transmission matrix is a table x with real inputs 0, $\pm s_{ij}$. Generalized real OSTBCs that provide maximum diversity gain, maximum code rate and ML decoding are simple for any number of transmitting antennas. The Alamouti code for two antennas and

presented in detail in the next section. Generalized Complex OSTBCs:

It orthogonal codes whose transmission matrix is an orthogonal matrix with complex-valued inputs $\pm s_{ij}$, $\pm s_{ij}^{*}$, $\pm s_{ij} j$, $\pm s_{ij}^{*} j$. Generalized Complex OSTBCs that provide maximum diversity gain, maximum code rate and simple ML decoding does not exist.

C. Alamouti Space-Time Block Code

Assume a telecommunications system with two transmitting and one receiving antenna. Two signals are emitted simultaneously from both antennas at a given time and encoded in space-time, as shown below:

Table 1.Transmission of symbols in Alamouti STBC

antenna	**0**	**1**
Time t	s_0	s_1
Timing (t + T)	$-s_1^*$	s_0^*

The block symbols take two moments. The first time emitted the modulated symbols s_0 and s_1 and second symbols $-s_1^*$ and s_0^*, Where the "*" denotes the conjugate of a complex number.

It is considered that the channel at time t is defined fading with $h_0(t)$ for the first antenna and $h_1(t)$ for the second antenna. Assume the fading is constant during two consecutive symbols, and the duration of a symbol is obtained:

$$\begin{aligned} h_0(t) &= h_0(t+T) = h_0 = a_0 e^{j\theta_0} \\ h_1(t) &= h_1(t+T) = h_1 = a_1 e^{j\theta_1} \end{aligned} \tag{11}$$

The signals received at the time points t and (t+T) are:

$$\begin{aligned} r_0 &= r_0(t) = h_0 s_0 + h_1 s_1 + n_0 \\ r_1 &= r_1(t+T) = -h_0 s_1^* + h_1 s_0^* + n_1 \end{aligned} \tag{12}$$

With n_0 and n_1 symbolized the noise at the receiver as complex random variable. Then create the following signals to the linear receiver (combiner) and sent to the maximum likelihood detector (maximum likelihood detector):

$$\begin{aligned} \tilde{s}_0 &= h_0^* r_0 + h_1 r_1^* \\ \tilde{s}_1 &= h_1^* r_0 - h_0 r_1^* \end{aligned} \tag{13}$$

Substituting the relations for r_0 and r_1 finally obtained:

$$\begin{aligned} \tilde{s}_0 &= (a_0^2 + a_1^2) s_0 + h_0^* n_0 + h_1 n_1 \\ \tilde{s}_1 &= (a_0^2 + a_1^2) s_1 - h_0 n_1^* + h_1^* n_0 \end{aligned} \tag{14}$$

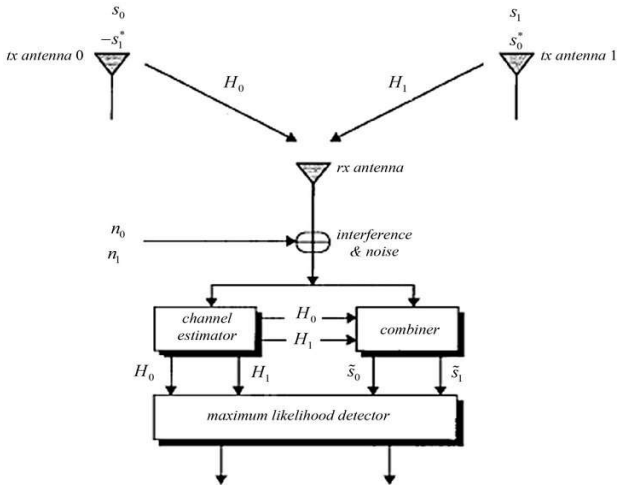


Fig 2 the transmission system with coding Alamouti space-time codes with two transmitting antennas and one receiving antenna.

D. The Alamouti Codes Scheme for 2xM

In some applications, it is desirable greater diversity gain is possible application of the Alamouti code for two transmitting antennas and M receive antennas, thus ensuring diversity gain 2xM. Below is the case of the 2x2 system, which generalizes easily to 2xM.

Define tables H and R:

$$\begin{aligned} H &= \begin{pmatrix} h_0 & h_2 \\ h_1 & h_3 \end{pmatrix} \\ R &= \begin{pmatrix} r_0 & r_2 \\ r_1 & r_3 \end{pmatrix} \end{aligned} \tag{15}$$

The element of the first row and the second column is the intermittency factor of the channel between the first transmitting antenna and the second receiving antenna. The table represents the received by the two antennas signals during the two moments which is the period of the block symbols. The lines concerning the times and the columns receiving antennas. Thus, the symbol of the second row and first column is the received symbol from the first antenna to the second time. The matrix of transmission is same as the case 2x1, i.e.:

$$S = \begin{pmatrix} s_0 & s_1 \\ -s_1^* & s_0 \end{pmatrix} \tag{16}$$

Finally, apply the received signals:

$$\begin{aligned} ro &= h_0 s_0 + h_1 s_1 + n_0 \\ r1 &= -h_o s_1^* + h_1 s_0^* + n_1 \\ r2 &= h_2 s_0 + h_3 s_1 + n_2 \\ r3 &= -h_2 s_1^* + h_3 s_0^* + n_3 \end{aligned} \tag{17}$$

Where n_o, n_1, n_2, n_3 is complex-valued random variables representing receiver noise and interference. The linear

receiver generates the following signals, which are then sent to the maximum likelihood detector:

$$\hat{s}_o = h_o^* r_o + h_1 r_1^* + h_2^* r_2 + h_3 r_3^*$$
$$\hat{s}_1 = h_1^* r_0 - h_0 r_1^* + h_3^* r_2 - h_2 r_3^* \quad (18)$$

Substituting r_0 and r_1 in the above relations follows:

$$\hat{s}_o = (a_0^2 + a_1^2 + a_2^2 + a_3^2)s_0 + h_0^* n_0 + h_1 n_1^* + h_2^* n_2 + h_3 n_3^*$$
$$\hat{s}_1 = (a_0^2 + a_1^2 + a_2^2 + a_3^2)s_1 - h_0 n_1^* + h_1^* n_0 - h_2 n_3^* + h_3^* n_2 \quad (19)$$

These signals are easily detected by the receiver maximum likelihood, and the system performance 2xM. This is shown below where given in detail the results of simulations.

III. RECEIVERS IN MIMO SYSTEMS

In most systems the complexity MIMO transmitter in terms of signal processing is low, and the bulk of the signal processing is performed at the receiver. The receptors are classified into the following two main categories:

A. Receivers Maximum Likelihood (Maximum Likelihood Detector - ML)

The maximum likelihood receivers provide better system performance (maximum diversity gain and better error probability curve), but using the most sophisticated detection algorithm (detection). The receiver calculates the maximum likelihood received signal for each of the elements of the modulation constellation that may be transmitted, knowing the channel and without taking into account the effect of noise. Then compare each received signal with each of the measured signals and calculate their distances. Then, deciding that the element of the constellation leading to the shortest distance is the signal that has been transmitted. The main disadvantage of maximum likelihood receiver is the computational complexity, which grows prohibitively for configurations with large constellation symbols.

B. Multiple Receive Antennas

Alamouti STBC can be used in MIMO communications. It benefits from additional diversity and array gain due to the presence of multiple receive antennas. However, it does not use MIMO multiplexing capabilities. Hence it is suboptimal as it does not achieve the highest possible throughput. The treatment with multiple receive antennas is very similar to the treatment with a single receive antenna except that we now manipulate vectors.
The Alamouti STBC has full rate and full diversity. Only for two transmit antenna can a space–time block code achieve both properties (except for real valued constellations). STBC designs for more than two transmit antennas can achieve (a) full rate but not full diversity or (b) full diversity but not full rate. Alamouti STBC transmission is equivalent to a SISO channel with SNR equal to:

$$SNR = \frac{\overline{P}}{2\sigma_n^2}(|h_1|^2 + |h_2|^2) \quad (20)$$

Where $\overline{p}$ total energy is transmitted from all antennas and σ_n^2 is noise power.
The transmission rate is equal to four bits per transmission. From the slopes of the curves at high SNR, the diversity gain of the Alamouti STBC for a 2×1 MISO is equal to two while the diversity gain of the Alamouti STBC for a 2×2 MIMO system is equal to four and outage probability is lower for the 2×2 MIMO.

C. STBC for More than Two Transmit Antennas

The code rate of an STBC is the number of symbols transmitted on average over a block. If the rate is equal to 1, then the STBC has full rate. The data to be transmitted is encoded, using the same encoder, into multiple code-words, or blocks, of same duration. Multiple copies of the same block are transmitted in space and in time. The STBC spreads over a number of T block transmissions and over all transmit antennas. Hence, decoding is based jointly on T blocks at the receiver. The main assumptions associated with STBC transmission are as follows.

- Slow fading channel
- Channel is constant over the duration of the STBC (transmission of symbol block).
- Unlike Alamouti STBC, the power is not always equally distributed across the transmit antennas (unequal power allocation might be necessary to guarantee orthogonality of the STBC).

D. Orthogonal and Quasi-orthogonal Designs

For more than two transmit antennas, two classes of STBC codes can be distinguished:
The class of orthogonal STBC and the class of quasi-orthogonal STBC.

• Orthogonal STBC: The lines of the STBC matrix are orthogonal. The advantage of orthogonal STBC are: (a) they have full diversity and (b) the optimal receiver is very simple as it is a simple matched filtering. The disadvantage is that those codes do not achieve full rate, with noticeable exception of the Alamouti STBC for two transmit antennas (as well as real valued constellations).
• Quasi-orthogonal STBC: The lines of the STBC matrix are not orthogonal. The orthogonality is sacrificed for rates that are higher than the orthogonal counterpart. However, the optimal receiver is more complex (ML receiver in general).

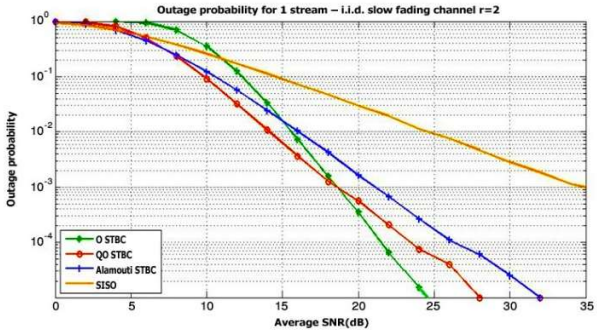
E. Comparison between Orthogonal STBC and Quasi-orthogonal STBC

The symbol error rate (SER) at the output of the receiver as a function of the SNR for the orthogonal STBC and that of the quasi-orthogonal STBC, assuming that a QPSK constellation is transmitted over an i.i.d. complex Gaussian (Rayleigh) fading channels. The SER for a SISO channel is

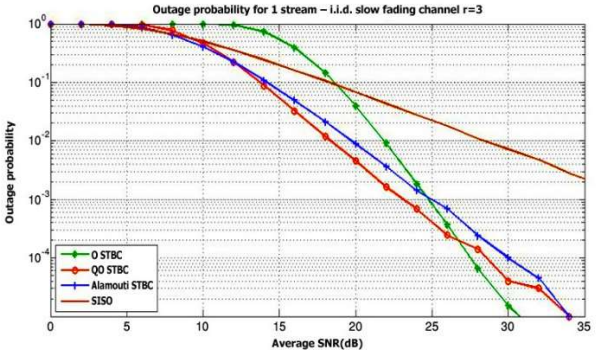
shown as a reference. QOSTBC has a worse SER than OSTBC. For a fixed input constellation, the SER of QOSTBC is degraded due to the inter-stream interference (or non-orthogonality of the QOSTBC). However, the OSTBC shown has half the rate of the QOSTBC.

IV. SIMULATION RESULTS

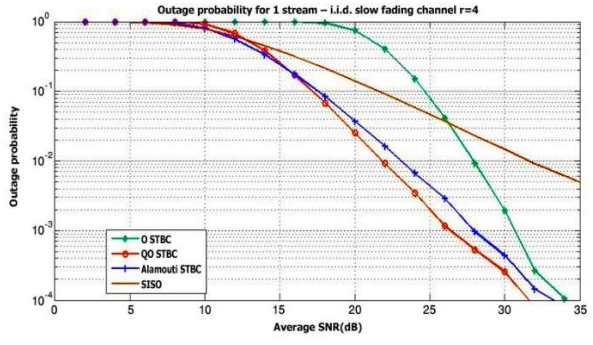
In this part, the simulation results are shown. The site (ground station) consist of some fixed objects like buildings and scaterrers around the receiver antenna. The number and location of the scaterrers are selected randomly somehow the site be similar to the reality. The simulation has done by MATLAB. The modulation is M-ary PSK. It is clear for some amounts of SNR, QOSTBC and OSTBC have better performance than Alamouti. Therefore with attention to the fig.3, depends on average SNR it is possible to change the coding method also factor (M) of the modulation for better Outage probability. For this purpose just needs that receiver with measurement of the SNR, sends a message to the UAV for changing modulation and coding scheme.



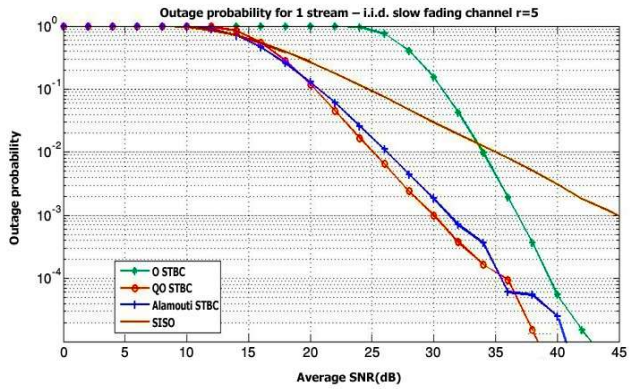
a



b



c



d

Fig 3. Outage probability of OSTBC, QOSTBC Alamouti and without coding (SISO) for a(r=2),b(r=3),c(r=4) and d(r=5) Rayleigh slow fading channel.

V.CONCLUSION

The most popular STBC is the Alamouti STBC. It is designed for a two-transmit antenna system. It is the only STBC code that achieves both full rate and full diversity (except for real constellation based STBC). Alamouti STBC minimizes the outage probability for an i.i.d. transmission.

For complex constellations and more than two transmit antennas, no STBC can be designed that achieve both full diversity and full rate. This coding method is suggested for error correction in many communication links with multipath fading [5].In this paper is showed that for some amounts of SNR,OSTBC and QOSTBC have lower outage probability and with combination with M-PSK as an adaptive method, a radio link for flat fading has better performance.

REFERENCES

[1] J.M. Torrence, L. Hanzo, "Upper bound performance of adaptive modulation in a slow Rayleigh fading channel," IEEE Electronics Letters, Vol. 32, April 1996.

[2] J. Pons, J. Dunlop, "Bit Error Rate Link Adaptation for GSM," The Ninth IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 1998, Volume: 3, Sep 1998, Page(s): 1530 -1534 vol.3.

[3] P. Bender, et al, "CDMA/HDR: A Bandwidth Efficient High Speed Wireless Data Service for Nomadic Users", Communications Magazine, IEEE, Vol. 38, No. 7, July 2000, pgs. 70-77.

[4] Andrea J. Goldsmith," Capacity of Downlink Fading Channels with Variable Rate and Power" IEEE transactions on vehicular technology, vol,4 no.3 Aug 1997.

[5] Chee Wei Tan; Dept. of Electr. Eng., Princeton Univ., Princeton, NJ, USA; Multiuser detection of Alamouti signals, Communications, IEEE Transactions on (Volume:57 , Issue: 7), Page(s):2080 – 2089, July 2009.

[6] C Evic and et.al "strategic and technology challenge for wireless communications beyond 3G" journal of communications and networks vol,4 no.4 Dec. 2003 pp:302-313.

[7] K. J. Hole, H. Holm, and G. E. Øien, "Adaptive multidimensional coded modulation over flat-fading channels," IEEE J. Select. Areas in Com., vol. 18, pp: 1153–1158, July 2000.

[8] S. T. Chung and A. J. Goldsmith, "Degrees of freedom in adaptive modulation: a unified view," IEEE Trans. Com., vol. 49, pp. 1561–1571, Sept. 2001.

[9] S. Hu and A. Duel-Hallen, "Combined adaptive modulation and transmitter diversity using long-range prediction for flat-fading mobile radio channels," in Proc. Global Telecommunications Conf., vol. 2, San Antonio, TX, Nov. 25–29, 2001, pp: 1256–1261.

[10] G. E. Øien, H. Holm, and K. J. Hole, "Channel prediction for adaptive coded modulation in Rayleigh fading," in Proc. Eur. Signal Processing Conf., Tolouse, France, Sept. 3–6, 2002.

[11] S. Ekman, M. Sternad, and A. Ahlen, "Unbiased power prediction on broadband channel," in Proc. IEEE Vehicular Technology Conf., vol. 1, Vancouver, BC, Canada, Sept. 2002, pp: 280–284.

[12] J. Kim I,. Kim, S. Ro, D. Hong"effercts of multipath diversity on adaptive QAM in frequency selective Rayleigh fading channels" IEEE trans comm.. Letters vol. 6 ,no.9, pp:1089-1091, Sep. 2002.

[13] S. Falahati, A. Svenson, T. Ekman, M .Sternad "adaptive modulation systems for predicted wireless channels: IEEE trans.on communications,vol.52,no.2 pp :307-316 Feb 2004.

[14] Andrea Goldsmith "An adaptive modulation scheme for simultaneous voice and data transmission over fading channel. IEEE journal on selected areas in communications, vol. 17, no. 5, May 1999.

BER Performance of 802.11p in SISO, MISO, and MIMO Fading Channels

Pavel Kukolev, Aniruddha Chandra, and Aleš Prokeš
Department of Radio Electronics, Brno University of Technology, Brno 61600, Czech Republic.
E-Mail: xkukol01@stud.feec.vutbr.cz

***Abstract*—The aim of this article is to simulate bit error rate (BER) of the IEEE 802.11p standard with different numbers of transmit and receive antennas. Assuming a simple flat Rayleigh fading channel, we have investigated three different antenna settings, namely, single input single output (SISO), multiple input single output (MISO), and multiple input multiple output (MIMO). When multiple antennas are considered at the transmitter end (MISO/ MIMO), simultaneous transmission is realized through Alamouti's space-time coding. For studying the effect of channel estimation on BER both the ideal case, i.e. perfect estimation, and the case when these estimators at receiver follow the least square (LS) algorithm, are examined. Effect of coding rate on the BER is also studied. Case specific simulation models are developed using MATLAB and results for all the different cases are compared. As expected, BER performance improved when more antennas are present at transmitter/ receiver or when the code rate is low. It was also found that the penalty of LS estimation can be compensated by lowering the code rate.**

***Index Terms*—802.11p, MIMO, BER, Rayleigh fading, LS estimation, Alamouti coding.**

I. Introduction

The IEEE 802.11p standard is used in wireless communication systems where the physical layer (PHY) channel parameters are rapidly time varying [1], such as in vehicular communications. This standard is included in wireless access in vehicular environments (WAVE) and is an approved amendment to the IEEE 802.11 [2]. A spectrum of 75 MHz bandwidth with a center frequency of 5.9 GHz was allocated by the U.S. Federal Communication Commission for vehicle-to-vehicle and infrastructure-to-vehicle communications under the dedicated short range communications (DSRC) regime [3]. The 802.11p standard operates in this band and features almost similar PHY parameters specified for 802.11a. The only exception is, 802.11p uses a 10 MHz bandwidth instead of the 20 MHz band used by 802.11a [1], [3].

In our earlier work [4], we have presented a MATLAB-SIMULINK model to assess the performance of 802.11p standard. However, the PHY model in [4] was limited to single transmit and receive antenna. It is well known that the use of multiple antennas at both the transmitter and the receiver can improve communication performance [5]. This paper extends the 802.11p MATLAB model to analyze the multiple input multiple output (MIMO) case and compare it with the single input single output (SISO) situation.

Another objective of the current text is to assess the effect of code rate on the BER of the 802.11p standard. As defined in [1], there are eight possible specifications for receiver performance enhancement for the 802.11p protocol, with each one achieving different data rates [6]. In this paper we will discuss the performance with BPSK-OFDM modulation for SISO and MIMO (2×2) systems using 2 different coding rates.

The paper is structured as follows. Section 2 presents PHY for 802.11p SISO and MIMO systems. In Section 3, a flat Rayleigh channel model for MIMO systems is described. Section 4 shows the results in terms of simulated bit error rate (BER) for different enhanced receiver performance requirements. Section 5 concludes the paper.

II. IEEE 802.11p PHY

In this chapter, we describe the main PHY parameters of the IEEE 802.11p [1]. The structures of the transmitter and the receiver are presented for both the SISO and MIMO systems.

A. The 802.11p SISO

The 802.11p protocol is analog to 802.11a and PHY is based on OFDM [1], [4]. The only significant difference being, as already stated in the introduction, the 10 MHz bandwidth in 802.11p is half of the bandwidth used by the 802.11a standard. The most important parameters and modulation schemes are shown in Table I and Table II.

TABLE I
Modulation schemes for 802.11p.

Modulation	Data bits per OFDM symbol	Coded bits per OFDM symbol	Coding rate	Data rate Mbit/s
BPSK	24	48	1/2	3
BPSK	36	48	3/4	4.5
QPSK	48	96	1/2	6
QPSK	72	96	3/4	9
16-QAM	96	192	1/2	12
16-QAM	144	192	3/4	18
64-QAM	192	288	2/3	24
64-QAM	216	288	3/4	27

The 802.11p transmitter and receiver with one transmitting and one receiving antennas are shown in Fig. 1. Data is encoded to correct random error. The coding rates are 1/2 and 3/4 and the constraint length is 7. To create a rate 3/4, the rate 1/2 and puncturing with vector [1 1 0 1 1 0] was used. For achieving desirable bit error distribution after demodulation, we used interleaving to distribute transmitted bits in time and frequency.

ISBN: 978-1-61804-318-4

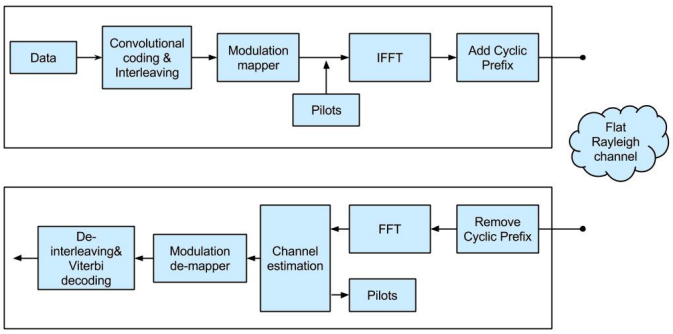


Fig. 1. The 802.11p SISO system model.

TABLE II
PARAMETERS OF 802.11P.

Data carriers	Pilot carriers	Symbol duration μs	Guard time μs	FFT period μs	Subcarrier spacing MHz
48	4	8.0	1.6	6.4	0.15625

The IFFT with $N = 64$ frequency subcarriers in a bandwidth of $Bw = 10$ MHz is used in the IEEE 802.11p. Frequency subcarrier spacing is 0.15625 MHz and is calculated as $\Delta f = Bw/N$. Data (48 subcarriers) and pilots (4 subcarriers) through orthogonal frequency subcarriers are transmitted. The 802.11p protocol uses a cyclic prefix (CP), which is added in the beginning of the signal. CP is the end part of the transmitting symbol and has a duration 1/4 of the symbol duration (6.4 μs). The duration of CP length is 1.6 μs. The final transmitting signal has a duration of 8 μs [1], [4], [7].

The receiver includes the following stages: removing CP for 16 points, FFT for 64 points, LS channel estimation and removing Pilots, modulation de-mapping, de-interleaving and Viterbi decoding using the hard decision algorithm.

The BER will attain its minimum value when perfect channel state information is available at receiver. The ideal case can be modelled easily by assigning the estimated channel coefficients to the simulated data stored in the channel matrix. The perfect estimation provides the lower bound for BER curves.

In practice these coefficients are not available a priori. To properly recover the transmitted signal in a randomly fluctuating wireless channel it is essential to estimate the channel transmission coefficient at receiver. In 802.11p standard pilot symbols are used to calculate the channel matrix. The least-square technique is widely used for channel estimation when pilot symbols are available [8]. The LS determines the difference between transmitted and received pilot symbols. In conventional comb-type pilot based channel estimation methods, the estimation of pilot signals is based on the LS method given by [9]:

$$\begin{aligned} \hat{H}_{p,LS} &= [H_{p,LS}(1)H_{p,LS}(2)\cdots H_{p,LS}(N_p)] \\ &= \left[\frac{Y_p(1)}{X_p(1)}\frac{Y_p(2)}{X_p(2)}\cdots\frac{Y_p(N_p)}{X_p(N_p)}\right] \end{aligned} \quad (1)$$

where N_p is the number of pilot symbols, Y_p are received pilot symbols, X_p are transmitted pilot symbols, and H_p is the estimated channel matrix.

B. The 802.11p MIMO

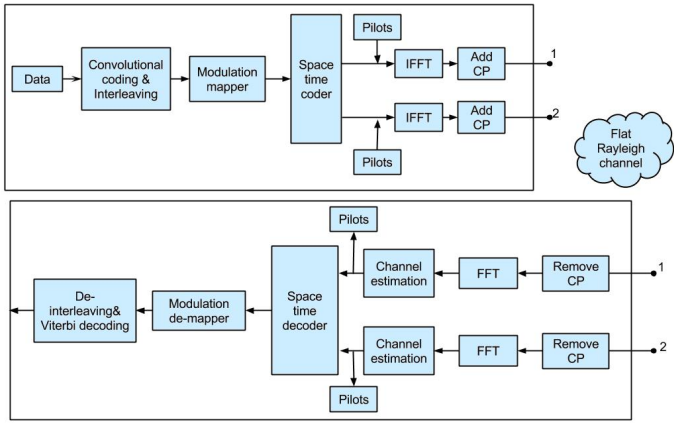


Fig. 2. The 802.11p SISO system model.

For creating the MIMO 802.11p model, Space-Time code is used. This technique allows to send data in two parts out of two transmit and two receive antennas. Probability of successfully recovering the receive signal will be higher at the expense of interleaving with transformed versions of the same information. Figure 2 describes the 802.11p MIMO system model.

III. RADIO CHANNEL

The radio channel is a part between the transmitter and receiver in wireless communication. A multipath channel is characterized by the time of the scattering signal which is affected by the changes of the signal propagation time, changes in the amplitude of the individual pulses, changing the relative delay and the number of pulses. The basic physical processes that determine the nature of the signal propagation in mobile communication systems are: reflection, diffraction and scattering [4].

The channel output can be described as [10]:

$$y_i = \sum_{j=1}^{N_t} h_{i,j} S_j + n_i \quad (2)$$

where $h_{i,j}$ complex number corresponding to the channel between antenna i and receive antenna j, N_t number from transmit antennas, S_j the channel input signal, n_i - Gaussian noise.

For the MIMO 802.11p model with two transmit and two receive antennas, and following 2×2 Alamouti coding, the received signal samples can be written [11], [12]:

$$\begin{aligned} y_{11} &= h_{11}s_1 + h_{12}s_2 + n_{11} \\ y_{12} &= h_{11}s_2^* - h_{12}s_1^* + n_{12} \end{aligned} \quad (3)$$

for the first antenna, and

$$y_{21} = h_{21}s_1 + h_{22}s_2 + n_{21}$$
$$y_{22} = h_{21}s_2^* - h_{22}s_1^* + n_{22} \quad (4)$$

for the second antenna. Symbols $*$ indicate a complex conjugate.

IV. Simulation Results

A simulation model was implemented using MATLAB 2012 and SIMULINK. The structures of transmitters and receivers correspond to blocks in Fig. 1(SISO) and Fig. 2 (MIMO). The transmitter generates the data for given energy per bit to noise power spectral density ratio (E_b/N_0) and the number of sub-frames. For getting accurate results, ten thousand data sub-frames are generated. The Simulation model includes modulation order BPSK and convolutional coding with rates 1/2 or 3/4. The generated data is transmitted over the radio channel. The channel estimator at receiver side follows LS estimation algorithm to retrieve the original transmitted signal. Finally, the signal is decoded by a hard decision Viterbi decoder.

For the implementation of simulation models have been developed simulation program Matlab. SISO and MIMO models were developed based on the software package Matlab functions with the necessary requirements to 802.11p protocol described in Section 2 and shown in Tab. III. Three channel models are implemented: a 2-tap flat Rayleigh SISO, MISO (2×1) and MIMO (2×2), as discussed in Section 3.

The outputs of the simulations are a set of BER vs. E_b/N_0 curves. The curve sets are plotted using a logarithmic scale for BER while a linear scale was used for the signal to noise ratio (E_b/N_0) as it is already in dB unit. Simple binary antipodal scheme, i.e. BPSK, is considered as the modulation scheme. We have compared MIMO (2×2), MISO (2×1), and SISO systems for different combinations of estimation techniques (perfect and LS) and coding rates (1/2 and 3/4). Details of the simulation parameters are enlisted in Table III.

TABLE III
Simulation Parameters

Parameters	Values
Bandwidth, MHz	10
FFT size	64
Number of data subcarriers	48
CP length	16
Number of pilots	4
Subcarrier spacing, MHz	0.15625
Channel model	2-tap Flat Rayleigh
Transmitting setting	SISO, MISO (2×1), MIMO (2×2)
Coding rate	1/2, 3/4
Modulation scheme	BPSK
Channel estimator	LS, Perfect

The results of simulations are shown in Fig. 3 to Fig. 8. The figures show that the best results are achieved when coding rate is 1/2, but it gives 3 Mbit/s data rate.

Figures 3, 4 and 5 present BER results for different transmitting settings and coding rates. In the three figures the margin

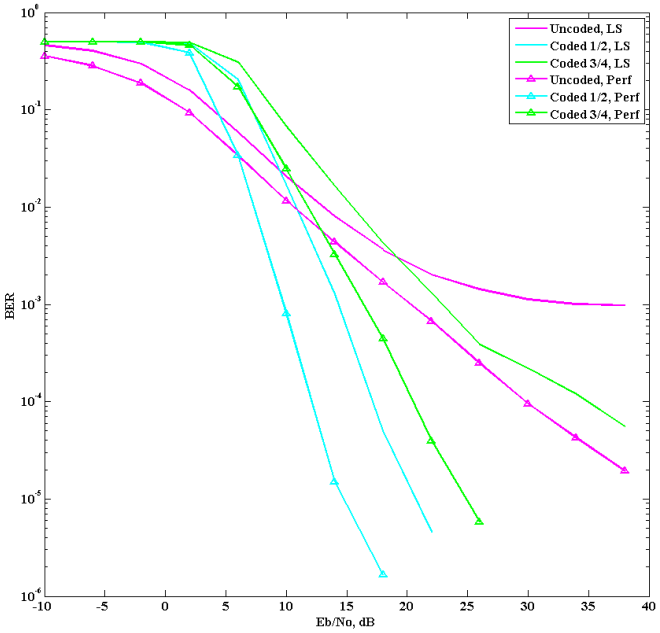


Fig. 3. BER vs E_b/N_0 of 802.11p SISO over Rayleigh channel.

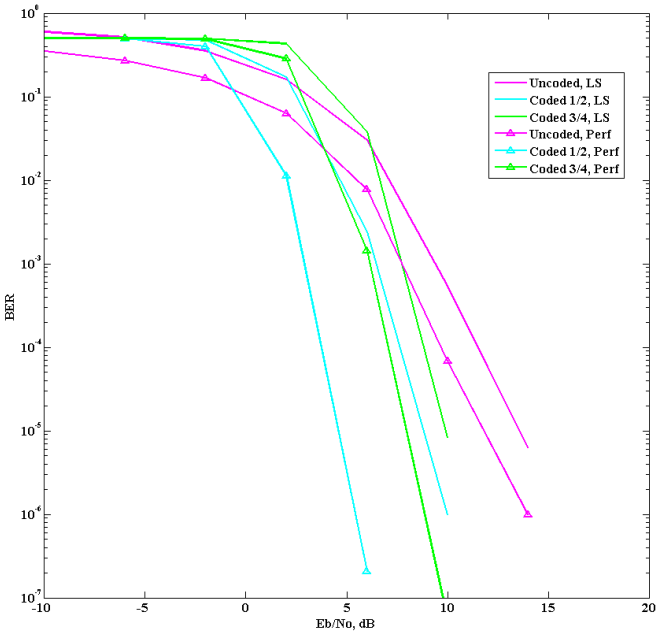


Fig. 4. BER vs E_b/N_0 of 802.11p MISO (2×1) over Rayleigh channel.

curves describe SISO, MISO and MIMO models without using convolutional coding with 2 different estimation techniques. The other curves show the performance of the simulation model with coding: blue – coding rate is 1/2, green – coding rate is 3/4.

As shown in Fig. 3, 4 and 5, results in terms of BER improve significantly with the use of coding. For example, for BER = 0.006 the required (E_b/N_0) differs by 6 dB for SISO, for BER $= 10^{-4}$ by 7-8 dB for MISO and by 4-5 dB for MIMO with LS estimation technique. The scheme with coding rate 1/2

ISBN: 978-1-61804-318-4

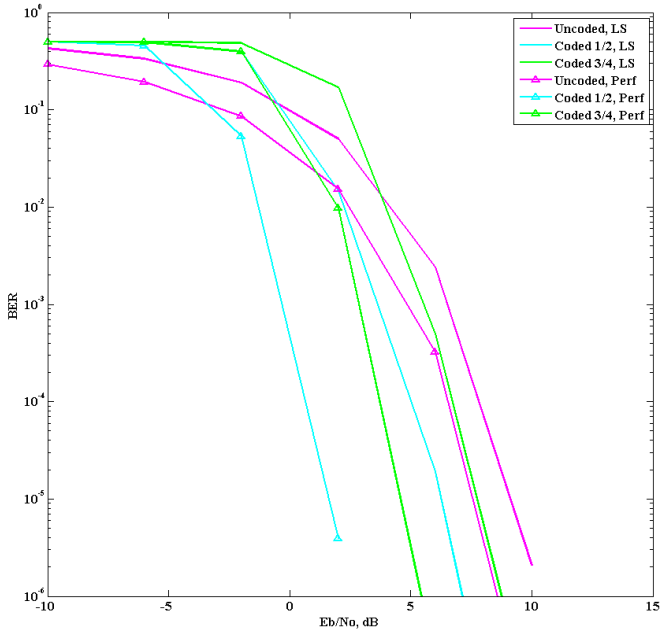


Fig. 5. BER vs E_b/N_0 of 802.11p MIMO (2×2) over Rayleigh channel.

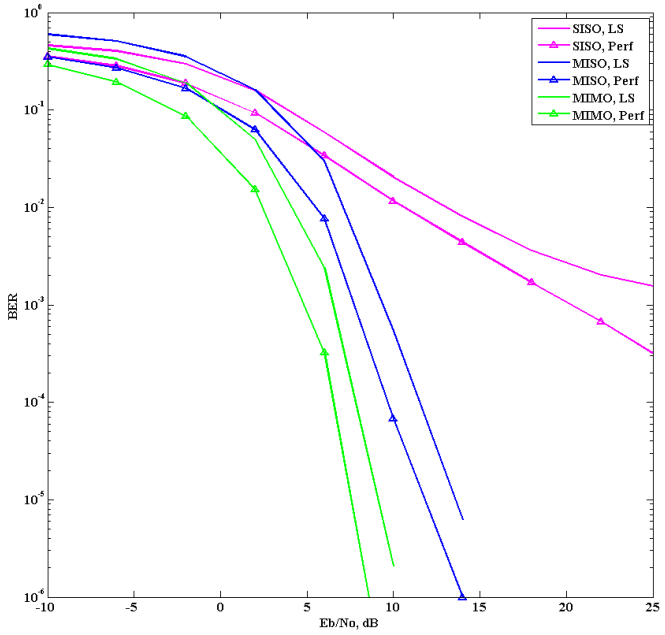


Fig. 6. BER vs E_b/N_0 of uncoded 802.11p over Rayleigh channel.

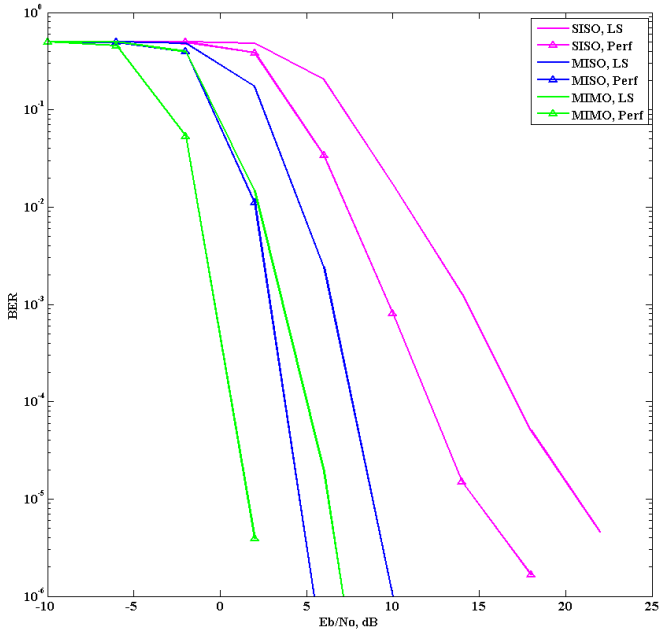


Fig. 7. BER vs E_b/N_0 of 802.11p with coding rate 1/2 over Rayleigh channel.

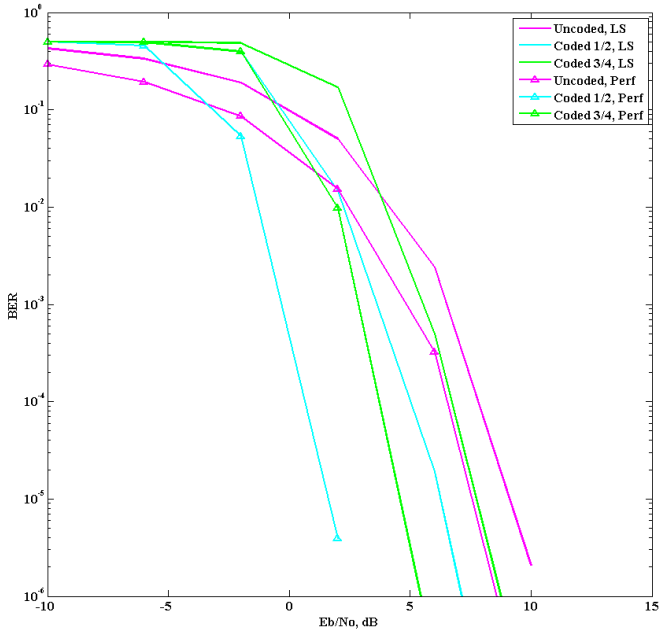


Fig. 8. BER vs E_b/N_0 of 802.11p with coding rate 3/4 over Rayleigh channel.

provides better results. However, in the scheme with coding rate 1/2 data rate is lower than in the scheme with coding rate 3/4. Obviously, the BER results for the Perfect channel estimation results are superior to LS estimation. It may be noted here that the curves for rate 3/4 coding with perfect estimation are quite close to that obtained with rate 1/2 and LS estimation.

Figures 6, 7 and 8 show the comparison of performances of the SISO, MISO and MIMO transmitting settings. Curves illustrate BER results with different coding parameters. In terms of BER, the scenario with 2 transmit antennas and 2 receive antennas (MIMO) slightly overcomes the performance of the 1 or 2 transmit antennas and 1 receive antenna for all coding settings.

Figure 6 compares the SISO, MISO and MIMO scenarios without coding. Increasing the number of transmit and receive antennas increases transmission performance, for example for (E_b/N_0) equals to 12dB the BER improves by a factor of 10.

The BER results for coding rate 1/2 are presented in Fig.

ISBN: 978-1-61804-318-4

7. The MIMO model shows slightly better performance than the MISO model, for example, for BER $= 10^{-4}$ the required (E_b/N_0) differs by 1 dB. For the same target BER $= 10^{-4}$ the SISO system requires 10 dB higher (E_b/N_0) than the system using 2 transmit and 2 receive antennas.

Finally, the performance of the SISO, MISO and MIMO models with coding rate 3/4 are compared and shown in Fig. 8. Results with the MIMO model is better than MISO or SISO, for example MIMO scheme requires approximately (E_b/N_0) = 7 dB to achieve BER $= 5 \times 10^{-4}$. Whereas for the same target of BER value, for SISO (E_b/N_0) = 19 dB and for MISO (E_b/N_0) = 10 dB.

V. Conclusion

This article provides a brief description of simulation of the IEEE 802.11p physical layer using SISO, MISO, and MIMO configurations. A 2-tap flat Rayleigh fading channel model was used to simulate the BER performance of the systems under different coding rates. Further, the influence of LS channel estimation with 4 pilots and Alamouti coding for MISO and MIMO schemes are presented. The performance depends on the used coding rate and the number of transmit and receive antennas. The article demonstrates that BER is significantly better in the MIMO (2×2) system in lower E_b/N_0 than in SISO. Results show that BER in MIMO (2×2) and MISO with coding rate 1/2 is very close. In future we plan to utilise these simulation models in conjunction with real world measurement data for predicting BER performance in those measurement scenarios.

Acknowledgment

The research is financed by the Czech Science Foundation, Project No. 13-38735S and by Czech Ministry of Education in frame of National Sustainability Program under grant LO1401. For research, infrastructure of the SIX Center was used.

This work was further supported by the SoMoPro II programme, Project No. 3SGA5720 *Localization via UWB*, co-financed by the People Programme (Marie Curie action) of the Seventh Framework Programme (FP7) of EU according to the REA Grant Agreement No. 291782 and by the South-Moravian Region.

References

[1] IEEE Computer Society, "IEEE Standard for Information TechnologyTelecommunications and Information Exchange between SystemsLocal and Metropolitan Area NetworksSpecific RequirementsPart 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 6: Wireless Access in Vehicular Environments," IEEE Std 802, July 2010.

[2] P. Kukolev, "BER performance of 802.11p standard over an ITU-R multipath channel," in *Proc. Radioelektronika*, Pardubice, Czech Republic, April 2013, pp. 391–396.

[3] D. Jiang and L. Delgrossi, "IEEE 802.11 p: Towards an international standard for wireless access in vehicular environments," in *Proc. IEEE Vehicular Technology Conference*, Singapore, May 2008, pp. 2036–2040.

[4] P. Kukolev, "Comparison of 802.11a and 802. 11p over fading channels," *Elektrorevue*, vol. 4, no. 1, pp. 7–11, Apr. 2013.

[5] Y. S. Cho, J. Kim, W. Y. Yang, and C. G. Kang, *MIMO-OFDM wireless communications with MATLAB*. Singapore: John Wiley and Sons, 2010.

[6] L. Bernado, N. Czink, T. Zemen, and P. Belanovic, "Physical layer simulation results for IEEE 802.11 p using vehicular non-stationary channel model," in *Proc. IEEE International Conference on Communications Workshops*, Cape Town, South Africa, May 2010, pp. 1–5.

[7] M. Muller, "WLAN 802.11p Measurements for Vehicle to Vehicle (V2V) DSRC," Application note Rohde & Schwarz: 1 MA152_Oe, Sep. 2009.

[8] J. J. Van De Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Borjesson, "On channel estimation in OFDM systems," in *Proc. IEEE Vehicular Technology Conference*, vol. 2, Chicago, IL, USA, July 1995, pp. 815–819.

[9] P. K. Pradhan, S. K. Patra, O. Faust, and B. K. Chua, "Channel estimation algorithms for OFDM systems," *International Journal of Signal and Imaging Systems Engineering*, vol. 5, no. 4, pp. 267–273, 2012.

[10] A. Shreedhar, T. S. Joshi, A. Rukmini, and H. M. Mahesh, "Space time block coding for MIMO systems using Alamouti method with digital modulation techniques," *World Journal of Science and Technology*, vol. 1, no. 8, 2011.

[11] A. R. Trivedi, S. B. Parmar, and S. B. Bhatt, "Comparison of different MIMO system using space time coding in Rayleigh channel," *International Journal of Emerging Technology and Advanced Engineering*, vol. 2, no. 4, pp. 625–630, April 2012.

[12] H. Jafarkhani, *Space-Time Coding: Theory and Practice*. Cambridge, UK: Cambridge University Press, 2005.

Power Adaptation for Opportunistic Incremental Relaying Systems in Rayleigh Fading Channels

Nam-Soo Kim
Department of Electronic Engineering
Cheongju University
Cheongju, Korea
nskim@cju.ac.kr

Ye Hoon Lee, Dong Ho Kim
Department of Electronic and IT Media Engineering
Seoul National University of Science and Technology
Seoul, Korea
y.lee@snut.ac.kr, dongho.kim@snut.ac.kr

***Abstract*—We consider an opportunistic incremental relaying (OIR) system with a simple and practical power adaptation scheme. We propose a modified truncated channel inversion (M-TCI) to adapt the transmission power at both a source and a relay node. The channel capacity and the outage performance are analytically derived and verified by Monte Carlo simulations. It is shown that the capacity of the proposed system increases with the number of available relays, and the capacity gain is substantial particularly at low signal-to-noise ratio (SNR) regime. It is also seen that the performance gain with the proposed system is saturated as the number of relays is increased. We further examine the effect of relay location on the outage probability of the proposed system. It is interesting to note that unlike the conventional relaying system with no power adaptation, the minimum outage probability with power adaptation does not occur at the mid-location between the source and the destination.**

Keywords—Incremental Relay, Opportunistic, Power Adaptation, Rayleigh Fading.

I. INTRODUCTION

Cooperative communications have been studied to improve the system performance in wireless fading channels [1][2]. The use of multiple active relays for a cooperative relaying system was considered in [3], [4], while a single relay in [5], [6]. Recently an opportunistic incremental relaying (OIR) system, which includes an additional relay only if the source-destination channel is of an unacceptable quality, was introduced [7], [8]. The OIR system has the advantage that the diversity gain can be obtained while utilizing fewer communication resources.

Meanwhile, it has been reported that the channel capacity and system performance of a cooperative relaying system can be improved by adaptive resource allocations [9]-[15]. There are two kinds of resource allocation schemes. The first scheme is to allocate the power and/or bandwidth to each transmitting nodes of a cooperative system [9]-[12]. However, such a power/bandwidth allocation to each node requires the global channel state information (CSI). It also requires a central control unit (CCU) to assign the power and/or bandwidth to individual nodes. Consequently, the first scheme incurs high overhead when the number of nodes in the network is large.

On the other hand, the second scheme adapts the power of each transmitting node individually without CCU for simple implementation and affordable improvements in the capacity and outage performance. Power adaptation at the source node only was considered for amplify-and-forward (AF) relaying systems [13]. Whereas power adaptation at the relay node only was considered for decode-and-forward (DF) relaying systems in [14] and for AF relaying systems in [15]. It is noted that the second scheme does not require the global CSI or the CCU.

In this paper we consider a power adapted OIR system, which belongs to the second scheme, in multiple DF relay environments. For the incremental relay selection, we adopt the maximum signal-to-noise ratio (SNR) rule, because it is simple and easy to implement compared to the other selection rules [3]. We propose a modified truncated channel inversion (M-TCI) policy for adapting the transmission power of both the source and the selected incremental relay. The M-TCI policy inverts the channel fading above a certain cutoff SNR and transmits constant power below the cutoff SNR. We derive the channel capacity and the outage probability of the proposed power adapted OIR system and verified the results by Monte Carlo simulations.

The rest of this paper is organized as follows: In Sect. II, the proposed OIR system, power adaptation strategy, and transmission phases are described. The average channel capacity and the outage probability are derived in Sect. III and Sect. IV, respectively. Numerical results are shown in Sect. V to demonstrate the derived capacity and the outage probability of the proposed system.

II. SYSTEM MODEL

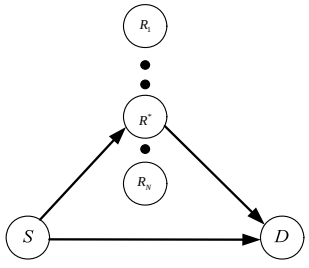


Fig. 1 Proposed OIR system model.

The OIR system considered in this paper consists of a source (S), a destination (D) and N relays (R_i, $i = 1, 2, \ldots, N$) as shown in Fig. 1. Let us denote by R^* a selected (opportunistic incremental) relay among N available relays. In our work, we assume the followings:

① The source S and the selected relay R^* know the CSI on $S-D$ path and R^*-D path, respectively. Accordingly the transmission power of S and R^* is adaptively adjusted based on each CSI. However, the S does not have the CSI on $S-R_i$ paths. Consequently, S transmits the information to R_i without power adaptation.

② Each channel has block Rayleigh fading, the CSI does not change during the information transmission; when the communication over the direct path success, the CSI constants during the direct ($S-D$ path) transmission. On the contrary, when the communication fails, the CSI constants during the indirect ($S-R^*-D$ path) transmission.

③ $S-D$ path and R_i-D paths have reciprocal CSI. D knows the CSIs on $S-D$ path and R^*-D path. D combines two signals from $S-D$ path and R^*-D path using MRC.

④ The location of the selected relay is normalized to the distance between the S and D as shown in Fig.2. We denote the distance of $S-R^*$ and R^*-D as d and $1-d$, where $0<d<1$, respectively. The received signal power at R^* and D is proportional to $d^{-\alpha}$ and $(1-d)^{-\alpha}$, respectively, where α is the path loss exponent between 3 and 4 in urban areas [16].

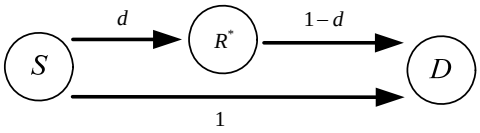


Fig. 2 Model for the relay location.

A. Power Adaptation

The applied power control technique for the proposed OIR systems is a modified form of the well-known TCI, which prevents the extreme power consumption at a deep fading. The modified TCI policy for the proposed OIR system are as follows:

When the received SNR is greater than the cutoff SNR γ_0, the transmission power is governed by the conventional TCI policy [17]. However, when the received SNR is less than the cutoff SNR γ_0, the transmitter send average transmission power to relays irrespective of the channel state. Because S do not have CSI of $S-R_i$ ($i=1,2,...,N$) paths, S transmits constant power. According to the modified TCI policy, the instantaneous transmission power $P(\gamma)$ at S subject to an average power constraint can be written by

$$P(\gamma)=\begin{cases}\dfrac{\sigma}{\gamma}\bar{P}, & \gamma \ge \gamma_0 \\ \bar{P}, & \gamma < \gamma_0\end{cases} \quad (1)$$

where γ, $\bar{P}$, γ_0, and σ denote the received SNR, average transmission power, cutoff SNR, and constant SNR which will be described in detail in next section, respectively. The instantaneous transmission power is constrained by [17]

$$\bar{P}=\int_{\gamma_0}^{\infty} P(\gamma) f_\gamma(\gamma) d\gamma \quad (2)$$

where the probability density function (pdf) $f_\gamma(\gamma)$ has an exponential distribution

$$f_\gamma(\gamma)=\frac{1}{\bar{\gamma}} e^{-\gamma/\bar{\gamma}} \quad (3)$$

where $\bar{\gamma}$ denotes average SNR. The outage probability that the received SNR is less than γ_0 is given by

$$P_0=\int_0^{\gamma_0}\frac{1}{\bar{\gamma}}e^{-\gamma/\bar{\gamma}}d\gamma = 1-e^{-\gamma_0/\bar{\gamma}} \ . \quad (4)$$

B. Transmit Protocol of the Proposed OIR System

The applied Transmit protocol of the proposed OIR system has two phases. S transmits and D, R_i are listen in phase 1. In phase 2, the selected relay R^* transmits to D. When the received SNR at D is greater than cutoff SNR (i.e., $\gamma \ge \gamma_0$) in phase 1, the transmission power of S is adaptively controlled by the TCI rule in (1). The indirect communication is not applied. However in the case of $\gamma < \gamma_0$, S transmits constant power which is average transmission power in (1). Also the OIR system selects an incremental relay. We adopt the max SNR algorithm which is simple and easy to implement compared to the other selection rules. Especially the max SNR rule does not require the global CSI for the relay selection. Consequently, it does not need a CCU which requires the heavy overhead as the increase of the number of relays in the system [19]. The max SNR rule selects the maximum SNR relay as an incremental relay R^* among N relays in $S-R_i$ paths. The index of the selected relay can be written by

$$k=\arg\max_{i=1,2,\ldots,N}\left(\gamma_{SR_i}\right). \quad (5)$$

For simplicity, we express $R_k = R^*$. If the transmit time of a relay is proportional to the inverse of the received SNR, the max SNR relay becomes the first transmitter. When a relay transmits, the other relays silent. Hence a CCU which assigns a selected relay does not required.[3]

In phase 2, the selected incremental relay decodes and forwards the received information from S. The transmission power of the selected relay is governed by the CSI of R^*-D path in (1). The signals both from the phase 1 and the phase 2 are MRC combined at D.

In practical mobile communication systems or broadcasting systems, a base station or a broadcasting station transmits a common pilot signal, whereas a power-limited mobile terminal or mobile equipment does not. The reverse

link in a mobile terminal or in mobile equipment can be a source or a mobile relay, and a base station or a broadcasting station becomes a destination. In that case, the source and the relay in the coverage area easily acquires CSI from the pilot signal of the destination. We would expect the transmission power adaptation based on the acquired CSI between the source-destination and relay-destination paths to improve the channel capacity and the system performance.

We assume independent and identically distributed (i.i.d) Rayleigh block fading and Maximal Ratio Combining (MRC) at the destination.

III. Channel Capacity of the OIR System

The channel capacity in bits/second/Hz (bps/Hz) of the OIR system in additive white Gaussian noise (AWGN) can be given by

$$C = \frac{1}{M}\log_2(1+\gamma) \tag{6}$$

where M denotes the whole time slots that require the information transfer in time division mode. And γ is the received SNR. As mentioned previous section, the transmission power of S is power controlled in phase 1. Hence the channel capacity C_{SD} of the direct path, $S-D$ path, can be written by

$$C_{SD} = \begin{cases} \log_2(1+\sigma_{SD}), & \gamma_{SD} \ge \gamma_1 \\ \log_2(1+\gamma_{SD}), & \gamma_{SD} < \gamma_1 \end{cases} \tag{7}$$

where γ_{SD} and γ_1 denote the received SNR of $S-D$ path and the cutoff SNR at D, respectively. And where σ_{SD} is the constant received SNR which can be maintained under the power constraint in (2), and satisfies

$$\int_{\gamma_1}^{\infty} \frac{P_{SD}(\gamma)}{\bar{P}_{SD}} f_{\gamma_{SD}}(\gamma_{SD}) d\gamma_{SD} = \int_{\gamma_1}^{\infty} \frac{\sigma_{SD}}{\gamma_{SD}} f_{\gamma_{SD}}(\gamma_{SD}) d\gamma_{SD} = 1. \tag{8}$$

Thus,

$$\sigma_{SD} = \frac{1}{\int_{\gamma_1}^{\infty} \frac{1}{\gamma_{SD}} f_{\gamma_{SD}}(\gamma_{SD}) d\gamma_{SD}} = \frac{\bar{\gamma}_{SD}}{E_1\left(\frac{\gamma_1}{\bar{\gamma}_{SD}}\right)} \tag{9}$$

where $P_{SD}(\gamma)$ and $\bar{P}_{SD}$ are the instantaneous transmission and the average power of S, respectively. $E_1(x)$ is the exponential integral given by [20]

$$E_1(x) = \int_1^{\infty} t^{-1} e^{-xt} dt, \quad \text{Re}\, x > 0. \tag{10}$$

While the indirect paths are composed of $S-R^*$ path and R^*-D path. Therefore an information transfer over the indirect path require two time slots ($M=2$). The power adaptation does not applied to $S-R^*$ path, hence, the channel capacity of $S-R^*$ path can be written by

$$C_{SR^*} = \frac{1}{2}\log_2(1+\gamma_{SR^*}) \tag{11}$$

where γ_{SR^*} is the received SNR in $S-R^*$ path.

In R^*-D path, however, the power control is applied similarly in $S-D$ path, and given by

$$C_{R^*D} = \begin{cases} \frac{1}{2}\log_2(1+\sigma_{R^*D}), & \gamma_{R^*D} \ge \gamma_2 \\ \frac{1}{2}\log_2(1+\gamma_{R^*D}), & \gamma_{R^*D} < \gamma_2 \end{cases} \tag{12}$$

where γ_2 denotes cutoff SNR in R^*-D path. We can obtain σ_{R^*D} by replacing γ_{R^*D} and $\bar{\gamma}_{R^*D}$ instead of γ_{SD} and $\bar{\gamma}_{SD}$ in (9).

The channel capacity after MRC at D are the sum of that of $S-D$ path and R^*-D path, ($C_{SD}+C_{R^*D}$). And the channel capacity of the indirect path is determined by the bottle neck of that of $S-R^*$ path and R^*-D path, and given by

$$C_{SR^*D} = \min\left(C_{SR^*},\ C_{SD}+C_{R^*D}\right). \tag{13}$$

Accordingly, the instantaneous channel capacity can be written by the maximum capacity between the capacity of direct path and that of indirect path:

$$C_{OIR} = \max\left(C_{SD},\ C_{SR^*D}\right). \tag{14}$$

As presented in Fig.2, the distance of $S-D$ path is normalized to 1. Then, the average received SNR of $S-D$ path becomes $\bar{\gamma}_{SD} = \bar{P}_{SD}/N_0$, where N_0 denotes the noise power and assumes equal at each receiver. Also the average SNR of $S-R^*$ path and R^*-D path can be written by $\bar{\gamma}_{SR^*} = \frac{\bar{P}_{SD}}{N_0}\frac{1}{d^{\alpha}}$ and $\bar{\gamma}_{R^*D} = \frac{\bar{P}_{SD}}{N_0}\frac{1}{(1-d)^{\alpha}}$, respectively.

IV. Outage Probability of the OIR System

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar. When the received SNR is less than the cutoff SNR, the outage is declared. Therefore the outage of the proposed OIR system is happened both of two cases: firstly, the received SNR from the direct path is less than the threshold. Secondly, minimum SNR between the selected relay and the combined SNR at D is below than the threshold, the outage probability can be written by the joint probability of [7],

$$\begin{aligned} P_{out} &= \Pr\left(\gamma_{sd} < \gamma_1,\ \gamma_{eq} < \gamma_2\right) \\ &= \int_0^{\gamma_1} f_u(u) \Pr\left(\gamma_{eq} < \gamma_2 \middle| \gamma_{SD} = u\right) du \end{aligned} \tag{15}$$

where the cutoff SNR γ_1 and γ_2 are given in (7) and (12), respectively. And we can write

$$\gamma_{eq} = \min\left\{\gamma_{SR^*},\ \gamma_{SD}+\gamma_{R^*D}\right\}, \tag{16}$$

where $\gamma_{SD}+\gamma_{R^*D}$ denote the combined SNRs of R^*-D path and $S-D$ path by MRC. While the conditional probability is written by

$$\begin{aligned} \Pr\left(\gamma_{eq} < \gamma_2 \middle| \gamma_{SD} = u\right) &= \Pr\left\{\min\left(\gamma_{SR^*},\ \gamma_{R^*D}+u\right) < \gamma_2\right\} \\ &= 1 - \Pr\left(\gamma_{SR^*} \ge \gamma_2\right)\Pr\left(\gamma_{R^*D} \ge \gamma_2 - u\right) \\ &= 1 - \left[1 - F_{SR^*}(\gamma_2)\right]\left[1 - F_{R^*D}(\gamma_2 - u)\right] \end{aligned} \tag{17}$$

ISBN: 978-1-61804-318-4

where $F_X(x)$ denotes cumulative distribution function (CDF) of the fading channel. According to the max SNR relay selection rule in (5), the CDF of the $S-R_i$ path in Rayleigh fading channel can be written by

$$F_{SR^*}(\gamma_2) = (1-e^{-\gamma_2/\overline{\gamma}_{SR^*}})^N = 1+\sum_{i=1}^{N}\binom{N}{i}(-1)^i \exp\left(-\frac{i\gamma_2}{\overline{\gamma}_{SR^*}}\right). \quad (18)$$

Replacing $F_{R^*D}(\gamma_2-u)=1-e^{-(\gamma_2-u)/\overline{\gamma}_{R^*D}}$ into (17), the outage probability in (15) can be rearranged by

$$P_{out} = 1-\mathrm{e}^{-\gamma_1/\overline{\gamma}_{SD}} + \sum_{i=1}^{N}\binom{N}{i}(-1)^i \mathrm{e}^{-\frac{i\gamma_2}{\overline{\gamma}_{SR^*}}} \frac{\overline{\gamma}_{R^*D}\,\mathrm{e}^{-\gamma_2/\overline{\gamma}_{R^*D}}}{\overline{\gamma}_{SD}-\overline{\gamma}_{R^*D}} \left\{ e^{-\gamma_1\left(\frac{1}{\overline{\gamma}_{SD}}-\frac{1}{\overline{\gamma}_{R^*D}}\right)} - 1 \right\}. \quad (19)$$

V. Numerical Results and Discussions

The channel capacity of $S-D$ path of (7) and R^*-D path of (12) are shown in Fig. 3 ($d=0.5$, $\alpha=3$, $\gamma_{SD}\ge\gamma_1$, $\gamma_{R^*D}\ge\gamma_2$). As we expected the channel capacity and the cutoff SNR are increasing with the average SNR of $S-D$ path. Also we noticed that C_{R^*D} is less than C_{SD}. We interpreted this to mean that the scaling factor 1/2 in (12) has affected the capacity.

Fig. 4 shows the average channel capacity of the OIR system, which averages the results from Monte Carlo simulation of the instantaneous channel capacity of (14). For the maximum capacity, we assume the threshold SNR γ_1 (γ_2) equals the cutoff SNR which maximizes the channel capacity of $S-D$ path (R^*-D path) in Fig. 3. In the case of the direct path only ($N=0$), the capacity coincides with that of Goldsmith at el [17]. It is noticed that the channel capacity of the OIR system, which utilizes the space diversity, always exceeds the capacity of a direct path only ($N=0$) system. The channel capacity with $N=5$ increases 93.5 %, 33.1 %, 12.1 %, and 5.8 % compared to that with the direct path only for the average SNR of 0 dB, 10 dB, 20 dB, and 30 dB, respectively. As the number of relay increases and the average SNR of $S-D$ path decreases, the improving rate of the channel capacity increases. Since the channel capacity is proportional to the logarithm of the received SNR, the substantial improvement can be expected in low SNR regions. Also the effect of the number of relays at high SNR is negligible.

Fig. 1(a) in [14] shows the channel capacity of the proposed cooperative relaying system which adapts a relay power according to the channel conditions. The average transmitting power $\overline{P}$ of 9 dB in that Fig. 1(a) is identical to the condition of $\overline{\gamma}_{SD}=0$ dB with $N=1$ in Fig. 4. Under this condition, the channel capacity of [14] is less than 1.2 bps/Hz, while it is 1.56 bps/Hz in our proposed system. We can reach about 30 % of the capacity increase.

Comparing the "Transmit SNR E_S/N_0 (dB)" of 9 dB in Fig. 2 of [13] to $\overline{\gamma}_{SD}=0$ dB with $N=1$ in Fig. 4, the transmitting power of the former is greater than 9 dB of that of a source and identical to that of a relay of the later. However, the capacity of the TCI upper bound in Fig. 2 of [13] is about 1.55 bps/Hz which is less than 1.56 bps/Hz in Fig. 4. We noticed that the channel capacity of the proposed OIR system, which adapts the transmitting power of a source and a relay, is greater than that of the system with a single node power adapting.

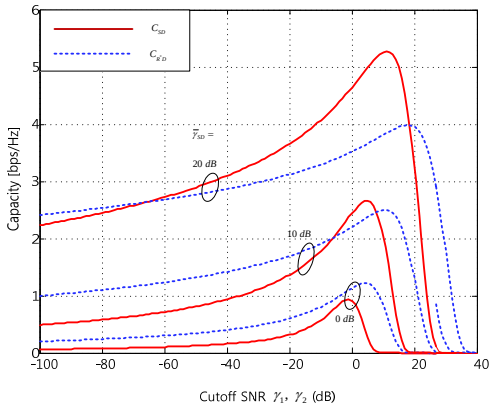


Fig. 3 Capacity of $S-D$ and R^*-D path ($\gamma_{SD}\ge\gamma_1$, $\gamma_{R^*D}\ge\gamma_2$, $d=0.5$, $\alpha=3$).

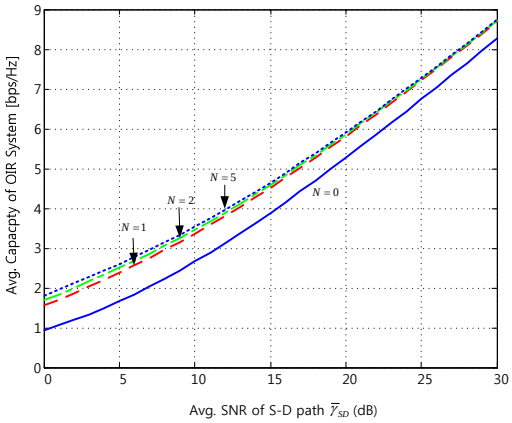


Fig. 4 Average channel capacity of the OIR system versus average SNR ($d=0.5$, $\alpha=3$, γ_1 and γ_2 at max capacity, 1×10^5 iteration).

Fig. 5 shows the outage probability of the OIR system, in which the analytic and Monte Carlo simulation results are almost identical. Especially, in the case of the direct path only ($N=0$), the outage probability is coincident with that of Alouini at el [18]. Clearly the SNR gain, which is the SNR savings to maintain the identical outage probability, is most substantial when going from $N=0$ to $N=1$, and decreases with the number of relays.

ISBN: 978-1-61804-318-4

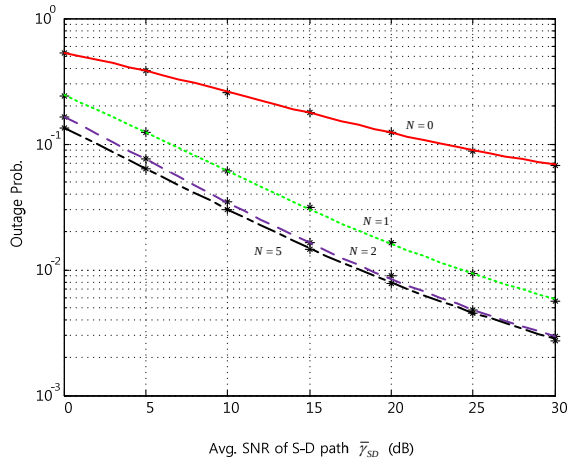


Fig. 5 Outage probability versus average SNR of S-D path ($d = 0.5$, $\alpha = 3$, γ_1 and γ_2 at max capacity).

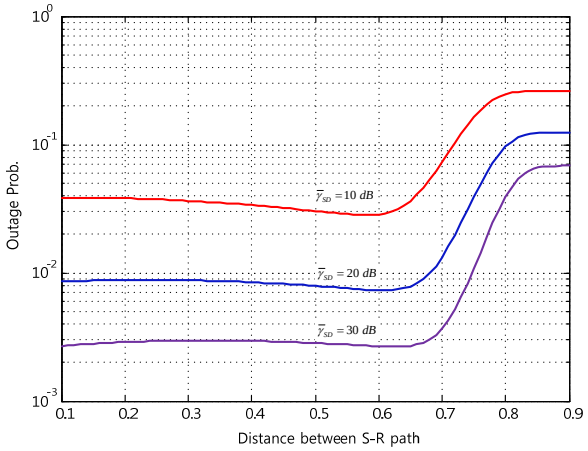


Fig. 6 Outage probability of OIR system versus distance ($N = 4$, $\alpha = 0.3$, γ_1 *and* γ_2 *at* max *capacity*)

Fig. 6 shows the outage probability of OIR system versus distance with the maximum cutoff SNR γ_1 and γ_2 in Fig. 6. A conventional OIR system which does not apply the power adaptation has the minimum outage probability at the center location ($d = 0.5$) between a source and a destination. Also the outage probability is symmetrically centered on the mid point [4]. However, the power control for the indirect path is applied on $R^* - D$ path only, and not on $S - R^*$ path, the minimum outage probability has not only occurred at the mid location but also symmetrically centered on that location. In Fig. 6, up to $d = 0.6$, there is slight variation of the outage probability, but the changes are substantial between $d = 0.6$ and $d = 0.8$. When the distance approaches $d = 0.9$, the differences are ignorable since the outage probability of $S - R^*$ path increases; this increase causes the outage probability of the indirect path, consequently the end-to-end outage probability of the OIR system is approaching to that of a direct path only. We can noticed that the outage probability of the direct path ($N = 0$) in Fig. 5 coincides with that of the different average SNR at $d = 0.9$ in Fig. 6.

REFERENCES

[1] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," IEEE Trans. Inf. Theory, vol.51, no.9, pp.3037-3063, Nov. 2005.

[2] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," IEEE Trans. on Inf. Theory, vol.50, no.12, pp.3062-3080, Dec. 2004.

[3] A. Bletsas, A. Khisti, D. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," IEEE J. Sel. Areas Commun., vol.24, no.3, pp.659-672, March 2006.

[4] A. Bletsas, H. Shin, and M. Win, "Cooperative communications with outage-optimal opportunistic relaying," IEEE Trans. Wireless Commun., vol.6, no.9, pp.3450-3460, Sep. 2007.

[5] N. C. Beaulieu and J. Hu, "A closed-form expression for the outage probability of decode-and-forward relaying in dissimilar Rayleigh fading channels," IEEE Commun. Lett., vol.10, no.12, pp.813-815, Dec. 2006.

[6] Y. Zhio, R. Adve, and T. Lim, "Outage probability at arbitrary SNR with cooperative diversity," IEEE Commun. Lett., vol.9, no.8, pp.700-702, Aug. 2005.

[7] K. Tourki, H.-C. Yang, and M.-S. Alouini, "Accurate outage analysis of incremental decode-and-forward opportunistic relaying," IEEE Trans. Wireless Commun., vol.10, no.4, pp.1021-1025, April 2011.

[8] M. Shaqfeh, F. Al-Qahtani, and H. Alnuweiri, "Optimal relay selection for decode-and-forward opportunistic relaying," Proceedings of International Conference on Communications, Signal Processing, and their Applications (ICCSPA), pp.1-4, Feb. 2013.

[9] Deniz Gunduz and Elza Erkip, "Opportunistic cooperation by dynamic resource allocation," IEEE Trans. Wireless Commun., vol.6, no.4, pp.1446-1454, April 2007.

[10] J. Luo, R. S. Blum, L. J. Cimini, L. J. Greenstein, and A. M. Haimovich, "Decode-and-forward cooperative diversity with power allocation in wireless networks," IEEE Trans. Wireless Commun., vol.6, no.3, pp.793-799, March 2007.

[11] Y.-R. Tsai and L.-C. Lin, "Optimal power allocation for decode-and-forward cooperative diversity under an outage performance constraint", IEEE Commun. Lett., vol.14, no.10, pp.945-947, Oct. 2010.

[12] I. Maric and R. Yates, "Bandwidth and power allocation for cooperative strategies in Gaussian relay networks," IEEE Trans. Inf. Theory, vol.56, no.4, pp.1880 -1889, April 2010.

[13] T. Nechiporenko, K. T. Phan, C. Tellambura, and H. Nguyen, "On the capacity of Rayleigh fading cooperative systems under adaptive transmission," IEEE Trans. Wireless Commun., vol.8, no.4, pp.1626-1631, April 2009.

[14] M. Chraiti, W. Ajib, and J.-F. Frigon, "Optimal long-term power adaption for cooperative DF relaying," IEEE Wireless Commun. Lett., vol.3, no.2, pp.201-204, April 2014.

[15] L. J. Rodríguez, N. H. Tran, A. Helmy, and T. Le-Ngoc, "Optimal power adaptation for cooperative AF relaying with channel side information," IEEE Trans. Veh. Technol., vol.62, no.7, pp.3164-3174, Sep. 2013.

[16] G. L. Stüber, Principles of Mobile Communication, 2nd Ed., Kluwer Academic Publishers, 2001.

[17] A. J. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," IEEE Trans. Inf. Theory, vol.43, no.6, pp.1986-1992, Nov. 1997.

[18] M.-S Alouini and A. J. Goldsmith, "Capacity of Rayleigh fading channels under different adaptive transmission and diversity-combining techniques," IEEE Trans. Veh. Technol., vol.48, no.4, pp.1165-1181, July 1999.

[19] M. Benjillali and M.-S Alouini, "Outage performance of decode-and-forward partial selection in Nakagami-m fading channels," Proceedings of International Conference on Telecommunications, pp. 71-76, April 2010.

[20] I. S. Gradshteyn and I. M. Ryzhik, Table of Integrals, Series, and Products, 6th Ed., pp.XXXV, Academic Press, 2000

Extended Filtering for Self-Localization over RFID Tag Grid Excess Channels – I

Moises Granados-Cruz, Yuriy S. Shmaliy, and Sanowar H. Khan

***Abstract*—High accuracy is often required for mobile robot self-localization in RFID tag information networks. In such networks, vehicle state can be observed over a big number of tags. Accordingly, the extended Kalman filter (EKF) algorithm is modified and a new extended unbiased finite impulse response (EFIR) filtering algorithm is developed. We show that redundant information captured from the tags allows increasing both the localization accuracy and system stability.**

***Keywords*—RFID self-localization, tag information grid, extended unbiased FIR filter, extended Kalman filter.**

I. Introduction

THE information grids organized using radio frequency identification (RFID) tags [1] are the most modern effective localization technique developed in recent years. Such grids have drawn researchers attention [2], [3] owing to several useful features. Each tag has its own identification (ID) number corresponding to unique coordinates of location and may be either active [4] or passive [5]. In order to increase awareness, information describing a local 2D or 3D surrounding can be programmed in each tag [2] and delivered to users by request. The method is low cost and available for any purpose, provided the communication between a target and the tags. However, the accessibility of a map-library organized and saved in such a way can be appreciated only if the target (vehicle, mobile robot, etc.) self-localization is guaranteed with a sufficient accuracy [6]. Otherwise, errors may lead to collisions. The problem is complicated by the required low cost measurement which typically imply large noise.

One meets usage of the RFID tags for vehicle localization in recent decades [7]–[9]. The most developed low-cost method of tag reading utilizes the received signal strength information (RSSI). The approach implies measuring distances between a vehicle and several tags where locations are precisely known. Multilateration algorithms based on averaging over some time interval are most widely used here [10], [11]. Other algorithms can also be employed, including the directional ones [12], [13]. There were also developed various hybrid structures in which information received from the tags is combined with information received from other sources. The support vector machine is used in [14] to analyze information received from the RFID tags and make a decision about a certain robot location. A single vision camera is exploited in [15] to correct the robot movements [16]. The Global Positioning System (GPS) has been included to the scheme in [17] in order to obtain autonomous travelling capability in the design of an intelligent wheelchair. Reviews of RFID tag-based localization algorithms are given in [18] and some other useful results can be found in [19], [20].

It is known that multilateration provides a straightforward localization in an "algebraic" way. However, noise reduction here is typically inefficient and optimal estimators are required. Most frequently, various modifications are exploited of the Kalman filter (KF) [5], [21]–[23], particle filter (PF) [8], [24], [25], and unscented KF (UKF) [26]. Algorithms utilizing the extended Kalman filter (EKF) require white noise approximation as well as known noise statistics, initial conditions, and initial error statistics in order for the EKF to be suboptimal. Otherwise, accuracy provided by EKF may be low [27] and unacceptable for information grids. Another flaw is that EKF can be unstable and demonstrate divergence under the uncertainties [28] and large nonlinearities with intensive noise [29]. The problem of not exactly known noise statistics also arises in UKF, although this filter demonstrates better performance than EKF for highly nonlinear systems.

The PF is free of many disadvantages peculiar to EKF. However, PF based on the Monte Carlo approach often requires large data and time and cannot always be used in real-time localization. On the other hand, it is known that the Gauss's least squares (LS) often give accuracy that is superior to the best available EKF [30]. Thus, methods of averaging implemented in LS and finite impulse response (FIR) filters may be more preferable. So, there is still room for discussion of the best estimator for RFID tag information grids.

The FIR filter has been under the development for decades [28], [31]–[37]. It has been shown that this filter is more robust than KF under the unbounded disturbances [35]. The FIR filter is also lesser sensitive to noise [36] and produces smaller round-off errors [31] owing to averaging. Of practical importance is that complex optimal FIR (OFIR) structures [31], [32] do not demonstrate essential advantages against simple unbiased FIR (UFIR) ones [36] which ignore the noise statistics [28], [32]. The effect is due to averaging leveling the difference between OFIR and UFIR on large averaging horizons. The latter has made the UFIR filter a strong rival to the Kalman filter.

Recently, the UFIR algorithm was developed in [37] to the extended UFIR (EFIR) algorithm following the same strategy

The results of these investigations were presented at the INASE International Conference on Systems, Control, Signal Processing and Informatics (SCSI 2015), Barcelona, Spain, April 7-9, 2015. This investigation was supported by the Royal Academy of Engineering under the Newton Research Collaboration Programme NRCP/1415/140.

Moises Granados-Cruz and Yuriy S. Shmaliy are with the Department of Electronics Engineering, Universidad de Guanajuato, Mexico (e-mail: shmaliy@ugto.mx).

S. H. Khan is with the Department of Electronics Engineering, City University London, London, UK, e-mail: S.H.Khan@city.ac.uk.

ISBN: 978-1-61804-318-4

as for the Kalman filter. First applications of the EFIR filter to localization problems [38], [39] have already shown some promising results. It was revealed [38] that the EFIR filter initiated by EKF can be much more successful in accuracy and stability in the triangulation-based localization. It was also noticed [39] that the EFIR filter has much stronger protection against divergency and instability than EKF in the RFID tag grid-based localization.

Summarizing, we notice that the RFID tag environment can be constructed such that a big number of tags are observed simultaneously by the reader. The commercially available GPS receivers are able to work with 8–12 satellites at once and the ground navigation accuracy becomes higher by increasing the number of satellites in a view. A similar effect was reported in [25] regarding the RFID tags grids.

Below, we propose a novel EFIR/Kalman algorithm for target self-localization in RFID tag-nested information grids. In the second part of this paper [40], we apply the algorithms to mobile robot self-localization over the RFID tag excess channels. Thereby we learn effect of redundant information captured from the tags on the localization accuracy.

II. Mobile Robot Model

In order to achieve highest localization accuracy in RFID tag information grids, we combine information captured from the tags with measurements of the target heading angle Φ obtained using a fiber optic gyroscope (FOG) [41]. A detailed diagram of such a scheme is sketched in Fig. 1. A vehicle travels in direction d and its trajectory is controlled by the left and right wheels. The incremental distances vehicle travels by these wheels are $d_{\rm L}$ and $d_{\rm R}$, respectively. The distance between the left and right wheels is b and the stabilized wheel is not shown. A vehicle moves in its own planar Cartesian coordinates $(x_{\rm r}, y_{\rm r})$ with a center at $M(x, y)$; that is, the vehicle direction always coincides with axis $x_{\rm r}$. A FOG measures Φ directly.

The floorspace boundaries are nested with L RFID tags $\mathrm{T}t(\chi_t, \mu_t)$, $t \in [1, L]$, (T1, ..., T12 in Fig. 1). Time-invariant coordinates (χ_t, μ_t) of the tags are supposed to be known. We assume that the vehicle reader can detect simultaneously $k_n \geqslant 2$ tags, where the number k_n is time-variant in discrete time index n, and measure a time-variant distance D_{in} to the ith tag, where $i \in [1, k_n]$. Note that the ith observed tag can be any of the nested tags. In Fig. 1, we illustrate this measurement strategy based on two tags $\mathrm{T}1(\chi_1 = 0, \mu_1 = 0)$ and $\mathrm{T}4(\chi_4 = 0, \mu_4 = Y_{\max})$. The distances between these tags and the reader are D_1 and D_2, respectively. Because altitudes are generally different of the points of installation of the reader and tags, the projections a_1 and a_2 to the vehicle plane are calculated following Fig. 1a and Fig. 1b for known c_1 and c_2.

Based on the vehicle odometry, the incremental distance d_n and the incremental change in heading ϕ_n can be found as

$$d_n = \frac{1}{2}(d_{{\rm R}n} + d_{{\rm L}n}), \quad (1)$$

$$\phi_n \cong \frac{1}{b}(d_{{\rm R}n} - d_{{\rm L}n}). \quad (2)$$

In turn, the vehicle coordinates x_n and y_n and heading Φ_n can be obtained by the vehicle kinematics with equations

$$f_{1n} = x_n = x_{n-1} + d_n \cos\left(\Phi_{n-1} + \frac{\phi_n}{2}\right), \quad (3)$$

$$f_{2n} = y_n = y_{n-1} + d_n \sin\left(\Phi_{n-1} + \frac{\phi_n}{2}\right), \quad (4)$$

$$f_{3n} = \Phi_n = \Phi_{n-1} + \phi_n, \quad (5)$$

in which x_{n-1}, y_{n-1}, and Φ_{n-1} are projected to time n by the time-variant incremental distances $d_{{\rm L}n}$ and $d_{{\rm R}n}$ via (1) and (2). Note that all of the values in (3)-(5) are practically not exact and have some additive random components.

Noticing that the minimum number of tags required to provide localization via (3)-(5) is equal to 2 and assuming that the reader antenna is able to detect simultaneously $k_n \geqslant 2$ tags with sufficient accuracy, we would like to investigate effect of redundant information delivered from more than 2 tags on the localization accuracy. We wish to solve this problem by developing and using the extended Kalman and FIR filtering techniques.

III. Self-Localization Problem in State Space

To solve the localization problem in state space, we follow (3)–(5) and introduce the state vector $\mathbf{x}_n = [\, x_n \;\; y_n \;\; \Phi_n \,]^T$ of unknown variables and an input vector $\mathbf{u}_n = [\, d_{{\rm L}n} \;\; d_{{\rm R}n} \,]^T$ of incremental distances. Random components in these values are additive and we suppose that they are zero mean, white Gaussian, and uncorrelated. Accordingly, we introduce the state noise vector $\mathbf{w}_n = [\, w_{xn} \;\; w_{yn} \;\; w_{\Phi n} \,]^T$ and the input noise vector $\mathbf{e}_n = [\, e_{{\rm L}n} \;\; e_{{\rm R}n} \,]^T$. We further unite equations (3)–(5) into the nonlinear state equation

$$\mathbf{x}_n = \mathbf{f}_n(\mathbf{x}_{n-1}, \mathbf{u}_n, \mathbf{w}_n, \mathbf{e}_n), \quad (6)$$

in which $\mathbf{f}_n = [\, f_{1n} \;\; f_{2n} \;\; f_{3n} \,]^T$ has components given by (3)–(5). The noise sources $\mathbf{w}_n$ and $\mathbf{e}_n$ are zero mean, $E\{\mathbf{w}_n\} = \mathbf{0}$ and $E\{\mathbf{e}_n\} = \mathbf{0}$, have the covariances

$$\mathbf{Q} = E\{\mathbf{w}_n \mathbf{w}_n^T\}, \quad (7)$$

$$\mathbf{L} = E\{\mathbf{e}_n \mathbf{e}_n^T\}, \quad (8)$$

and a property $E\{\mathbf{w}_n \mathbf{e}_v^T\} = \mathbf{0}$ for all integer n and v. Hereinafter, $E\{x\}$ means averaging of x.

In the RFID tag environment such as that shown in Fig. 1, k_n time-variant distances D_{in}, $i \in [1, k_n \geqslant 2]$, can simultaneously be measured between the reader and the tags $\mathrm{T}t(\chi_t, \mu_t)$. Along with the measurements of Φ_n, the observation equations can thus be written as

$$\begin{aligned} D_{1n} &= \sqrt{(\bar{\mu}_1 - y_n)^2 + (\bar{\chi}_1 - x_n)^2 + c_1^2}, \\ &\vdots \\ D_{k_n n} &= \sqrt{(\bar{\mu}_{k_n} - y_n)^2 + (\bar{\chi}_{k_n} - x_n)^2 + c_{k_n}^2}, \\ \Phi_n &= \Phi_n, \end{aligned}$$

where the coordinates $\bar{\chi}_i$, $\bar{\mu}_i$, and c_i belong to the ith detected tag which is one of the nested tags $\mathrm{T}t(\chi_t, \mu_t)$.

ISBN: 978-1-61804-318-4

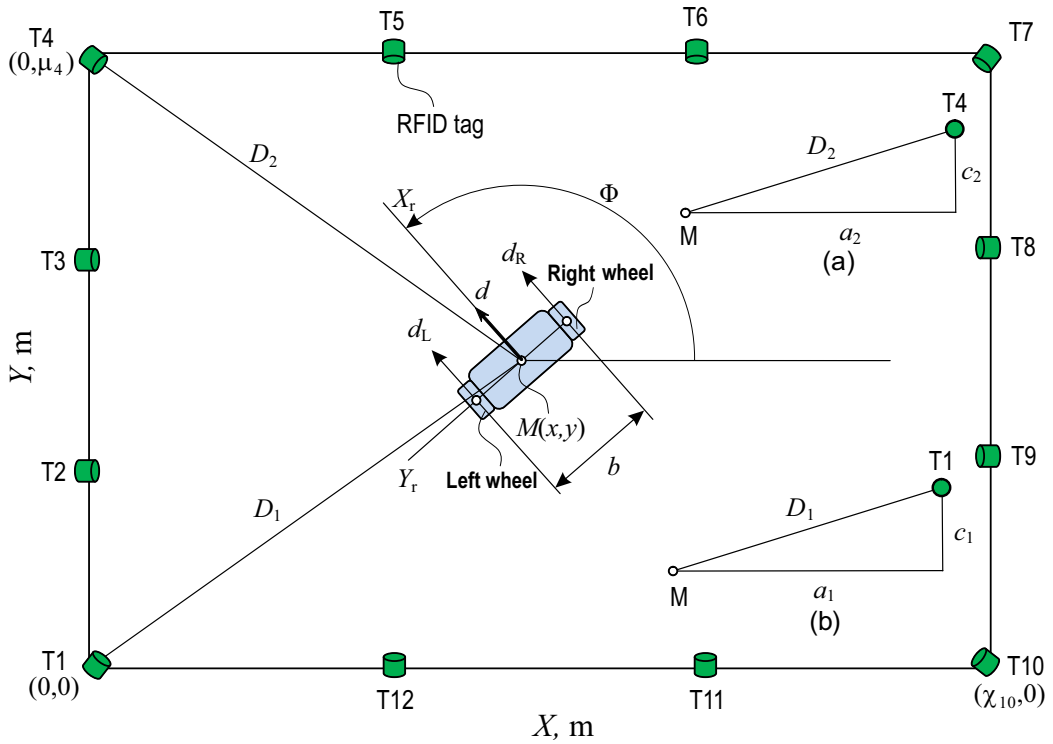


Fig. 1. 2D schematic geometry of a vehicle travelling on an indoor floorspace nested with RFID tags, T1, . . . , T12, having exactly known coordinates. The vehicle reader measures distances to RFID tags and Φ is measured using a FOG: (a) and (b) represent situations when a vehicle and the tags T1 and T4 are not in the same plane.

If we now introduce the observation vector $\mathbf{z}_n = [\, z_{1n} \ \ldots \ z_{k_n n} \ z_{\phi n} \,]^T \in \mathbb{R}^{k_n+1}$, the nonlinear function vector $\mathbf{h}_n(\mathbf{x}_n) = [\, D_{1n} \ \ldots \ D_{k_n n} \ \Phi_n \,]^T \in \mathbb{R}^{k_n+1}$, and the measurement additive noise vector $\mathbf{v}_n = [\, v_{1n} \ \ldots \ v_{k_n n} \ v_{\phi n} \,] \in \mathbb{R}^{k_n+1}$, then the state observation equation can be written as

$$\mathbf{z}_n = \mathbf{h}_n(\mathbf{x}_n) + \mathbf{v}_n \,, \tag{9}$$

where noise $\mathbf{v}_n$ is white Gaussian with zero mean, $E\{\mathbf{v}_n\} = \mathbf{0}$, the covariance

$$\mathbf{R}_n = E\{\mathbf{v}_n \mathbf{v}_n^T\} \in \mathbb{R}^{(k_n+1)\times(k_n+1)} \,, \tag{10}$$

and the properties $E\{\mathbf{v}_n \mathbf{w}_v^T\} = \mathbf{0}$ and $E\{\mathbf{v}_n \mathbf{e}_v^T\} = \mathbf{0}$ for all integer n and v. The vehicle dynamics is thus represented with the state-space model (6) and (9), in which (9) has time-variant both the coefficients and dimensions. Note that (10) also has time-variant dimensions.

A. Extended State-Space Model

Extension of linear filters to nonlinear problems is usually provided using the first-order Taylor series expansion, because the second-order expansion has no definitive advantages in state space [37]. Below, we arrive at the extended state-space model taking into account specifies of the RFID tag networks. The standard procedure applied to $\mathbf{f}_n \triangleq \mathbf{f}_n(\mathbf{x}_{n-1}, \mathbf{u}_n, \mathbf{w}_n, \mathbf{e}_n)$ yields

$$\begin{aligned} \mathbf{f}_n &\cong \mathbf{f}_n(\hat{\mathbf{x}}_{n-1}, \mathbf{u}_n, \mathbf{0}, \mathbf{0}) + \mathbf{F}_n(\mathbf{x}_{n-1} - \hat{\mathbf{x}}_{n-1}) \\ &\quad + \mathbf{W}_n \mathbf{w}_n + \mathbf{E}_n \mathbf{e}_n \\ &= \mathbf{F}_n \mathbf{x}_{n-1} + \bar{\mathbf{u}}_n + \mathbf{W}_n \mathbf{w}_n + \mathbf{E}_n \mathbf{e}_n \,, \end{aligned} \tag{11}$$

where $\hat{\mathbf{x}}_n = [\, \hat{x}_n \ \hat{y}_n \ \hat{\Phi}_n \,]$ is the estimate[1] of $\mathbf{x}_n$, $\bar{\mathbf{u}}_n = \mathbf{f}_n(\hat{\mathbf{x}}_{n-1}, \mathbf{u}_n, \mathbf{0}, \mathbf{0}) - \mathbf{F}_n \hat{\mathbf{x}}_{n-1}$ is a known input of the extended model and $\mathbf{F}_n$, $\mathbf{W}_n$, and $\mathbf{E}_n$ are Jacobian. Assuming relatively slow vehicle movement and control, we suppose that an increment $\mathbf{u}_n - \mathbf{u}_{n-1}$ in the input signal is insignificant on a unit time step. Besides, the noise components are supposed to be zeros at an initial point, $\mathbf{w}_{n-1} = \mathbf{0}$ and $\mathbf{e}_{n-1} = \mathbf{0}$.

By simple transformations, the Jacobian matrix $\mathbf{F}_n = \frac{\partial \mathbf{f}_n}{\partial \mathbf{x}}\Big|_{\hat{\mathbf{x}}_{n-1}}$ becomes

$$\mathbf{F}_n = \begin{bmatrix} 1 & 0 & -d_n \sin(\hat{\Phi}_{n-1} + \frac{1}{2}\phi_n) \\ 0 & 1 & d_n \cos(\hat{\Phi}_{n-1} + \frac{1}{2}\phi_n) \\ 0 & 0 & 1 \end{bmatrix} . \tag{12}$$

Because noise $\mathbf{w}_n$ is additive with respect to the components of $\mathbf{x}_n$ in (3)–(5) and $\mathbf{w}_{n-1} = \mathbf{0}$, we also have

$$\mathbf{W}_n = \frac{\partial \mathbf{f}_n}{\partial \mathbf{w}}\bigg|_{\hat{\mathbf{x}}_{n-1}} = \mathbf{F}_n \,. \tag{13}$$

[1] $\hat{\mathbf{x}}_{n|v}$ means the estimate at n via measurements from the past to v. We use the following notations: $\hat{\mathbf{x}}_n \triangleq \hat{\mathbf{x}}_{n|n}$ and $\hat{\mathbf{x}}_n^- \triangleq \hat{\mathbf{x}}_{n|n-1}$.

ISBN: 978-1-61804-318-4

Finally, the Jacobian matrix $\mathbf{E}_n = \frac{\partial \mathbf{f}_n}{\partial \mathbf{u}}\big|_{\hat{\mathbf{x}}_n^-}$ can be written as

$$\mathbf{E}_n = \frac{1}{2b}\begin{bmatrix} be_{cn} + d_n e_{sn} & be_{cn} - d_n e_{sn} \\ be_{sn} - d_n e_{cn} & be_{sn} + d_n e_{cn} \\ -2 & 2 \end{bmatrix}, \qquad (14)$$

where $e_{cn} = \cos\left(\hat{\Phi}_n^- + \frac{\phi_n}{2}\right)$ and $e_{sn} = \sin\left(\hat{\Phi}_n^- + \frac{\phi_n}{2}\right)$.

Unlike the state model (6), expansion of the observation model (9) needs some care. Basically, the nonlinear function $\mathbf{h}_n(\mathbf{x}_n)$ can be expanded at n similar to (11) as

$$\begin{aligned} \mathbf{h}_n(\mathbf{x}_n) &\cong \mathbf{h}_n(\hat{\mathbf{x}}_n^-) + \frac{\partial \mathbf{h}_n}{\partial \mathbf{x}}\bigg|_{\hat{\mathbf{x}}_n^-}(\mathbf{x}_n - \hat{\mathbf{x}}_n^-) \\ &= \mathbf{H}_n\mathbf{x}_n + \bar{\mathbf{z}}_n\,, \end{aligned} \qquad (15)$$

where $\bar{\mathbf{z}}_n = \mathbf{h}_n(\hat{\mathbf{x}}_n^-) - \mathbf{H}_n\hat{\mathbf{x}}_n^-$ is known and $\mathbf{H}_n = \frac{\partial \mathbf{h}_n}{\partial \mathbf{x}}\big|_{\hat{\mathbf{x}}_n^-}$ is Jacobian. However, the Jacobian matrix $\mathbf{H}_n$ has here time-variant dimensions,

$$\mathbf{H}_n = \begin{bmatrix} \frac{\hat{x}_n^- - \bar{\chi}_1}{\nu_{1n}} & \frac{\hat{y}_n^- - \bar{\mu}_1}{\nu_{1n}} & 0 \\ \vdots & \vdots & \vdots \\ \frac{\hat{x}_n^- - \bar{\chi}_{k_n}}{\nu_{(k_n n)}} & \frac{\hat{y}_n^- - \bar{\mu}_{k_n}}{\nu_{(k_n n)}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \in \mathbb{R}^{(k_n+1)\times 3}, \qquad (16)$$

where $\nu_{in} = \sqrt{(\bar{\mu}_i - \hat{y}_n^-)^2 + (\bar{\chi}_i - \hat{x}_n^-)^2 + c_i^2}$, $i \in [1, k_n]$. Such a structure of $\mathbf{H}_n$ requires modifications of the standard extended algorithms that we will do in the next sections.

The first-order extended state-space model is thus given by

$$\mathbf{x}_n = \mathbf{F}_n\mathbf{x}_{n-1} + \bar{\mathbf{u}}_n + \tilde{\mathbf{e}}_n + \tilde{\mathbf{w}}_n\,, \qquad (17)$$

$$\mathbf{z}_n = \mathbf{H}_n\mathbf{x}_n + \bar{\mathbf{z}}_n + \mathbf{v}_n\,, \qquad (18)$$

where the zero mean noise vectors $\tilde{\mathbf{w}}_n$ and $\tilde{\mathbf{e}}_n$ have the covariances $\tilde{\mathbf{Q}}_n = \mathbf{F}_n\mathbf{Q}\mathbf{F}_n^T$ and $\tilde{\mathbf{L}}_n = \mathbf{E}_n\mathbf{L}\mathbf{E}_n^T$ and $\mathbf{Q}$ and $\mathbf{L}$ are specified by (7) and (8), respectively.

IV. Extended Filtering Techniques

Below, we modify the EKF algorithm and develop a new EFIR filtering algorithm adapted to specifics of the extended model (17) and (18).

A. Extended Kalman Filter Algorithm

In order to modify the EKF algorithm for (17) and (18), we first specify the prior estimation error as

$$\mathbf{P}_n^- = E\{(\mathbf{x}_n - \hat{\mathbf{x}}_n^-)(\mathbf{x}_n - \hat{\mathbf{x}}_n^-)^T\} \qquad (19)$$

$$= \mathbf{F}_n\mathbf{P}_{n-1}\mathbf{F}_n^T + \tilde{\mathbf{Q}}_n + \tilde{\mathbf{L}}_n \qquad (20)$$

and the estimation error as

$$\mathbf{P}_n = E\{(\mathbf{x}_n - \hat{\mathbf{x}}_n)(\mathbf{x}_n - \hat{\mathbf{x}}_n)^T\} \qquad (21)$$

$$= (\mathbf{I} - \mathbf{K}_n\mathbf{H}_n)\mathbf{P}_n^-\,. \qquad (22)$$

where $\mathbf{K}_n$ is the bias correction gain (Kalman gain) to be defined in Table I.

Provided (20) and (22), the EKF algorithm for time-variant dimensions in $\mathbf{H}_n$ can be coded as in Table I. Here $\sigma_{v(ii)}^2$, $i \in [1, k_n]$, represents the noise variance of the ith observed tag and $\sigma_{v(ij)}^2$, $j \in [1, k_n]$, represents the noise cross covariance between the ith and jth observed tags. Typically, in the localization problems, the cross covariances are put to zero. Note that k_n is time-variant and thus the measurement vector $\mathbf{z}_n \in \mathbb{R}^{k_n+1}$ goes to the input with time-variant dimensions. Due to this, the measurement noise covariance matrix $\mathbf{R}_n$ has time-variant dimensions as well.

TABLE I. EKF Pseudo Code for Over Observed Target State (18)

Input: $\mathbf{z}_n$, k_n, $\hat{\mathbf{x}}_0$, $\mathbf{P}_0$, $\mathbf{Q}$, $\mathbf{L}$

for $n = 1 : \infty$ **do**

$\hat{\mathbf{x}}_n^- = \mathbf{f}_n(\hat{\mathbf{x}}_{n-1}, \mathbf{u}_n, \mathbf{0}, \mathbf{0})$

for $i = 1 : k_n$ **do**

$\nu_{in} = \sqrt{(\bar{\mu}_i - \hat{y}_n^-)^2 + (\bar{\chi}_i - \hat{x}_n^-)^2 + c_i^2}$

end for

$$\mathbf{H}_n = \begin{bmatrix} \frac{\hat{x}_n^- - \bar{\chi}_1}{\nu_{1n}} & \frac{\hat{y}_n^- - \bar{\mu}_1}{\nu_{1n}} & 0 \\ \vdots & \vdots & \vdots \\ \frac{\hat{x}_n^- - \bar{\chi}_{k_n}}{\nu_{(k_n n)}} & \frac{\hat{y}_n^- - \bar{\mu}_{k_n}}{\nu_{(k_n n)}} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{R}_n = \begin{bmatrix} \sigma_{v(11)}^2 & \cdots & \sigma_{v(1k_n)}^2 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \sigma_{v(k_n 1)}^2 & \cdots & \sigma_{v(k_n k_n)}^2 & 0 \\ 0 & \cdots & 0 & \sigma_\phi^2 \end{bmatrix}$$

$\mathbf{P}_n^- = \mathbf{F}_n(\mathbf{P}_{n-1} + \mathbf{Q})\mathbf{F}_n^T + \mathbf{E}_n\mathbf{L}\mathbf{E}_n^T$

$\mathbf{K}_n = \mathbf{P}_n^-\mathbf{H}_n^T(\mathbf{H}_n\mathbf{P}_n^-\mathbf{H}_n^T + \mathbf{R}_n)^{-1}$

$\hat{\mathbf{x}}_n = \hat{\mathbf{x}}_n^- + \mathbf{K}_n[\mathbf{z}_n - \mathbf{h}_n(\hat{\mathbf{x}}_n^-)]$

$\mathbf{P}_n = (\mathbf{I} - \mathbf{K}_n\mathbf{H}_n)\mathbf{P}_n^-$

end for

Output: $\hat{\mathbf{x}}_n$

B. Extended FIR and FIR/Kalman Filtering Algorithms

Unlike the minimization of the mean square error (MSE) in KF, an idea behind the UFIR filter [37] is to satisfy only the unbiasedness condition $E\{\hat{\mathbf{x}}_n\} = E\{\mathbf{x}_n\}$; that is, an average of the estimate at the filter output is equal to that of the model at each time index n. By virtue of this, the UFIR filter and its extended version EFIR completely ignore the noise statistics but require an optimal averaging interval of N_{opt} points in order to minimize the MSE.

Following [37], the EFIR filter can be developed similarly to EKF. The EFIR estimate appears iteratively as

$$\hat{\mathbf{x}}_l = \hat{\mathbf{x}}_l^- + \mathbf{K}_l[\mathbf{z}_l - \mathbf{h}_l(\hat{\mathbf{x}}_l^-)]\,, \qquad (23)$$

where an auxiliary variable l ranges from $m + K$ to n with $m = n - N - 1$. Here, K is the dimension of the state vector (number of the target states) and N is the averaging horizon length. The output is taken when $l = n$ in each iteration cycle. The bias correction gain is defined only by the model matrices,

$$\mathbf{K}_l = \mathbf{G}_l\mathbf{H}_l^T\,, \qquad (24)$$

ISBN: 978-1-61804-318-4

via the generalized noise power gain (GNPG)

$$\mathbf{G}_l = [\mathbf{H}_l^T\mathbf{H}_l + (\mathbf{F}_l\mathbf{G}_{l-1}\mathbf{F}_l^T)^{-1}]^{-1} \tag{25}$$

in which the inverse exists when $l \geqslant m+K$.

To run the EFIR filter iteratively, K initial estimates are required. For the localization problem implying three states, $K=3$, the initial estimates at $s=m+K-1$ can be found in batch forms as [36]

$$\hat{\mathbf{x}}_s = \mathbf{F}_s\mathbf{F}_{s-1}(\mathbf{H}_{s,m}^T\mathbf{H}_{s,m})^{-1}\mathbf{H}_{s,m}^T\mathbf{Y}_{s,m}\,, \tag{26}$$

$$\mathbf{G}_s = \mathbf{F}_s\mathbf{F}_{s-1}(\mathbf{H}_{s,m}^T\mathbf{H}_{s,m})^{-1}\mathbf{F}_{s-1}\mathbf{F}_s\,, \tag{27}$$

$$\mathbf{Y}_{s,m} = \left[\,\mathbf{y}_{m+2}^T\ \mathbf{y}_{m+1}^T\ \mathbf{y}_m^T\,\right]^T, \tag{28}$$

$$\mathbf{H}_{s,m} = \begin{bmatrix} \mathbf{H}_{m+2}\mathbf{F}_{m+2}\mathbf{F}_{m+1} \\ \mathbf{H}_{m+1}\mathbf{F}_{m+1} \\ \mathbf{H}_m \end{bmatrix}, \tag{29}$$

where $\mathbf{y}_n$ is assumed to be a vector of linear measurements of $\mathbf{x}_n$. Because linear measurement $\mathbf{y}_n = \mathbf{H}_n\mathbf{x}_n + \mathbf{v}_n$ may be either available, by nonlinear-to-linear conversion applied to (9), or not, there are two feasible options:

- If $\mathbf{y}_s$ is available, then use it instead of $\hat{\mathbf{x}}_s$ given by (26).
- If $\mathbf{y}_s$ is unavailable, then the output of EKF (with even roughly set noise statistics) can be used instead of $\mathbf{y}_s$ to initiate the EFIR filter on a horizon of first $N_{\rm opt}$ points. We call such an algorithm the EFIR/Kalman algorithm.

There is an important specific of the scheme shown in Fig. 1. Because noise reduction is negligible with $N=K$ and GNPG can thus be associated with unity at time index s, one may let $\mathbf{G}_s = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. Such a simplification ignoring (26)–(29) does not affect the localization accuracy essentially, at least for the localization problem in question.

The EFIR filtering algorithm can now be coded as in Table II. Provided the number k_n of the detected tags along with measurements $\mathbf{z}_n$ and $\mathbf{y}_n$ at each n, the algorithm requires only N and K to start computing and updating iteratively all the vectors and matrices. No noise statistics are involved that is an important advantage against EKF. Although the algorithm (Table II) admits $\mathbf{G}_s = \mathbf{I}$, one may substitute $\mathbf{G}_s$ with (27) expecting for some increase in accuracy which, in our case, was insignificant. On the other hand, the EFIR filter is unbiased, thus it does not guarantee optimality. To minimize the MSE in the EFIR estimate, $N_{\rm opt}$ has to be found [32], [37], [42]. If test measurements can be organized to track a target trajectory using precise equipment, the minimization of MSE using (21) is most straightforward [42]. We shall follow this approach in the second part of this paper [40].

TABLE II. EFIR PSEUDO CODE FOR OVER OBSERVED TARGET STATE (18)

Input: $\mathbf{z}_n$, k_n, $\mathbf{y}_n$, K, N

for $n = N-1:\infty$ **do**

  $m = n-N+1, \quad s = m+K-1$

  $\tilde{\mathbf{x}}_s = \begin{cases} \mathbf{y}_s\,, & \textbf{if} \quad s < N-1 \\ \hat{\mathbf{x}}_s\,, & \textbf{if} \quad s \geqslant N-1 \end{cases}$

  $\mathbf{G}_s = \mathbf{I}$

  for $l = m+K:n$ **do**

    $\tilde{\mathbf{x}}_l^- = \mathbf{f}_l(\tilde{\mathbf{x}}_{l-1}, \mathbf{u}_l, \mathbf{0}, \mathbf{0})$

    for $i = 1:k_n$ **do**

      $\nu_{il} = \sqrt{(\bar{\mu}_i - \hat{y}_l^-)^2 + (\bar{\chi}_i - \hat{x}_l^-)^2 + c_i^2}$

    end for

    $\mathbf{H}_l = \begin{bmatrix} \frac{\hat{x}_l^- - \bar{\chi}_1}{\nu_{1l}} & \frac{\hat{y}_l^- - \bar{\mu}_1}{\nu_{1l}} & 0 \\ \vdots & \vdots & \vdots \\ \frac{\hat{x}_l^- - \bar{\chi}_{k_n}}{\nu_{(k_n l)}} & \frac{\hat{y}_l^- - \bar{\mu}_{k_n}}{\nu_{(k_n l)}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \in \mathbb{R}^{(k_n+1)\times 3}$

    $\mathbf{G}_l = [\mathbf{H}_l^T\mathbf{H}_l + (\mathbf{F}_l\mathbf{G}_{l-1}\mathbf{F}_l^T)^{-1}]^{-1}$

    $\mathbf{K}_l = \mathbf{G}_l\mathbf{H}_l^T$

    $\tilde{\mathbf{x}}_l = \tilde{\mathbf{x}}_l^- + \mathbf{K}_l[\mathbf{z}_l - \mathbf{h}_l(\tilde{\mathbf{x}}_l^-)]$

  end for

  $\hat{\mathbf{x}}_n = \tilde{\mathbf{x}}_n$

end for

Output: $\hat{\mathbf{x}}_n$

Remark: use the EKF estimate as $\mathbf{y}_s$ if linear measurement $\mathbf{y}_n$ is not available.

V. CONCLUDING REMARKS

Situation awareness in RFID tag information networks critically depends on the localization accuracy. With an insufficient localization accuracy, information about local 2D or 3D surroundings delivered to a target by request can be useless and even dangerous – may provoke collisions. We have shown in this paper that the target state observation over RFID tag-nested grid excess channels *increases* both the *localization accuracy* and *system stability*. It also *prevents divergence in EKF*. For such grids, we modified the EKF and developed a new EFIR/Kalman algorithm which are studied in the second part of this paper.

REFERENCES

[1] K. Finkenzeller, *RFID Handbook: Radio-Frequency Identification Fundamentals and Applications*, Wiley: New York, 2000.

[2] S. Willis and S. Helal, "RFID information grid for blind navigation and wayfinding," in *Proc. IEEE Int. Symp. on Wearable Computers*, 2005, pp. 34–37.

[3] M.-S. Jian and J.-S. Wu, "RFID Applications and Challenges," in *Radio Frequency Identification from System to Applications*, InTech, 2013.

[4] M. N. Lionel, Y. Liu, Y. C. Lau, and A. P. Patil, "LANDMARC: Indoor location sensing using active RFID," *Wireless Networks*, vol. 10, no. 6, pp. 701–710, Nov. 2004.

[5] S. S. Saab and Z. S. Nakad, "A standalone RFID indoor positioning system using passive tags," *IEEE Trans. Ind. Electron.*, vol. 58, no. 5, pp. 1961–1970, May 2011.

[6] M. Luimula, K. Sääskilahti, T. Partala, S. Pieskä, and J. Alaspää, "Remote navigation of a mobile robot in an RFID-augmented environment," *Personal and Ubiquitous Comput.*, vol. 14, no. 2, pp. 125–136, Feb 2010.

[7] R. Krigslund, P. Popovski, G. F. Pedersen, and K. Olesen, "Interference helps to equalize the read range and reduce false positives of passive RFID tags," *IEEE Trans. Ind. Electron.*, vol. 59, no. 12, pp. 4821–4830, Dec. 2012.

[8] S. Park and H. Lee, "Self-recognition of vehicle position using UHF passive RFID tags," *IEEE Trans. Ind. Electron.*, vol. 60, no. 1, pp. 226–234, Jan. 2013.

[9] J. Pomarico-Franquiz, M. Granados-Cruz, and Y. S. Shmaliy, "Self-localization over RFID tag grids excess channels using extended filtering techniques," *IEEE J. of Selected Topics in Signal Process.*, vol. 9, no. 2, pp. 229–238, Mar. 2015

[10] A. Smaliagic, and D. Kogan, "Location sensing and privacy in a context-aware computing environment," *IEEE Wireless Commun.*, vol. 9, no. 5, pp. 10–17, Oct. 2002.

[11] D.G. Seo and J.M. Lee, "Localization algorithm for a mobile robot using iGS," in Proc. *17th World Congress of the Int. Fed. of Autom. Control*, Seoul, Korea, July 6–11, 2008, pp. 742–747.

[12] Y. Zhang, M. G. Amin, and S. Kaushik, "Localization and Tracking of Passive RFID Tags Based on Direction Estimation," *Int. J. Antennas Propag.*, vol. 2007, art. 17426, pp. 1–9, 2007.

[13] A. Papapostolou and H. Chaouchi, "RFID-assisted indoor localization and the impact of interference on its performance," *J. Network Comput. Appl.*, vol. 34, no. 3, pp. 902–913, May 2011.

[14] K. Yamano, K. Tanaka, M. Hirayama, E. Kondo, Y. Kimuro, and M. Matsumoto, "Self-localization of mobile robots with RFID system by using support vector machine," in Proc. *IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2004, pp. 3756–3761.

[15] T. Tsukiyama, "Navigation system for mobile robots using RFID tags," in Proc. *Int. Conf. on Advanced Robotics (ICAR)*, Coimbra, Portugal, 2003, pp. 1130–1135.

[16] H. Chae and K. Han, "Combination of RFID and vision for mobile robot localization," in Proc. *IEEE Int. Conf. Intell. Sensors, Sensor Netw. Inf. Process*., 2005, pp. 75–80.

[17] O. Matsumoto, K. Komoriya, T. Hatase, H. Nishimura, K. Toda, and S. Goto, "Autonomous traveling control of the "TAO Aidce" intelligent wheelchair," in Proc. *2006 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, Oct. 9–15, 2006, Beijing, China, pp. 4322–4327.

[18] J. Zhou and J. Shi, "RFID localization algorithms and applications – a review," *J. Intell. Manuf.*, vol. 20, no. 6, pp. 695–707, Dec. 2009.

[19] B.-S. Choi, J.-W. Lee, J.-J. Lee, and K.-T. Park, "A Hierarchical Algorithm for Indoor Mobile Robot Localization Using RFID Sensor Fusion," *IEEE Trans. Ind. Electron.*, vol. 58, no. 6, pp. 2226–2235, Jun. 2011.

[20] W. Gueaieb and M. S. Miah, "An intelligent mobile robot navigation technique using RFID technology, *IEEE Trans. Instrum. Meas.*, vol. 57, no. 9, pp. 1908–1917, Sep. 2008.

[21] E. DiGiampaolo and F. Martinelli, "A passive UHF-RFID system for the localization of an indoor autonomous vehicle," *IEEE Trans. Ind. Electron.*, vol. 59, no. 10, pp, 3961–3970, Oct. 2012.

[22] V. Savic, A. Athalye, M. Bolic, and P. M. Djuric, "Particle filtering for indoor RFID tag tracking," in Proc. *IEEE Statist. Signal Process. Workshop (SSP)*, 2011, pp. 193–196.

[23] M. Boccadoro, F. Martinelli, and S. Pagnotelli, "Constrained and quantized Kalman filtering for an RFID robot localization problem," *Auton. Robots*, vol. 29, no. 3-4, pp. 235–251, Nov. 2010.

[24] A. Howard, "Multi-robot simultaneous localization and mapping using particle filters," *Int. J. of Robotics Research*, vol. 25, no. 12, pp. 1243–1256, Dec. 2006.

[25] E. DiGiampaolo and F. Martinelli, "Mobile robot localization using the phase of passive UHF-RFID signals," *IEEE Trans. Ind. Electron.*, vol. 61, no. 1, pp. 365–376, Jan. 2014.

[26] F. Martinelli, "Robot localization: comparable performance of EKF and UKF in some interesting indoor settings," in Proc. *16th Mediterranean Conf. on Contr. Autom.*, Ajaccio, France, June 25-27, 2008, pp. 499–504.

[27] B. Gibbs, *Advanced Kalman Filtering, Least-Squares and Modeling*, New York: Wiley, 2011.

[28] Y. S. Shmaliy, "An iterative Kalman-like algorithm ignoring noise and initial conditions," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2465–2473, Jun. 2011.

[29] R. J. Fitzgerald, "Divergence of the Kalman filter," *IEEE Trans. Autom. Control*, vol. AC-16, no. 6, pp. 736–747, Dec. 1971.

[30] F. Daum, "Nonlinear filters: beyond the Kalman filter," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 20, no. 8, pp. 57–69, Aug. 2005.

[31] W. H. Kwon and S. Han, *Receding Horizon Control: Model Predictive Control for State Models*. London: Springer, 2005.

[32] Y. S. Shmaliy, "Linear optimal FIR estimation of discrete time-invariant state-space models," *IEEE Trans. Signal Process.*, vol. 58, pp. 3086–3096, Jun. 2010.

[33] A. M. Bruckstein and T. Kailath, "Recursive limited memory filtering and scattering theory," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 3, pp. 440–443, May 1985.

[34] W. H. Kwon, Y. S. Suh, Y. I. Lee, and O. K. Kwon, "Equivalence of finite memory filters," *IEEE Trans. Aerospace Electron. Syst.*, vol. 30, no. 8., pp. 968–972, Jul. 1994.

[35] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, New York: Academic, 1970.

[36] Y. S. Shmaliy, "Unbiased FIR filtering of discrete time polynomial state space models," *IEEE Trans. Signal Process.*, vol. 57, no. 4, pp. 1241–1249, Apr. 2009.

[37] Y. S. Shmaliy, "Suboptimal FIR filtering of nonlinear models in additive white Gaussian noise," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5519–5527, Oct. 2012.

[38] J. Pomarico-Franquiz, S. Khan, and Y.S. Shmaliy, "Combined extended FIR/Kalman filtering for indoor robot localization via triangulation," *Measurement*, vol. 50, pp. 236–242, Apr. 2014.

[39] J. Pomarico-Franquiz and Y.S. Shmaliy, "Accurate self-localization in RFID tag information grids using FIR filtering," *IEEE Trans. Ind. Inform.*, vol. 10, no. 2, pp. 1317–1326, May 2014.

[40] M. Granados-Cruz and Y. S. Shmaliy, "Extended filtering for self-localization over RFID tag grids excess channels – II," in *Recent Advances on Electroscience and Computers:* Proc. *Int. Conf. on Systems, Control, Signal Process. Informatics (SCSI 2015), Int. Conf. on Electronics and Communication Systems (ECS 2015)*, Barcelona, Spain, April 7-9, 2015, pp. 159–164.

[41] K. Komoriya and E. Oyama, "Position estimation of a mobile robot using optical fiber gyroscope," in Proc. *IEEE/RSJ/GI Int. Conf. Intell. Robots Syst.*, vol. 1, 1994, pp. 143–149.

[42] F. Ramirez-Echeverria, A. Sarr, and Y. S. Shmaliy, "Optimal memory for discrete-time FIR filters in state space," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 557–561, Feb. 2014.

Array factor directivity for interference scenarios

M.A. Lagunas, A. Perez-Neira, X. Artiga.

***Abstract*—** The concept of array factor directivity is revised in order to propose a new well-suited definition for scenarios dominated by interference. A novel directivity measure is proposed including the interference that is caused to un-intended receivers. The new measure is properly bounded and, at the same time, it introduces the notion of retro-directivity in the design of antenna arrays. In addition, the beamformer that maximizes the new directivity measure is reported, proving that it never presents retro-directivity.

***Keywords*—**Array factor, beamforming, directivity, interference, retro-directivity.

I. Introduction

THE directivity of a planar aperture at a given direction, which depends on the elevation and azimuth angels (θ,φ), is defined as the field intensity in such direction at a distance r divided by the power density of an omnidirectional antenna radiating the same power at the same distance [1].

$$D(\theta,\varphi)=\frac{P(\theta,\varphi,r)}{P_{RAD}/4\pi r^2} \quad (1)$$

Note that under this definition, the directivity has dimension of surface, i.e. m^2. Also, note that, although there is not any problem for its use both in the far as well as in the near field, developing this definition in the far field is easier than in the near field. In fact, for the far field, the directivity is always proportional to a factor depending on the field in the aperture as it is shown in (2).

$$D(\theta,\varphi)\propto\frac{\left|\int_A E(x,y)e^{j2\pi(xa1+ya2)/\lambda}\,dx\,dy\right|^2}{\int_A\left|E(x,y)\right|^2dx\,dy}$$

where (2)

$$a1=\sin(\theta)\cos(\varphi)\quad a2=\sin(\theta)\sin(\varphi)$$

This work was supported in part by the Genralitat of Catlunya (Grant 2014SGR1567) and the EU (Projects SANSA) and ESA (SATNEX), and from the Spanish Ministry of Economy and Competitiveness TEC2014-59255-C3-1-R.

M. A. Lagunas (corresponding author) is with the department of Antenna and Multichannel Signal processing (AMSP) of the Centre Tecnologic de Telecomunications de Catalunya (CTTC), 08860 Castelldefels SPAIN. Phone: +34 93 6452921; fax: +34 93 6452929, e-mail: m.a.lagunas@cttc.es. He is also professor at Universitat Politecnica de Catalunya UPC-TSC.

A. Perez-Neira is with the AMSP department of CTTC, e-mail: ana.perez@cttc.es. She is also professor at UPC-TSC

X. Artiga is with the AMSP department of CTTC, e-mail: xavier.artiga@cttc..es.

The effective area of the aperture surface bounds this last factor.

$$D(\theta,\varphi)\le\frac{4\pi A_0}{\lambda^2} \quad (3)$$

Since A_0 denotes the area of the aperture, it can be said that the maximum of the directivity is bounded by 4π times the area of the aperture in square wavelengths.

Focusing on the case of planar antenna arrays, the factor affecting the directivity is shown in (4), where: b(m) are the beamformer coefficients controlling the current at every antenna element, d_m is the radius, with respect to the phase center of the aperture in wavelengths, and φ_m is the azimuth location of antenna element m within the aperture [1]-[2].

$$D(\theta,\varphi)\propto\frac{\left|\sum_m b(m)\exp\left(j2\pi d_m\cos(\varphi-\varphi_m)\sin(\theta)\right)\right|^2}{\sum_m|b(m)|^2} \quad (4)$$

This factor is denoted as the array factor directivity. The array factor depends only on the geometry and the relative excitation to every antenna element in the aperture.

The rest of this paper focuses on how to modify (4) for scenarios where unintended receivers are present.

The main objective of the paper is to put emphasis on the fact that any of the existing definitions for array factor directivity are done disregarding the interference caused to other receivers in the scenario (see [3]-[5] as examples). However, currently, most communication scenarios are moving from noise dominated to interference dominated. Since proper transmitter antenna beamforming, or design, implies the optimization of the array factor directivity, it is evident that for scenarios dominated by interference the traditional definition is becoming obsolete. Note that to control the interference, the transmitter should have information about the location or spatial signatures for the unintended receiver locations. This information or knowledge is denoted as channel state information at the transmitter (CSIT).

The structure of the paper is: Section II.A reviews the traditional notion of directivity described in this section. Section II.B describes how directivity has been defined for scenarios dominated by interference, as well as the optimum beamforming policy to maximize the newly defined directivities. Section III introduces the notion of retro-directivity for interference-dominated scenarios. Section IV describes a new array factor directivity based on retro-directivity that overcomes the problems of the measures described in the previous section. Finally, Section V includes

simulations to provide evidence of the superiority of the proposed directivity. Section VI concludes the paper.

II. Array Factor Directivity

This section will revisit the two current definitions of antenna array factor directivity, with emphasis on its suitability for interference scenarios. Along the paper, vectors are underlined and matrixes doubled underlined. Vector $\underline{b}$ will denote the beamformer. The sub-indexes "d" and "i" will refer to desired location and interfered location respectively; in consequence $\underline{h}_d$ and $\underline{h}_{iq}$ represent the channel response from the transmitter array to the desired or intended location and the channel to the un-intended or interfered location "q" respectively. Note that the above channel vectors are the spatial signatures of the transmitting array in a given location and they will coincide with the steering vector in a pure line of sight (LOS) propagation scenario. The angular dependency of the array factor on the elevation (linear arrays) or elevation/azimuth (2-D dimensional or 3-Dimensional arrays) will be omitted in the formulas for the sake of presentation.

A. Traditional Array Factor directivity

The most popular array factor directivity measure assumes LOS propagation and it is defined as the quotient between the response of the array factor at a giving direction, i.e. the intended or desired direction characterized by the steering vector $\underline{S}_d$, divided by the power density (power/surface) radiated by a single omni-directional that uses the same power that the antenna array. The latter power is given by the norm of the beamformer in use. This popular definition of array factor directivity is shown in (5).

$$D_{ANT} = \frac{\left|\underline{b}^H \underline{S}_d\right|^2}{\left|\underline{b}\right|^2} \tag{5}$$

Note that this definition refers to direction and not location and, of course doe not reflects the possibility of interference to un-intended locations. When refereeing to a desired location it is usual to mimic (5) just replacing the steering vector by the spatial signature of the desired location as it is shown in (6).

$$D_{TRA} = \frac{\left|\underline{b}^H \underline{h}_d\right|^2}{\left|\underline{b}\right|^2} \tag{6}$$

It is easy to check that this array factor directivity is always in the range between zero and the squared norm of the location vector $\underline{h}_d$.

This array factor directivity does not consider the negative impact of un-intended locations, i.e. interference dominated scenarios.

The beamformer that maximizes (6) is the so-called matched beamformer (MF), which is the desired location vector with norm equal to one.

$$\underline{b} = \underline{h}_d / \left|\underline{h}_d\right| \tag{7}$$

Finally, it is worth comment that an omnidirectional antenna does not exist in practice. Nevertheless, note that the denominator in (6) is not merely a normalization factor.

B. The Virtual Signal to Interference plus Noise ratio (VSNR) directivity factor

The traditional alternative to the previous definition D_{TRA} of array factor directivity is the so-called VSNR directivity. The definition is shown in (8), where NI unintended locations contribute to the denominator term with their respective location vectors, grouped in a single matrix in the second term of this formula.

$$D_{VSNR} = \frac{P_T \left|\underline{b}^H \underline{h}_d\right|^2}{\left|\underline{b}\right|^2 + P_T \sum_{q=1}^{NI} \left|\underline{b}^H \underline{h}_{iq}\right|^2} = \frac{P_T \left|\underline{b}^H \underline{h}_d\right|^2}{\left|\underline{b}\right|^2 + P_T \underline{b}^H \underline{\underline{R}}_i \underline{b}} \tag{8}$$

The justification of this formula is somehow quite artificial. Basically, this justification is based on the definition of signal to noise ratio (SNR) at intended location plus a "virtual" interference term. The so-called virtual SNR that we define bellow.

The desired receiver experiences the power from the transmitter and its own noise power. This SNR is transformed in by the addition in the denominator of the virtual interference term (it does not exists in practice). This term is formed by the interference caused to the un-intended locations accumulated in a VSINR as show in (9). Where P_T is the available power at transmission referred to the noise power of the receiver. Note that the definition of P_T entails that the norm of the beamformer has to be one.

$$VSINR = \frac{P_T \left|\underline{b}^H \underline{h}_d\right|^2}{1 + P_T \underline{b}^H \underline{\underline{R}}_i \underline{b}} \tag{9}$$

After (9), which is really a virtual SINR, since any directivity measure cannot change with the beamformer norm; the one in the denominator of (9) is just changed by the norm of the beamvector. Finally D_{VSNR} takes its usual form as (10).

$$D_{VSNR} = \frac{\left|\underline{b}^H \underline{h}_d\right|^2}{\left(1/P_T\right)\left|\underline{b}\right|^2 + \underline{b}^H \underline{\underline{R}}_i \underline{b}} \tag{10}$$

Several conceptual changes are in (9) with respect to the traditional version of directivity D_{TRA}. First, note that D_{VSNR} depends on the used power for transmission since there is no other way to reflect the amount of interference caused to the un-intended locations. As a consequence, this power, being referred to the noise of the intended receiver depends also on the receiver that is used. In summary, we may say that all the propagation-rooted arguments to derive D_{TRA} have disappeared in favor of communications arguments. This is the case of the beamformer norm that in (9) is more normalization factor than before in (6).

The beamformer that maximizes the Rayleigh quotient (10) is the so-called VSNR beamformer, which is the maximum generalized eigenvalue of the quadratic form in the numerator and in the denominator respectively. This beamvector is shown in (11.a), where parameter ϕ normalizes to one the norm of the beamvector. When the numerator is a quadratic form of a rank-one matrix, the generalized eigenvalue has the closed form in (11.a).

$$\underline{b}_{VSNR} = \phi\left(\left(1/P_T\right)\underline{\underline{I}} + \underline{\underline{R}}_i\right)^{-1} \underline{h}_d \tag{11.a}$$

$$D_{VSNR}^{MAX} = \underline{h}_d \left(P_T^{-1}\underline{\underline{I}} + \underline{\underline{R}}_i\right)^{-1} \underline{h}_d \tag{11.b}$$

ISBN: 978-1-61804-318-4

The bounds of this directivity measure are cero and (12), which is directly derived from (11.b).

$$D_{VSNR}^{MAX} \le \left|\underline{h}_d\right|^2 \left(P_T^{-1} + \lambda_{\min}\left(\underline{\underline{R}}_i\right)\right)^{-1} \tag{12}$$

When the number of antennas n_T is greater than the number of interfered location NI, the interfered matrix is rank deficient and its minimum eigenvalue is zero, thus (13) is the upper bound of this array factor directivity. This maximum is achieved when the desired is in the null-subspace of the interference matrix.

$$D_{VSNR}^{MAX} \le P_T \left|\underline{h}_d\right|^2 \tag{13}$$

Before living this section it is important to remark that D_{VSNR} does not penalize the fact that the interference power could be higher than the power delivered to the desired. Furthermore, for the scenario where all the interference locations coincide with the desired one, the directivity is given by (14), which evidently, is greater than zero.

$$D_{VSNR}\big|_{NI=1,\, at\ \underline{h}_d=\underline{h}_i} = \left(P_T \left|\underline{h}_d\right|^2\right) / \left(P_T n_T \left|\underline{h}_d\right|^2 + 1\right) \tag{14}$$

This problem is linked to the retro-directivity concept, as it will be shown in the next section.

It is also worth remark that in order to compare this directivity, which depends on the used power P_T, with the traditional definition (6), it is necessary to set the power equal to one. Doing this the bounds of the two directivities are the same.

III. Retro-Directivity

The notion of retro-directivity or negative directivity comes from the assumption that a proper array factor should devote more power to the intended location than the interference promoted to the un-intended locations.

The first attempt to reflect retro-directivity is to compare the difference in terms of traditional directivity to the intended and un-intended locations. Since directivity is expressed in dB, the difference reduces to the quotient of the two directivities, as it is shown in (15).

$$D_{RDA} = 10\log_{10}\left(\frac{\left|\underline{b}^H \underline{h}_d\right|^2}{\underline{b}^H \underline{b}}\right) - 10\log_{10}\left(\frac{\underline{b}^H \underline{\underline{R}}_i \underline{b}}{\underline{b}^H \underline{b}}\right)$$
$$D_{RDA} = 10\log_{10}\left(\left|\underline{b}^H \underline{h}_d\right|^2 \Big/ \underline{b}^H \underline{\underline{R}}_i \underline{b}\right) \tag{15}$$

When the number of unintended locations is greater than the number of antenna, the beamformer that maximizes this directivity is given by (16.a). On the other hand, when the number of unintended locations is less than the number of antennas for transmission, the solution is the so-called Zero-Forcing (ZF) beamforming (16.b). The beamformer (16.b), for rank deficient interference matrix, nulls out the interference in all unintended locations. Parameter ϕ ensures that the norm of the beamformer is one for both cases.

$$\underline{b} = \phi \underline{\underline{R}}_i^{-1} \underline{h}_d \tag{16.a}$$

$$\underline{b}_{ZF} = \phi\left(\underline{\underline{I}} - \underline{\underline{R}}_i \left(\underline{\underline{R}}_i^H \underline{\underline{R}}_i\right)^{-1} \underline{\underline{R}}_i^H\right)\underline{h}_d \tag{16.b}$$

Note that this directivity is independent of the used power as it is the case for the traditional definition.

It is important to check the bounds of this directivity. There is no upper bound for D_{RDA} since, for rank deficient interference matrix, this directivity is unbounded. On the other hand, when the unintended locations are all of then at the intended location this directivity is negative. It seems reasonable that for this case any directivity instead of negative should be zero. This implies to define the directivity in terms of the average interference power instead of the global power, as it is indicated in (17). The number of unintended locations is N_I.

$$D_{RDA} = 10\log_{10}\left(N_I \left|\underline{b}^H \underline{h}_d\right|^2 \Big/ \underline{b}^H \underline{\underline{R}}_i \underline{b}\right) \tag{17}$$

The major problem for the ZF beamformer is that for closed intended/unintended locations, forcing the null to the unintended precludes significant levels of power at the intended location. This results in a waste of the used power for transmission. This is a serious problem for limited available power for transmission. In addition, the unbounded character of this measure should be considered as a problem.

Next section suggests a new directivity measure, which overcomes both problems.

IV. New Array Factor Directivity

The new definition is rooted in the VSNR and it introduces the retro-directivity at its numerator. Solving, as it will be shown hereafter, it solves the problems associated with the D_{VSNR} definition. The new directivity is called D_{RD} and it is defined as (18).

$$D_{RD} = \frac{\left|\underline{b}^H \underline{h}_d\right|^2 - \underline{b}^H \underline{\underline{R}}_i \underline{b}}{P_T^{-1} \underline{b}^H \underline{b} + \underline{b}^H \underline{\underline{R}}_i \underline{b}} \tag{18}$$
$$with \quad \underline{\underline{R}}_i = \frac{1}{N_I} \sum_{q=1}^{N_I} \underline{h}_q \underline{h}_q^H$$

The interest of this new definition of the array factor directivity can be observed in (19.a). In consequence, the new array factor directivity is derived from the maximum eigenvalue of the generalized singular value decomposition of the numerator and the denominator of (19.b) minus one.

$$D_{RD} + 1 = \frac{P_T^{-1} \underline{b}^H \underline{b} + \left|\underline{b}^H \underline{h}_d\right|^2}{P_T^{-1} \underline{b}^H \underline{b} + \underline{b}^H \underline{\underline{R}}_i \underline{b}} \tag{19.a}$$

$$\left(P_T^{-1} \underline{\underline{I}} + \underline{h}_d \underline{h}_d^H\right)\underline{b} = \lambda\left(P_T^{-1} \underline{\underline{I}} + \underline{\underline{R}}_i\right)\underline{b} \tag{19.b}$$

This last expression (19.a), reads, with $\log_2(.)$, as the mutual information delivered to a receiver at the intended location, minus the mutual information delivered to NI receivers cooperating at the unintended locations with used power at the transmitter equal to P_T divided by the number of receivers. Thus, maximizing (18) entails to minimize the information delivered to the unintended locations with respect to the mutual information towards the desired one. In summary, the new definition is rooted on maximum achievable rates instead of on virtual SINR.

In order to get more insight, let us assume that N_I is equal to one and the channel to the unintended location is $\underline{h}_i$, the

generalized eigenvalues λ of (19.b) satisfy the following equation (see equation (A.7) in the appendix):

$$1=\Phi(\lambda)=\frac{\lambda P_T h_i^2 \rho}{P_T h_d^2+(1-\lambda)}+\frac{\lambda P_T h_i^2(1-\rho)}{(1-\lambda)}$$
$$with \quad h_d^2=\underline{h}_d^H\underline{h}_d \quad h_i^2=\underline{h}_i^H\underline{h}_i \quad \rho=\frac{\left|\underline{h}_d^H\underline{h}_i\right|^2}{h_d^2 h_i^2} \tag{20}$$

This function is always increasing with λ, has zeros at λ equal to cero and $1+E_T h_d^2 (1-\rho)$, and a single pole at $1+E_T h_d^2$. An example of the shape of this function is shown in Figure 1.

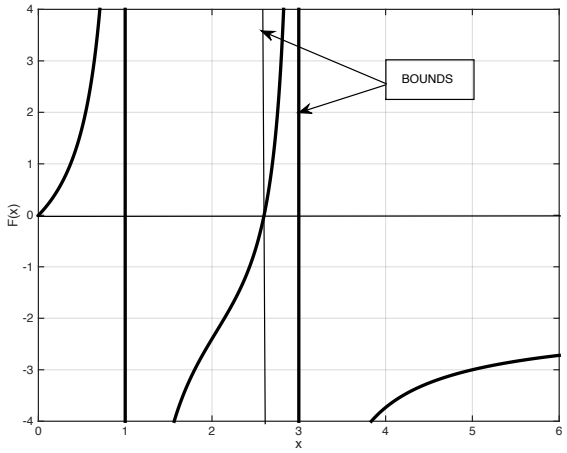


Figure 1 Function $\Phi(\lambda)$, showing the two zeros and two poles. Bound for the maximum eigenvalue between the greatest zero and pole, i.e. including the largest value of λ for which $\Phi(\lambda)$ is equal to one.

From (20), as it is depicted in Figure 1, the maximum eigenvalue of (19), which is equal to the directivity (18) plus one, is bounded as it is shown in (21).

$$1+h_d^2 P_T(1-\rho)\lambda_{\max}\le 1+P_T h_d^2$$
$$P_T h_d^2(1-\rho)\le D_{RD,\max}\le P_T h_d^2 \tag{21}$$

Note that the maximum of this directivity is bounded by the same value of D_{VSNR}. At the same time, it can be observed that the maximum directivity, achieved when using the maximum generalized eigenvector of the Rayleigh quotient (19), is always positive, i.e. no retro-directivity is present for the optimum beamformer. In addition, it is easy to check that for coincident locations of the intended and the unintended locations the directivity is cero. The maximum value in (21) is achieved when the intended channel is orthogonal to the unintended one.

The previous bounds, derived for a single interfered location can be extended for any number n_I of locations. In this case, function $\Phi(\lambda)$ is as (22), where d_q are the eigenvalues of the average interference matrix and n_T is the number of antenna elements of the array.

The function is always decreasing with λ and the greatest pole is located at $1/(P_T d_{min}+1)$. Above this pole is the value of λ for which the function is equal to one, i.e. the optimum λ. In consequence the directivity is greater than this values minus one.

$$\Phi(\lambda)=\sum_{q=1}^{n_T}\frac{\left|v_q\right|^2 P_T}{(P_T d_q+1)\lambda-1}$$
$$with \tag{22}$$
$$\underline{v}=\begin{bmatrix}1 & .. & v_q & .. & v_{n_T}\end{bmatrix}^T=\underline{\underline{U}}\,\underline{h}_d$$
$$\underline{\underline{R}}_i=(svd(\underline{\underline{R}}_i))=\underline{\underline{U}}\,\underline{\underline{D}}\,\underline{\underline{U}}^H$$

In addition, the function $\Phi(\lambda)$, for λ greater than one, can be bounded as (23).

$$\Phi(\lambda)\le\sum_{q=1}^{n_T}\frac{\left|v_q\right|^2 P_T}{(E_T d_{\min}+1)\lambda-1}= \tag{23}$$
$$=\frac{P_d^2 E_T}{(\lambda-1)+\lambda P_T d_{\min}}\le\frac{h_d^2 P_T}{\lambda-1}$$

Therefore, by using λ equal to $1+P_T h_d^2$ in (23), it can be concluded that the function is below one. Since the function, for λ greater than one, is monotonically decreasing (24) results.

$$\Phi\left(\lambda=\frac{1}{1+P_T d_{\min}}\right)\ge\Phi(\lambda_{opt})=1\ge\Phi(1+P_T h_d^2)$$
$$\frac{1}{1+P_T d_{\min}}\le\lambda_{opt}=D_{RD,\max}+1\le 1+P_T h_d^2 \tag{24}$$
$$\frac{-P_T d_{\min}}{1+P_T d_{\min}}\le D_{RD,\max}\le P_T h_d^2$$

The bounds shown in the last expression of (24) generalize the bounds shown in (21). Note that when the number of interferers is lower than the number of antennas of the array, the minimum eigenvalue of the interference matrix will be zero. This indicates that for rank deficient interference matrix there is not retro-directivity when using the optimum beamformer, i.e. the maximum eigenvector of (19.b). On the other hand, when the number of unintended locations is greater than the number of antennas of the array, then retro-directivity may appear making the maximum directivity negative, i.e. the average power experienced at the unintended locations can be greater than the power experienced at the desired location. The upper bound does not change with the number of interfered locations in the scenario. It is worth remark that to achieve this upper bound it is necessary that the null-space of the interference matrix is not empty, i.e. the number of antennas have to be greater that the number of unintended locations.

It is remarkable that this new array factor directivity effectively goes to zero when the location of the intended coincides with the un-intended one. This was not the case for D_{VSNR}.

Finally, the dependence of the new directivity factor on the used power for transmission can be questionable, since traditional array factor directivity was independent of it, as expected for a directivity factor. Second, it is intuitive that directivity cannot depend of the path loss factor, in other words, array factor directivity should not change if the path loss factor is multiplied for the same value at all the locations. These two comments force to re-write the proposed directivity as (25); where α^2 is the path loss from the transmitter, with a single element to the desired location. Note that this

modification precludes changes on the directivity when all locations increase or decrease simultaneously their path loss with respect to the transmitter site.

$$D_{RD} = \frac{\left|\underline{b}^H \underline{h}_d\right|^2 - \underline{b}^H \underline{\underline{R}}_i \underline{b}}{\alpha^2 \underline{b}^H \underline{b} + \underline{b}^H \underline{\underline{R}}_i \underline{b}} \tag{25}$$

Furthermore, the factor α^2 can be defined in terms of the quotient between the power received P_{Rd}, at the intended location divided by the used power P_T for a single antenna, as it is shown in (26). This notion of the path-loss was used in [6] for a different scenario dealing with decentralized beamforming for regulated interference channel.

$$P_T \left|\underline{h}_d\right|^2 = P_R \quad \alpha^2 = \left|\underline{h}_d\right|^2 / n_T = P_{Rd} / P_T \tag{26}$$

V. Simulations

This section provides some simulations that support the superiority of the proposed array factor directivity.

The tested definitions are basically the D_{VSNR} and the D_{RDE}. Traditional directivity is not included in the graphics, since it does not includes any penalty for interfering pre-defined locations. Also, the D_{RD} cannot be properly compared due to its unbounded character. The two directivities included in the simulation are re-written below for the sake of clarity.

$$D_{RD} = \frac{\left|\underline{b}^H \underline{h}_d\right|^2 - \underline{b}^H \underline{\underline{R}}_i \underline{b}}{\alpha^2 \underline{b}^H \underline{b} + \underline{b}^H \underline{\underline{R}}_i \underline{b}}$$

$$D_{VSNR} = \frac{\left|\underline{b}^H \underline{h}_d\right|^2}{\alpha^2 \underline{b}^H \underline{b} + \underline{b}^H \underline{\underline{R}}_i \underline{b}} \tag{27}$$

$$with \quad \alpha^2 = \left|\underline{h}_d\right|^2 / n_T$$

The scenario considered will be the LOS scenario and, in consequence the channel vectors will be the steering-vectors of the locations with respect the phase center of the transmitting array. This implies that α^2 parameter will be always one.

In the directivity plots, the beamformers under test are the optimizers of D_{RDE} and D_{VSNR}, together with the zero-forcing beamformer, which is the optimizer of D_{RD}. These beamformers are reproduced in (28) as the solutions of a generalized eigenvalue problem, regardless VSNR and ZF have a closed form.

$$\begin{aligned}
&\left(\underline{\underline{I}} + \underline{h}_d \underline{h}_d^H\right) \underline{b}_{RDE} = \lambda_{1\max} \left(\underline{\underline{I}} + \underline{\underline{R}}_i\right) \underline{b}_{RDE} \\
&\left(\underline{h}_d \underline{h}_d^H\right) \underline{b}_{VSNR} = \lambda_{2\max} \left(\underline{\underline{I}} + \underline{\underline{R}}_i\right) \underline{b}_{VSNR} \\
&\left(\underline{h}_d \underline{h}_d^H\right) \underline{b}_{ZF} = \lambda_{1\max} \left(\underline{\underline{R}}_i\right) \underline{b}_{ZF}
\end{aligned} \tag{28}$$

Figure 2 shows the major differences among these beamformer for a scenario with a single unintended location at -1.5 degrees of the desired location. The desired location is the broadside of a 10 antennas uniform linear array (ULA).

Since all the beamformers behave similar at locations that are well separated, in order to observe differences it is necessary to use scenarios with at least one unintended location very close to the intended one. In these scenarios it can be observed that the VSNR is unable to remove the interference. Meanwhile, the ZF beamformer removes completely the interference at the expense of reduced power delivered at the intended location. The RDE beamfomer shows a tradeoff between the VSNR and the ZF beamforming, i.e. more interference attenuation to the interference than the VSNR, but more power delivered to the intended location than the ZF. It is worth remind that RDE maximizes the mutual information for the desired location versus the interfered location.

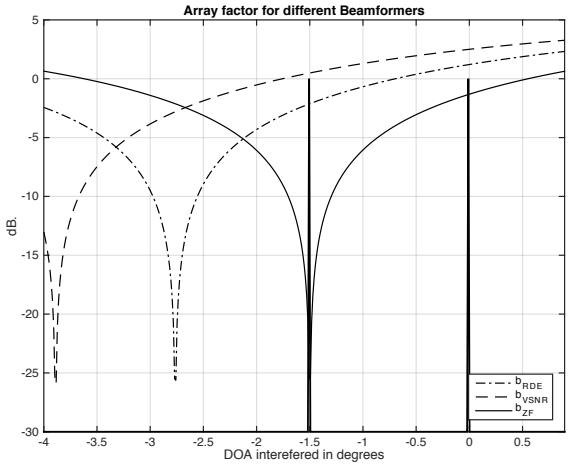


Figure 2 Beamformers' responses for a 10-antenna ULA array. Desired signal is at 0º and the un-intended location is at -1.5º.

Figure 2 shows the evolution of D_{VSNR}, for the same ULA array used in the previous figure, versus the location of the un-intended location. The intended is located at the broadside of the aperture. The unintended location varies from -3 degrees up to -0.1 degrees. Note that the VSNR beamformer, the optimizer of this directivity shows the best performance.

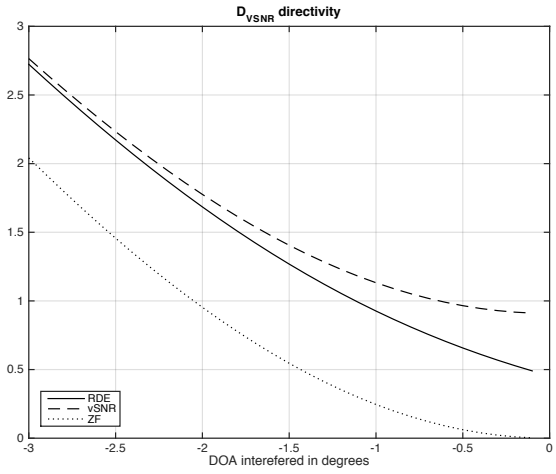


Figure 3 D_{VSNR} performance for 3 beamformers (VSNR,ZF and RDE), versus the separation of the unintended and intended location for a 10-antenna ULA.

This figure evidences the major drawback of this directivity, which is that it does not go to zero when the angles of departure are the same, i.e. when the intended and

unintended locations coincide. This come from the fact that D_{VSNR} is not a proper measure of directivity.

Finally Figure 3, shows the performance of D_{RDE}, for the same scenario that was used in Figure 2. Now, all the beamformers converge to zero directivity when both positions coincide, and, of course the maximum performance is experienced for the optimizer beamformer.

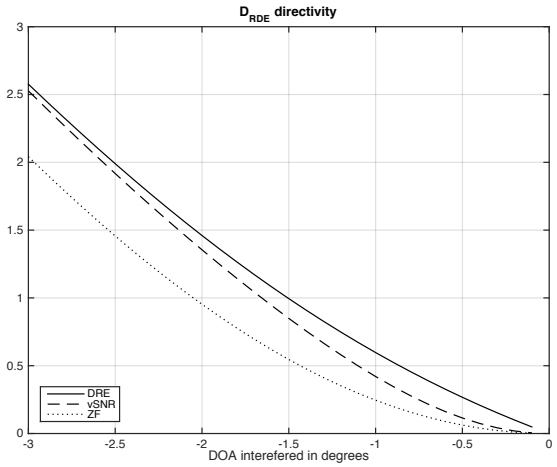


Figure 4 D_{RDE} performance for 3 beamformers (VSNR,ZF and RDE), versus the separation of the unintended and intended location for a 10-antenna ULA array.

VI. Conclusions

A new definition for array factor directivity has been introduced. With this new definition proper beamforming for transmission can be obtained from its maximization. The new definition is based, differently from the traditional virtual signal to noise plus interference alternative, in the maximization of the mutual information for the intended location with respect the mutual information to the unintended ones. This basis is more solid than that of the so-called virtual SNR. In addition, the new directivity includes the notion of retro-directivity, i.e. negative directivity. Retro-directivity occurs when the average power delivered to the unintended locations is greater than the power delivered to the intended one. As a consequence, the proposed directivity goes to zero when desired and undesired locations coincide. To sum up, the new directivity reported stays above the traditional directivity for interference- dominated scenarios in all respects.

Appendix

The generalized eigenvalues of (A.1) can be bounded thanks to the structure of the matrix of the left side of the equation.

$$\left(P_T^{-1}\underline{\underline{I}}+\underline{h}_d\underline{h}_d^H\right)\underline{b}=\lambda\left(P_T^{-1}\underline{\underline{I}}+\underline{h}_i\underline{h}_i^H\right)\underline{b} \tag{A.1}$$

The eigenvectors matrix and eigenvalues of the left term matrix are:

$$\begin{aligned}&\left(P_T^{-1}\underline{\underline{I}}+\underline{h}_d\underline{h}_d^H\right)=\underline{\underline{U}}\,\underline{\underline{D}}\,\underline{\underline{U}}^H\\&\underline{\underline{U}}=[\underline{h}_d/h_d^2,\underline{\underline{P}}_{\perp\underline{h}_d}]\\&D=diag\left(h_d^2+P_T^{-1},P_T^{-1},...,P_T^{-1}\right)\end{aligned} \tag{A.2}$$

Note that the maximum eigenvector is the interfered location vector, and the rest of eigenvectors are orthogonal to it. In addition, the desired location vector can be written as (A.3).

$$\begin{aligned}&\underline{h}_i=\underline{\underline{U}}\,\underline{v}\quad or\quad \underline{v}=\underline{\underline{U}}^H\underline{h}_i\\&\left|\underline{v}\right|^2=h_i^2\\&\left|v(1)\right|^2=h_{di}^2/h_d^2\Rightarrow\sum_{q=2}^{n_T}\left|v(q)\right|^2=h_i^2-\left(h_{di}^2/h_d^2\right)\\&with\quad h_d^2=\underline{h}_d^H\underline{h}_d\quad h_{di}^2=\left|\underline{h}_d^H\underline{h}_i\right|^2\end{aligned} \tag{A.3}$$

Using (A.2) and (A.3) in (A.1), (A.4) results.

$$\begin{aligned}&\underline{\underline{U}}\,\underline{\underline{D}}\,\underline{\underline{U}}^H\underline{b}=\lambda\underline{\underline{U}}\left(P_T^{-1}\underline{\underline{I}}+\underline{v}\underline{v}^H\right)\underline{\underline{U}}^H\underline{b}\\&\underline{\underline{D}}\,\underline{e}=\left(\lambda P_T^{-1}\underline{\underline{I}}+\lambda\underline{v}\underline{v}^H\right)\underline{e}\quad with\quad \underline{e}=\underline{\underline{U}}^H\underline{b}\end{aligned} \tag{A.4}$$

Since all the matrixes involved in (A.4) are diagonal, it is easy to arrange terms in (A.4) as indicated in (A.5).

$$\begin{aligned}&\lambda\underline{v}\underline{v}^H\underline{e}=\left(\underline{\underline{D}}-\lambda P_T^{-1}\underline{\underline{I}}\right)\underline{e}\\&\lambda\left(\underline{\underline{D}}-\lambda P_T^{-1}\underline{\underline{I}}\right)^{-1}\underline{v}\underline{v}^H\underline{e}=\underline{e}\\&\underline{v}^H\lambda\left(\underline{\underline{D}}-\lambda P_T^{-1}\underline{\underline{I}}\right)^{-1}\underline{v}=1\end{aligned} \tag{A.5}$$

The last equation in (A.5) can be written in terms of the components of vector $\underline{v}$ and the diagonal entries of the matrix as (A.6).

$$1=\sum_{q=1}^{n_T}\frac{\lambda\left|v(q)\right|^2}{d_q-\lambda P_T^{-1}} \tag{A.6}$$

Finally, using that the eigenvalues after the first one are equal, together with that the first component of vector $\underline{v}$ and its norm can be expressed as it is shown in (A.2), the functional relationship that the generalized eigenvalues satisfy is (A.7).

$$\begin{aligned}&1=\Phi(\lambda)\\&\Phi(\lambda)=\frac{\lambda P_T h_i^2\rho}{P_T h_d^2+(\lambda-1)}+\frac{\lambda P_T h_i^2(1-\rho)}{(\lambda-1)}\\&\rho=\frac{h_{id}^2}{h_d^2h_i^2}\end{aligned} \tag{A.7}$$

References

[1] Tapan K. Sarkar, Magdalena Salazar-Palma, Eric L. Mokole, Physics of Multiantenna Systems and Broadband Processing, Wiley Interscience 2008.

[2] Warren L. Stutzman, "Estimating Directivity and Gain of Antennas, " IEEE Antennas and Propagation Magazine, Vol. 40, no. 4, August 1998.

[3] Osama N. Alrabadi, Elpiniki Tsakalaki, Howard Huang, Gert F. Pedersen "Beamforming via Large and Dense Antenna Arrays Above a Clutter," Selected Areas in Communications, IEEE Journal on Vol.31 , Issue: 2 , February 2013.

[4] Karim Y. Kabalan, Ali El-Hajj, Mohammed Al-Husseini, Elias Yaacoub, "Directivity and Interference Tradeoffs with Cylindrical

Antenna Arrays," Wireless Communications and Mobile Computing Conference, 2008. IWCMC '08. International.

[5] Yifan Chen, Zhenrong Zhang, Vimal K. Dubey "Effect of Antenna Directivity on Angular Power Distribution at Mobile Terminal in Urban Macrocells: A Geometric Channel Modeling Approach," Journal Wireless Personal Communications, Volume 43 Issue 2, October 2007, Pages 389-409.

[6] Miguel Ángel Vázquez, Ana I. Pérez Neira, Miguel Ángel Lagunas, "Generalized Eigenvector for Decentralized Transmit Beamforming in the MISO Interference Channel". IEEE Transactions on Signal Processing, vol. 61, no. 4, pp. 878 - 882, ISSN 1053-587X, April 2013.

M.A. Lagunas (IEEE S'73-M'78-SM'89-F97) was born in Madrid, Spain, in 1951. He received the Telecommunications Engineer degree in 1973 from UPM, Madrid, and the Ph.D. degree in Telecommunications from UPB, Barcelona. During 1971-1973, he was a Research Assistant at the Semiconductor Lab ETSIT, Madrid. He joined UPC as Teacher Assistant in 1973. Since 1983, he has been a Full Professor at UPC. He was Project Leader of high-speed SCMA (1987-1989) and ATM (1994-1995) cable network. He is also Co-director of the first projects for the European Spatial Agency and the European Union, providing engineering demonstration models on smart antennas for satellite communications using DS/FH systems (1986) and antenna arrays for mobile communications GSM (1997). His research activity is devoted to spectral estimation, DSP on communications and array processing. His technical activities are in advanced front-ends for digital communications combining spatial with frequency-time and coding diversity. Prof. Lagunas was Vice-President for Research of UPC from 1986-1989 and Vice-Secretary General for Research, CICYT, Spain from 1994-1996. He was Member of the NATO scientific committee (97-01). Currently, he is Director of the Telecommunications Technological Center of Catalonia (CTTC) in Barcelona (http://www.cttc.es). He is an elected member of the Royal Academy of Engineers of Spain and of the Academy of Science and Arts of Barcelona. He is elected Correspondent of the Member of the Nordrhein-Wesfälische Akademie der Wissenchaften und der Künste (Dusseldorf, Germany) 2009. He received the Technical Achievement Award from Eurasip 2010 and he is Fellow Eurasip (2014). He was a Fulbright scholar at the University of Boulder, CO.(USA).

A. Perez-Neira is full professor at UPC (Technical University of Catalonia) in the Signal Theory and Communication department. Her research topics are in: multi-antenna signal processing for satellite communications and wireless and in physical layer scheduling for multicarrier systems. She has been the leader of 18 projects and has participated in over 50 (7 for ESA). She is author of 50 journal papers (20 related with Satcom) and more than 200 conference papers (20 invited). She is co-author of 4 books and 5 patents (1 on satcom). Since 2008 she is member of EURASIP BoD (European Signal Processing Association) and since 2010 of IEEE SPTM (Signal Processing Theory and Methods). She has been guest editor in 5 special issues and currently she is editor of IEEE Transactions on Signal Processing and of Eurasip Signal Processing and Advances in Signal Processing. She has been the general chairman of IWCLD'09, EUSIPCO'11, EW'14 and IWSCS'14. She has participated in the organization of ESA conference 1996 and SAM'04. She has been in the board of directors of ETSETB (Telecom Barcelona) from 2000-03 and Vicepresident for Research at UPC (2010-13). She created UPC Doctoral School (2011). Currently, she is Scientific Coordinator at CTTC. She is the coordinator of the Network of Excellence on satellite communications, financed by the European Space Agency: SatnexIV.

X. Artiga (Barcelona, 1979) obtained his M.Sc. degree in Telecommunications Engineering at Universitat Politècnica de Catalunya (UPC) in July 2006. He did his Master Thesis (with honors) at CTTC) He joined CTTC in September 2007 as a Research Engineer. Prior this, he worked as a Management Analyst at Bearingpoint Consultants (November 2006- August 2007) and in 2002, he did an undergraduate training at Nokia Spain. At CTTC, he has participated in the development of the strategic Test-bed ULAND, and in European funded projects such as FP7-BuNGee or H2020-SANSA, and in several industrial contracts (being the leader in two of them). He is a usual reviewer of IEEE Transactions on Antennas and Propagation and IEEE Antennas and Wireless Propagation Letters. His research interests include reflect-array antennas, reconfigurable antennas, Multi-antenna and Massive MIMO systems, UWB systems and radiofrequency transceivers.

Efficient Resolution Enhancement Algorithm for Compressive Sensing Magnetic Resonance Image Reconstruction

Osama A. Omer[1] and Ken'ichi Morooka[2]

[1]Department of Electrical Engineering, Aswan University, Aswan 81542, Egypt
[2]Graduate School of Information Science and Electrical Engineering, Kyushu University
744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan

Abstract **-** This paper presents the reconstruction of high resolution (HR) magnetic resonance (MR) image from very limited samples. The proposed algorithm is based on compressed sensing, which combines wavelet sparsity with the sparsity of image gradients, where the MR images are generally sparse in wavelet and gradient domain. The main goal of the proposed algorithm is to reconstruct the HR MR image directly from a few measurements. Unlike the compressed sensing (CS) MRI reconstruction algorithms, the proposed algorithm uses multi measurements to reconstruct HR image. Also, unlike the resolution enhancement algorithms, the proposed algorithm perform resolution enhancement of MR image simultaneously with the reconstruction process from few measurements. The proposed algorithm is compared with three state-of-the-art CS-MRI reconstruction algorithms in sense of signal-to-noise ratio, and full-with-half-maximum values.

Keywords **—** MRI, Wavelet Transform, Sparsity, Resolution Enhancement

1. INTRODUCTION

Sparsity has been demonstrated to be a powerful tool in several problems in last years [1]. It has been recognized that sparsity is an important structure in MR image reconstruction techniques [2-5]. It is well known that sparse signals require fewer samples than required by the Shannon-Nyquist sampling theorem. Therefore, to shorten magnetic resonance imaging (MRI) scanning time, compressed sensing is widely applied in the MRI reconstruction.

On the other hand, there are several approaches for increasing the resolution of MR images. Among these approaches, hardware improvements can directly increase the resolution of the MR images [6]. For example, for a similar signal-to-noise-ratio (SNR) value, scanners with a high value of magnetic field and a high number of coil receiver channels will produce images with higher spatial resolution and contrast. However, high magnetic field's strength affect human bodies [6]. Nowadays, most MRI scanners used for medical purposes have magnetic field value of 1.5 or 3 Tesla. Another approach to enhance the resolution of MRI is the post-acquisition image processing techniques that is super-resolution (SR) techniques.

Although there is doubt that SR is not achievable in MRI [7, 8, 9], since the Fourier encoding scheme excludes aliasing in frequency and phase encoding directions, simulation results show that SR techniques can achieve resolution enhancement in MRI [10-14]. Moreover, it is shown that for a given acquisition time the SR reconstructed images presented higher SNR when compared to images directly acquired at the same resolution [13, 14] This is of great interest for practical applications, because it offers the possibility of decreasing the acquisition time, which is often a critical parameter.

Reconstruction of HR MR image from a few samples is still challenging task in MRI reconstruction. This paper proposes a new method for enhancing the resolution for MRI using resolution enhancement technique using multi measurements. Like the work done in [12], the resolution enhancement is done simultaneously with the reconstruction process rather than being done as a post-process. However, in this paper the simultaneous resolution enhancement and reconstruction is adopted with the compressed sensing to shorten the acquisition time.

2. SPARSITY OF MRI RECONSTRUCTION

Compressed sensing focuses on reconstructing an unknown signal from a very limited number of samples. Because information such as boundaries of organs is very sparse in most MR images, compressed sensing makes it possible to reconstruct the same MR image from a very limited set of measurements while significantly reducing the MRI scan duration. In the literature, compressed sensing MRI algorithms minimize a linear combination of total variation and wavelet sparsity constrains [3, 4, 5].

TVCMRI: In [3], Ma et al. proposed the method jointly minimizing the L1 norm of the image, total variation (TV) of the wavelet coefficients, and the least squares of the error as a solution for CS-MRI. This algorithm is based upon an iterative operator-splitting framework. The cost function proposed in [3] is formulated as

$$J(y) = \|\mathbf{R}y - b\|_2^2 + \alpha\|\mathbf{\Phi}y\|_{TV} + \beta\|y\|_1 \qquad (1)$$

where **R** is a matrix representing the partial Fourier transform, y is the MR image to be reconstructed, b is the measured data in k-space, **Φ** is a matrix representing the

wavelet transformation β, α are positive weighting parameters and $\|y\|_{TV} = \sum_i \sum_j \sqrt{(\nabla_1 y_{ij})^2 + (\nabla_2 y_{ij})^2}$, where ∇_1, ∇_2 are the forward difference operators, of a variable *y*, on the first and second coordinates, respectively.

FCSA: In [4], Huang et al. proposed to jointly minimizes the L1 norm of the wavelet coefficients, total variation (TV) of the image, and a least squares of the error as a solution the for CS-MRI. The cost function proposed in [4] is formulated as

$$J(y) = \|\mathbf{R}y - b\|_2^2 + \alpha\|y\|_{TV} + \beta\|\mathbf{\Phi}y\|_1 \quad (2)$$

The minimization of TV(*y*) leads to sparsity of the gradient of y, which is the case of MR images, while minimizing $\|y\|_1$ leads to sparsity of y, which is not the case of MR images. Therefore, minimization of (2), which leads to sparsity of image gradient and sparsity of wavelet coefficients of the image, leads to better results compared to minimization in (1), which leads to sparsity of gradients of wavelet coefficients and sparsity of images values, as will be shown in simulation results.

WaTMRI: In [5], the quad-tree sparsity constraint is combined with the sparsity of wavelet coefficients and sparsity of gradient image. The cost function of this algorithm is formulated as

$$J(y) = \|\mathbf{R}y - b\|_2^2 + \alpha\|y\|_{TV} + \beta\left(\|\Phi y\|_1 + \sum_{g\in G}\|\Phi y_g\|_2\right)$$

where G indicates the set of all parent-child groups and y_g is the data belonging group G.

CS-MR imaging is interested in low sampling ratio. In [3,4,5], authors follow the sampling strategy that is randomly choose more Fourier coefficients from low frequency and less on high frequency.

3. Sparsity-Based HR MRI reconstruction

Inspired by the success of the minimization of L1-norm and TV in CS-MRI reconstruction, we design the reconstruction of HR CS-MRI by fusing multi measurements in the proposed HR CS-MRI reconstruction model. In the proposed model called CS-MRISR, we propose to penalize the least square of error measure, sparsity of wavelet coefficients and sparsity of gradient image. The proposed cost function is formulated as

$$J_2(y) = \sum_{k=1}^{N}\left[\|\mathbf{RDBF}_k y - b_k\|_2^2\right] + \alpha\|y\|_{TV} + \beta\|\mathbf{\Phi}y\|_1 \quad (3)$$

where **D** is the sampling operator, **B** is the blurring operator and $\mathbf{F}_k$ is the warping operator for k-th image. It is commonly assumed that the point spread function (PSF) induced by the MRI acquisition process is space-invariant, so that we used the same operator **B** for all images.

To fasten the proposed algorithm, we utilize the composite splitting algorithm [15]; 1) Splitting variable y into two variables x and z, 2) Performing operator splitting over each of the two variables independently, and 3) Obtaining the solution y by linear combination of z and x. Therefore, the optimization problem can be divided into three sub-problems that alternatively solved;

1) Minimize least square problem:

$$\hat{y} = \arg\min_y \sum_{k=1}^{N}\frac{1}{2}\left[\|\mathbf{RDBF}_k y - b_k\|_2^2\right] \quad (4)$$

2) De-noising:

$$\hat{x} = \arg\min_x \left\{\frac{1}{2}\|x - \hat{y}\|_2^2 + \alpha\|x\|_{TV}\right\} \quad (5)$$

3) Sparsity constraint in the wavelet domain

$$\hat{z} = \arg\min_z \left\{\frac{1}{2}\|z - \hat{y}\|_2^2 + \beta\|\mathbf{\Phi}z\|_1\right\} \quad (6)$$

The reconstructed MR image is the weighted sum of the de-noised term and the constrained wavelet coefficients

$$y = \frac{\hat{z} + \hat{x}}{2} \quad (7)$$

Finally, at each iteration, values of *y* are projected in the reasonable range of MR images which is [0,255] for 8-bit MR images.

4. Simulation

4.1 Setup

In the simulation, we used 4 low-resolution (LR) measurement data that are sensed from 128 × 128 positions. The resolution enhancement factor is used as 2 in each direction. The relative shift of the simulated object to generate LR measurements is assumed to be known. The fewer measurements we samples, the less MR scanning time is need. So MR imaging is always interested in low sampling ratio cases. The sampling ratio is fixed to be approximately 20%. We follow the sampling strategy of previous works [3, 4, 5]. All measurements are mixed with 0.01 white Gaussian noise. We conduct experiments on two images, namely, "Synthetic Image" and "Brain Image".

4.2 Simulation results

The proposed CS-MRISR is compared with the following methods; 1) total variation L1 Compressed MRI (TVCMRI [3]) using system matrix that exhibits LR grid (128 × 128), 2) TVCMRI using system matrix that exhibits HR grid (256 × 256), 3) Fast Composite Splitting Algorithm (FCSA [4]) using system matrix that exhibits LR grid (128 × 128), 4) FCSA using system matrix that exhibits HR grid (256 × 256), 5) Wavelet Tree Sparsity MRI (WaTMRI [5]) using system matrix that exhibits LR grid (128 × 128), and 6) WaTMRI using system matrix that exhibits HR grid (256 × 256). To evaluate these algorithms SNR, full-width-half-maximum (FWHM) and visual results are used.

4.2.1 Visual Results

Figure 1a shows the original phantom image. The reconstructed MR image using the proposed algorithm is shown in Fig. 1b. The reconstructed MR images using

ISBN: 978-1-61804-318-4

algorithms FCSA, TVCMRI and WaTMRI are shown in Figs. 1c, 1e and 1g, respectively. The reconstructed MR images on HR grid using algorithms FCSA, TVCMRI and WaTMRI are shown in Figs. 1d, 1f and 1h, respectively. From these figures we can see that the proposed resolution enhancement algorithm improves the quality of the CS-MR image (see Figs. 1b and 2b) compared to the conventional CS-MRI algorithms (see Figs. 1c, 1e, 1g, 2c, 2e and 2g). Also, the reconstructed HR MR image using proposed method has better quality than the reconstructed MR image with HR grid obtained by TVCMRI. On the other hand, the proposed method has as quality as the reconstructed MR image with HR grid obtained by FCSA and WaTMRI.

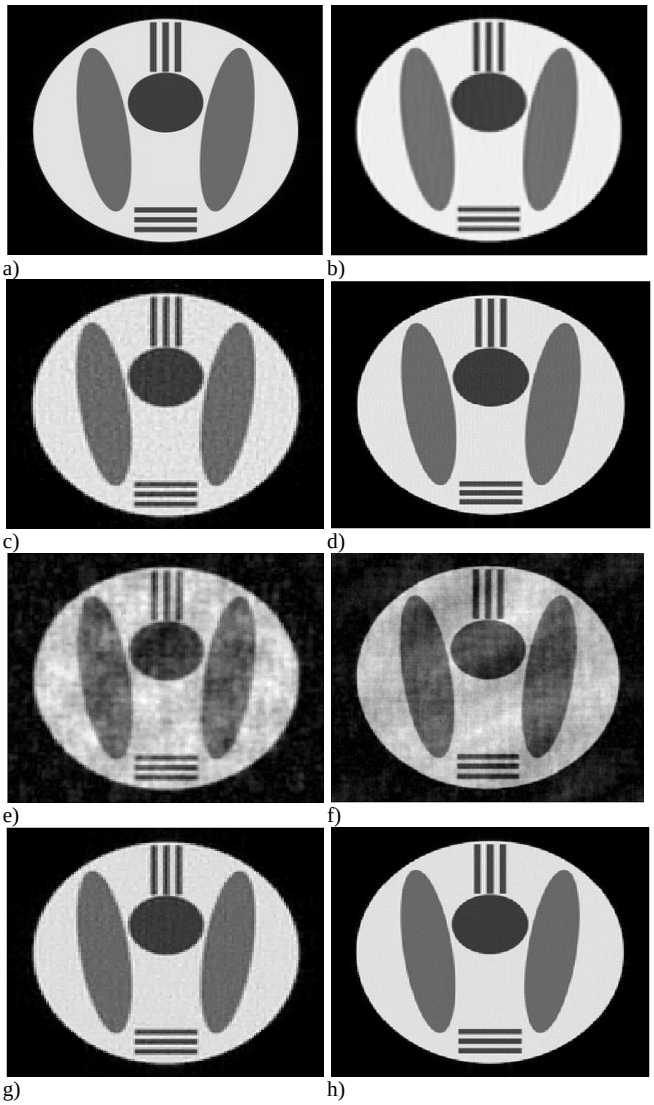


Figure 1. original test image #1, b) Proposed CS-MRI reconstruction c) LR FCSA-based MRI reconstruction, d) LR FCSA-based MRI reconstruction on HR grid, e) LR TVC -based MRI reconstruction, f) LR TVC-based MRI reconstruction on HR grid, g) LR WaTMRI reconstruction, h) LR WaTMRI reconstruction on HR grid

The results of the other experiment are shown in Fig. 2. This example confirm the results in the first example, that is the proposed algorithm can enhance the quality of the CS-MRI compared to the reconstructed LR MR images using algorithms FCSA, TVCMRI and WaTMRI. The proposed algorithm can reconstruct a similar quality as that obtained by algorithms [4] and [5].

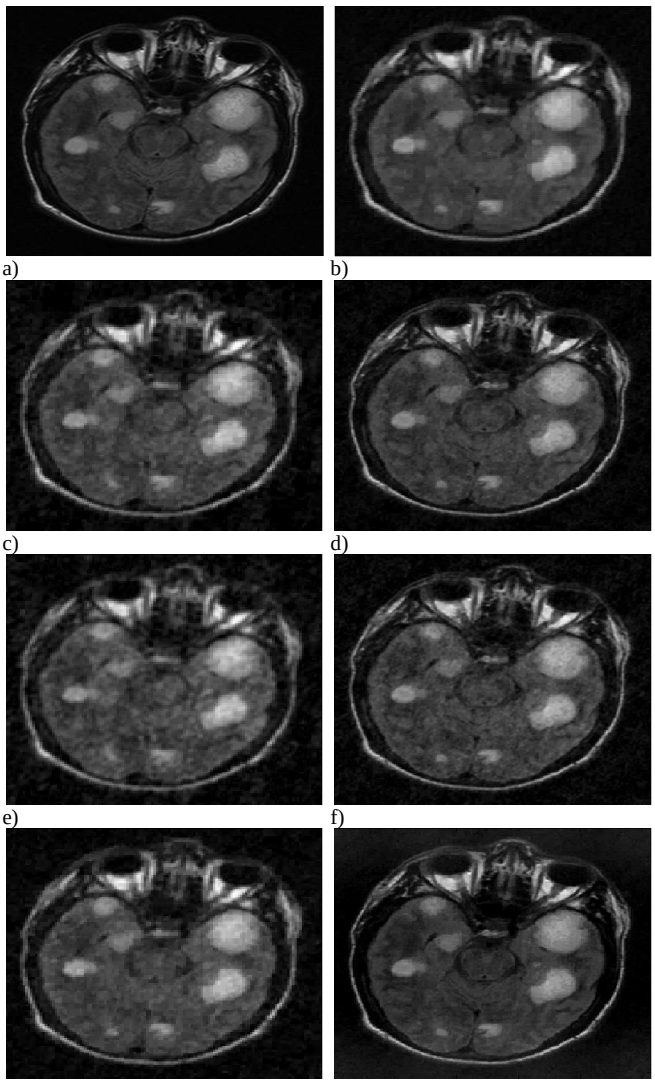


Figure 2. a) original test image #2, b) Proposed CS-MRI reconstruction c) LR FCSA-based MRI reconstruction, d) LR FCSA-based MRI reconstruction on HR grid, e) LR TVC -based MRI reconstruction, f) LR TVC-based MRI reconstruction on HR grid, g) LR WaTMRI reconstruction, h) LR WaTMRI reconstruction on HR grid

4.2.2 Objective results

The FWHM values for the PSF function of the reconstructed MR images is shown in Fig. 3. From this figure it can be shown that the proposed algorithm results in low FWHM value which indicate higher resolution compared CS-MRI reconstruction algorithms proposed in [3], [4] and [5].

Another measure for the quality that can demonstrate the efficiency of the proposed algorithm is shown in Table 1. This table can show the higher SNR for the proposed algorithm compared to CS-MRI algorithms. For fair comparison, the reconstructed MR images by using algorithms in [3], [4] and [5] are interpolated to be compared with the original HR images. The plot of cost

function versus iteration number is shown in Fig. 4 which can show the convergence of the proposed algorithm.

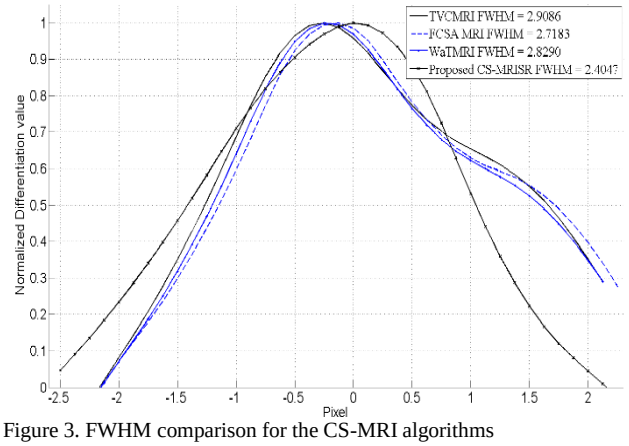


Figure 3. FWHM comparison for the CS-MRI algorithms

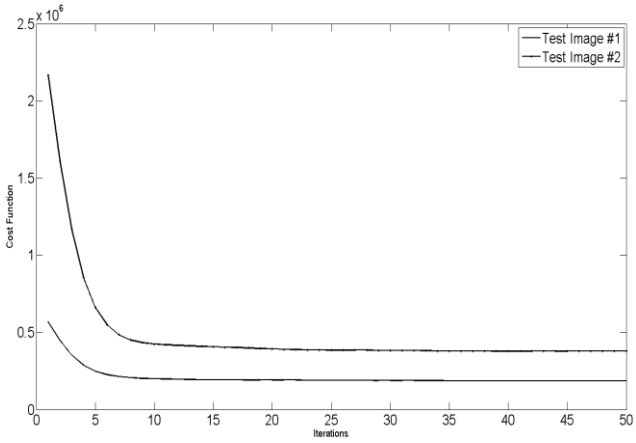


Figure 4. Convergence of the proposed algorithm

Table 1: SNR comparison for different CS-based MRI reconstruction algorithms

	TVCMRI	FCSA	WaTMRI	Proposed
Image#1	12.0783	15.7975	15.9910	16.3744
Image#2	10.3166	11.4029	11.9797	12.4043

5. CONCLUSION

We proposed a CS-MRI reconstruction algorithm that reconstructs a HR MR image from multi LR measurements. The proposed algorithm adopts the idea of compressed sensing with the resolution enhancement algorithm. The proposed algorithm reconstructs the HR MRI directly from the LR measurements. Based on the simulation results, the proposed algorithm can efficiently reconstruct MR images from very low samples, with sampling ratio about 20%. The proposed algorithm outperforms three state-of-the-art CS-MRI reconstruction algorithms in sense of SNR, FWHM and visual results.

6. REFERENCES

[1] Donoho, D. "Compressed sensing," *IEEE Trans. on Information Theory* 52(4):1289-1306, 2006.

[2] Lustig, M., Donoho, D. & Pauly, J. "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine* 58(6):1182-1195, 2007.

[3] Ma, S., Yin, W., Zhang, Y. & Chakraborty, A. "An efficient algorithm for compressed MR imaging using total variation and wavelets. In In Proc. of the *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, 2008.

[4] Huang, J., Zhang, S. & Metaxas, D. "Efficient MR Image Reconstruction for Compressed MR *Imaging," Medical Image Analysis* 15(5):670-679, 2011.

[5] Chen Chen and Junzhou Huang, "Compressive Sensing MRI with Wavelet Tree Sparsity", In Proc. of the 26th *Annual Conference on Neural Information Processing Systems* (NIPS), Nevada, USA, December 2012

[6] Eric Van Reeth, Ivan W. K. Tham, Cher Heng Tan and Chueh Loo Poh, "Super-resolution in magnetic resonance imaging: A review," *Concepts in Magnetic Resonance Part A*, Vol. 40A, Issue 6, pages 306–325, November 2012,

[7] K. Scheffler, "Superresolution in MRI?" *Magnetic Resonance in Medicine*, vol.48, p.408, 2002.

[8] S. Peled and Y. Yeshurun, "Superresolution in MRI – Perhaps sometimes," *Magnetic Resonance in Medicine*, vol. 48, p. 409, 2002.

[9] Uecker M, Sumpf TJ, Frahm J. "Reply to: MRI resolution enhancement: how useful are shifted images obtained by changing the demodulation frequency?," *Magnetic Resonance in Medicine* 66:1511–1512, 2011.

[10] Tieng QM, Cowin GJ, Reutens DC, Galloway GJ, Vegh V., "MRI resolution enhancement: how useful are shifted images obtained by changing the demodulation frequency?," *Mag. Res. in Medicine*, 65:664–672, 2011.

[11] S. Peled and Y. Yeshurun, "Superresolution in MRI: Application to human white matter fiber tract visualization by diffusion tensor imaging," *Magnetic Resonance in Medicine*, vol. 45, pp. 29-35, 2001.

[12] O. A. Omer, "High Resolution Magnetic Resonance Image Reconstruction in K-Space," *ICIC Express Letters, Part B: Applications*, vol.5, No. 6, pp. 1659 – 1666, Dec. 2014.

[13]Plenge E, Poot D H J, Bernsen M, Kotek G, Houston G, Wielopolski P, et al., "Super-resolution reconstruction in MRI: better images faster?" In: Haynor DR, Ourselin S, eds., *SPIE Medical Imaging*, Vol. 8314. Bellingham, WA: SPIE Press, P83143V, 2012.

[14] Scherrer B, Gholipour A, Warfield SK. "Superresolution reconstruction to increase the spatial resolution of diffusion weighted images from orthogonal anisotropic acquisitions," *Medical Image Analysis*, No. 16, pp. 1465–1476, 2012.

[15] Beck, A. and Teboulle, M., "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences* 2(1):183-202, 2009.

Analysis of cloud computing usability for teleworking

H. Mohelska, J. Ansorge

***Abstract*—** Cloud computing in recent years is a very important means of ICT. Its technical and especially economic characteristics can change the view of the future in an organisation. In this paper, these benefits are examined and correlated with teleworking. It represents an important opportunity - flexibility for an organisation as well as employees. The benefits of implementing teleworking affect not only employees and employers, but the whole company. It will reduce the burden posed by the transport operator. It also decreases the burden of the high concentration of people in urban centres. This creates employment opportunities for people with a remote residence and also opportunities for mothers with children and for people with disabilities. The aim of this paper is to describe the selected resources (from the field of soft and hard factors) for the introduction of telework and to show the current status of the use of teleworking in the Czech Republic on the results of the research. The paper should serve as a basic information tool in decision-making on whether to allow a new or different way of working in a company.

***Keywords*—** Benefit of implementing, Cloud computing, Telecommuting, Teleworking

I. Introduction

Teleworking means working off-site, generally via a computer, and the results are transmitted via communication technologies or work on assignments is carried out directly on a remote server using an Internet connection. The definitions of teleworking do not always fully correspond. According to a major IDC American company, we can speak about teleworking when an employee works at a distance for three or more days a month. (O´Brien, 2011). The European Telework Development Program supported by the European Commission uses the following definition "Teleworking is work that utilises computing and communication technologies for the opportunity to work off-site" (ETSI, 2014).

The claim that teleworking is working from home using a computer frequently occurs in the Czech media. However, this is not entirely accurate, this description fits more for the so-called homeworking (sometimes also called "home office"). Homeworking is work which is directly in the residence of an employee. Teleworking has multiple forms and homeworking is just one of them. Limitation of physical presence in the workplace may be partial or almost full, depending on the type of employer and the work performed. In this case we are talking about partial or full teleworking. Typical examples of telework include research and development work, consultancy, processing of various analyses and projects, work of sales representatives, auditors and journalists, hotline services, accounting, engineering and design work, costing, order processing, invoicing and statistics, proof-reading, scientific and literary activities, editing, informatics services, graphic and managerial work.

Many scientific papers have been published on the topic of cloud computing, but the definition and delimitation of cloud differs. As a basic document the paper used a special publication (Badger, Grance, Corner, & Voas, 2014) from the US National Institute of Standards and Technology (abbreviated as NIST), which very accurately and precisely defines the basic properties of cloud computing, and also very often other authors reference to it. They deal with cloud computing from the perspective of business as well as the security and deployment itself; namely (Hurwitz, 2010), (Mc Donald, 2010), (Hugos & Hulitzky, 2010). For the definition of teleworking, primarily the book by (Martoch,J., 2013) was used, which is exposed on the portal www.pracenadalku.cz, and other sources.

II. METHODOLOGY

Based on a detailed analysis of the possibilities of cloud computing recommendations for implementation in small and medium-sized enterprises, which is considering the introduction of telework, are described. These recommendations are applied to the example of a small company, which is currently considering whether to invest in the rehabilitation of an existing infrastructure or to go through cloud computing. The paper focuses only on technical and economic aspects. The so-called soft factors such as employees' motivation, morale and so on, are not addressed in this paper, although marginally mentioned.

III. Theoretical Framework and Approaches to Cloud computing

Cloud computing, often abbreviated to only cloud, is currently still a hotly debated concept. The rise in popularity of this concept began roughly in the second half of 2007. The rise of the term of cloud computing was discussed at such a rate that over time it began to be regarded as a buzzword. (Houser, 2008) Over the last few years a lot has changed and the concept of cloud computing has naturalised in the parlance of ICT professionals. Even though it is a distinct area of ICT,

even today, it is not easy to grasp and clearly explain this concept. Therefore, there is still a need to further define the meaning of those words. As an evidence of this need may also be the slightly ironic term of cloud washing, which emerged mainly due to overuse of the term of cloud computing and its unclear significance. It is an analogy to the English notion of brain washing. Cloud washing mocks the situation on the market as the marketing departments of ICT companies began to use the attribute of cloud for such products/services which from the perspective of the fundamental principles of cloud computing cannot be identified as such. (BABCOCK, 2015)

IV. Service Oriented Architecture (SOA)

In the past, business processes often adapted to the information systems design. Although it seemed logical and such an approach yielded the desired results, business processes became more static, because they were conserved by capabilities of these information systems. Any modification caused by a change in the business process entailed additional cost. In the long term, however, a company must adapt to the market, not its information systems. (Mohelska, Kozel, 2010). The plural is deliberate, since a company usually uses more than only one system. Other software is used e.g. in production, another for accounting and one for customer support. Since it is important that these systems mutually exchange information, they must be connected to each other. These connections are then often dedicated and gradually growing to the actual requirements. SOA is one of the founding fathers of cloud computing. Cloud computing service also arises from the needs of the business, it is standardised and for communication purposes it uses platform-independent means. Despite these common features, it is not true that cloud computing has replaced SOA or completely integrated it. The most fundamental characteristics of cloud computing is that it is always a form of service thanks to which ICT can perform business needs better and easier. Companies use this service in their strategic plans. "Service is any act or performance that one party may provide to another. By its essence, a service is intangible and does not result in the ownership of anything. It may or may not be tied to any physical product." (Kotler & Keller, 2007). For a basic breakdown of cloud computing, the definition by NIST is used. The NIST organisation itself, which falls under the US Department of Commerce, emphasises that due to the continuous development it is not possible to fully define the area which covers cloud computing. (Badger, Grance, Corner, & Voas, 2014) Still, there are some generally recognised fundamental characteristics that cloud computing solutions must meet. These characteristics are closely described by NIST in their review (Badger, Grance, Corner, & Voas, 2014):

- On-demand self-service – the user can unilaterally adjust the level of services provided (e.g. the size of storage space, number of user accounts, CPU performance), without the need of this requirement to demand direct intervention of the provider.
- Broad network access - a cloud-based solution is available through a telecommunication network using standard mechanisms that allow the user to use a diverse plethora of devices. In practice, this means that it is possible to use any device with Internet access - starting with e.g. camera, desktop computer and other devices.
- Resource pooling - users share the computational capacity of the provider of cloud solution between themselves. The user does not know the exact location of the provider's service – they do not know in what housing the server with data is located. However, they have information about the state of which service is offered, if applicable, and in which data centre the servers are located. The example of shared resources is storage space for data, memory space, network connectivity, computing power.
- Rapid elasticity - the ability to instantly adapt a service to demand (ideally automatically). Resources that (intentionally) appear to be unlimited are reserved by users to virtually any extent and whenever they need it.
- Measurability of service - cloud-based systems automatically control and adapt the use of resources by appropriate measures (typically the size of storage space, processor time, bitrate, number of users). This monitoring, controlling and evaluating of use, in turn ensures the transparency of the service for both parties - user and provider. A slightly different angle is provided by (Hugos & Hulitzky, 2010), by referring to these three basic characteristics of cloud computing:

1. Virtually unlimited computing resources - resources in the form of computing power, data storage space and adding additional users on request indicates a high degree of agility and scalability according to the needs of business.
2. No long-term commitments - computing resources are available immediately and can be used as long as they are needed. Once they are not needed, they are "returned" in monthly, daily or even minute intervals.
3. Pay-as-you-go cost structure - because there are no long-term commitments, the price for cloud computing services are billed flexibly depending on how the service is actually used. Cloud computing with its principles causes the user's abstraction from the physical levels (telecommunication networks, hardware) as they are solely provided by a cloud solution provider.

V. Theoretical Framework and Approaches to Teleworking

Work is commonly understood as an activity for which the person, who has done the work, is rewarded. This person -

worker - is motivated by the reward because through it they wish to meet their desires and needs. The reward is in the form of financial and non-financial. Organisations hire a labour force for a salary or wage (or financial reward) to achieve their goals. For this purpose and their own interest they have to ensure appropriate conditions for workers. The most basic requirement is the place where the work is performed for the organisation - the workplace. In a large number of organisations (especially in the tertiary sector), with the advent of computer technology, the workplace has gradually shifted to the office desk with a computer and a telephone. If the work can be performed just by a computer and a phone, nothing prevents a worker from carrying out their work outside the office, i.e. remotely. If a job allows, then telework also takes the form of non-financial benefits, when the worker can more efficiently fulfil their personal goals and still continue to receive financial reward. In September 2010, the study of "ICT and competitiveness of the Czech Republic" was published. The findings of this study indicate that part of the infrastructure of competitiveness is among others, labour market efficiency. The basic parameters by (Voříšek & Novotný, 2010) include:

- availability of labour with appropriate qualifications,
- acceptable labour cost,
- labour mobility.

When these parameters are juxtaposed, allowing teleworking seems to be an acceptable solution for their fulfilling. Another factor is the effort of organisations to more efficiently use the funds expended as in Europe (not only) the pressure to reduce costs is even bigger due to the protracted economic crisis.

VI. Teleworking - definition

Telework, teleworking, telecommuting and working online are synonyms that express the way of performing work, when individual workers are not physically present at one workplace (typically in office). They communicate and collaborate in real time, but remotely using information and telecommunication technologies. The place, from which they can work, can be completely arbitrary. (Martoch, 2012) Often telework occurs without the worker being able to effectively use this fact. In multi-national corporations it is normal that the teams are international and the staff never have to physically meet. Such a situation can be used so that a worker is allowed to do their work outside of the office or other premises of the company, i.e. to work remotely.

VII. Working from home – definition of the concept

Concepts of working from home, homeworking and home office are a sub-set of telework, with the difference that it is temporary or permanent work at home of an employee (home office, garden, terrace, work area, etc.). The work from home in the concept, which is described here, does not include the traditional crafts or activities such as knitting, crocheting, collation of envelopes or other similar house chores. The main prerequisite of working from home is performation in real time with the help of information technologies (Martoch J., 2013)

VIII. Results and Discussion

The main objective of this research was to determine the extent of the use of alternative forms of working in companies on the Czech market. Furthermore, the benefits and barriers to the use of these forms of working were examined. The research tried to find out the degree of formalisation in the use of alternative forms of working and their targeting to different groups of employees. Furthermore, planned development of alternative forms of working in companies was charted. (Flexibilní trh práce, 2011)

The research was carried out in March 2011 on a sample of 855 respondents, who were selected by quota sampling, the target person was the person responsible for human resources. The research was conducted by telephone (CATI). The results were sorted by the size of companies:

1. size category - up to 19 employees
2. size category - 20-99 employees
3. size category - 100-249 employees
4. size category - 250 and more employees

The distribution of the sample corresponds to the natural distribution of companies in the Czech Republic, in four main segments:

1. public administration
2. services
3. business
4. industry

IX. The use of flexible work arrangements

Flexible forms of work are used by 78% of the companies surveyed. The best known form of flexible employment is part-time (60%), followed by flexible working hours (45%), third place in the rating belongs to the combination of working at the workplace and working from home, which is used in a surprisingly small percentage (14%). (Flexibilní trh práce, 2011)

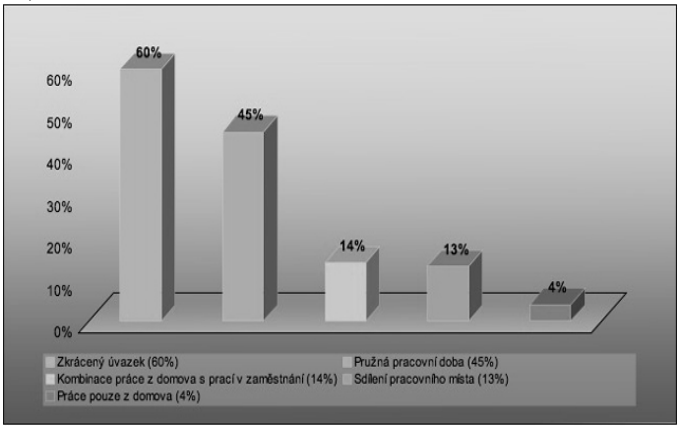


Fig. 1: Use of flexible employments in the Czech Republic

A surprising discovery is that part-time work is widely used in the public sector, where it is used by up to 86% of the surveyed firms. Followed by the service sector, business sector and companies operating within the industry. However, the differences are not so high. The service sector is slightly at the forefront in the use of the combination of working from home and from workplace. The combination of workplace and working from home is not used much on the Czech market.

The most recent survey of telework was published by the portal of www.pracenadalku.cz (Martoch J., 2013) It took place from January to April 2013 and was attended by 708 respondents, 39% men and 61% women aged 21-60 years. More than half of the respondents were recruited from the executive ranks (a total of 345, of which 57% were men and 43% women). In terms of size, 16% of the respondents were self-employed and 57% of the respondents were from organisations with 1-200 employees and 24% from organisations with more than 200 employees. The survey approached the aforementioned website visitors; a more precise description of the addressed sample is not available.

The survey unfortunately focused solely on the utilisation rate of means for telework and did not deal with using more and more frequent and sometimes very necessary personal contact. Perhaps this fact mostly reflected in the means used to communicate with customers and suppliers, where personal appointments are completely absent in the options. Quite obviously, the survey then found that phone is the most used for communication (93%), followed by email (90%). The result of communication within an organisation was also very similar.

Specifically, working from home is also a concern for the CSO Statistical Yearbook. (Czech Statistical Office, 2014), which summarises statistical reports covering all sectors of the national economy.

The chart shows noticeable slow growth of the organisations up to 2011 that allows this kind of work. Although working from home is possible overall in 25% of the monitored companies, only approximately 3% of employees work in this way.

X. Conclusion

Organisations with their focus are often suitable for the introduction of telework, but they do not use this option too often. Survey results also show that the opportunity for telework grows with the size of the company, which does not mean that the absolute number of such workers corresponds to the ratio. On the contrary, it is not clear what the biggest obstacle is for organisations for greater expansion of telework. The most frequent reasons why they do not support more of this kind of work, is the fear of lower efficiency of workers. (Mohelska, 2010)

Technical resources that are needed for teleworking are not inherently difficult. However, if they are to be used for the purposes of a private company, which moves in a competitive market, or a national organisation that needs to protect sensitive data, the entitlements to their application are huge. In order to make telework possible, it is necessary to ensure (Martoch M., 2014):

- Communication - individual needs to pass the outputs of their activities.
- Flexibility - the availability of resources from anywhere, anytime; timeliness of data.
- Security - protection of systems and data stored therein from loss, destruction, theft or wiretapping.
- Checking - means for monitoring and measurability.

Checking depends to a large degree on the ability of a manager. For the purposes of monitoring and measurability of performance of a worker, however, the resources that are also offered by cloud computing can be used. While teleworking, the costs for both the worker and the organisation are reduced. From the perspective of the worker, there are savings in funds spent on commuting to workplace. A worker also reduces the inefficiently used time that must be devoted to this activity on a daily basis. (Mohelska, 2012)

Organisations can save on office rental and associated costs (cleaning, furniture etc.). Organisations offer flexible jobs and provide workers with more autonomy, and simultaneously expand the range of potential candidates for a job. Indirectly, it is possible to occupy better qualified person or it increases the chance of finding more people within the same budget. In order to maintain trust between worker and organisation, it is necessary to introduce supervision and control system. This is also related to the change in the way of staff management and reward system. With this change, costs for the organisation can arise, but they are largely one-off in nature. If telework is associated with changes, it is also important to think about the awareness that workers are familiar with the terms of the modified relationship with the organisation and the operation of the means that must be used for job performance.

Czech legislation does not restrict telework in any way, if both parties agree. Otherwise, the employee must perform their work in the place that the employer caters for this purpose.

Acknowledgment

The paper was written with the support of the specific project grant "Determinanty Ovlivňující Pracovní Spokojenost" granted by the University of Hradec Králové, Czech Republic.

References

[1] Babcock, C. (2015). Worst Cloud Washers Of 2011 - Cloud-computing - Infrastructure as a Service [online]. [cit. 2015-03-02].

[2] Badger, L., Grance, T., Corner, R. P., & Voas, J. (2014). Cloud Computing Synopsis and Recommendations. Načteno z The National Institute of Standards and Technology: Dostupné z:http://csrc.nist.gov/publications/nistpubs/800- 146/sp800-146.pdf

[3] Český statistický úřad. (2014). Získáno: Enterprises and employees using selected information technologies:http://www.czso.cz/csu/2014edicniplan.nsf/kapitola/320198-14-r_2014-2100

[4] ETSI. (2014). Získáno 1. February 2015, z Initial analysis of standardization requirements for Cloud

services:http://www.etsi.org/website/document/tr_102997v010101p.pdf
[5] Flexibilní trh práce. (2011). Načteno z Flexibini.cz: http://www.flexibilni.cz/aktuality/
[6] Hugos, M. H., & Hulitzky, D. (2010). Business in the Cloud: What Every Business Needs to Know About Cloud Computing. 1. vyd. B.m.: Wiley.
[7] Hurwitz, J. (2010). Cloud computing for dummies. Hoboken: Wiley Publishing.
[8] Kotler, P., & Keller, K. L. (2007). Marketing Management. Prague: Grada Publishing.
[9] Martoch, J. (2013). Práce na dálku v ČR. Načteno z Průzkum Práce na dálku: www.pracenadalku.cz
[10] Martoch, M. (2014). Práce na dálku - Jak chytře zvýšit konkurenceschopnost organizace. Načteno z Práce na dálku.
[11] Mc Donald, K. T. (2010). Above the clouds managing risk in the world of cloud computing. Ely: IT Governance Publication.
[12] Mohelska, H. Options of access to internal systems in companies. AWER Procedia information technology and computer science. 2012, 2/2012(Nov. 2012), s. 92-95. ISSN 2147-5105.
[13] Mohelska, H. Effect of Technological Innovation with Changes on the Internal Environment of an Organization. In Recent Researches in Applied Economics: Proceedings of the 3rd World Multiconference on Applied Economics, Business and Development (AEBD11). Iasi, Romania, 2011, s.13-16. ISBN 978-1-61804-009-1
[14] Mohelska, H., Kozel, T. Modely firem s mobilně orientovanou architekturou. E+M Ekonomie a Management, 2010, roč. 13, č. 4, s. 135 – 142. ISSN 1212-3609.
[15] Mohelska, H. The use of mobile ICT devices in small and medium-sized companies in the Czech Republic. In Communication and management in technological innovation and academic globalization. Athens: World scientific and engineering academy and society, 2010, s. 113-116. ISBN 978-960-474-254-7.
[16] O´Brien, A. (May 2011). Security First: Buckle Up Before Entering the Telework On-Ramp. Načteno z IDC Community: https://idc community.com/government/smart_government/security-first-buckle-up-before-entering-the-telew
[17] Voříšek, J., & Novotný , O. (2010). Studie ICT a konkurenceschopnost. Načteno z B.m. Česká společnost pro systémovou integraci: Získáno z: http://www.cssi.cz/cssi/system/files/cssi/Studie_ICT_a_konkurencescho pnost_CR_201010 03.pdf. In: [online]. [cit. 2015-03-02].

Information technology in insolvency proceedings

Jan Plaček, Luboš Smrčka, Jaroslav Schönfeld

***Abstract*—**The work focuses on the potential implementation of information technologies in the area of insolvency proceedings and the expected impact of prospective implementation of modern technologies on the efficiency of insolvency proceedings. The foundation of the work is an analysis of the situation in specific countries and own surveys realized in the economic environment of the Czech Republic. The authors further demonstrate the possibilities which information technology and public sharing of data offer for increasing the efficiency of insolvency proceedings. The study also provides information on existing projects on this area in the Czech Republic and on certain surveys which generally focus on the possibilities of expanding the field of activity of information technologies in the area of enforcing receivables as a whole.

***Keywords*—**Enforcing receivables, Forfeiture, Insolvency, Insolvency register.

I. ENFORCEMENT OF RECEIVABLES AS AN ECONOMIC PROBLEM

It is usually inferred that enforcement of receivables is primarily the province of the area of general enforceability of law and this area is therefore not considered to be a space for economic research, but rather the natural territory of law – the theory of law in this connection. However, it in fact applies that enforcement of receivables is primarily an economic problem, although it is handled in a specific legal environment. [1] This is given by the fact that modern states decided to regulate the enforcement of receivables as early as several hundred, and in many cases, even several thousand years ago. It is clear that societal and social reasons led to this, inasmuch as it was necessary, on the one hand, to create mechanisms for creditors which could serve as a support in the enforcement of their receivables, as the absence of clear legislative regulation would in principle leave creditors with no other alternative of recourse against creditors than resorting to violent means. On the other hand, however, it was necessary to secure for the debtor fundamental protection of their mental and bodily integrity, for in the event of non-regulated enforcement numerous excesses or pressure on the debtor based on threats of physical violence would clearly occur,[1] to say nothing of the high probability of a situation in which enforcement would indeed be accompanied by violence – whether physical or psychological. We now consider the regulation of enforcement of receivables to be an integral aspect of the most fundamental regulations by which the behaviour of every individual in society is adjusted.

Possible and permitted means of enforcement are thus defined by the state power, and two principle methods are at issue in developed countries. The first is individual enforcement, where a lawsuit between a single creditor and a debtor is concerned, and this creditor lays claim at a court (or arbitration) to recognition of the right to enforce a receivable further. This usually means that it acquires a forfeiture title, i.e. confirmation of the fact that forcible means may be used against the debtor – usually forfeiture of its property. Theoretically, several creditors could, in mutual agreement, utilize this individual method, although this is not standard in practice. However, cases where several distrainments against a debtor's property (initiated by a larger group of creditors) are in progress simultaneously are more frequent – these distrainments are realized concurrently, but not in a coordinated manner.[2]

The second is collective enforcement – the method in which two or more creditors proceed together. This collective method is enforced by law, for one of the creditors will file an insolvency proposal when it is of the impression that its receivable has not been covered and there is, from its perspective, the danger that the creditor's viable property would be acquired by a limited number of other creditors. Further individual enforcement is impossible after the proposal is filed, although within the bounds of insolvency proceedings creditors are satisfied according to the character of their receivables, and given the identical character thereof, in an essentially proportional manner.

The article is processed as one of the outputs of the research project *"Research of insolvency practice in the CR, with the aim of forming proposals for changes in the legislation that would enable increased yields from insolvency proceedings for creditors, which would contribute towards increasing the competitiveness of the Czech economy"*, registered at the Technological Agency of the Czech Republic (TA CR) under the registration number TD020190.

J. Plaček is an assistant at the University of Economics, Prague, Faculty of Business Administration (phone: 420- 224-098656; fax: 420- 224 098 649; e-mail: jan.placek@gmail.com).

L. Smrčka is an associate professor at the University of Economics, Prague, Faculty of Business Administration (phone: 420- 224-098656; fax: 420- 224 098 649; e-mail: smrckal@vse.cz).

J. Schönfeld is an assistant at the University of Economics, Prague, Faculty of Business Administration (phone: 420- 224-098656; fax: 420- 224 098 649; e-mail: jaroslav.schonfeld@vse.cz).

[1] We could, for the purposes of placement into the general cultural context, here call to mind William Shakespeare's *The Merchant of Venice.*

[2] This is a standard state of affairs among natural persons who are not entrepreneurial subjects.

ISBN: 978-1-61804-318-4

II. THE STATE OF COMPUTERIZED ENFORCEMENT OF RECEIVABLES

Modern information technologies have, in the area of enforcing receivables, led to an increase in the efficiency of these processes and to a state of affairs where a relatively small number of (suitably qualified) workers are able to secure all legal and further actions connected with enforcement in a significant number of specific cases. We are not, however, going to focus on software and other aids for administration and enforcement, but rather on the computerization of proceedings *per se* (forfeiture and insolvency). It is here highly surprising that although a truly marked number of highly sophisticated methods using modern information technology are utilized in the given field, the proceedings as such lag very much behind in this sense, and most importantly, the potential of the high public monitoring thereof on the part of the public is not adequately utilized; the space for extremely effective control of proceedings on the state's part has also been neglected.

It is here apposite to assert that both types of proceeding are socially demanding. They reach, on the one hand, the very existence of entrepreneurial subjects and therefore also such issues as employment or the social sensitivity of the enforceability of rights and the justice of the system, and then on the other hand, the standard of living and societal perspectives or the dignity of natural persons. It is unnecessary to elaborate on the extent to which this is a risky situation. It is for this reason that all developed societies are unusually sensitive to cases in which unauthorized or unreasonable enforcement, misuse of forfeiture or insolvency proceedings in the sense of damaging debtors' rights, excessive inherence in their individual freedom and so forth occur. It is therefore clearly in the public's interest that maximum state supervision over these proceedings is made possible, regardless of the fact that proceedings that usually take place between private subjects are at issue. However, states usually determine all aspects and methods of this enforcement to a highly detailed degree, and proceedings regulated by the state power are at issue; societies in developed countries expect adequate inherence of the state in supervisory activities over these proceedings.

From this perspective, however, the usage of modern information technologies is highly inadequate, which applies to the situation in the Czech Republic, with which we are here concerned, but also to the situation in developed countries (we can, for instance, speak of the OECD countries).

In the part of the study to follow, we will describe the state of affairs in the Czech Republic for 2008-2015, whilst this state of affairs can, in the general sense of the word, be applied to other developed countries.[3]

In the area of forfeiture proceedings, individual forfeiture authorities for the most part keep files electronically in relatively sophisticated systems. Of course, a further classical component is kept containing originals of correspondence, individual court distrainer rulings and other written material. The reciprocal content of the electronic file and this paper file is not usually at a 1:1 proportion, primarily due to the fact that the majority of forfeiture proceedings take place with natural non-entrepreneurial persons and the electronic version of these documents cannot be as valid as a paper original. Files are not accessible to the public and there is not even a publically accessible list of forfeitures in progress or previously in progress. It is possible to ascertain whether a specific subject or natural person is undergoing forfeiture, but not by means of a public source that is fully accessible without limitation. Neither state not supervisory bodies have remote access to files (access via internet); they can, however, force access to a file within the bounds of supervisory activity.

Statistics on forfeiture proceedings are kept, at least theoretically, by the Chamber of Distrainers of the Czech Republic, which is by law the association of court distrainers with mandatory membership, and it is also an autonomous body of distrainers. It has, however, transpired that the quality of statistical data is low and the Chamber of Distrainers releases only general data (especially the number of ordered forfeitures, the number of the liable (debtors) and certain other figures).

Insolvency proceedings in the Czech Republic are public in the sense that all insolvency proceedings are announced in the insolvency register, which is accessible freely and free of charge to any party interested in these proceedings. Practically all documents inserted into a classical file (all court rulings, insolvency administrator's reports and numerous others) are made public in this register - usually in pdf. format. This then means that, in the course of insolvency proceedings, it is possible to monitor and check these proceedings, although in the bounds of individual proceedings it is in substance impossible to search for connections or successions of individual documents, as this is demanding both technically and in terms of time.

The course of insolvency proceedings is made public by statistics, although these only record partial (if general from the perspective of volume) data. This concerns especially the number of insolvency proceedings, the methods of settling the debtor's bankruptcy, duration of proceedings – but no figures on real results of insolvency proceedings or the costs of these proceedings and other important circumstances are available.

If we were somehow to characterize generally this state of affairs in both areas of proceedings leading to the enforcement of receivables, the legal state of affairs and legal circumstances are relatively precisely recorded, although it can declared overall that our knowledge on the economic aspects of forfeiture and other proceedings are practically nil. In addition to this, one must add that, due to inadequately precise expressions in legislation, the situation in the area of forfeiture proceedings is fundamentally poorer than in insolvency proceedings, for neither forfeiture authorities nor even the Chamber of Distrainers of the Czech Republic have a definition of statistical obligation in a manner that would enable data collection and the meaningful interpretation thereof.

[3] Bearing in mind, of course, that neither forfeiture nor insolvency legislation are, in their parameters, regulated either by European Union law or by any international agreements. It thus applies that there are, in numerous countries, marked divergences and specific regulations or circumstances, although these proceedings are quite similar in given states.

ISBN: 978-1-61804-318-4

III. CERTAIN DATA ON PROCEEDINGS THAT COULD BE ASCERTAINED

Thanks to the activity of the Insolvency Research scientific team, which has for several years been working with the support of the Technological Agency of the Czech Republic, certain steps have been taken towards the rectification of the above-mentioned situation.

First and foremost, statistical research of comprehensive samples of insolvency proceedings has been undertaken, thanks to which it has been possible to ascertain – at least partially – numerous important data on these events from the economic point of view. This especially applies to such data as the recoverability of receivables, i.e. the usual fulfilment acquired by a creditor in insolvency proceedings, costs of proceedings, duration of individual phases of proceedings and numerous others. These statistical surveys have focused on those debtors which are entrepreneurial subjects.[4]

As regards natural non-entrepreneurial persons, the situation is still unclear; nevertheless, in this case it is a problem that is more social than national-economic. Despite this, it would he highly apposite if similar research took place in the area of individual debtors (this area is usually also labelled with the term *personal bankruptcy*). Therefore, at the present time (first quarter 2015), we have no more substantial information from this area of insolvency proceedings that would be adequately relevant and which could, with its quality, measure up to data on insolvency proceedings with entrepreneurial subjects.

It was found that the standard yield of non-secured creditors (against debtors – entrepreneurial subjects) is lower than five percent of their claimed and substantiated receivables (including costs). Among secured creditors, the ratio of satisfaction reaches roughly 25 percent of the demanded sum. Receivables beyond property (bankruptcy costs), which are 80 percent covered,[5] are relatively well satisfied; receivables placed on a par with receivables beyond property are repaid to almost a half. Substantially more information on results is contained in professional literature [2]–[5].

In the area of forfeiture proceedings, the above-mentioned scientific team has acquired data from certain major creditors and creditor groups, as it has transpired that there is little interest in releasing such data on the parts of forfeiture authorities and the Chamber of Distrainers of the Czech Republic. For instance, the Chamber of Distrainers refuses to release or merely does not possess data on the success of individual forfeiture authorities; figures on the general utilization percentage acquired in the bounds of forfeiture proceedings are also problematic. However, data from creditors (authorized persons) show, for instance, that between individual forfeiture authorities, marked differences have been detected in the utilization percentage; similarly, highly divergent costs are also demanded on their parts, yet this demand on costs is not in any fundamental relation to the utilization percentage. For instance, data on the number of motions in individual forfeiture proceedings have transpired to be entirely undetectable; these would be data by which (given appropriate analysis thereof) data on accounted costs could be gauged. Therefore, while it would be possible in the area of insolvency proceedings (with the aid of public sources) to acquire information on the real economic results of the process of collective enforcement of receivables, this has proven entirely impossible in forfeiture proceedings, for the only data which can be used further originate from participants of proceedings – authorized persons (creditors) in the given case.

In both cases, acquired data were subjected to analyses with the aid of various methods of approach. Pertinent professional literature is available also in this area [6]–[7].

IV. EXPRESSION OF THE LEGISLATIVE RECTIFICATION OF THE SITUATION

The main sense of the work of the Insolvency Research scientific team, however, is not the actual collection of partial data on the development and state of efficiency of insolvency and forfeiture proceedings, on costs connected therewith and other fundamental national-economic facts, but rather the definition of methods and approaches by which it would be possible to ensure fundamentally greater openness of these proceedings, the transparency and especially the improved monitoring thereof.

In this area, the team has arrived at certain conclusions which assume fundamentally greater inclusion of modern information methods into processes of conducting proceedings and processing the results thereof. The team's basic hypothesis, then, is that the information asymmetry which is present in these proceedings either increases the costs of authorized persons (creditors) or leads to their passivity. We usually refer to this state as rational apathy.[6]

A necessary conclusion stems from the preceding hypothesis – that removal of information asymmetry has to lead to a reduction of transaction costs of participants of proceedings (especially creditors), which would be a

[4] Over 3,200 legitimately closed insolvency proceedings were analyzed in the above-mentioned surveys; in the given period, this sample represents roughly twenty percent of the entire number of legitimately closed insolvency proceedings. One could of course lead a discussion as to whether such a sample is adequate in the specific case of insolvency proceedings. This, however, is not the subject of this study.

[5] This, however, at the same time means that bankruptcy is declared even among a number of debtors whose property transpires to be so low that it does not even remotely suffice for the coverage of the costs of the proceedings. In the majority of proceedings, the plaintiff (i.e. the creditor who filed the insolvency proposal) has to place a deposit on the costs of the proceedings; in proceedings where the property found and monetized is insufficient, the deposit is expended, and if the insolvency administrator too has a receivable from the title of their work during insolvency proceedings, this receivable is then covered by the state.

[6] Rational apathy arrives at the moment when the potential of profit from a certain activity is so low that the entrepreneurial subject (or anyone else) senses no economic gain from exerting any effort, for the costs of such activities are probably higher than the gain which could arise therefrom. This is a situation which is unfortunate from the perspective of public interest, if understandable from the perspective of economic rationality. It would be extremely useful for creditor subjects to demand in a truly consistent manner the repayment of their receivables; if, however, the probability is low (see, for instance, the data on the standard yield from non-secured receivables in the Czech Republic in insolvency proceedings), we can hardly demand from private subjects the fulfilment of public interest. This, however, necessarily leads to the strengthening of moral hazard on the parts of debtors, for they can anticipate rational apathy of creditors.

ISBN: 978-1-61804-318-4

development fully corresponding to public interest. The rduction of transaction costs will reduce the level of rational apathy in the economic environment. The transparency and public nature of proceedings (in the given context, this applies especially to insolvency proceedings) becomes an economic circumstance. One could also accept a further working hypothesis, that a higher level of transparency of proceedings would elicit consequent impacts (of which the reduction of the level of rational apathy of creditors and an increase of the space for effective monitoring of processes are especially fundamental); these impacts would then lead, on the one hand, to growth of yields from proceedings (it could be asserted yet again – especially from insolvency proceedings), and on the other hand, to a generally higher level of enforceability of rights, and in its final consequence, to a higher level of trust in the legal system as a whole on the part of the public.

Numerous steps present themselves in the area of insolvency proceedings; with the aid of greater usage of modern information technologies, these could benefit the general level of enforceability of receivables. This especially includes the implementation of electronic registration of receivables, and this should be in succession to electronic filing of an insolvency proposal. Furthermore, it would be possible to implement legislative and other communication between participants of proceedings and the court or the insolvency administrator in completely electronic form, with utilization of data boxes, which all entrepreneurial subjects (with the exception of tradesmen) in the Czech Republic are already obliged to have. In the case of other participants (of the non-entrepreneurial natural person or tradesmen types), it would be possible to enact the utilization of electronic methods at least when filing an insolvency proposal and when filing claims for receivables, and this by means of specialized insolvency courts workplaces. These measures would also entail significant limitation of space for so-called bullying proposals, which are an accompanying phenomenon of the present state of insolvency proceedings and the conducting thereof in the Czech Republic. These bullying proposals significantly misuse the situation where pertinent information technology is utilized only partially and inconsistently.

In the area of forfeiture proceedings, the Insolvency Research team approaches with the demand to establish a central forfeiture register which would be more accessible to the public than is the case with the current information sources.

In the area of forfeiture proceedings, the Insolvency Research team approaches with the demand to establish a central forfeiture register which would be more accessible to the public than the current information sources. Furthermore, there are here proposals for specific electronic keeping of a forfeiture file and especially for structurally divided keeping of a financial account belonging to the file in such a way that it would enable highly effective monitoring not only of the real activity of the distrainer's authority, but especially the payment morale of these authorities in the direction of authorized persons (i.e. to the creditors). This then assumes that this specific financial account will be accessible not only to supervisory bodies, but also to authorized persons.

In both cases it applies that the primary goal is not the actual supervision of processes by state bodies, but pressure on their increased activity and therefore on the higher economic efficiency of the system as a whole.

V. THE PROBLEM OF ECONOMIC EFFICIENCY

The problem of economic efficiency is of course far more comprehensive and deeper; it is impossible to solve it through simple pressure on the transparency of insolvency or forfeiture proceedings. The enforceability of receivables in insolvency proceedings has already been mentioned; the enforceability of receivables in forfeiture proceedings is roughly 30-40 percent of the demanded sum – this data is, however, confusing to some extent, as statistics from these proceedings are not based on particularly high-quality surveys and, most importantly, data provided by individual major creditors do not cover the entire problem of forfeiture proceedings and there is also the issue of various methodologies for acquiring a total figure in individual sources.

VI. CONCLUSION

The work provides information on the progress of the Insolvency Research scientific team in defining legislative changes which, in the bounds of the Czech Republic, are to lead to increasing the efficiency of insolvency and forfeiture proceedings. As it has transpired, it has been asserted, first and foremost, that the team considers the broadening of transparency of individual proceedings to be an appropriate method. In this sense, we see a significant space especially for a new approach in the utilization of information technologies and in the expansion of the possibility of supervision over enforcement processes, especially on the parts of creditors (authorized persons) and state bodies designated for this purpose. Our starting point is that the utilization of information technologies will enable the reduction of transaction costs for participants of individual proceedings, which will enable suppression of the effect of rational apathy.

REFERENCES

[1] T. Richter, *Insolvenční právo.* Prague, ASPI Walters Kulvert, 2008, ch. 1 – 3.

[2] L. Smrčka, J. Schönfeld, Insufficient utilisation of information technology in the state administration. The example of insolvency proceedings and the insolvency register in the Czech Republic. In: A. Rocha, A. M. Correira, F. B. Tan, K. A. Stroetmann (eds.). New Perspectives in Information Systems and Technologies, 2014, Heidelberg, Springer. Available: http://link.springer.com/chapter/10.1007/978-3-319-05951-8_1

[3] L. Smrčka, J. Schönfeld, P. Ševčík, Czech Insolvency Law after for years. WSEAS Transactions on Business and Economics,2013,(3), 10, pp. 204–214. Available: http://wseas.org/cms.action?id=6931

[4] E. Kislingerová, E. (2014): Struktura *potřebných legislativních změn v insolvenčním zákonu.* In E. Kislingerová, J. Špička (eds.): *Insolvenční praxe: Věřitelé a dlužníci.* Prague, Oaconomica, 2014, pp. 15-25.

[5] J. Schönfeld, L. Smrčka, E. Kislingerová, J. Erlitz, Assessment of statistical research of insolvency proceedings realized in the Czech Republic. In: L. Smrčka, J. Plaček (eds.). *Insolvency 2014: Current Problems and Experiences.* London, The London School of Economics and Political Science (LSE), 2014, pp. 11–22.

[6] L. Smrčka, M. Arltová, J. Hnilica, An attempt to analyze the relationship between the performance of the economy and certain results of insolvency proceedings in selected countries. In: *The 8th International Days of Statistics and Economics.* Prague, Melandrium, 2014, pp. 1375–1385. Available: http://msed.vse.cz/msed_2014/article/361-Smrcka-Lubos-paper.pdf.

[7] L. Smrčka, J. Schönfeld, M. Arltová, J. Plaček, The significance of insolvency statistics and the regression analysis thereof – the example of the Czech Republic. *WSEAS Transactions on Business and Economics*, 2014, (11), pp. 227–241. Available: http://www.wseas.org/multimedia/journals/economics/2014/a025707-091.pdf.

Assoc. Prof. Luboš SMRČKA, M.Sc., PhD. In 1984, he graduated from the Czech University of Life Sciences in Prague. After 1993, he left the Institute of Experimental Botany at the Czechoslovak Academy of Sciences to start business. He gradually acquired several professional specializations: tax advisor (1993), broker (1996), certified balance accountant, (1998), forensic expert in economy, prices, and valuation specialized in the valuation of securities, RM-S and stock exchange and business valuation (2000, extended in 2003), accounting and tax expert (2001).

In the last 6 years, he has worked as a Lecturer in the Department of Business Economics at the Faculty of Business Administration of the University of Economics in Prague. In 2013, he gained the title of Associate Professor. He focuses primarily on the area of insolvency proceedings, their macroeconomic impacts and issues of insolvency law and the problem of personal finances.

Assoc. Prof. Smrčka is the author of numerous books and articles in professional publications.

Jaroslav Schönfeld, M.Sc. PhD. graduated from the University of Economics in Prague in 2007. He works at the Czech Savings Bank, Inc., in the department of restructuring and recovery. Since 2008, he has also been active at the University. He deals with financial management, restructuring, insolvency and pricing. He is the author of the monograph *Modern View on the Valuation of Receivables* (CH Beck 2011) and numerous articles in professional journals.

Security and Countermeasures against SIP-Message-based Attacks on the VoLTE

Bonmin Koo, Sekwon Kim, Hwankuk Kim

***Abstract*—**With the migration of the mobile communication service environment from 3G to 4G, the VoLTE service environment, including the 4G service subscriber base, is continuously expanding. Introduction of All-IP networks enables operators to provide their mobile communication services with various protocols. Operators also adopt various security measures, including firewall, to protect their mobile networks against various types of security threats. However, firewall, which is an IP-based security solution, cannot provide countermeasures against security threats on the 4G mobile network. This study outlines the security vulnerabilities where an attacker may acquire IP of a VoLTE user by utilizing a message requesting state of the user, and provides the targeted attack scenarios in which the attacker abuses the acquired IPs of the users. Furthermore, this study suggests the technology against illegal acquisition of user equipment IPs which might cause targeted attacks.

***Keywords*—VoLTE, SIP, 4G, Security**

I. Introduction

With the recent popularization of smartphones, the 4G mobile communication service market is growing rapidly. In the 4G regime, the data network integrates the mobile networks which have been divided into the voice network and the data network. The VoLTE service utilizes Session Initiation Protocols (SIPs), and the IMS network, which is configured with SIP-handling CSCF servers, provides higher-quality voice and video services than the existing 3G voice network.

VoLTE and data services on the 4G network are already under various security threats [1]-[3]. The data-service-based security threats may cause various kinds of loss for user equipments, such as unexpectedly high phone bills. The known security threats, however, targeted unspecified individuals, and the attackers could not specify the targets.

This study outlines the VoLTE service environment where an attacker may acquire IP of a user by utilizing the security vulnerability of the CSCF server, and provides the security threat scenario in which the attacker specifies the target of attack with the acquired IPs of the users.

Section II deals with the 4G mobile network, the GTP protocols used on the 4G mobile network, the IMS network for the VoLTE services, and the SIP Subscribe messages used by the attackers to acquire IP of a user equipment. In Section III, the targeted attack utilizing SIP messages is described. Section IV suggests the method to detect abnormal SIP Subscribe messages utilizing the vulnerabilities of the CSCF server. In Section V, the result of the study and the future research plan are provided.

II. Related work

A. 4G Mobile Network

A 4G mobile network consists of User Equipments (UEs), wireless access networks, and Evolved Packet Cores (EPCs). Figure 1 illustrates the architecture of the 4G mobile network, where, UE indicates the user equipment that attempts access to the mobile network. E-UTRAN is the network that supports the 4G wireless access technology, and is composed of eNodeBs. An eNodeB provides the user with the wireless interface, and manages wireless communication resources. An EPC consists of Mobility Management Entities (MMEs), Serving Gateways (S-GWs) and PDN Gateways (P-GWs). MME provides the user authentication and roaming functions, and manages EPS bearers.

S-GW is the end point of E-UTRAN and EPC. It becomes the anchoring point for inter-eNodeB handover. P-GW provides IP routing/forwarding, and UE IP allocation.

B. GTP (GPRS Tunneling Protocol)

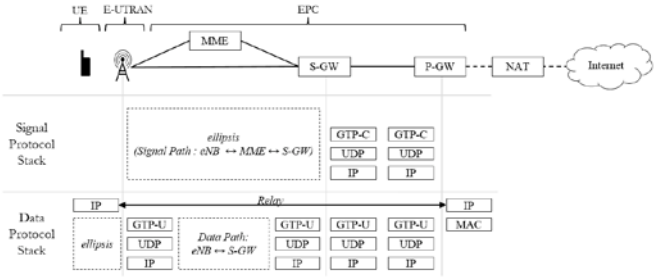


Fig. 1 4G mobile network architecture

This research was funded by the MSIP(Ministry of Science, ICT & Future Planning), Korea in the ICT R&D Program 2015.

F. Bonmin Koo is with the Korea Internet&Security Agency, Seoul, Korea(South)(e-mail:bmkoo@kisa.or.kr).

S. Sekwon Kim is with the Korea Internet&Security Agency, Seoul, Korea(South)(e-mail:heath82@kisa.or.kr).

T. Hwankuk Kim is with the Korea Internet&Security Agency, Seoul, Korea(South)(e-mail:rinyfeel@kisa.or.kr).

GTP is the protocol used in the 4G core network. In the GTP protocol, GTP control plane messages (GTP-C) are used for tunnel management functions, such as tunnel information exchange, and creation/update/deletion of tunnel. GTP user data messages (GTP-U) are used to send T-PDUs through the GTP tunnel. A GTP protocol contains, in its header, a simplex Tunnel Endpoint Identifier (TEID) that identifies the user. In the S11 section (MME↔ S-GW), for example, S11 SGW TEID is the uplink control TEID allocated by the S-GW, while S11 MME TEID is the downlink control TEID allocated by MME. In the S1-U section where GTP-U is transmitted, S1-U eNodeB TEID is the downlink data TEID allocated by eNodeB, while S1-U S-GW TEID is the uplink data TEID allocated by S-GW. These TEID values are inserted in the GTP-C messages for communication between EPC equipments [4].

C. IMS Network

The IMS network performs session control, routing and UE registration for the VoLTE service, and contains P/I/S-CECF servers performing the roles. P(Proxy)-CSCF delivers SIP messages from EPC to other CSCFs depending on the type of the message. I(Interrogating)-CSCF is the gateway to other mobile IMS networks or outside networks. And lastly, S(Serving)-CSCF interworks with Registrar, the DB of the subscribers registered in the IMS network, providing services, such as user registration, depending on the type of messages [1]. CSCF servers provide the VoLTE service by handling the SIP request messages and transmitting SIP responses to UEs [5].

D. SIP Subscribe

SIP SUBSCRIBE is the message that provides the SIP handling servers with information on registration of VoLTE user and state of VoLTE call [6].

In the networks of some of the mobile service operators in Korea, when a VoLTE user is registered initially, the SIP SUBSCRIBE message showing whether the UE supports VoLTE calls or not is sent to the IMS network. Figure 2 illustrates the procedures for VoLTE user registration, including SIP SUBSCRIBE message.

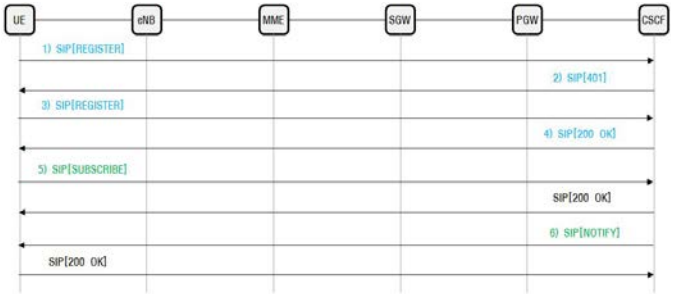


Fig. 2 VoLTE registration procedure

1) For initial registration of VoLTE service, UE sends the SIP REGISTER message to the CSCF server, requesting the nonce value for creation of response value.
2) The CSCF server sends the 401 Authentication message attached with the nonce value to the UE.
3) UE combines the nonce value with the K value in USIM to generate the response value, and returns the REGISTER message with the response value to the CSCF server.
4) The CSCF server sends the 200 OK message to UE, notifying that VoLTE user registration has been completed.
5) UE sends the SIP SUBSCRIBE message to the CSCF server, providing information on activation and preparation of VoLTE call.
6) The CSCF server sends the NOTIFY message to check the UE state for activation and preparation of VoLTE call.

Steps 1~4 in the above Figure 2 are the basic VoLTE user registration procedures. Whether to send the SIP SUBSCRIBE message or not varies between the UE manufacturers and the operators.

III. Security Vulnerabilities to SIP-Message-Based Attack

A. Acquisition of UE IP address by weakness of CSCF server

The SIP message for VoLTE is sent to the CSCF server on the IMS network via EPC. Most of the mobile networks have no security equipment on the mobile network, conducting no abnormal SIP message check. If an attacker abuses the weakness of the CSCF server and sends a forged SIP message, there is a high chance of security threat as the CSCF server will handle the traffic with the forged SIP message. SIP is a text-based message of which sample can be easily acquired on the Internet. The sample itself, however, cannot create the SIP message field required by the CSCF server. As mentioned in Section II, some operators in Korea sends the SIP SUBSCRIBE message when registering VoLTE users. Therefore, if you can collect the VoLTE user registration procedure performed in UE, you can also collect the SIP SUBSCRIBE message that can be handled in the CSCF server.

In order to acquire the IP of the victim's UE, you need the normal SIP SUBSCRIBE message and MSISDN (phone number) of the victim. By replacing the MSISDN in the normal SIP SUBSCRIBE message with the MSISDN of the victim, and sending the message to CSCF server, you can acquire IP of the victim from the SIP NOTIFY message. Figure 3 provides the procedure to acquire IP of the victim UE.

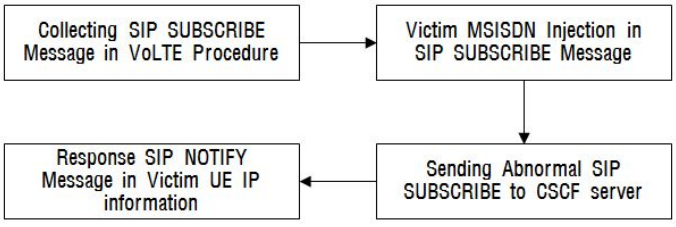


Fig. 3 UE IP acquisition procedure

The attacker collects the SIP SUBSCRIBE message through

the VoLTE user registration procedure, and replaces MSISDN

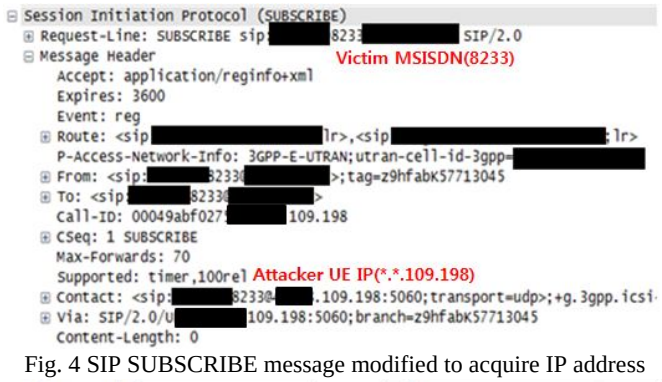


Fig. 4 SIP SUBSCRIBE message modified to acquire IP address

Fig. 5 Acquisition of Victim UE IP through SIP NOTIFY message

with that of the victim as shown in Figure 4. For the forged SIP SUBSCRIBE message as MSISDN of the victim UE, the CSCF server sends the reply without checking forgery. As shown in Figure 5, the CSCF server sends the SIP NOTIFY message containing the IP of the victim UE to the attacker UE.

B. Attack with the acquired UE IP

The mobile network is vulnerable to abnormal traffic with forged IP address of UE (IP spoofing) [8]. Because the 4G mobile networks of 3 Korean operators are configured based on the NAT+ firewall, no external traffic can be received by the network unless it is triggered by UE. However, the attacker may send abnormal traffic into the mobile network by forging the IP address of UE. The abnormal traffic generated by the attacker through IP spoofing will be received by the victim UE. By abusing the victim IP acquisition method provided in this study, the attacker can send a large amount of abnormal traffic targeting a specific user into the network. Figure 6 illustrates the IP spoofing attack with the acquired UE IP.

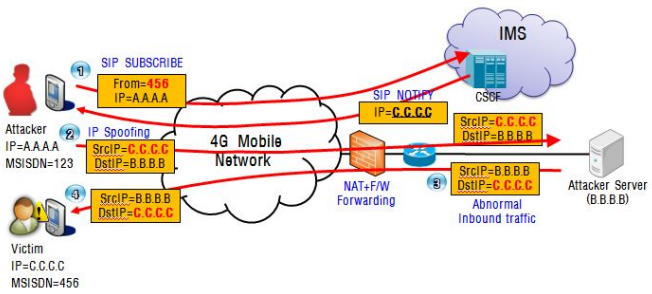


Fig. 6 IP spoofing attack with the acquired victim UE IP

By sending the SIP SUBSCRIBE messages with the victim's MSISDN continuously to the CSCF server, the attacker may easily acquire the changed victim IP, and the attacker can conduct targeted attack to the victim.

If an attacker keeps sending the data traffic that meets the billing policy of an operator by using IP spoofing, the amount of data used by the victim UE increases in proportion to the received traffic volume, resulting in an abnormal billing. In general, mobile operators mirror in/out traffic of P-GW, estimate the amount of data used by each UE, and make billing for each UE based on the estimated amount of data. The amount of data used is calculated normally based on the accumulated volume (byte) of T-PDU, but the accumulation reference varies between operators, and special bulling policies can be applied. Therefore, the victim of the continuous attack might have to pay considerable amount of fee.

As described above, attackers may acquire victim IP address by utilizing the security vulnerability of the CSCF server in handling the SIP messages. By utilizing the acquired IP address, attackers may attempt targeted IP spoofing attack, sending a large amount of abnormal traffic to the victim and causing the victim to pay a large amount of bill.

IV. How to Detect Abnormal SIP SUBSCRIBE Messages

Acquisition of UE IP utilizes the SIP SUBSCRIBE message sent by the CSCF server in the course of handling VoLTE user registration requests. On the 4G mobile network, a GTP-C message for user authentication is transmitted on the S11 section, and a GTP-U message for the user data is transmitted in the S1-U section. Therefore, it is possible to detect abnormal SIP SUBSCRIBE messages by collecting and analyzing all the user control/data (GTP-C/U) traffic on the EPC S11 section and the S1-U section (eNodeB↔ S-GW). GTP-C, the control traffic, contains TEID for identification of users, which can be used to detect abnormal SIP SUBSCRIBE messages from the GTP-U traffic on the S1-U section. Figure 7 illustrates the flow of detecting abnormal SIP SUBSCRIBE messages.

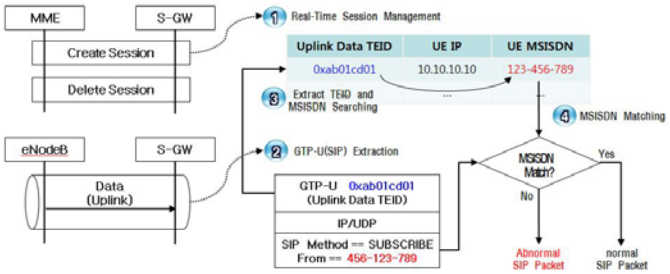


Fig. 7 Detection Technique for Abnormal SIP SUBSCRIBE Message

V. Conclusion and Future Work

VoLTE uses SIP and RTP protocols to provide audio/video services through the IMS network where the SIP handling equipment (CSCF server) is located. The SIP messages sent by UEs for the VoLTE services are sent to the IMS network via EPC without filtering, and on the IMS network, CSCF server handles the SIP messages and provide the VoLTE service to the UEs. Attackers can acquire UE IPs by using the weakness of

ISBN: 978-1-61804-318-4

CSCF server that does not check forgery of SIP messages. When an attacker sends an abnormal SIP SUBSCRIBE message to acquire the victim IP, the victim cannot recognize the fact that his/her IP address has been exposed. Exposed UE IP is subject to various security threats, including abnormal billing caused by IP spoofing. This study described how attackers can utilize weakness of the CSCF server, illegally acquire UE IPs, and cause security threats. In addition, this study suggested the method to detect abnormal SIP SUBSCRIBE messages by using the real-time session management method. Further studies are required to find the security threats on the VoLTE environment and the security vulnerabilities on the 4G mobile network where evolvement to IPv6 is in progress.

REFERENCES

[1] Seongmin Park, Sekwon Kim, Joohyung Oh, Myoungsun Noh, Chaetae Im, "Threats and countermeasures on a 4G Mobile Network", International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, July 2014

[2] Sekwon Kim, Seongmin Park, Hwankuk Kim, "Security Vulnerabilities and Threats of LTE Networks", on Advances in Computing, Control and Networking, Feb 2015

[3] Joohyung Oh, Sekwon Kin, Myoungsun Noh, Chaetae Im, "Phone Number Spoofing Attack in VoLTE", World Academy of Science, Engineering and Technology, Dec 2014

[4] 3GPP TS 29.274 Version 12.0.0 Release 11

[5] 3GPP, TS 23.228, IP Multimedia subsystem(IMS)

[6] RFC 3265 Session Initiation Protocol (SIP)-Specific Event Notification

[7] Dong W.Kang, Joo H, Oh, Chae T, Im, Wan S. Yi, Yoo J, Won, "A Practical Attack on Mobile Data Network Using IP Spoofing", Applied Mathematics & Information Sciences, Nov 2013.

Modeling Machine to Machine Vehicular Safety Communication

Yen-Hung Chen
Institute for Information Industry
Taipei, Taiwan
pplong@gmail.com

Yuan-Cheng Lai
Department of Information Management
National Taiwan University of Science and Technology
Taipei, Taiwan
laiyc@cs.ntust.edu.tw

Ching-Neng Lai
Department of Information Technology
Hsing Wu University of Science and Technology
New Taipei City, Taiwan
093062@mail.hwc.edu.tw

Pi-Tzong Jan
Institute for Information Industry
Taipei, Taiwan
i750712@iii.org.tw

***Abstract*—Machine to Machine (M2M) vehicular safety communication, which shares safety information of road traffics between vehicles autonomously without human intervention, is introduced for the automobile safety in order to reduce the high injuries and deaths of road traffic accident. For supporting the M2M vehicular safety communication, the IEEE Standards Association develops the IEEE 1609/802.11p standards for rapidly exchanging vital safety information between moving vehicles. However, there is still no sufficient mathematical model to evaluate the performance of IEEE 1609/802.11p standards and other related M2M vehicular safety researches. This study therefore proposes a mathematical model to analyze the performance of M2M vehicular safety communication in IEEE 1609/802.11p environment. The mathematical model can benefit academic researchers and industrial developers to discover new ideas of M2M vehicular safety message communication and verify their designed algorithms/mechanism more quickly and easily.**

***Keywords*—WAVE, IEEE 802.11p, IEEE 1609, Markov chain model, M2M, Vehicle Safety**

I. INTRODUCTION

THE non-fetal injuries and deaths of road traffic accidents have been decreased in last decade. This achievement is largely ascribed to the improved automobile safety including the active safety technologies, e.g., the anti-lock brake system and electric power steering, and positive safety technologies, e.g., the airbag and seatbelt. However, the injuries and deaths remain relatively flat because current safety technologies are only based on the information within the driver's Line of Sight (LoS) [1]. In order to overcome the LoS limitation for further reducing accident, applying Machine to Machine (M2M) vehicular safety communication, which shares information between electronic system autonomously without human intervention, in automobile safety is inescapably necessary [2]. With the help of M2M vehicular safety communication, safety information from emergency-detecting sensors (such as accelerometers and the braking system) can be automatically disseminated to other vehicles, when an emergency occurred. This safety information enables vehicles to grasp the traffic condition which is not visible to them, and thus it helps vehicles to make proper decisions in a timely manner, leading that the occurrence and consequence of automobile accidents can be further minimized. Therefore, the M2M vehicular safety communication is widely regarded as the next step of automobile safety.

In order to support the M2M vehicular safety communication, the IEEE Standards Association (IEEE-SA) develops the IEEE 1609/802.11p standards to provide a wireless communication mechanism for rapidly exchanging vital safety information between moving vehicles [3-7]. The IEEE 802.11p/1609 standards simplify the time-consuming IEEE 802.11 carrier sense multiple access (CSMA) scheme and disseminate the safety information in broadcast manner in order to provide the ability of the instantaneous safety information exchange [8]. Moreover, the IEEE 802.11p/1609 standards divide the radio spectrum into one control channel (CCH) and six service channels (SCHs). The CCH is used for the exchange of management frames and emergent data, while the SCHs are applied for the specific application service data transmission. The standards also define corresponding multichannel operations for vehicles to hop between CCH and SCHs. Therefore, the enormous connection setup time, such as channel scanning, can be relieved by simply monitoring the CCH. Moreover, lots of researchers proposed their own solutions to enhance the M2M communication performance of IEEE 802.11p/1609 including the bandwidth regulation mechanisms to minimize the packet error rate and to improve the throughput [9-11], the multi-antenna assignment algorithm to maximize the channel usage [12], and the scheme to manage the multi-hop traffics [13].

In order to evaluate the performance of IEEE 802.11p/1609 and the related M2M vehicular safety researches, a mathematical analysis model is inescapably needed. With regard to the mathematical analysis, the conventional IEEE 802.11 analysis model is not applicable for the broadcast behavior of IEEE 802.11p/1609. Moreover, the IEEE 802.11p/1609 model derived in [14] assumes the buffer size of each vehicle is one and does not consider the multichannel operations of IEEE 802.11p/1609 standards, thus this model is also not suitable.

This study therefore proposes a mathematical model for M2M vehicular safety researches to evaluate the

ISBN: 978-1-61804-318-4

performance of IEEE 802.11p/1609 standards. The mathematical model applies the Markov modulated on-off process (MMOP) model which allows the researchers to analyze the performance under different traffic loads rather than merely the saturated situation. This model also considers the multichannel operations of IEEE 1609/802.11p standards.

The organization of the paper is as follows. Section II reviews the IEEE 802.11p/1609 standards. In section III, the mathematical analysis on the performance of IEEE 802.11p/1609 standards is derived. In section IV, the performance of the IEEE 802.11p/1609 standards is demonstrated by analytic results and simulation results. Finally, conclusions are given in section V.

II. IEEE 1609/802.11p Standards

The IEEE 802.11p and 1609 draft standards are proposed to provide seamless and interoperable services for M2M wireless access in vehicular environments (WAVE). These standards adopt the multi-channel cooperation to cope with the requirement that emergent safety messages must be disseminated fast enough to allow the driver to make sufficient decision.

The multi-channel cooperation contains two mechanisms as shown in Figure 1. First, the radio spectrum is divided into one control channel (CCH) and six service channels (SCHs). The CCH is used for the exchange of management frames and emergent data, while the SCHs are applied for the specific service data transmission. The vehicles can establish communications and exchange emergent messages on the CCH to relieve the enormous setup time such as channel scanning. The second mechanism is to divide the channel time into synchronization intervals, where a synchronization interval consists of a CCH interval (CCHI) and a SCH interval (SCHI). This enables a single-antenna equipped vehicle to monitor the CCH during a common time interval, i.e., CCHI, for communicating management frames and emergent data. After the CCHI, the vehicle can hop to the chosen SCH to exchange service data during SCHI.

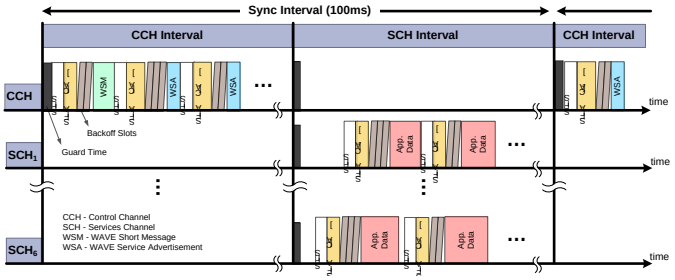


Figure 1 WAVE channel coordination

The IEEE 802.11p/1609 network accommodates two protocol stacks, WAVE Short Message Protocol (WSMP) and the conventional TCP/IP protocol suit (e.g., IPv6 and TCP/UDP), for optimizing the operation in WAVE. WSMP is designed for the application demanding time-sensitive data exchange such as emergent safety and traffic information. In order to enhance the transmission efficiency, WSMP omits the multiple handshakes defined in IEEE 802.11 standard including association, authentication, RTS/CTS, and ACK functionalities. That is, WSMP composes the application data as a WAVE short message (WSM), addresses it to the broadcast MAC address, and transmits it directly on the CCH without any connection establishment procedure. A receiving vehicle then accepts the frame and passes it up to the upper protocol stack.

On the other hand, the application with less time-sensitive data demanding, e.g., digital map updating, still requires efficient connection setup because of the limited contact time between vehicles and roadside units. The IEEE 802.11p/1609 standards therefore modified the IEEE 802.11 channel access scheme for the conventional TCP/IP protocol suit. That is, a service provider broadcasts WAVE service advertisement (WSA) frame on the CCH during CCHI to announce the availability of the service. This WSA contains the detailed service information such as IP and MAC address of the service provider, port number of the application, and SCH number on which the channel this provider offers this service. If other vehicles receive this WSA and decide to acquire the service, they will hop to the designated SCH during SCHI to receive the service data. This mechanism enables the vehicles to simply monitor the CCH for acquiring the service setup information without the channel scanning and other handshake procedures which cause intolerant connection establishment delay in WAVE.

III. Mathematical Analysis

This section derives a mathematical model to analyze the performance of IEEE 802.11p/1609 safety transmission during CCHI including the throughput, goodput, and frame error rate. As shown in Figure 2, the performance analysis consists of two parts, which are (1) the performance of the transient state and (2) the performance of the steady state. Table I shows the notations used in this section.

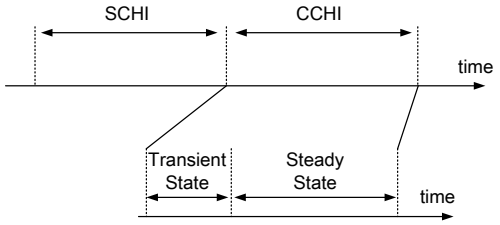


Fig. 2 Performance analysis of CCHI

Figure 3 shows the Markov modulated on-off process (MMOP) model applied in this section. In this model, the "on" period means the number of frames generated in each burst follows the geometric distribution, and the "off" period means the silent time between two bursts which follows the exponential distribution. The reasons to apply MMOP are: (1) MMOP allows this study to analyze the performance under different traffic loads rather than merely

the saturated situation, (2) the conventional IEEE 802.11 analysis model is not applicable for the broadcast transmission of the IEEE 802.11p/1609 safety message, and (3) the IEEE 802.11p model derived in [14] assumes the buffer size of each vehicle is one and does not consider the channel switching behavior of IEEE 802.11p/1609, thus it is not suitable to this study.

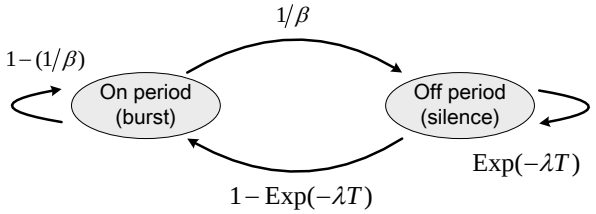


Figure 3 Markov modulated on-off process

TABLE I. USED NOTATIONS FOR PERFORMANCE ANALYSIS MODEL

Notation	Definition
m	The buffer size of a vehicle
n	$\text{CW}_{\min}$
N	Number of vehicles in the same coverage range network
BR	Channel bit rate
$S_{payload}$	Payload length of a WSM frame
S_{header}	Header length of a WSM frame header including PHY and MAC headers
S_{ratio}	$S_{ratio}=\frac{S_{payload}}{S_{header}+S_{payload}}$
T_s	Length of time in which a vehicle transmits a WSM frame
T_σ	Length of a slot in which no vehicle transmits any WSM frame
T_{SCHI}	Length of SCHI
T_{CCHI}	Length of CCHI
τ	Probability that a vehicle transmits a WSM frame
P_{tx}	Probability that at least one of other vehicles transmits a WSM frame. $P_{tx}=1-(1-\tau)^{N-1}$
β	Average number of WSM frames in a burst which follows the geometric distribution
P_β	Probability that no next WSM frame in current burst. $P_\beta=1/\beta$
$1/\lambda$	Average length of the off period between two arriving bursts follows the exponential distribution.
P_{bz}	Probability of arrival of a burst during T_s. $P_{bz}=1-\exp(-\lambda T_s)$
P_σ	Probability of arrival of a burst during T_σ. $P_{bz}=1-\exp(-\lambda T_\sigma)$
R	WSM traffic arrival rate of a vehicle. $R=\frac{S_{payload}\times\beta}{(1/\lambda)+T_s\times\beta}$

A. Transient state performance

There is no numerical model, to our best knowledge, built for the transient state of IEEE 802.11p/1609 network. Therefore, we assume that the transient state ends when all vehicles finish their first round transmissions during CCHI, given that the safety message data arrival rate of each vehicle is not high. This assumption bases on the fact that IEEE 802.11p/1609 standards omit all the collision avoidance mechanisms including RTS/CTS and ACK functionalities, and thus the vehicles would transmit all frames in their buffers in their first transmission opportunity. Afterward, the vehicles will quickly enter the steady state, when the data arrival rate of each vehicle is not high.

Based on the above assumption, the throughput (TH_{ts}) of all vehicles in the transient state (TS) during CCHI approximates to the data size queued in their buffers in their first transmission opportunities, that is:

$$TH_{ts}\approx R\times T_{SCHI}\times N \tag{1}$$

where $R\times T_{SCHI}$ is the expected data size queued in the buffer of the vehicle *i* during SCHI.

The frame error rate in the TS (FER_{ts}) can be also derived by

$$FER_{ts}=1-\Pr\{\text{Succesful Transmission}\}$$
$$\approx 1-\frac{\sum_{d=1}^{n}E[CW_{d,1}]}{\sum_{d=1}^{n}\sum_{l=1}^{N}E[CW_{d,l}]} \tag{2}$$

where $CW_{d,l}$ means that there are *l* vehicles which choose *d* as their CW value, and $E[CW_{d,l}]$ is the expectation that *l* vehicles which choose *d* as their CW value. Thus, $\sum_{d=1}^{n}E[CW_{d,1}]$ is the expected number of CW value that only one vehicle chooses it, and in such cases this vehicle will not suffer any collision during its transmission. Moreover, $\sum_{d=1}^{n}\sum_{l=1}^{N}E[CW_{d,l}]$ means the expected number of CW value that one or more vehicles choose it. Therefore,

$$\frac{\sum_{d=1}^{n}E[CW_{d,1}]}{\sum_{d=1}^{n}\sum_{l=1}^{N}E[CW_{d,l}]}$$

is the probability that no collision occurs when a vehicle chooses a CW value. In order to calculate $E[CW_{d,l}]$, let

$$E[CW_{d,l}]=1\times\Pr\{CW_{d,l}\}+0\times(1-\Pr\{CW_{d,l}\}) \tag{3}$$

where $\Pr\{CW_{d,l}\}$ is the probability that *l* vehicles choose *d* as their CW, that is

$$\Pr\{CW_{d,l}\}=\left[\binom{N}{l}\times(\frac{1}{n})^{l}\right]\times\left[(\frac{n-1}{n})^{N-l}\right] \tag{4}$$

where $\binom{N}{l}$ is the set of all *l* combinations of the set of overall vehicles in this network, and *N* is the size of this set.

Hence, $\binom{N}{l}\times(\frac{1}{n})^{l}$ is the probability that l vehicles choose d as their CW value, and $(\frac{n-1}{n})^{N-l}$ is the probability that remaining vehicles choose other CW values.

The goodput (GP_{ts}) of all vehicles during TS can be calculated by

$$GP_{ts} = TH_{ts}\times(1-FER_{ts}) \quad (5)$$

In order to obtain the length of TS (T_{ts}), let $\sum_{d=1}^{n}\sum_{l=1}^{N}E[CW_{d,l}]$ be the number of groups of vehicles, and a group mentioned here is a set of vehicles choose the same CW value and transmit their frames in the same time. Assume the transmission time of each member in each group is the same, the average transmission time length of each group in TS can be calculated by $\frac{(TH_{ts}/N)}{BR\times S_{ratio}}$. Therefore, the length of T_{ts} can be estimated by

$$T_{ts}\approx\frac{(TH_{ts}/N)}{BR\times S_{ratio}}\times\sum_{d=1}^{n}\sum_{l=1}^{N}E[CW_{d,l}] \quad (6)$$

B. Steady state performance

The overview of the proposed MMOP model for the steady state (SS) performance analysis is demonstrated in Figure 4. The state $W_{h,j}$ represents the vehicle has j frames in its buffer for transmission and its backoff counter value is h. The vehicle is only allowed to change its state under one of the following two conditions: (1) when a busy period ends with a channel idle period of DCF Interframe Space (DIFS), and (2) when an idle slot time ends given that the previous slot time was also idle. Noted that $W_{0,0,0}$ means the state that the vehicle's transmission buffer is empty and other vehicles in the same coverage range have no intention for data transmission; $W_{0,0,1}$ means the state that the vehicle's transmission buffer is also empty and one or more other vehicles transmit data in the same coverage range. The reason of this special case is that $W_{0,0}$ contains the situations that the channel is idle and the channel is busy, leading that the time and probability that the vehicle stays in $W_{0,0}$ is difficult to estimate. Therefore, separating $W_{0,0}$ into $W_{0,0,0}$ and $W_{0,0,1}$ is easier for this model to estimate the time length and probability that the vehicle's buffer is empty.

Figure 4 also shows that this MMOP model consists of three phases including transmission phase (TP), backoff phase (BP), silence phase (SP). The states in TP (i.e., $W_{0,j}$ where j>0) imply the vehicle's backoff counter equals 0 and start transmitting its data queued in the buffer; the states in BP (i.e., $W_{h,j}$ where h>0 and j>0) specifies that the vehicle's buffer is not empty and its backoff counter is bigger than 0; the states in SP (i.e., $W_{0,0,0}$ and $W_{0,0,1}$) means the vehicle's buffer is empty. Figures 5, 6, and 7 present the transition probabilities of the states in SP, TP, and BP, respectively.

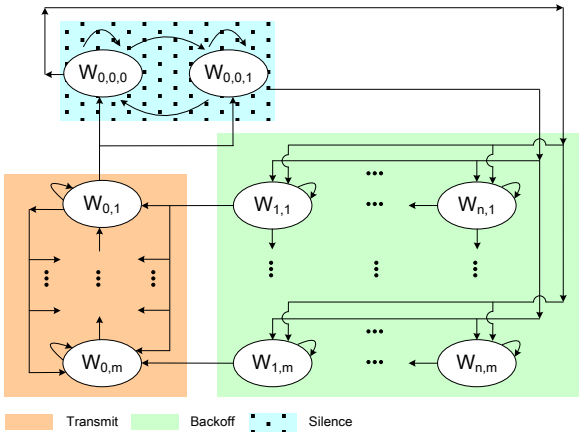


Figure 4 An overview of the proposed MMOP model

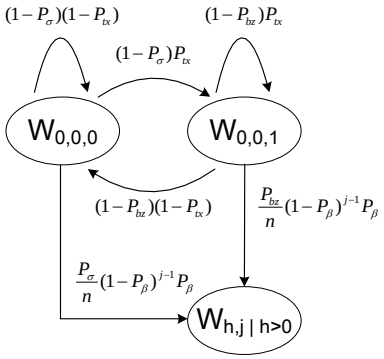
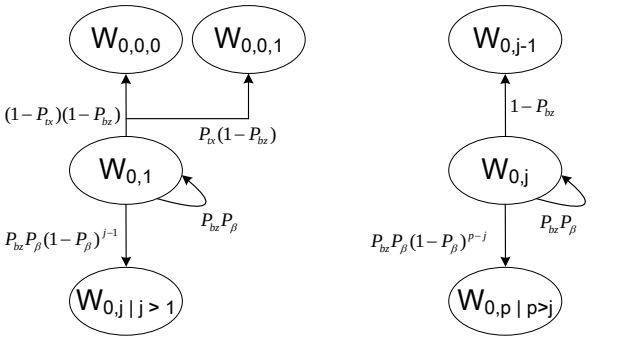


Figure 5 The transition probabilities of $W_{0,0,0}$ and $W_{0,0,1}$



(a) The transition probabilities of $W_{0,1}$ (b) The transition probabilities of $W_{0,j}$ where j>1

Figure 6 The transition probabilities of $W_{0,j}$ where j>0

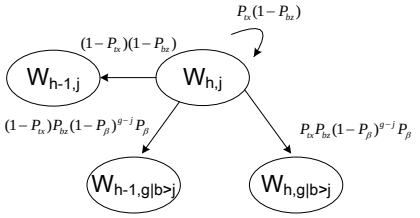


Figure 7 The transition probabilities of $W_{h,j}$ where $h > 0$ and $j > 0$

Let P denotes the single-step transition probability matrix described in Figure 5 ~ Figure 7. Let P^s denote the steady state probability matrix of the proposed MMOP, that is

$$P^s = \{b_r \mid 0 \le r \le 1+(n+1)\times m\} \tag{7}$$

where b_r denotes the steady-state probability of the state r of the proposed MMOP, that is

$$\begin{cases} b_0\text{: the steady state probability of } W_{0,0,0} \\ b_1\text{: the steady state probability of } W_{0,0,1} \\ b_c\text{: the steady state probability of Transmit Phase } (W_{0,j}) \\ \quad \text{, where } 1 < c \le m+1 \text{ and } 0 < j \le m \\ b_q\text{: the steady state probability of Backoff Phase } (W_{h,j}) \\ \quad \text{, where } m+1 < q \le 1+(n+1)\times m,\ 0 < h, \text{ and } 0 < j \end{cases} \tag{8}$$

P^s can be obtained by repeating the Chapman-Kolmogorov Eq. (9) until the new values and the previous values are sufficiently close.

$$P^k = P^{k-1}P \tag{9}$$

where P^k is the probability of each state after k transition.

It should be noted that τ, which is the probability that a vehicle transmits a WSM frame described in Table I, is required to build P. However, τ is not a constant value and is continuously changed in each transition. The following approximation is proposed to estimate τ_k for the single-step transition probability matrix in the k-th transition $P_{(k)}$. For $P_{(1)}$, τ_1 is set to $1/N$ because the probability that each vehicle acquires the transmission opportunity is nearly proportional to the number of the vehicles in the same coverage range if the network is not extremely congested. For $P_{(k)}$, where k>1, τ_k is obviously the probability that the vehicle stays in TP after the $(k-1)$-th transition. Therefore, let $\pi_{(k-1)}$ be the matrix of the expected probability that the vehicle stays in each state d in the $(k-1)$-th transition, that is:

$$\pi_{(k-1)} = [\pi_{(k-1)}(0), \cdots, \pi_{(k-1)}(d), \cdots, \pi_{(k-1)}(1+(n+1)\times m)] \tag{10}$$

According to Ergodicity Theorem that the expected probability matrix that a vehicle stays in each state will converge to a unique value, which means the matrix multiplication of $\pi_{(k-1)}$ and $P_{(k-1)}$ would be equal to $\pi_{(k-1)}$. Therefore, $\pi_{(k-1)}$ can be obtained by

$$\begin{cases} \pi_{(k-1)} = \pi_{(k-1)}P_{(k-1)} \\ 1 = \sum_{d=0}^{1+(n+1)\times m} \pi_{(k-1)}(d) \end{cases} \tag{11}$$

Based on Eq. (5), τ_k can be obtained by

$$\tau_k = \sum_{d \in TS} \pi_{(k-1)}(d) \tag{12}$$

After obtaining the steady state probability matrix P^s, the frame error rate in SS (FER_{ss}) which is the probability that a frame transmission fails due to collision can be calculated by

$$FER_{ss} = 1-(1-(\sum_{r \in TP} b_r))^{N-1} \tag{13}$$

where $(1-(\sum_{r \in TP} b_r))^{N-1}$ is the probability that no other vehicles in the states of TP.

The throughput in the SS of CI (TH_{ss}) can be calculated by

$$TH_{ss} = \left[\frac{\sum_{l \in TP}(S_{payload} \times b_l)}{T_\sigma \times b_0 + (\sum_{u=1}^{1+(n+1)\times m} T_s \times b_u)} \right] \times N \tag{14}$$

where $T_\sigma \times b_0 + (\sum_{u=1}^{1+(n+1)\times m} T_s \times b_u)$ means the average time length that a vehicle stays in a state, and $\sum_{l \in TP}(S_{payload} \times b_l)$ is the expected data size that a vehicle transmits in a state. Therefore

$$\left[\frac{\sum_{l \in TP}(S_{payload} \times b_l)}{T_\sigma \times b_0 + (\sum_{u=1}^{1+(n+1)\times m} T_s \times b_u)} \right]$$

is the average throughput of one vehicle during SS.

The goodput in SS (GP_{ss}) can thus be estimated by

$$GP_{ts} = TH_{ts} \times FER_{ts} \tag{15}$$

C. Overall performance

Based on the above analysis, the overall throughput (TH) and goodput (GP) of safety message transmission during CCHI can be calculated by

$$\begin{aligned} TH &= TH_{ts} + TH_{ss} \\ GP &= GP_{ts} + GP_{ss} \end{aligned} \tag{16}$$

and the frame error rate (FER) of the safety message transmission during CCHI can be obtained by

$$FER = \frac{GP}{TH} \tag{17}$$

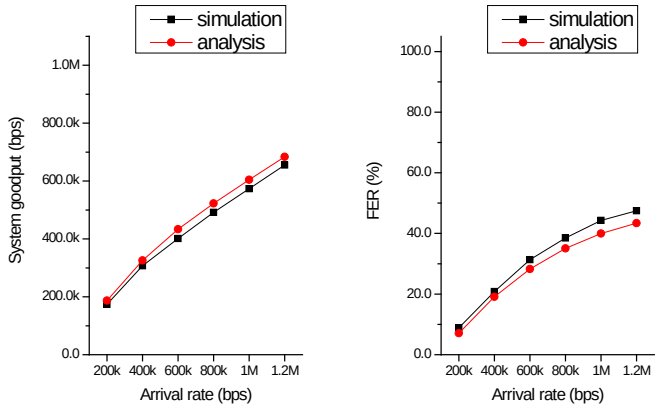
IV. Analytical and Simulation Results

To validate the proposed mathematical model, one scenario is performed in terms of system goodput and frame error rate (FER), where the former is the goodput of all WSM traffics that a vehicle can receive on CCH during CCHI while the latter means the ratio of frames received with errors to total frames received.

The environment is an IEEE 802.11p/1609 OFDM system with a 10MHz frequency band. The detailed parameter settings are listed in Table II. The arriving traffic of each connection follows the Poisson distribution. To simplify the simulation scenario, each vehicle owns one application service and one traffic applying WSM protocol (named WSM traffic), where the former is transmitted on one SCH during SCHI and the latter is transmitted on CCH during CCHI. The default arrival rate of the application service is 100kbps, and that of the WSM traffic is 100kbps.

TABLE II Simulation environment parameter settings

Parameter	Value
Number of CCH	1
Number of SCH	6
Channel bandwidth	10 MHz
CCH Interval	50ms
SCH Interval	50ms
Slot time	13μs
SIFS	32μs
GTI	58μs (SIFS + two slot time)
MTI	45μs (SIFS + one slot time)
WSM payload length	512 bytes
Application payload length	512 bytes
Modulation coding scheme	QPSK 1/2



(a) System goodput (b) FER

Figure 8. The effect of different safety message arrival rates

The number of vehicles is varied from 2 to 12, which means the overall arrival rates is varied from 200 kbps to 1.2Mbps. As shown in Figure 8, the system goodput and the FER increase as the safety message traffic arrival rate increases. The high and increased FER is due to that IEEE 1609/802.11p omits several CSMA/CA mechanisms including association, authentication, RTS/CTS, and ACK functionalities, and thus the vehicles cannot detect and avoid the collision, leading that enormous FER and inferior system goodput are suffered. This means that current IEEE 1609/802.11p standards do not work very well for the M2M vehicular safety message communication even the traffic load is not heavy. On the other hand, the traffic load on SCHs during SCHI is relatively light because there are six SCHs to deal with the application traffics, when the safety messages experienced high FER are disseminated in CCH during CCHI. This implies that the SCH bandwidth is not fully used, and thus the unused SCH bandwidth should be well utilized in order to minimize the high FER on CCH.

V. Conclusion

This paper presents a mathematical model for the M2M vehicular safety researches. The analytical results match the simulation results very well and thus the proposed mathematical model is validated.

References

[1] M. J. Booysen, S. Zeadally, and G.-J. van Rooyen, "Survey of media access control protocols for vehicular ad hoc networks," IET Communications, vol. 5, no. 11, pp. 1619–1631, 2011.

[2] M. J. Booysen, J. S. Gilmore, S. Zeadally, and G.-J. v. Rooyen, "Machine-to-machine (M2M) communications in vehicular networks," KSII Transactions on Internet & Information Systems, vol. 6, no.2, pp. 529-546, 2012.

[3] IEEE std. 1609.1-2006, "IEEE trial-use standard for wireless access in vehicular environments (WAVE) — resource manager," 2006.

[4] IEEE std. 1609.2-2006, "IEEE trial-use standard for wireless access in Vehicular Environments — Security Services for Applications and Management Messages," 2006.

[5] IEEE std. 1609.3-2010, "IEEE standard for wireless access in vehicular environments (WAVE) — networking services," 2010.

[6] IEEE std. 1609.4-2010, "IEEE standard for wireless access in vehicular environments (WAVE) — multi-channel operation," 2011.

[7] IEEE Std 802.11p-2010, "IEEE standard for information technology--telecommunications and information exchange between systems--local and metropolitan area networks--Specific requirements Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 6: wireless access in vehicular environments," 2010.

[8] D. Jiang and L. Delgrossi, "IEEE 802.11p: towards an international standard for wireless access in vehicular environments," in Proc. IEEE Vehicular Technology Conference, pp. 2036-2040, 2008.

[9] Y. Zang, L. Stibor, B. Walke, H. J. Reumerman, and A. Barroso, "A novel MAC protocol for throughput sensitive applications in vehicular environments," in Proc. Vehicular Technology Conference, pp. 2580-2584, 2007.

[10] N. Lu, Y. Ji, F. Lin, and X. Wang, "A dedicated multi-channel MAC protocol design for VANET with adaptive broadcasting," in Proc. Wireless Communications and Networking Conference, pp. 1-6, 2010.

[11] Q. Wang, S. Leng, H. Fu, and Y. Zhang, "An IEEE 802.11p-based multichannel MAC scheme with channel coordination for vehicular ad hoc networks," IEEE Transactions on Intelligent Transportation Systems, vol. 13, no. 2, pp. 449-458, 2012.

[12] Y. Y. Chen, S.C. Liu, and C. Chen, "Channel assignment and routing for multi-channel wireless mesh networks using simulated annealing," in Proc. Global Telecommunications Conference, pp. 1-5, 2006.

ISBN: 978-1-61804-318-4

[13] S.-Y. Wang, C.-C Lin, W.-J. Hong, and K.-C. Liu, “On the performances of forwarding multihopunicast traffic in WBSS-based 802.11(p)/1609networks,” Computer Networks, vol. 55, no. 14, pp. 2592-2607, 2011.

[14] J. R. Gallardo, D. Makrakis, and H. T. Mouftah, “Mathmetical analysis of EDCA’s performance on the control channel of an IEEE 802.11p WAVE vehicular network,” EURASIP Journal on Wireless Communications and Networking, vol. 2010, article ID 489527, 2010.

Cooperative Non-linear Stochastic Wireless Channel Modeling using State Space Analysis

Ankumoni Bora[1], Kandarpa Kumar Sarma[2] and Nikos Mastorakis[3]

Abstract— Stochastic wireless channels are abundant with nonlinear contents. Most of the traditional channel modeling approaches adopts a linearized representation which removes a sizable portion of the contents. It results in lowering of the quality of service (QOS) and degrades system performance. To consider the non-linear components in the channel, better analytical methods are required. The state space modeling is a viable approach as it encompasses higher order terms in the representation and thereby prevents discarding of important channel components. Here, we discuss the application of state space modeling to cooperative wireless channels and provide an analytical treatment of such a propagation medium configured for high data rate cooperative communication.

Keywords— State Space model, path gain, delay spread and nonlinear channel.

I. Introduction

Nonlinearity of a channel is an important aspect in wireless communication which when discarded lowers the reception quality of a signal. Most of the traditional approaches of channel modeling adopt a linearized representation which removes a sizable portion of the contents. It results in lowering of the quality of service (QOS) and degrades system performance. To consider non-linear components in the channel, better analytical methods are required. The state space modeling is a viable approach as it encompasses higher order terms in the representation and thereby prevents discarding of important channel components. Some related works are discussed here. In [1], authors discuss a model of a wireless channel using state space analysis. Rayleigh and Rician state space channels are modeled. Known parameters for the state space matrices are considered in this work.

Ankumoni Bora is with the Department of Electronics and Communication Technology , Gauhati University, Guwahati- 781014, Assam, India (e-mail: ankuele@ gmail.com).

Kandarpa Kumar Sarma is with the Department of Electronics and Communication Technology , Gauhati University, Guwahati- 781014, Assam, , India (e-mail: kandarpaks@ gmail.com).

Nikos Mastorakis is with the Technical University Sofia, Bulgaria, (e-mail: mastor@tu-sofia.bg).

In [2], the state space model for MIMO channel is analyzed. Authors compare state space model with FIR model and obtain an improved results in comparison to FIR model. In [3], a state space based channel equalizer is modeled. The symbol error rate of this equalizer is smaller than FIR based one. In all the above cases authors considered linear channel. But in real time applications, all channel coefficients vary nonlinearly. As the fading of a wireless channel increases, the normal propagation behaviour also changes. Due to this, the traditional tools related to the channel, requires lot of reference symbols during recovery of signals. With increase in nonlinear behaviour, the dependence on reference symbols increases which reduces bandwidth availability. Therefore, non-linear behaviour in channels is required to be considered during designing data recovery systems. There are several methods to analyze channel nonlinearities. These are Volterra model [4], state space model [5] etc. Among these, state space modeling is a viable approach as it encompasses higher order terms in the representation and thereby prevents discarding of important channel components. State space modeling is better since it considers initial and zero state conditions and allows contributions from higher order terms in the calculation. It can be applied to non linear, time invariant systems and multiple input multiple output (MIMO) systems unlike the traditional transfer function based analysis. Further, with state space analysis, the internal state of the system becomes observable and controllable.

Though state space analysis involves elaborate mathematical treatment, its use in nonlinear channels continues to be an evolving area of research. In wireless communication, the properties of a channel depend on the channel parameters. The primary parameters are path gain, delay spread, path loss, coherence bandwidth, input signal quality etc. Analytical modeling is the basis of all practical works. Mathematical expression enables a detailed analysis of the steps required to solve a given problem. For any system, we can design a system state. State of a system means a set of the system variables. These variables are named as state variable. According to [5], a system can be defined as a set of state variables, x_i (t), i=1…n. Here, the variables consist of the system variables with initial time t_0 and the system input. Depending on this state condition, the system output also can be modeled with the input and system variables. The state of a system represents a set of mathematical equations that are called state equations. The state equations are actually differential equations where the derivatives are the differentiation of state variables and the input of the system with respect to time In our work, we consider the nonlinear aspects of the channel. This is done in case of the propagation space existing between source and relay and relay and destination. In this wireless medium, non-linear aspects are

ISBN: 978-1-61804-318-4

prominent. Nonlinear channel modeling becomes less tedious with state space analysis. Here, we derive a state space channel model for cooperative wireless channel. We provide an analytical treatment of a stochastic wireless set-up configured for high data rate cooperative communication. The nonlinear state spaced based channel is linearized by using Taylor series expansion. After linearization, that channel is compared with a Rayleigh frequency selective channel. The Rayleigh channel is estimated by using a zero forcing equalizer. The state space based channel is identified by Numerical algorithm for Subspace State space System Identification (N4SID) and Multivariable Output Error State Space (MOESP).

II. THEORITICAL BACKGROUND

A. Path Gain

In the free space propagation, the path gain in dB is the ratio of receiver (P_r) and transmitter power (P_t) [6]. This is written as

$$\text{Path gain} = 10\log\frac{P_r}{P_t} \dots (1)$$

In the Line Of Sight propagation the transmitter and receiver are located in a minimum distance. Therefore the effects in propagation like diffraction, scattering, reflection etc can be ignore. According to [6], the LOS path gain for two isotropic Tx-Rx antenna is expressed as

$$PG_{LOS} = \left[\frac{\lambda}{4\pi R}\right]^2 \left|2\sin\left(2\pi\frac{h_t h_r}{\lambda R}\right)\right|^2 \dots (2)$$

Here, λ *is the transmission wavelength,*

h_t, h_r *are Tx, Rx antenna height respectively*

and R is the distance between antennas.

In case of NLOS environment, the rays of propagation are effected with diffraction, reflection etc. Because of these effects, the path gain model also changes. The path gain equation for NLOS environment is given as [6]

$$PG_{LOS} = \left[\frac{\lambda}{4\pi(R_0+R_1)}\right]^2 \left|2\sin\left(2\pi\frac{h_t h_r}{\lambda(R_0+R_1)}\right)\right|^2 \frac{D^2}{R_0 R_1}(R_0+R_1) \dots (3)$$

Here, D is the diffraction coefficient,

R_0 is the distance from Tx antenna to the diffraction point and

R_1 is the distance from diffraction point to the Rx antenna.

In case of CMIMO, there are two main channel part. So the path gain also changes in each path. Due to this fact, the path gain equation is modified for each path.

$$PG_{LOS1} = \left[\frac{\lambda 1}{4\pi(R_0+R_1)}\right]^2 \left|2\sin\left(2\pi\frac{h_t h_{r1}}{\lambda 1(R_0+R_1)}\right)\right|^2 \frac{{D_0}^2}{R_0 R_1}(R_0+R_1) \dots (4)$$

$$PG_{LOS2} = \left[\frac{\lambda 2}{4\pi(R_2+R_3)}\right]^2 \left|2\sin\left(2\pi\frac{h_r h_{r1}}{\lambda 1(R_2+R_3)}\right)\right|^2 \frac{{D_1}^2}{R_2 R_3}(R_2+R_3) \dots (5)$$

where,

D_0 is the diffraction coefficient for channel 1,

D_1 is the diffraction coefficient for channel 2,

R_0, R_1 are distances from Tx antenna to diffraction point and diffraction point to the relay respectively and

R_2, R_3 are distances from Relay to diffraction point and diffraction point to the Rx antenna respectively.

B. Delay spread

The wireless channel properties are characterized by the channel parameters of that channel. Delay spread is a one kind of channel parameter which depends on the difference of time of arrival of the channel LOS component and multipath channel component. According to [7], the delay spread can be measured by the root mean square (RMS) delay spread. The mean excess delay equation of the channel is given as [7]

$$\overline{\tau} = \frac{\sum {a_k}^2 \tau_k}{\sum {a_k}^2} \dots (6)$$

where a_k is the amplitude of the nth path

$and\ \tau_k\ is\ the\ \ delay\ time.$

The RMS delay spread of the channel is given as

$$\sigma_\tau = \sqrt{\overline{\tau^2} - (\overline{\tau})^2} \dots (7)$$

where, $\overline{\tau^2} = \frac{\sum {a_k}^2 {\tau_k}^2}{\sum {a_k}^2}$

In cooperative communication, for the two channels the amplitude of the nth path and time delay are varied. Therefore, for the two channel paths, delay spread also will vary.

C. State space modeling

The simple State Space model of a wireless channel is shown in fig 1.

This is the block diagram for a state space model to realize a proper channel.

The state equations for a channel

$$\dot{X} = Ax + Bu$$
$$Y = Cx + Du \dots (8)$$

where, x is the state of the system, u is the input signal of the system, Y is the output signal of the system, A is the state matrix, B is the input matrix, C is the output matrix and D is the direct feed through matrix

The State Space modeling set-up for a wireless channel is shown in Figure 1.

As describe earlier, the state space modeling is a standard mathematical modeling to explain a system. In control system state space model gives the mathematical form of a physical system with the set of input, output and state variables. The state equations with state variables $x_1(t) \dots x_n(t)$ and inputs $u_1(t) \dots u_n(t)$ are represented as follows:

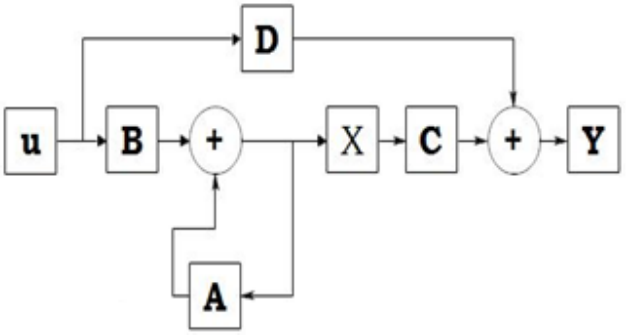


Figure 1: State Space model of a system

$$\dot{x}_1 = f(x_1, u_1, t)$$
$$\dot{x}_2 = f(x_2, u_2, t)$$
$$\vdots$$
$$\dot{x}_n = f(x_n, u_n, t) \ldots (9)$$

where, $\dot{x} = \frac{dx}{dt}$.

The linear state space model can be represented in a matrix form. Let us consider three state variables x_1, x_2 and x_3 and inputs u_1, u_2 and u_3 of a system. The state equations are,

$$\dot{x}_{1=} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + b_{11}u_1 + b_{12}u_2 + b_{13}u_3$$

$$\dot{x}_{2=} a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + b_{21}u_1 + b_{22}u_2 + b_{23}u_3$$

$$\dot{x}_{2=} a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + b_{31}u_1 + b_{32}u_2 + b_{33}u_{3\ldots(10)}$$

Here, a_{ij} and b_{ij} are coefficients of state variable and input respectively.

The matrix-vector form of these equations as follows

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

The vector form of this is as shown in eq (11)

$$\dot{X} = AX + BU \ldots (11)$$

Similarly the output equation also can be written in state space form. The output state equation depends on the state variables and input of the system. In similar way, the vector form of output equation is as follows

$$\dot{Y} = CX + DU \ldots (12)$$

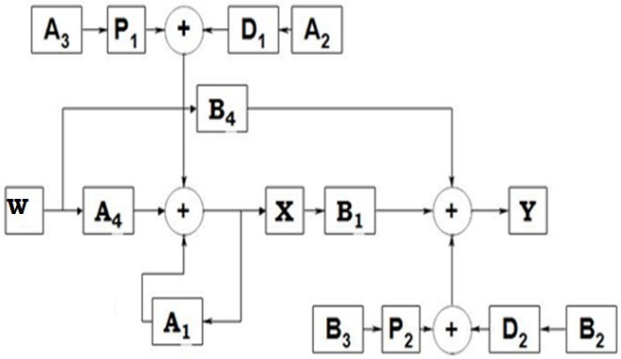


Figure 2: The state space representation of the channel between base station and relay and relay and receiver

III. PROPOSED STATE SPACE MODELING OF A COOPERATIVE WIRELESS CHANNEL

The state space representation of the channel between base station and relay and relay and receiver set-up is shown in figure 2.

The cooperative severely faded channel is assumed as a nonlinear. Because every real time application is nonlinear. Therefore the channel response also should be nonlinear. The cooperative communication has mainly two channels. Source to relay and relay to destination. Each channel is considered as Rayleigh frequency selective channel. The NLOS Rayleigh channel posses Inphase and Quadrature phase component.

The received signal is defined as

$$Y(t) = \sum_{1}^{N} \left[\{I_n(t, \tau_n(t))\}\cos(\omega_{ct}) - \{Q_n(t, \tau_n(t))\}\sin(\omega_{ct})\right] s(t - \tau_n(t)) + V_{I(t)}\cos(\omega_{ct}) - V_{Q(t)}\sin(\omega_{ct}) \ldots . (13)$$

State space model for the channel (BS-RS): For the channel between base station (BS) and Relay station (RS), the path gain and delay spread are considered as channel variables. The state equations for the two state nonlinear channel,

$$\dot{x}_1 = f_1(x_1, x_2, D, P, w, t) \ldots (14)$$
$$\dot{x}_2 = f_2(x_1, x_2, D, P, w, t) \ldots (15)$$

The Taylor series expansion of nonlinear function of equation (31)

$$f_1(x_1, x_2, P, D, w, t) = f_1(x_{1s}, x_{2s}, P, D, w, t) + \frac{\partial f_1}{\partial x_1}(x_1 - x_{1s}) + \frac{\partial f_1}{\partial x_2}(x_2 - x_{2s}) + \frac{\partial f_1}{\partial P}(P - P_s) + \frac{\partial f_1}{\partial D}(D - D_s) + \frac{\partial f_1}{\partial w_0}(w_0 - w_s) + \frac{1}{2}[\frac{\delta^2 f_1}{\delta x_1{}^2}(x_1 - x_{1s})^2 + \frac{\delta^2 f_1}{\delta x_2{}^2}(x_2 - x_{2s})^2 + \frac{\delta^2 f_1}{\delta P^2}(P - P_s)^2 + \frac{\delta^2 f_1}{\delta D^2}(D - D_s)^2 + \frac{\delta^2 f_1}{\delta w_0{}^2}(w_0 - w_{0s})^2] + \frac{\delta^2 f_1}{\delta x_1 \delta x_2}(x_1 - x_{1s})(x_2 - x_{2s}) + \frac{\delta^2 f_1}{\delta x_1 \partial P}(x_1 - x_{1s})(P - P_s) + \frac{\delta^2 f_1}{\delta x_1 \partial D}(x_1 - x_{1s})(D - D_s) + \frac{\delta^2 f_1}{\delta x_1 \delta w_0}(x_1 - x_{1s})(w_0 - w_{0s}) + higher\ terms \ldots (16)$$

To make it linearization neglect the higher terms

$$f_1(x_1,x_2,P,D,w,t) = f_1(x_{1s},x_{2s},P,D,w,t) + \frac{\partial f_1}{\partial x_1}(x_1 - x_{1s}) + \frac{\partial f_1}{\partial x_2}(x_2 - x_{2s}) + \frac{\partial f_1}{\partial P}(P - P_s) + \frac{\partial f_1}{\partial D}(D - D_s) + \frac{\partial f_1}{\partial w_0}(w_0 - w_s) \dots (17)$$

Similarly for equation (15)

$$f_2(x_1,x_2,P,D,w,t) = f_2(x_{1s},x_{2s},P,D,w,t) + \frac{\partial f_2}{\partial x_1}(x_1 - x_{1s}) + \frac{\partial f_2}{\partial x_2}(x_2 - x_{2s}) + \frac{\partial f_2}{\partial P}(P - P_s) + \frac{\partial f_2}{\partial D}(D - D_s) + \frac{\partial f_2}{\partial w_0}(w_0 - w_s) \dots (18)$$

The state space model of these equations (17) and (18)

$$\frac{d}{dt}\begin{bmatrix}x_1\\x_2\end{bmatrix} = \begin{bmatrix}\frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2}\\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2}\end{bmatrix}\begin{bmatrix}x_1 - x_{1s}\\x_2 - x_{2s}\end{bmatrix} + \begin{bmatrix}\frac{\partial f_1}{\partial P} & 0\\ 0 & \frac{\partial f_2}{\partial P}\end{bmatrix}[P - P_s \quad 0] + \begin{bmatrix}\frac{\partial f_1}{\partial D} & 0\\ 0 & \frac{\partial f_2}{\partial D}\end{bmatrix}[D - D_S \quad 0] + \begin{bmatrix}\frac{\partial f_1}{\partial w_0}\\ \frac{\partial f_2}{\partial w_1}\end{bmatrix}[\omega_0 - \omega_s] \dots (19)$$

Therefore,

$$\dot{X}_I = A_1X + A_2P + A_3D + A_4W \dots (20)$$

For the output inphase component

$$Y_{I1} = g_1(x_1, x_2, D, P, w, t) \dots (21)$$
$$Y_{I2} = g_2(x_1, x_2, D, P, w, t) \dots (22)$$

After Taylor series expansion and linearization the output state space model is as follows

$$\frac{d}{dt}\begin{bmatrix}x_1\\x_2\end{bmatrix} = \begin{bmatrix}\frac{\partial g}{\partial x_1} & \frac{\partial g}{\partial x_2}\\ \frac{\partial g}{\partial x_1} & \frac{\partial g}{\partial x_2}\end{bmatrix}\begin{bmatrix}x_1 - x_{1s}\\x_2 - x_{2s}\end{bmatrix} + \begin{bmatrix}\frac{\partial g_1}{\partial P} & 0\\ 0 & \frac{\partial g_2}{\partial P}\end{bmatrix}[P - P_s \quad 0] + \begin{bmatrix}\frac{\partial g_1}{\partial D} & 0\\ 0 & \frac{\partial g_2}{\partial D}\end{bmatrix}[D - D_S \quad 0] + \begin{bmatrix}\frac{\partial g_1}{\partial w_0}\\ \frac{\partial g}{\partial w_1}\end{bmatrix}[w - w_s] \dots (23)$$

The output Inphase state equation

$$Y_I = B_1X + B_2P + B_3D + B_4W \dots (24)$$

The quadrature components for both state input and output equations are

$$\dot{X}_Q = A_{Q1}X + A_{Q2}P + A_{Q3}D + A_{Q4}W \dots (25)$$
$$Y_Q = B_{Q1}X + B_{Q2}P + B_{Q3}D + B_{Q4}W \dots (26)$$

Now applying Laplace transforms,

$$Y_I(s) = B_1[sI - A]^{-1}[A_2P(s) + A_3D(s) + A_4W(s)] + B_2P(s) + B_3D(s) + B_4W(s) \dots (27);$$

$$Y_I(S) = H_I(S) + C_1P(S) + C_2D(S) \dots (28)$$

And

$$Y_Q(S) = H_Q(S) + C_3P(S) + C_4D(S) \dots (29)$$

Here, $H_I(S) = B_1[sI - A]^{-1}A_4 + B_4 \dots$ (30)

$H_Q(S) = B_{Q1}[sI - A]^{-1}A_{Q4} + B_{Q4} \dots$ (31)

The same type of state space channel model equation we can designed for the next channel path i,e from relay to destination. The previous output is considered as input in this case.

IV. EXPERIMENTAL DETAILS AND RESULTS

The proposed state space model for a wireless channel is shown in figure 2. The same channel model can be applied for both transmitter to the relay station and relay to the destination. The nonlinear state spaced based channel is linearized by using Taylor series expansion. After linearization, that channel is compared with a Rayleigh frequency selective channel. The Rayleigh channel is estimated by using zero forcing equalizer.

The state space based channel is identified by state space system identification algorithm such as Numerical algorithm for Subspace State space System Identification (N4SID) and Multivariable Output Error State Space (MOESP) [8]. After that the simulated output is compared with measured output. The figure 3 shows the measured and simulated output for 1000 samples. The best fitness value for this system is 71.68% based on N4SID. The figure 4 shows the N4SID based Singular Value Decomposition (SVD). The best order is 2 in this case. For 500 samples, the modeled and simulated output is shown in figure 5. In this case the best fit value is 62.53%.

The SVD model of this system for 500 samples is shown in figure 6. In this case also the best order is 2.

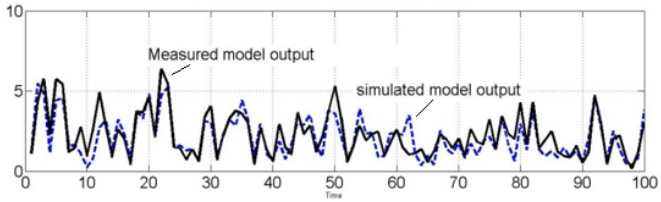


Figure 3: measured and simulated output of state space model for 1000 samples.

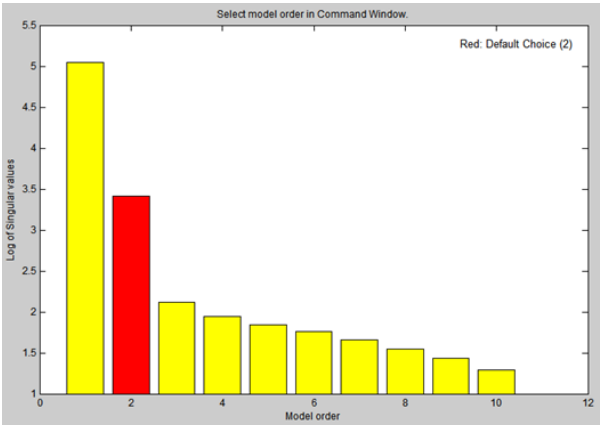


Figure 4: N4SID based SVD for 1000 samples.

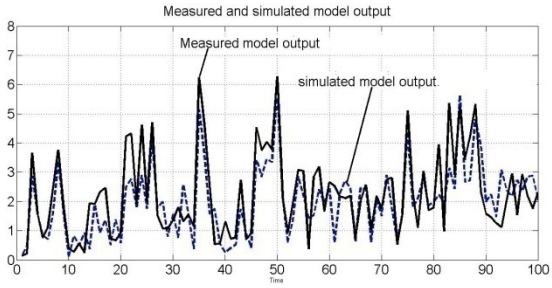


Figure 5: modeled and simulated output with best fit value 62.53% of 500 samples for the system.

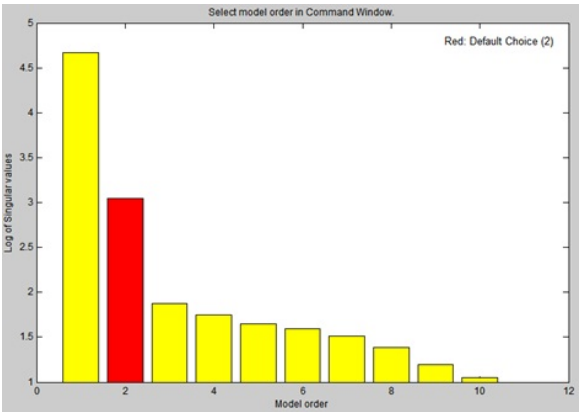


Figure 6: N4SID based SVD for 500 samples

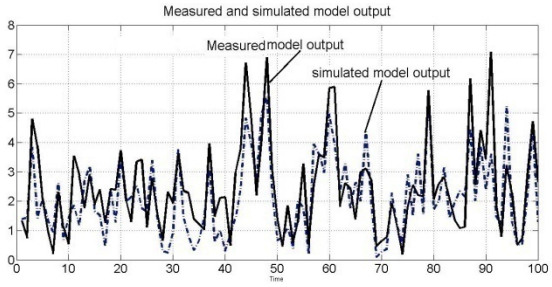


Figure 7: measured and simulated output of state space model for 100 samples.

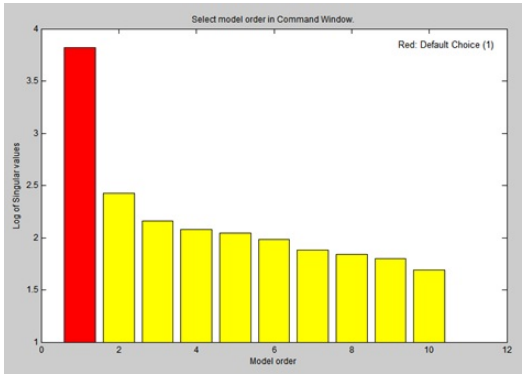


Figure 8: N4SID based SVD for 100 samples

Similarly for 100 samples, the best fit value is 61.25%. The modeled and simulated output is shown in figure 7. The SVD model of this system for 500 samples is shown in figure 8. Figure 11 shows the unit circle of this system for 1000 samples. The stability of a system can be determined by the position of poles and zeroes. According to the figure 11, all poles and zeroes are inside the unit circle. Therefore this system can be treated as stable system.

But in case of 100 and 500 samples, the unit circles show unstable result.

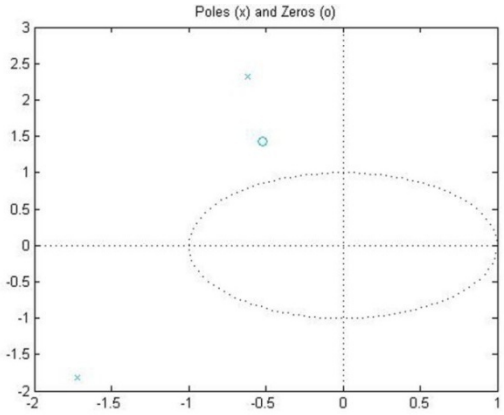


Figure 9: unit circle for 100 sample based on N4SID algorithm

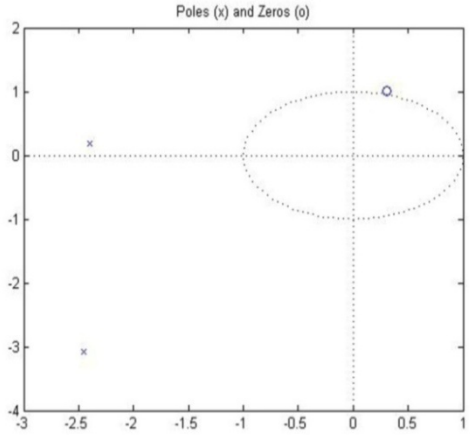


Figure 10: unit circle for the N4SID based system for 500 samples.

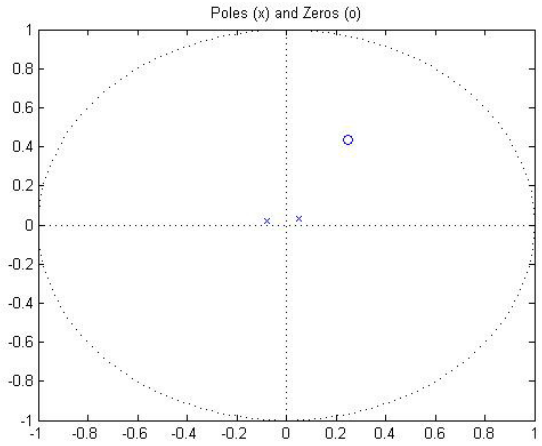


Figure 11: unit circle for the N4SID based system for 1000 samples.

From this analysis it is clear that the system with 1000 samples is a stable one. The state matrix values of the state equation are obtained from the Rayleigh channel coefficient. The parameters for the Rayleigh channel estimation as shown in table 1.

Table 1: Parameters for Rayleigh channel estimation

Parameters	values
No of frames	100
Length of frame	1000
Length of pilot	100
modulation	BPSK
channel	Rayleigh channel
SNR	-20 to 20dB
noise	AWGN
Equalizer	Zero forcing

IV. CONCLUSION

This paper presents a state space model of a stochastic non-linear channel between BS-RS as part of a high data rate cooperative communication setup. It considers Rayleigh fading. System identification has been done using N4SID and MOESP algorithms. The proposed model has been simulated for a frequency selective channel and compared with the measured signal. This validates the proposed state-space model for a non-linear channel.

ACKNOWLEDGMENT

The authors express their thanks and gratitude to the Ministry of Communication and Information Technology (MCIT), Govt. of India for their support in executing the work.

References

[1] X. Li, "State Space Estimation for Wireless Fading Channels", MS Thesis, Dept. of System Science, University of Ottawa, Canada, 2002.

[2] C. Zhang, R.R. Bitmead, "State Space Modeling for MIMO Wireless Channels," IEEE International Conference on Communications (ICC 2005),vol. 4, pp. 2297 - 2301, 16-20 May, 2005.

[3] C. Zhang, R.R. Bitmead, "MIMO Equalization with State-Space Channel Models", IEEE Trans. on signal processing, vol. 56, No. 10, pp 5222-5231, October 2008.

[4] G. Mileounis, P. Koukoulas, N. Kalouptsidis, "Input–output identification of nonlinear channels using PSK, QAM and OFDM inputs" Published in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008), pp 3589-3592, Las Vegas, Nevada, March 31-April 4, 2008.

[5] D. G. Schultz and J. L. Melsa, "State Functions and Linear Control Systems", McGraw-Hill, New York, 1967.

[6] J.S Lu, Bertoni, L Henry, C. Chrysanthou and J. Boksiner, "Simplified Path Gain Model For Mobile to Mobile Communications in an Urban High -Rise Environment", Published in Sarnoff Symposium, 2010 IEEE, pp 1-6, 12-14 April, 2010.

[7] T. S. Rappaport, "Wireless Communications: Principles and Practice", 2nd Ed., Pearson Education, New Delhi, 2004.

[8] I. W Jamaludin, N. A Waheb, N.S. Khalid, S. Sahlan, Z Ibrahim, M. F. Rahmat, "N4SID and MOESP Subspace Identification Methods", 2013 IEEE 9th International Colloquium on Signal Processing and its Applications, pp 140-145, Kuala Lumpur, Malaysia, 8-10 March, 2013.

Adaptive NARMA Equalization of Nonlinear ITU Channels

Murchana Baruah, Aradhana Misra, Kandarpa Kumar Sarma and Nikos Mastorakis

***Abstract*—Wireless transmission in severe fading environments require suitable channel equalization methods for true recovery of the transmitted signal. Designing an equalizer, however, depends on the channel conditions. Traditional methods of channel modeling rely upon certain tools including tapped delay line (TDL) aided by Autoregressive (AR), Moving Average (MA) and Autoregressive Moving Average (ARMA) models where non-linear aspects are disregarded. But the actual characteristics of a wireless channel can be more accurately defined by considering nonlinear channel contents. In this paper, ITU channels are approximated using Nonlinear ARMA (NARMA) process. At the receiver end, adaptive nonlinear equalization is performed using least mean square (LMS) algorithm. The specifications of the NARMA channel and the adaptive filter parameters are varied for different channel lengths, for both ITU pedestrian and vehicular conditions and a range of experiments performed.**

Keywords—*adaptive equalization, channel modeling, ITU channels, least mean square, NARMA*.

I. Introduction

Fading is a significant aspect in wireless communications. It mainly occurs due to multiple reflections, refraction, scattering and diffraction of the transmitted signal from different objects present in the medium [1]. Fading results in a destructive effect on the original signal. The channel impulse response (CIR) for a wireless system can be considered to be a linear time invariant system. But in a practical scenario, a wireless channel is highly volatile due to the relative motion of the transmitter-receiver pair [1]. Shadowing also causes variation of the channel characteristics [1]. These factors, in the real scenario, contribute towards nonlinearity of the channel during severe fluctuations which otherwise remains passive in normal fading conditions. The nonlinearity of the channel, therefore, should be considered during channel modeling. Autoregressive (AR), Moving Average (MA) and Autoregressive Moving Average (ARMA) models are different conventional methods used for representing the stochastic behavior in wireless channels [2]-[4]. These processes also enable time series modeling. ARMA with its feedback structure provides a more stable process than AR and MA. However, these models use tapped delay line (TDL) structures that fail to track the nonlinear variations of a wireless channel effectively. Also, they show high levels of dependence on reference symbols during recovery. Fading effects either severe or normal must be mitigated due to its deleterious effect on the original signal for proper recovery of the message bits. Various techniques are explored to minimize fading. Different coding and modulation schemes, multiplexing techniques, diversity methods are implemented to reduce errors during transmission [5]. In this regard, diversity technique is very effective in fading scenarios. Different diversity schemes like spatial diversity, time diversity and antenna diversity are employed in this regard [1] and [5]. Apart from such techniques, channel equalization must be effectively used at the receiver end to remove the effects of the fading channel. Both linear and nonlinear equalizers are applied in this regard. Different channel equalization methods have been adopted for removing the detrimental effects of fading. Linear equalizers like Least Square (LS) [6], zero forcing [7]-[9], minimum mean square error [7]-[9] etc. does not give satisfactory results under severe fading conditions. Decision Feedback Equalizer (DFE), a nonlinear class of equalizer in this regard provides satisfactory results [1] and [9]. However, the disadvantage lies in its ability to propagate noise in due course of time. In this paper, we explore a non-linear ARMA (NARMA) process for generating the complex channel coefficients of an ITU channel to represent the actual scenario with a proper real time set-up. Here, an adaptive nonlinear equalizer using least mean square (LMS) algorithm is designed for equalization at the receiver end [10]. The performance of the system is compared to that of a conventional TDL and DFE based equalizer. A range of experiments are performed. Results show that the NARMA approximation of the stochastic wireless channel in combination with the proposed adaptive equalization is effective in mitigating adverse effects of fading. The rest of the paper is organized as follows. Section 2 gives a description of the theoretical considerations involved, Section 3 shows the experimental details and results obtained and Section 4 concludes.

M. Baruah is with the Electronics and Communication Engineering Department, Gauhati University, Guwahati, 781014, Assam, India (e-mail: murchanabaruah@gmail.com).

A. Misra is with the Electronics and Communication Engineering Department, Gauhati University, Guwahati, 781014, Assam, India (e-mail: aradhana66@gmail.com).

K. K. Sarma is with the Electronics and Communication Technology Department, Gauhati University, Guwahati, 781014, Assam, India (e-mail: kandarpaks@gmail.com).

N. Mastorakis is with the Military Institutions of University Education, Hellenic Naval Academy, Terma Hatzikyriakou, 18539, Piraeus, Greece (e-mail: mastor@ieee.org).

II. Proposed Method Of Nonlinear Equalization In NARMA Approximated ITU Channels

This section discusses in detail the considerations employed in the design of the proposed non-linear adaptive equalizer for adaptation in stochastic wireless channels approximated using NARMA models. It is divided into three subsections. The first subsection provides an overview of the system model. The second subsection describes the proposed adaptive NARMA filter used as a system identifier (channel) and the third subsection provides a description of the proposed NARMA equalization method.

A. System Model

The channel coefficients for an ITU specified channel required for channel modeling in designing the system are generated using Jakes' tap gain approach considering the Third Generation Partnership Project (3gpp) specifications and verified using data resembling practical conditions. A fading channel being nonlinear in characteristics [11] and [12] and approximated by an ARMA specified structure [4], can be expressed using a NARMA process. The mathematical expression of a NARMA process is given as [13].

$$h(n) = \sum_{i=0}^{p} a(i)v(n-i) + \sum_{j=1}^{q} b(j)h(n-j) + \sum_{i=0}^{p}\sum_{j=1}^{q} a(i,j)v(n-i)v(n-j) + \sum_{i=0}^{p}\sum_{j=1}^{q} b(i,j)h(n-i)h(n-j) + \sum_{i=0}^{p}\sum_{j=1}^{q} c(i,j)v(n-i)h(n-j) \quad (1)$$

Here, $a(i)$, $b(j)$, $a(i,j)$, $b(i,j)$, $c(i,j)$ are the MA, AR, non-linear MA, non-linear AR, non-linear cross terms respectively and p and q denotes the order of the MA and AR process respectively. Also, $h(n)$ and $v(n)$ represents the complex channel gain and the additive white gaussian noise (AWGN) respectively.

Here, we assume a 1 X 1 single input single output (SISO) system. We consider frequency selective block fading Rayleigh channel and Binary Phase Shift Keying (BPSK) modulation scheme. Fig. 1 shows the schematic diagram of the proposed system.

In this system, we represent the channel, $h(n)$ as shown in Fig. 1 in terms of a NARMA model. This process is similar to that of system identification by an adaptive filter as in [14] and [15]. A suitable length of training sequence is considered from the original data $d(n)$ which is fed to the adaptive NARMA filter block.

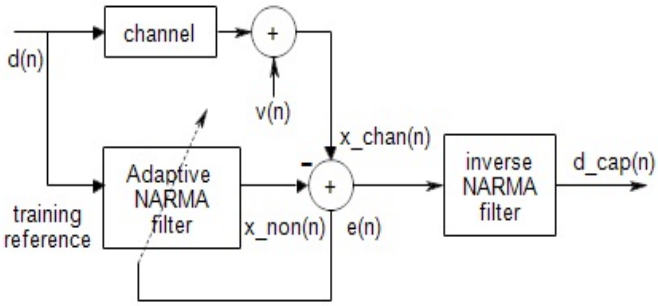


Fig. 1: System Model

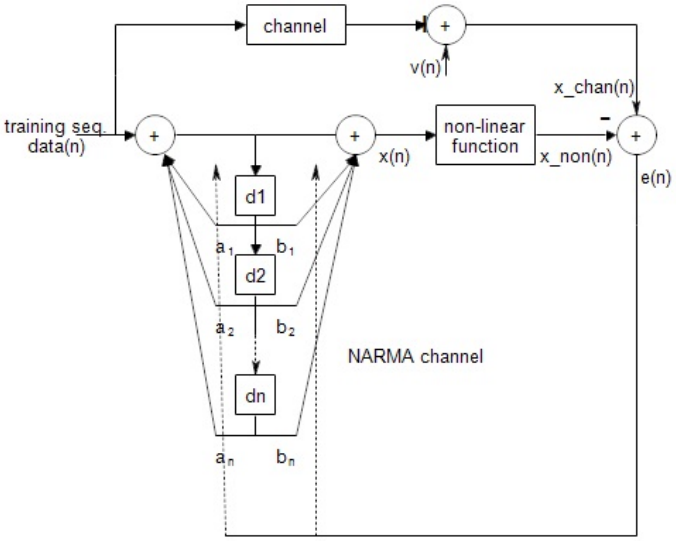


Fig. 2: System Identification using an Adaptive NARMA Filter

The initial coefficients of the NARMA filter are set to zero. The coefficients of the NARMA filter are then updated using LMS algorithm from error obtained from the channel output $x\_chan(n)$ and the adaptive NARMA filter output $x\_non(n)$. As this error minimizes to zero, the final coefficients obtained from the adaptive NARMA filter are used to design an inverse NARMA filter which acts as an equalizer.

B. System Identification using Adaptive NARMA Filter

The block diagram of the adaptive NARMA filter representing the nonlinear channel is shown in Fig. 2. The blocks $d1, d2, \ldots dn$ in Fig. 2 represents the delay units of the NARMA block. The poles and zeros of the system are denoted as $(b_1, b_2, \ldots b_l)$ and $(a_1, a_2, \ldots a_m)$ respectively where (l, m) is the order of the NARMA process. If $data(n)$ is the known training sequence which is random in nature, eqs. (2) and (3) gives the ARMA and NARMA process outputs $x(n)$ and $x\_non(n)$ respectively for time instant n. These are as follows:

$$x(n) = -\sum_{i=1}^{l} b_l x\_non(n-i) + \sum_{j=0}^{m} a_m data(n-j) \quad (2)$$

$$x\_non = scalar * \frac{1-e^{-x(n)}}{1+e^{-x(n)}} \quad (3)$$

Here, we have considered the nonlinear Gaussian function as $\frac{1-e^{-x(n)}}{1+e^{-x(n)}}$ as seen from eq. 3. This function gives suitable results while tracking the nonlinear channel by the adaptive NARMA system. The factor $scalar$ is a real valued number which must be adjusted such that the adaptive NARMA filter tracks the channel suitably. The error $e(n)$ is given as the difference between the received signal, x_chan(n) and the output of the adaptive NARMA filter, x_non(n).

$$e(n) = x\_chan(n) - x\_non(n) \quad (4)$$

The coefficients of the adaptive NARMA filter are initially set to zero which are then updated according to the LMS algorithm [10] as given in eqs. 5 and 6.

$$b_l(n+1) = b_l(n) + \mu * x\_non(n) * e(n) \quad (5)$$

$$a_m(n+1) = a_m(n) + \mu * data(n) * e(n) \quad (6)$$

Here, μ is the step size which is an important factor for deciding the convergence of the system. The step size must be adjusted accordingly to give proper convergence of the mean square error (MSE) obtained from the error, $e(n)$. As the LMS error minimizes, the resulting coefficients can be used for obtaining the inverse NARMA filter coefficients and subsequently for determining the original signal.

C. Proposed NARMA Equalizer

The block diagram of the proposed NARMA equalizer is shown in Fig. 3. The structure of the NARMA equalizer is same as that of the adaptive NARMA filter used for identifying the channel described in the previous subsection. An addition of a decision device (DD) is used in predicting the output signal to prevent the flow of error. In this case, we use an inverse nonlinear logarithmic function to obtain $x\_inv(n)$ from the received data $x\_chan(n)$ which is passed through the ARMA filter. If (a_m, b_l) are the final coefficients obtained from the adaptive NARMA filter, the equalized signal $d\_cap(n)$ is obtained as shown in the following eqs.

$$x\_inv(n) = -\log\left(\frac{1-\text{x\_chan(n)}}{1+\text{x\_chan(n)}}\right) \quad (7)$$

$$d\_cap(n) = -\sum_{i=0}^{l} b_l x\_inv(n-i) + \sum_{j=1}^{m} a_m d\_cap(n-j) \quad (8)$$

After BPSK demodulation of the received signal, we obtain the decoded bits, $d\_dec(n)$.

III. Experimental Details And Results

Here, we consider a block fading frequency selective rayleigh fading channel with a training sequence overhead of 40%. Two cases of channel conditions are analysed using the proposed system. The cases are mentioned as follows.

Case 1: ITU pedestrian case with a speed of 10 $kmph$ and 4 channel taps.

Case 2: ITU vehicular case with a speed of 120 $kmph$ and 6 channel taps.

Experiments are performed for both the cases. The channel simulation parameters for the two cases are shown in Table 1.

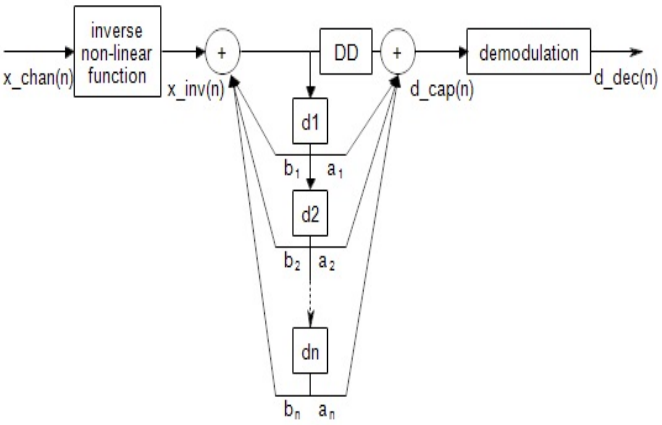


Fig. 3: NARMA Equalizer

Table 1: Simulation parameters

Parameters	Case 1	Case 2
Speed	$10kmph$	$120kmph$
Maximum Doppler Shift	$19.9Hz$	$238.89Hz$
No. of channel taps	4	6
Chip Rate	$3.84 * 10^6\ Hz$	$3.84 * 10^6\ Hz$
Carrier Frequency	$2.15 * 10^9\ Hz$	$2.15 * 10^9\ Hz$
Wavelength	$0.139535\ m$	$0.139535\ m$

The results of the proposed system are analysed and compared with a conventional LS, DFE and an ARMA equalizer.

Figs. 4 and 5 compares the real and imaginary value of the output signal respectively obtained from the fading channel and the one tracked by the adaptive NARMA filter for a Signal to Noise Ratio (SNR) of 10 db. Slight variations are observed due to the effect of AWGN on the transmitted signal.

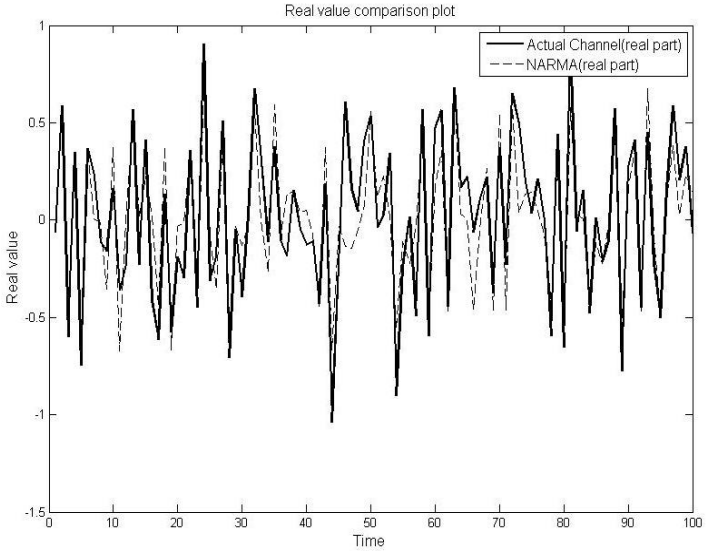


Fig. 4: Real value comparison plot for case 2

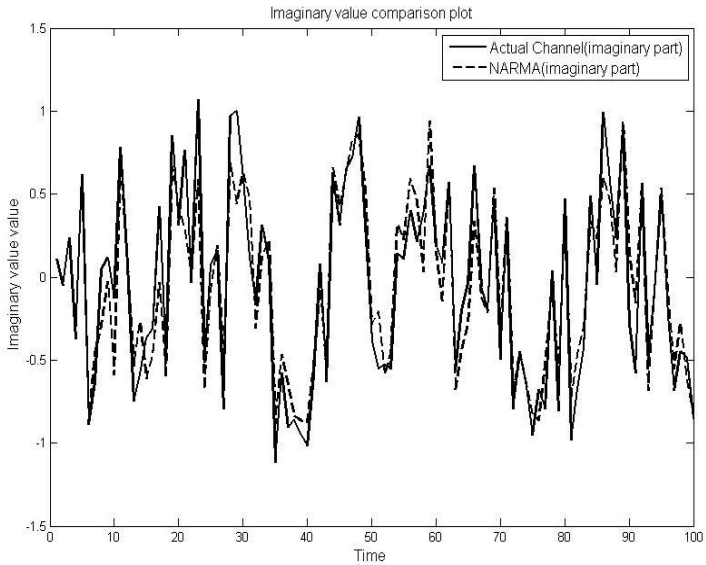


Fig. 5: Imaginary value comparison plot for case 2

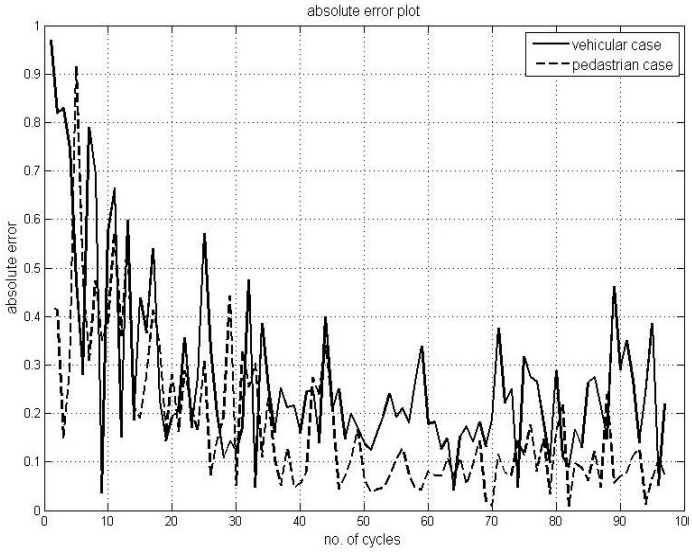


Fig. 6: Absolute error plot from LMS Algorithm

The average mean square error (MSE) convergence plots for both NARMA pedastrian case and NARMA vehicular case for a step size of 0.05 (Case 1) and 0.005 (Case 2) are shown in Fig. 6. The vehicular case shows a poor convergence when compared to the pedestrian one due to critical fading conditions. From Fig. 6, it is also seen that the convergence of the curve takes place in 50 cycles for the pedestrian case and at about 70 cycles for the vehicular case.

Figs. 7 and 8 shows the Bit Error Rate (BER) plot against the SNR for the proposed equalizer for Cases 1 and 2 respectively. The results are compared with conventional methods of LS, DFE and ARMA equalizer. The results show that the proposed system gives better results than the conventional methods. This is due to the consideration of the nonlinear terms in the proposed method. It is seen that the performance of the vehicular case with higher channel length is poor compared to the pedastrian case with lower channel length. LS shows the worst performance followed by DFE and ARMA.

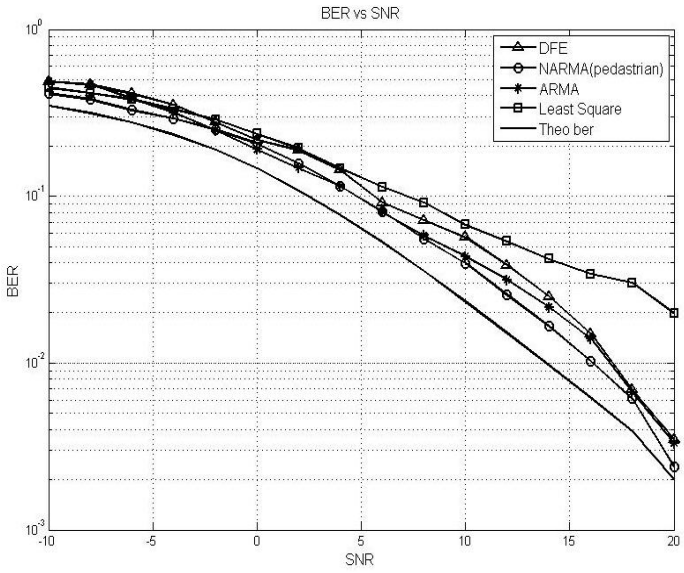


Fig. 7: BER-SNR comparison plot for case 1

Table 2: Equalizer Parameters

Parameters	Case 1	Case 2
NARMA(order)	(1,4)	(3,6)
NARMA(step size)	0.05	0.005
DFE(order)	(4,2)	(6,4)
DFE(step size)	0.05	0.005

Table 3: Average value of the MSE obtained after 50 numbers of random trial for SNR = 20db.

Equalization method	Case1 (pedastrian)	Case 2 (vehicular)
LS	0.0166	0.3144
DFE	0.0036	0.2542
ARMA	0.0034	0.0874
NARMA	0.0020	0.0054

The order of the adaptive NARMA filter coefficients obtained after a number of cycles from LMS algorithm and subsequently used for channel equalization are specified in Table 2. The order of the DFE equalizer is also shown in Table 2. The order of the NARMA and DFE equalizer is found to be greater for Case 2 due to increase in the number of channel taps and Doppler shift.

Table 3 shows a comparison of the average MSE for the different conventional methods with both cases of the proposed system. The results were averaged for some cycles obtained from random trails. The results are shown for an SNR of 20 db.

The results verify the performance and reliability of the proposed adaptive NARMA system. It can be inferred that in severe fading conditions as in Case 2, the proposed system gives satisfactory results compared to conventional methods of LS, DFE or ARMA. This is due to the inclusion of a nonlinear term in the proposed system which gives required nonlinear characteristics to the system for tracking the channel coefficients effectively.

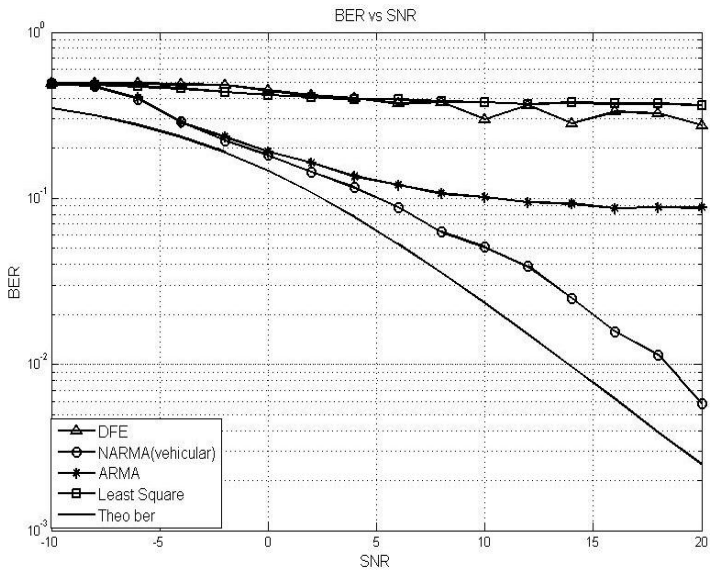


Fig. 8: BER-SNR comparison plot for case 2

IV. Conclusion

The nonlinear nature of a frequency selective block fading Rayleigh channel is explored in this paper and an adaptive NARMA equalizer is proposed for equalizing the channel. Experiments are performed for a lower fading scenario with lower channel taps and a severe fading scenario with higher channel taps to justify the feasibility of the proposed system. Performance evaluation is done from which it can be concluded that the proposed adaptive NARMA equalizer provides better performance in severe fading scenarios. The proposed model is verified to give good results when compared to conventional schemes.

References

[1] T. S. Rappaport, *Wireless Communications-Principles And Practice*, 2nd ed.ition, PHI, New Delhi, 2008.

[2] W. Turin, R. Jana, C. Martin and J. Winters J., "Modeling Wireless Channel Fading", *Proceedings of Vehicular Technology Conference*, Atlantic City, NJ, October 2001.

[3] K. E. Baddour and N. C. Beaulieu, "Autoregressive Modeling for Fading Channel Simulation", *IEEE Transactions on Wireless Communications*, vol. 4, no. 4, July 2005, pp. 1650-1662.

[4] H. Mehrpouyan and S. D. Blostein, "ARMA Synthesis of Fading Channels", *IEEE Transactions on Wireless Communications*, vol. 7, no. 8, August 2008, pp. 2846-2850.

[5] T. M. Duman and A. Ghrayeb, *Coding for MIMO Communication Systems*, 2nd edition, John Wiley & Sons Ltd, England, 2007.

[6] A. Khlifi and R. Bouallegue, "Performance Analysis of LS and LMMSE Channel Estimation Techniques for LTE Downlink Systems", *International Journel of Wireless & Mobile Networks (IJWMN)*, vol. 3, no. 5, October 2011, pp. 141-149.

[7] V. J. Naveen, K. M. Krishna and K. RajaRajeswari, "Performance analysis of equalization techniques for MIMO systems in wireless communications", *International Journel of Smart Home*, vol. 4, no. 4, October 2010, pp. 47-63.

[8] V. Dawar and R. Sharma, "Reduction in Bit Error Rate from Various Equalization Techniques for MIMO Technology", *International Journel of Soft Computing and Engineering (IJSCE)*, vol. 2, no. 4, September 2012, pp. 66-70.

[9] G. Singh and P. Sharma, "BER Comparison of MIMO Systems using Equalization Techniques in Rayleigh Flat Fading Channel", *International Journel of Engineering Research and Applications (IJERA)*, vol. 2, no. 5, October 2012, pp. 2008-2015.

[10] M. H. Hayes, *Statistical Digital Signal Processing And Modelling*, 1st edition, John Wiley & Sons, Inc, USA, 1996.

[11] Bhuyan M and Sarma K. K. (2014), "Nonlinear Model based Prediction of Time Varying SISO-MIMO Channels using FANN-DFE Combination", *Proceedings of IEEE 1st International Conference on Emerging Trends and Applications in Computer Science*, Shillong, India.

[12] Q. Min, Mouyannan and Z. Xiuping, "The Modeling and Equalization Technique of Nonlinear Wireless Channel", *The Open Cybernatics and Systemics Journel*, vol. 2014, no. 8, December 2014, pp. 297-301.

[13] M. Bhuyan and K. K. Sarma, "MIMO-OFDM Channel Tracking using a Dynamic ANN Topology", *Journal of World Academy of Science*, Engineering and Technology, vol. 2012, no. 71, November 2012, pp. 1321-1327.

[14] S. C. Douglas, "Introduction to Adaptive Filters", in *The DSP Handbook*, V. J. Madisetti and D. Williams, eds. (Boca Raton, FL: CRC/IEEE Press, 1998), Chapter 18.

[15] I. Dornean, M. Topa, B. S. Kirei and E. Szopos, "System Identification with Least Mean Square Adaptive Algorithm", *Interdisciplinarity in Engineering: Scientific International Conference*, TG. Mures-Romania, November 2007.

Language, communication and society: a gender based linguistics analysis

P. Cutugno, D. Chiarella, R. Lucentini, L. Marconi and G. Morgavi

***Abstract*—**The purpose of this study is to find evidence for supporting the hypothesis that language is the mirror of our thinking, our prejudices and cultural stereotypes. In this analysis, a questionnaire was administered to 537 people. The answers have been analysed to see if gender stereotypes were present such as the attribution of psychological and behavioural characteristics.

In particular, the aim was to identify, if any, what are the stereotyped images, which emerge in defining the roles of men and women in modern society. Moreover, the results given can be a good starting point to understand if gender stereotypes, and the expectations they produce, can result in penalization or inequality. If so, the language and its use would create inherently a gender bias, which influences evaluations both in work settings both in everyday life.

***Keywords*—**computational linguistics, language and gender studies, communication studies, gender stereotypes

I. Introduction

LANGUAGE transmits information, in number and variety, far more than appears on the surface. There is a direct relationship between reality, language and thought. In particular, the language expresses our vision of reality: it does not reflect the world in itself, but the way it is interpreted by us. The language reflects the culture of a society and influences its behaviours. Supporting the gender language, in order to sensitize society on the proper use of the Italian language in a respectful perspective of both genders, is an important step to address the problem of violence against women. Violence against women is an atrocious violation of fundamental rights. In Italy, in 2011, one hundred thirty-seven women were murdered; in 2012 one hundred twenty-four and, in 2013, there were one hundred thirty-four victims of femicide. In Italy, sex crimes have been thirty-six in the first six months of 2014. In the majority of cases, women were killed by their husbands, partners or former partners.

The first thematic report on femicide was made in Italy by Rashida Manjoo and was presented to ONU in 2012 [1]. In this report it was pointed out that most of the violence is not reported because domestic violence is not recognized as a crime and victims are economically subordinate than the perpetrators of violence. In particular, the report pointed out that, in Italy, gender stereotypes are deeply rooted and predetermine the roles of men and women in society.

To deal with these issues, the report "on eliminating gender stereotypes in the EU" [2] highlights the need to:

- Emphasise the need for education programmes/curricula focusing on equality between men and women, respect for others, respect amongst young people, respectful sexuality and rejection of all forms of violence, as well as the importance of training teachers in this subject;
- Emphasise the need for a gender mainstreaming process in schools and therefore encourages schools to design and implement awareness training exercises and practical exercises designed to promote gender equality in the academic curriculum;
- Point out that, although a majority of countries in the EU have gender-equality policies for higher education, almost all the policies and projects are focused on young women; calls, therefore, on the Member States to draw up general national strategies and initiatives combating gender stereotyping in higher education and targeting young men.

Although any specific course focused on respect for gender and gender equality is not envisaged by the Italian government, in Italy we find countless projects implemented by regions [3], provinces [4], schools [5] and community services sector [6]. All these projects are aimed at raising awareness of these problems among teenagers and the public.

This preliminary work aims the same objectives by providing a basic survey on gender stereotypes [7] [8] in Italian language.

II. The analysis

A. Data set

In order to reflect on "being men and women" and "building positive relationships", the City of Genoa (Italy) has distributed a brief questionnaire (see Figure 1 Questionnaire submitted to the citizens), in the period from November 2013 to February 2014, where it was asked to reflect about stereotypes and prejudices and how they are accomplices of violence.

D. Chiarella is with the National Research Council of Italy, Institute of Intelligent Systems for Automation – Genoa branch, Genova, Italy (phone: +39-010-6475220; fax: +39-010-6475207; e-mail: chiarella@ge.issia.cnr.it or davide.chiarella@ilc.cnr.it).

P. Cutugno, R. Lucentini, L. Marconi are with the National Research Council of Italy, Institute of Computational Linguistics "A. Zampolli" – Genoa branch, Genova, Italy (e-mail: paola.cutugno | roberta.lucentini | lucia.marconi @ilc.cnr.it).

G. Morgavi is with the National Research Council of Italy, Institute of Electronics, Computer and Telecommunication Engineering – Genoa branch, Genova, Italy. (e-mail: giovanna.morgavi@ieiit.cnr.it)

ISBN: 978-1-61804-318-4

In Fig. 1, the structure of the questionnaire given to the citizens of the City of Genoa is shown. The questionnaire was made up of the following questions:

1. three words about what you think is important in a loving relationship;
2. three adjectives to define "masculine";
3. three adjectives to define "female";
4. three things that hurt in a loving relationship;
5. free thoughts.

Municipio.....
Iniziativa .....
Data .....

RelAZIONI IN CORSO

Uomini e Donne - rel AZIONI senza violenza
che GENERE di cuore vuoi?

Riflettiamo insieme sulla conoscenza dell'altro/a e sulle relazioni affettive per contrastare la violenza di genere

Il Comune sta svolgendo una campagna di sensibilizzazione sul tema. Le chiediamo di dedicare alcuni minuti del suo tempo per compilare questa breve scheda, assolutamente anonima. I risultati verranno restituiti alla città l'8 marzo 2014.

1) tre parole su cosa ritieni importante in una relazione affettiva di coppia
2) tre aggettivi per definire "maschile"
3) tre aggettivi per definire "femminile"
4) tre cose che fanno male in una relazione affettiva di coppia

spazio per un pensiero libero

paese di provenienza .....

genere ☐ M ☐ F ☐.....

età: ☐18/29 ☐30/39 ☐40/49 ☐50/69 ☐70/99

Figure 1 Questionnaire submitted to the citizens

The short questionnaire has been proposed to the following age groups: 18/29 - 30/39 - 40/49 - 50/59 - 70/99. It was compiled in manuscript form by 334 people and in web form by 203 people. The questionnaire has been completed by 459 women, 75 men and 3 transgender for a total of five-hundred thirty-seven people.

Table I Groups of respondents divided by age

18/29	30/39	40/49	50/59	70/99
220	66	68	155	17
***11 people cannot be classified because they didn't provide their age**				

There is a big difference in numbers between women who completed the questionnaire, compared to men.

This big difference, of course, leads us to reflect and to ask: what does this mean?

Perhaps it is true that women are more numerous in Italy, but it is necessary to point out that women are more sensitive and willing to deal with these problems than men are.

This work, based on the answers given to the questionnaire, tries to understand if stereotypes were present in the answers of the interviewees. In particular, the main aim is to identify what are the "stereotypical" images to define the roles of male and female. The answers have been analysed from a linguistic point of view, in order to verify if the language reveals the perception of reality through images, concepts and beliefs that exist in our minds, and if the language is revealing of stereotypes.

B. What is a stereotype?

A stereotype is a set of beliefs, simplistic representations, and oversimplified views of reality rigidly connected to each other, that a social group associated with another group. Stereotypes are born by the behaviour of a group of people or a gender (masculine or feminine). Therefore, gender stereotypes are a subclass of stereotypes.

The examples seem trivial, but it is not so, because stereotypes not only affect the ideas of groups of individuals, but also have implications in our actions and in our society.

For example, there are stereotypes associated with the rigid division of roles in the family and in the social and professional sphere.

A key feature of the language is to convey information in quantity and variety much bigger than we imagine. It is, indeed, through language that our vision of reality and society is transmitted to others; the spoken or written language is therefore an important vehicle of common sense [9].

C. First question

The first question was "three words about what you think is important in a loving relationship". The most frequent answered words for group 18-29 (the largest group) have been:

- "Fiducia" ["trust"]: 127 occurrences.
- "Rispetto" ["respect"]: 97 occurrences.
- "Amore" ["love"]: 80 occurrences.
- "Sincerità" ["sincerity"]: 68 occurrences.

Given these most frequent words, triplets referring to the same age group (i.e. 18-29) containing at least one of them were extracted in order to see if and how people in the same age group had used the same set or subset of words.

The same procedure was conducted for all age groups in order to identify the triplets of most frequent words. Then, the most frequent words of each age group were compared to see if:

- There was a same set of words used in different age groups.
- Words were the same or were similar.

Fig. 2 shows the set of words that contain at least one of the four most frequent words of Table II related to the age group 18 -29; e.g., "fiducia", "amore", "rispetto", "sincerità" ["trust", "love", "respect", "sincerity"].

ISBN: 978-1-61804-318-4

Figure 2: triplets for first question, age group 18-29

The greater or lesser font size of characters indicates the frequency degree of use of the triplet. Therefore, Fig. 2 shows the more or less common use of the same words used in response to the first question by people of the same age.
Note that the most common triplet is “amore, fiducia, rispetto” ["love, trust, respect"] and then, in descending order of frequency, “fiducia, rispetto, sincerità” ["trust, respect, sincerity"] and “amore, fiducia, sincerità” ["love, trust, sincerity"] and so on.
Furthermore, the most frequent word is “fiducia” ["trust"] that is linked with the words: “amore e rispetto” ["love and respect"], “rispetto e sincerità” ["respect and sincerity"], “amore e onestà” ["love and honesty"], “complicità e rispetto” ["complicity and respect"], “comprensione e rispetto” ["comprehension and respect"], “amore e complicità” ["love and complicity"], “amore e fedeltà” ["love and faithfulness"], etc.

Figure 3: top rated single words for first question for age group 18-29 (lighter) and 50-69 (darker), male respondents

Fig. 3 and 4, which refer to the first question of the questionnaire, show the words used in the two age groups 18-29 and 50-69 from the only male gender (M) and female gender and male (FM). In both figures, the words written with a darker font refer to the age group 50-69.
In Fig. 3, the words quoted by the male gender in the age group 18-29 are: “accoglienza, condivisione, correttezza, fare l’amore, impegno, ironia, libertà, omissione, onestà, quiete, soldi, stima, trasparenza” ["acceptance, sharing, fairness, making love, commitment, humor, freedom, omission, honesty, peace, money, respect, transparency"].

Figure 4: single words in common for first question between age group 18-29 (lighter) and 50-69 (darker), male and female respondents

Instead, the words quoted by the male gender for the age group 50-59 are: “bella presenza, coabitazione, compartecipazione, cortesia, dolcezza, franchezza, gentilezza, intesa, la buona cucina, passione, stare vicini, voglia” ["beautiful presence, cohabitation, sharing, kindness, gentleness, honesty, kindness, understanding, good food, passion, stay close, desire"]. As it can be seen, in different age groups, the male gender has used different words that can show the evolution of thinking through aging in men.
Fig. 4 contains the top rated words related to both genders (FM) for the same age groups. The words used are the following for the age group 18-29: “affetto, amore, attrazione, complicità, comprensione, condivisione, dialogo, fedeltà, felicità, fiducia, onestà, reciprocità, rispetto, sincerità, stima” ["affection, love, attraction, complicity, comprehension, sharing, dialogue, faithfulness, happiness, trust, honesty, reciprocity, respect, sincerity, esteem"]. The words of the age group 50-59 are: “affetto, amicizia, amore, ascolto, attrazione, complicità, comprensione, condivisione, dialogo, disponibilità, empatia, fiducia, intesa, onestà, rispetto, sesso, sincerità, sintonia, stima, tolleranza” ["Affection, friendship, love, listening, attraction, complicity, comprehension, sharing, dialogue, willingness, empathy, trust, understanding, honesty, respect, sex, sincerity, harmony, esteem, tolerance"].
In this case, there are some words in common to the two age groups. The words “felicità, libertà, reciprocità” [“happiness, freedom, reciprocity”] are present only in the age group 18-29 and the words “ascolto, disponibilità, empatia, intesa, pazienza, sintonia, stima, tolleranza” ["listening, willingness, empathy, understanding, patience, harmony, respect, tolerance"] are present only in the age group 50-69. The diversity of words is explanatory of what is considered important in a loving relationship from two different generations both females both males.

Table II - First question: highest frequency words per age groups

18-29	30-39	40-49	50-69	70-99
Trust	Respect	Respect	Respect	Respect
Respect	Faithfulness	Love	Love	Sincerity
Love	Love	Trust	Trust	Comprehension
Sincerity	Complicity	Sharing	Complicity	Love
Faithfulness	Sincerity	Complicity	Sharing	Feelings

ISBN: 978-1-61804-318-4

If we analyse the most answered words to the first question, it can be observed that there is a certain homogeneity between the age groups (see Table II). The most frequent words, in each age group, are the words “rispetto” and “amore” ["respect" and "love"]. The word “fiducia” ["trust"] appears in almost all the age groups, except in the group of 70-99 years old that, however, being a very small sample, it cannot be considered representative for the study.

D. Second question

In the second question, the word on which the attention is focused is an adjective (male); the adjectives are used to give meaning and to classify the words; in general, they are a sample of our thinking [9] [10].

Adjectives used by interviewees were 1475; the ones given by those people who have not declared either gender or age were excluded. The largest sample was of 601 adjectives related to the age group 18 - 29. Some of the answers, unfortunately, do not belong to the grammatical category requested (i.e. the adjective), anyway it has been chosen to analyse all the answers even if they were not adjectives.

Figure 5: triplets for second question, age group 18-29

The adjective with the highest number of occurrences is "forte" [“strong”]. The second word was "protettivo" [“protective”] and the third "possessivo" [“possessive”] at the same frequency with "egoista" [“selfish”]. The analysis of the age group 30-39 has revealed 187 adjectives; the largest number of occurrences was reserved for the word “protettivo” ["protective"] followed by “forte” ["strong"]. The age group 40-49 used 22 adjectives. In this age group, the female gender expressed a preference for the adjective “forte” ["strong"] followed by “protettivo” ["protective"] and “egoista” ["selfish"]. The age group 50-69 expressed a preference for the word “forte” ["strong"], followed by “intelligente” ["intelligent"] and “protettivo” ["protective"]. The lexeme with the largest number of occurrences attributed to the concept of masculinity is the adjective “forte” ["strong"]. It is used especially for the female gender in almost all age classes, showing how the stereotype of the "stronger sex" is rooted in our society.

The use of gender stereotypes lead to a rigid and distorted perception of reality, which is based on what we mean by "feminine" and "masculine" and what we expect from women and men. Through this way of thinking, a priori expectations on the roles, which men and women should take in society, are established only because of being biologically male or female. For example, a woman is considered quieter, less aggressive, good listener; a woman loves to take care of others, while man has a strong personality, great logical skills, spirit of adventure and leadership.

Fig. 6 shows the words used in the two age groups 18-29 and 50-69 by the female gender and male. The words written with the darker character refer to the age range 50-69.

The words used by female and male gender in the age group 18-29 are: “aggressivo, arrogante, coraggioso, concreto, dolce, egoista, forte, forza, fragile, galante, geloso, indeciso, infedele, intelligente, introverso, istintivo, megalomane, menefreghista, passionale, possessivo, protettivo, protezione, rozzo, sicuro, stabile, superficiale, virile” ["aggressive, arrogant, brave, pragmatic, sweet, selfish, strong, strength, fragile, gallant, jealous, hesitant, unfaithful, smart, introverted, instinctive, megalomaniac, uncaring, passionate, possessive, protective, protection, rude, self-confident, superficial, manly"].

Figure 6: Words for age groups 18-29/50-69 (both genders) which define a male

The words used by female and male gender in the age group 50-69 are: “affettuoso, autoritario, attento, colto, complementare, comprensione, comprensivo, coraggio, coraggioso, determinato, forza, forte, generoso, gentile, intelligente, lavoratore, muscoloso, non femminile, protettivo, sicurezza, sicuro, solido, trasparente, virile” ["loving, authoritarian, careful, cultured, complementary, comprehension, understanding, bravery, brave, determined, strenght, strong, generous, kind, smart, hardworking, muscular, not feminine, protective, safety, self-confident, tough, transparent, manly"].

Some words are common to the two age groups and genders: in particular "courageous, strenght, strong, smart, protective, protection, self-confident, manly". These words are circled in blue in the Fig. 6 (dashed line for group 18-29, continuous for 50-69).

E. Third question

As for the third question (i.e. three adjectives to define "female"), we can observe that the lexemes more used were those shown in Table III represented in the different age groups.

Since the female gender is the most represented in the questionnaire, it is our interest to focus our attention on the self-stereotype of women.

ISBN: 978-1-61804-318-4

Table III - Third question: highest frequency words per age groups

18-29	30-39	40-49	50-69
Problematic	Comfortable	Comfortable	Strong
Comprehension	Sensitive	Strong	Comfortable
Efficient	Sweet, Maternal	Sensitive	Sweet/sweetness Maternal
Jealous	Patient	Maternal	Sensitive
Comprehensive Elegant	Creative Unselfish	Dedicated/loving Sweet Empathic	Willing

Fig. 7 shows, for the most representative groups, how the self-stereotype of women is to be "comprensiva, di occuparsi degli altri, sensibile, dolce e materna" ["comprehensive, to take care of others, sensitive, gentle and maternal"].

Figure 7: Words for age groups 18-29 50-69 (both genders) which define a female

Considering the set of the two questions related to the concepts of "male" and "female", a benevolent sexism in its peculiarity is revealed. In fact, in the first question, men are in charge of protecting women and are responsible for their protection and, in the second one, the image of the female in her private and emotional dimension is given: a social context, where women must provide and ensure the welfare of man.

F. Fourth question

The fourth question asked to identify three things that hurt in a loving relationship. As for the previous questions the most representative groups were examined.

In the age group 18-29, the answers related to the fourth question consist of 197 sets of words used by the female gender and 19 sets of words used by the male gender.

As shown in Fig. 8, the sets of three words most frequently used to characterize the three things that hurt in a loving relationship are "falsità, tradimento, violenza" ["deceitfulness, unfaithfulness, violence"] followed by "bugie, tradimento, violenza" ["lies, unfaithfulness, violence"], etc.

Figure 9 shows the words used by age group 50-69. It's interesting to note that there were no particular variations on what is considered hurting by different age groups in a loving relationship.

In Fig. 9, indeed, we find the same sets of words: "falsità, tradimento, violenza" ["deceitfulness, unfaithfulness, violence"] followed by "bugie, tradimento, violenza" ["lies, unfaithfulness, violence"], etc.

Figure 8: triplets for fourth question, age group 18 -29

Figure 9: triplets for fourth question, age group 50 -69

In particular, in Table IV, it can be seen that the first word of each line is the one with the highest frequency within each age group. Then we find the words in descending order according to their number of occurrences; the last line of Table IV, characterized by a line more evident, contains the words close ,as number of occurrences, to the last ones for each age class.

Table IV - Fourth question: highest frequency words per age groups

18-29	30-39	40-49	50-69	70-99
Unf.*	Jealousy	Selfishness	Violence	Jealousy
Jealousy	Selfishness	Apathy	Apathy	Lack of respect
Violence	Incomprehension	Silence	Jealousy	Unf.*
Lies	Violence, distrust, lack of respect	Violence Jealousy	Selfishness	Apathy
	Unf.*	Lies, unf.*	Lies, unf.*	Violence, lies

Where Unf.* stays for Unfaithfulness

The words "gelosia" ["jealousy"] and "violenza" ["violence"] are present in all age groups. "Infedeltà" ["infidelity"] is synonymous with "tradimento" ["unfaithfulness"], for this reason we can add to the words "gelosia" ["jealousy"] and "violenza" ["violence"] the word "unfaithfulness" that is constant in the different age groups.

ISBN: 978-1-61804-318-4

Therefore, data seem to indicate that “gelosia” ["jealousy"], “violenza” ["violence"] and “tradimento” ["unfaithfulness"] are among the things considered the most harmful in a loving relationship.

Figure 10: top rated single words for fourth question for age group 18-29 (lighter) and 50-69 (darker), male respondents

The words of Fig. 10, referring to the fourth question, were written only by male gender and illustrate the thought of men in two age groups 18-29 and 50-69. Inside the word cloud it can be seen the following concepts/words: “troppa sincerità” [“too much sincerity”], “troppa gelosia” [“too much jealousy”], “mancanza di maturità” [“lack of maturity”], “mancanza di condivisione” [“lack of sharing”], “insoddisfazione” [“dissatisfaction”], “distacco” [“apathy”], “falsità” [“deceitfulness”], “avidità” [“greed”], “mancanza di figli” [“childlessness”], “invidia” [“envy”], “inganno” [“deceit”], “la suocera” [“the mother-in-law”] etc.

G. Fifth question:”free thought”

Figure 11: top rated bag of words for fifth question

In Fig. 11, it can be seen the answer to the fifth and last question. In this question, the interviewee was given the possibility to write a "free thought": the word cloud contains some groups of words that were among the most frequent.
Not all people, who have completed the questionnaire, responded to the fifth question. The phrases were analyzed and word sequences were extracted. The most frequent words were: “relazione” ["relationship"], “prevaricazione” ["abuse"], “violenza” ["violence"], “amore ["love"], “rispetto” ["respect"], “rapporto” ["relationship"], “sensazione” ["feeling"]. Each of these words is linked to other forming sequences that give rise to fundamental concepts in the man-woman relationship.
Some examples referring to the above words follow:

- “rapporto” [“relationship”]
  - Experienced relationship
  - important emotional relationship
  - loving relationship
  - wrong relationship
  - good relationship
  - the value of the relationship.
- “prevaricazione” [“abuse”]: psychological abuse.
- “violenza” [“violence”]
  - violence
  - gender violence
  - violence perpetrated by women
  - family violence
  - child abuse
  - violence against women;
- “amore” [“love”]
  - mature love
  - true love
  - love with maturity
  - love of coexistence
  - love in affection
  - love of self
  - love relationship.
- “rispetto” [“respect”]
  - culture of compliance
  - together with respect
  - respecting the autonomy
  - mutual respect in love
  - respect for rhythms of life
  - respect for the opinions
  - respect for coexistence
  - respect for diversity.
- “rapporto” [“relationship”]
  - love relationship
  - stable relationship
  - equal relationship
  - rapport of relationship
  - things in a relationship.
- “sensazione” [“feeling”]
  - feeling of loneliness
  - feeling of insecurity.

III. CONCLUSION

The language is acquired from early childhood, unconsciously, through social interaction and the environment; it is a communication and information tool. It is unavoidable to the society in which one lives. Language contributes to the definition of the meanings, both individual and collective, of the thoughts and values that we associate with words.
In the "Global Gender Gap Report 2013" compiled by the World Economic Forum, Italy is in 71st place with regard to gender equality.
It is fundamental the use of a language respectful of gender equality to gain an effective overcoming of the woman-man discrimination. Careful reflection on linguistic stereotypes in the media, in textbooks and in daily communication in general, could be able to develop a positive and equal gender identity.
The analysis of language can be a source of reflection on simplifications of the world and on how we perceive it.

ISBN: 978-1-61804-318-4

This brief analysis on gender differences and relationships, referring to the words used in the questionnaire, supports the hypothesis that language is, surely, the mirror of our thinking, our prejudices and cultural stereotypes and evidence of this work shows that gender stereotypes are still deeply rooted in Italian society.

ACKNOWLEDGMENT

All the authors wish to thank "Comune di Genova" for having made available the data to analyse.

REFERENCES

[1] Rashida Manjoo. "Report of the Special Rapporteur on violence against women, its causes and consequences, Rashida Manjoo". United Nations General Assembly A/HRC/20/16, 2012.
[2] Kartika Tamara Liotard. "On eliminating gender stereotypes in the EU (2012/2116(INI))". Committee on Women's Rights and Gender Equality, 6 December 2012.
[3] "Violenza sulle donne. I giovani come la pensano?", Regional commission for equal opportunity, Veneto Region, April 2011.
[4] "Rappresentazioni di genere e violenza privata", report of "Azione e contrasto della violenza sulle donne" project, Province of Parma, January 2009.
[5] "A scuola di pari opportunità", report of "…così diversi, così uguali…" project, Istituto Scolastico Liceo Scientifico Statale "G.P. Vieusseux".
[6] "Educare alle relazioni di genere", Consorzio di solidarietà Con.Sol. Soc. Coop, Chieti.
[7] S. A. Basow, "Stereotypes and roles", Belmont, CA, US: Thomson Brooks/Cole Publishing Co., 1992.
[8] G. N. Powell, D. A. Butterfield, J. D. Parent, "Gender and Managerial Stereotypes: Have the Times Changed?", Journal of Management vol. 28 no. 2 177-193, April 2002.
[9] F. Sabatini, "More than a preface", in "Sexism in Italian Language" by A. Sabatini, Presidency of the Council of Ministers of the Italian Republic, Department of Information and Publishing, Rome, 1993.
[10] Gough, H. G., & Heilbrun, A. B., "The Adjective Check List manual". Consulting Psychologists Press Inc., Palo Alto, CA, 1983.
[11] I. Briggs Myers, M. H. McCaulley, N. Quenk, and A. Hammer. "MBTI Handbook: A Guide to the development and use of the Myers-Briggs Type Indicator" Consulting Psychologists Press, 3rd edition, Palo Alto, CA, 1998.

Performance of Macrodiversity System with Two SC Microdiversity Receivers in the Presence of Rician Fading

Dragana S. Krstic, Mihajlo C. Stefanovic, Danijela A. Aleksic, Ivica Marjanovic, and Goran Petkovic

***Abstract*—**In this work, macrodiversity system consisting of macrodiversity selection combining (SC) receiver and two microdiversity SC receivers in the presence of shadowing and multipath fading is considered. Communication channel is subjected to Gamma long term fading and Rician short term fading. Probability density function (PDF), cumulative distribution function (CDF) and average level crossing rate (LCR) of macrodiversity SC receiver output signal envelope are calculated. The obtained expressions for PDF, CDF and LCR converge for any values of Gamma long term fading parameters and Rician short term fading parameters. Also, the influence of Gamma shadowing parameters and Rician multipath fading parameters on PDF, CDF and LCR are analyzed and discussed.

***Keywords*—**Gamma shadowing; level crossing rate, macrodiversity and microdiversity, Rician fading.

I. Introduction

Long term fading and short term fading are present in communication channel simultaneously, resulting in system performance degradation [1]. Reflections and refractions cause multipath propagation resulting in signal envelope variation and large obstacles cause shadowing resulting in signal envelope average power variation. Macrodiversity system enables simultaneously reduction of long term fading effects and short term fading effects on system performance in wireless communication channels. Macrodiversity system has macrodiversity receiver and two or more microdiversity receivers. Macrodiversity receiver mitigates shadowing effects fading effects and microdiversity receivers mitigate multipath fading effects. There are more distributions that can be used model signal envelope variation and signal envelope average power variation in communication channels in the presence of short term fading and long term fading. Mathematical model for describing short term fading channel depend on existence the line-of-site components, the number of clusters in propagation environment, non-homogenous of environment and inequality of quadrature components powers [2].

Rician distribution can describe multipath fading channel in the presence of one strong dominant component and more scattering components in propagation environment with one cluster. This distribution has parameter k. Parameter k is Rician factor. Rician factor is defined as ratio of dominant component power to scattering component power. When Rician factor goes to infinity, Rician fading channel becomes no fades channel and when Rician factor goes to zero, Rician fading channel becomes Rayleigh fading channel. Rician fading model has application in cellular mobile radio channel and land mobile satellite environment.

Long term fading channel can be modeled by using long-normal distribution or Gamma distribution. When large scale fading is described with Gamma distribution, expressions for PDF and CDF of receiver output signal can be derived in closed form. In this case, performance analysis of wireless communication systems is simpler.

There are more works considering outage probability and bit error probability of wireless communication systems with macrodiversity reception in the presence of long term fading and short term fading. In [3], average level crossing rate (LCR) and average fade duration (AFD) of wireless communication system with macrodiversity selection combining (SC) receiver and two microdiversity maximal ratio combining (MRC) receivers operating over Gamma shadowed Nakagami-m multipath fading channel are evaluated.

The paper [4] also considers second-order statistics of wireless communication system with micro- and macrodiversity reception in correlated gamma shadowed Nakagami-m fading channels. Here, macrolevel is of selection combining (SC) type and consists of two base stations (dual diversity), while N-branch receiver employing maximal ratio combining (MRC) is implemented on microlevel. Rapidly converging infinite-series expressions for LCR and AFD are derived.

Macrodiversity system including macrodiversity SC

This work has been funded by the Ministry of Education, Science and Technological Development of Republic of Serbia under projects III-44006 and TR-33035.

D. S. Krstic is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Nis, Serbia (corresponding author, fax: +38118588399; e-mail: dragana.krstic@ elfak.ni.ac.rs).

M.C. Stefanovic, I. Marjanovic, G. Petkovicwas were with Faculty of Electronic Engineering, University of Niš, Serbia (e-mails: misa.profesor@gmail.com; ivicabeograd@gmail.com; goran.petkovic969@gmail.com).

D.A. Aleksic is with the College of Applied Technical Sciences, Aleksandra Medvedeva 14, 18000 Nis, Serbia (e-mail: danijela.aleksic@vtsnis.edu.rs).

receiver and two microdiversity SC receivers is considered in [5]. Received signal experiences long term Rayleigh fading and short term Gamma shadowing. Closed form expressions for level crossing rate of microdiversity SC receivers output signals envelopes are calculated. This expression is used for evaluation of level crossing rate of macrodiversity SC receiver output signal envelope.

Effect of microdiversity and macrodiversity on average bit error probability in Gamma-shadowed Rician fading channels are investigated in [6]. The second order statistics of macrodiversity system working over Gamma shadowing and Rician multipath fading channel are calculated in [7]. The probability density function, cumulative distribution function and moments of macrodiversity output signal are computed.

In [8], a wireless communication system with a wireless communication system with two *L*-branch MRC receivers at the micro level and a dual-branch SC receiver at the macro level in gamma-shadowed Rician fading channels is considered. Exact and rapidly converging infinite-series expressions for the average level crossing rate and average fade duration at the output of the system are provided.

In this paper, macrodiversity system with macrodiversity SC receiver and two microdiversity SC receivers in the presence of shadowing and multipath fading is considered. The received signal experiences Rician short term fading resulting in signal envelope variation and Gamma long term fading resulting in signal envelope average power variation. Closed form expressions for probability density function, cumulative distribution function and average level crossing rate of macrodiversity SC receiver output signal are evaluated.

Probability density function can be used for calculation the important performance measures of wireless system such as outage probability and bit error probability. The obtained results are analyzed to calculate the influence of shadowing parameters and multipath fading parameters on system performance. To the best authors' knowledge, the performance of macrodiversity system with two microdiversity SC receivers operating over correlated Gamma shadowed Rician multipath fading is not reported in the available technical literature.

II. Rician Random Variable Level Crossing Rate

Rician random variable follows distribution:

$$p_x(x)=\frac{2x}{\Omega}e^{-\frac{x^2}{\Omega}}\cdot I_0\left(\frac{2Ax}{\Omega}\right)=$$

$$=\frac{2x}{\Omega}e^{-\frac{x^2}{\Omega}}\cdot\sum_{i_1=0}^{\infty}I_0\left(\frac{Ax}{\Omega}\right)^{2i_1}\cdot\frac{1}{(i_1!)^2}=$$

$$=\frac{2}{\Omega}e^{-\frac{x^2}{\Omega}}\cdot\sum_{i_1=0}^{\infty}\frac{A^{2i_1}}{\Omega^{2i_1}}\cdot\frac{1}{(i_1!)^2}\cdot x^{2i_1+1} \qquad (1)$$

where Ω is average square value of *x*, A is dominant component and $I_0(x)$ is modified Bessel function of the first kind, zero order and argument *x*. The cumulative distribution function of Rician random variable is:

$$F_x(x)=\int_0^x p_x(t)dt=$$

$$=\frac{2}{\Omega}e^{-\frac{x^2}{\Omega}}\cdot\sum_{i_1=0}^{\infty}\frac{A^{2i_1}}{\Omega^{2i_1}}\cdot\frac{1}{(i_1!)^2}\cdot x^{2i_1+1}\int_0^x dt\ e^{-\frac{t^2}{\Omega}}t^{2i_1+1}=$$

$$=\frac{2}{\Omega}e^{-\frac{x^2}{\Omega}}\cdot\sum_{i_1=0}^{\infty}\frac{A^{2i_1}}{\Omega^{2i_1}}\cdot\frac{1}{(i_1!)^2}\cdot\frac{1}{2}\Omega^{i_1}\gamma\left(i_1,\frac{x^2}{\Omega}\right) \qquad (2)$$

The joint probability density function of Rician random variable and its first derivative is

$$p_{x\dot{x}}(x\dot{x})=p_x(x)\cdot p_{\dot{x}}(\dot{x})=$$

$$=\frac{2}{\Omega}e^{-\frac{x^2}{\Omega}}\cdot\sum_{i_1=0}^{\infty}\frac{A^{2i_1}}{\Omega^{2i_1}}\cdot\frac{1}{(i_1!)^2}\cdot x^{2i_1+1}\frac{1}{\sqrt{2\pi}\beta}e^{-\frac{\dot{x}^2}{2\beta^2}} \qquad (3)$$

where variance of $\dot{x}$ is:

$$\beta^2=\pi^2 f_m^2\Omega,$$

with f_m being maximal Dopler frequency.

Average level crossing rate of Rician random process can be calculated as average value of the first derivation of Rician random process:

$$N_x=\int_0^{\infty}d\dot{x}\ \dot{x}\ p_{x\dot{x}}(x\dot{x})=$$

$$=\frac{2}{\Omega}e^{-\frac{x^2}{\Omega}}\cdot\sum_{i_1=0}^{\infty}\frac{A^{2i_1}}{\Omega^{2i_1}}\cdot\frac{1}{(i_1!)^2}\cdot x^{2i_1+1}\frac{1}{\sqrt{2\pi}}\pi f_m\Omega^{1/2}=$$

$$=\frac{f_m\sqrt{2\pi}}{\Omega^{1/2}}e^{-\frac{x^2}{\Omega}}\cdot\sum_{i_1=0}^{\infty}\frac{A^{2i_1}}{\Omega^{2i_1}}\cdot\frac{1}{(i_1!)^2}\cdot x^{2i_1+1} \qquad (4)$$

Probability density function of dual SC receiver output signal operating over identical, independent Rician multipath fading channel is:

$$p_y(y)=2p_{y_1}(y)\cdot F_{y_2}(y)$$

where y_1 and y_2 are signal envelope at input of SC receiver and *y* is SC receiver output signal envelope, as it is shown in Fig.1. $p_{y_1}(y)$ is given by (1) and $F_{y_1}(y)$ is given by (2) .

Fig. 1 SC receiver model

Cumulative distribution function of SC receiver output signal is:

$$F_y(y) = F_{y_1}(y)F_{y_2}(y) = \left(F_{y_1}(y)\right)^2 . \quad (5)$$

Average level crossing rate of SC receiver output signal is:

$$N_y = 2F_{y_1}(y)\cdot N_{y_1} \quad (6)$$

where N_{y_1} is given by (4).

III. Performance of Macrodiversity System

Macrodiversity system with macrodiversity SC receiver and two microdiversity SC receivers operating over Gamma shadowed Rician multipath fading channel is considered. Model of macrodiversity system is shown in Fig. 2.

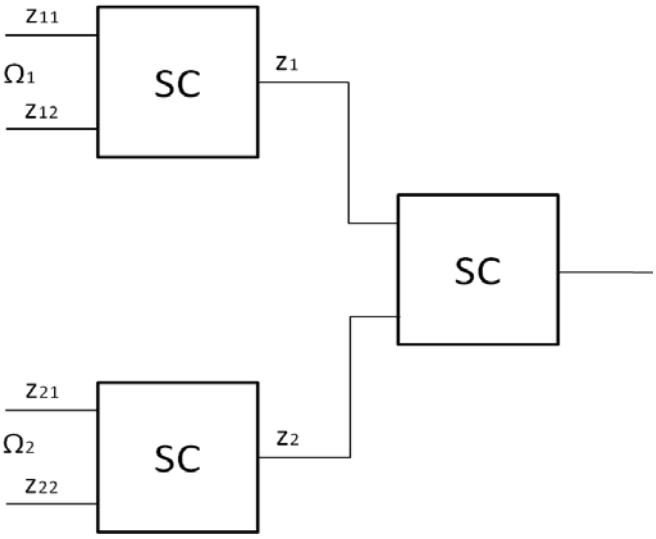


Fig. 2 System model

Signal envelopes at inputs of the first SC receiver are denoted with z_{11} and z_{12}, and z_{11} and at the second with z_{21} and z_{22}. Signal envelopes at outputs of microdiversity receivers are z_1 and z_2, and macrodiversity SC receiver output signal is z.

Signal envelope average power of inputs of microdiversity receivers, Ω_1 and Ω_2, follows joint Gamma distribution [9]:

$$p_{\Omega_1\Omega_2}(\Omega_1\Omega_2) = \frac{1}{\Gamma(c)\left(1-\rho^2\right)\rho^{\frac{c-1}{2}}\Omega_0^{c+1}}\cdot$$

$$\cdot e^{-\frac{\Omega_1+\Omega_2}{\Omega_0\left(1-\rho^2\right)}} I_{c-1}\left(\frac{2\rho}{\Omega_0\left(1-\rho^2\right)}\Omega_1^{1/2}\Omega_2^{1/2}\right) =$$

$$= \frac{1}{\Gamma(c)\left(1-\rho^2\right)\rho^{\frac{c-1}{2}}\Omega_0^{c+1}}\cdot\sum_{i_2=0}^{\infty}\left(\frac{\rho}{\Omega_0\left(1-\rho^2\right)}\right)^{2i_2+c}\frac{1}{i_2!\,\Gamma(i_2+c)}\cdot$$

$$\cdot\Omega_1^{2i_2+c-1}\Omega_2^{2i_2+c-1}\cdot e^{-\frac{\Omega_1+\Omega_2}{\Omega_0\left(1-\rho^2\right)}} \quad (7)$$

Macrodiversity SC receiver selects microdiversity SC receiver with the highest signal envelope average power at inputs to provide service to user. Therefore, probability density function of macrodiversity SC receiver output signal is:

$$p_z(z) = \int_0^{\infty} d\Omega_1 \int_0^{\Omega_1} d\Omega_2\, p_{z_1}(z/\Omega_1)p_{\Omega_1\Omega_2}(\Omega_1\Omega_2) +$$

$$= \int_0^{\infty} d\Omega_2 \int_0^{\Omega_2} d\Omega_1\, p_{z_2}(z/\Omega_2)p_{\Omega_1\Omega_2}(\Omega_1\Omega_2) =$$

$$= 2\int_0^{\infty} d\Omega_1 \int_0^{\Omega_1} d\Omega_2\, p_{z_1}(z/\Omega_1)p_{\Omega_1\Omega_2}(\Omega_1\Omega_2) =$$

$$= 4\sum_{i_1=0}^{\infty}\frac{A^{2i_1}}{(i_1!)^2}\cdot x^{2i_1+1}\cdot\sum_{i_2=0}^{\infty}\frac{A^{2i_2}}{(i_2!)^2}\cdot$$

$$\cdot\frac{1}{i_2}x^{2i_2}\sum_{j_1=0}^{\infty}\frac{1}{(i_2+1)(j_1)}x^{2j_1}\cdot$$

$$\cdot\frac{1}{\Gamma(c)\left(1-\rho^2\right)\rho^{\frac{c-1}{2}}\Omega_0^{c+1}}\cdot\sum_{i_3=0}^{\infty}\left(\frac{\rho}{\Omega_0\left(1-\rho^2\right)}\right)^{2i_3+c}\frac{1}{i_3!\,\Gamma(i_3+c)}\cdot$$

$$\cdot\frac{1}{i_3+c}\cdot\frac{1}{\left(\Omega_0\left(1-\rho^2\right)\right)^{i_3+c}}\sum_{j_2=0}^{\infty}\frac{1}{(i_3+c+1)(j_2)}\left(\frac{1}{\Omega_0\left(1-\rho^2\right)}\right)^{j_2}\cdot$$

$$\cdot\int_0^{\infty} d\Omega_1\cdot\Omega_1^{-1-1-2i_1-i_2-i_2-j_1+i_3+c+j_2+i_3+c-1}\cdot e^{-\frac{2x^2}{\Omega_1}-\frac{\Omega_1}{\Omega_0\left(1-\rho^2\right)}} =$$

$$= 4\sum_{i_1=0}^{\infty}\frac{A^{2i_1}}{(i_1!)^2}\cdot x^{2i_1+1}\cdot\sum_{i_2=0}^{\infty}\frac{A^{2i_2}}{(i_2!)^2}\cdot\frac{1}{i_2}x^{2i_2}\sum_{j_1=0}^{\infty}\frac{1}{(i_2+1)(j_1)}x^{2j_1}\cdot$$

$$\cdot\frac{1}{\Gamma(c)\left(1-\rho^2\right)\rho^{\frac{c-1}{2}}\Omega_0^{c+1}}\cdot\sum_{i_3=0}^{\infty}\left(\frac{\rho}{\Omega_0\left(1-\rho^2\right)}\right)^{2i_3+c}\frac{1}{i_3!\,\Gamma(i_3+c)}\cdot$$

$$\cdot\frac{1}{i_3+c}\cdot\frac{1}{\left(\Omega_0\left(1-\rho^2\right)\right)^{i_3+c}}\sum_{j_2=0}^{\infty}\frac{1}{(i_3+c+1)(j_2)}\left(\frac{1}{\Omega_0\left(1-\rho^2\right)}\right)^{j_2}\cdot$$

$$\cdot\left(2x^2\Omega_0\left(1-\rho^2\right)\right)^{-1-i_1-i_2-j_1/2+i_3+c+j_2/2}\cdot$$

$$\cdot K_{-2-2i_1-2i_2-j_1+2i_3+2c+j_2}\left(2\sqrt{\frac{2x^2}{\Omega_0\left(1-\rho^2\right)}}\right) \quad (8)$$

Cumulative distribution function of macrodiversity SC receiver output signal is:

$$F_z(z) = \int_0^{\infty} d\Omega_1 \int_0^{\Omega_1} F_{z_1}(z/\Omega_1)p_{\Omega_1\Omega_2}(\Omega_1\Omega_2)d\Omega_2 +$$

$$= \int_0^{\infty} d\Omega_2 \int_0^{\Omega_2} d\Omega_1\, F_{z_2}(z/\Omega_2)p_{\Omega_1\Omega_2}(\Omega_1\Omega_2) =$$

$$= 2\int_0^{\infty} d\Omega_1 \int_0^{\Omega_1} d\Omega_2 F_{z_1}\left(z/\Omega_1\right) p_{\Omega_1\Omega_2}\left(\Omega_1\Omega_2\right) =$$

$$= \sum_{i_1=0}^{\infty} \frac{A^{2i_1}}{\left(i_1!\right)^2} \cdot \frac{1}{i_1} x^{2i_1} \sum_{j_1=0}^{\infty} \frac{1}{\left(i_1+1\right)\left(j_1\right)} x^{2j_1} \cdot$$

$$\cdot \sum_{i_2=0}^{\infty} \frac{A^{2i_2}}{\left(i_2!\right)^2} \cdot \frac{1}{i_2} x^{2i_2} \sum_{j_2=0}^{\infty} \frac{1}{\left(i_2+1\right)\left(j_2\right)} x^{2j_2} \cdot$$

$$\cdot \frac{1}{\Gamma(c)\left(1-\rho^2\right)\rho^{\frac{c-1}{2}}\Omega_0^{c+1}} \cdot \sum_{i_3=0}^{\infty} \left(\frac{\rho}{\Omega_0\left(1-\rho^2\right)}\right)^{2i_3+c} \frac{1}{i_3!\,\Gamma\left(i_3+c\right)} \cdot$$

$$\cdot \frac{1}{i_3+c} \cdot \left(\frac{1}{\Omega_0\left(1-\rho^2\right)}\right)^{i_3+c} \sum_{j_3=0}^{\infty} \frac{1}{\left(i_3+c+1\right)\left(j_3\right)} \left(\frac{1}{\Omega_0\left(1-\rho^2\right)}\right)^{j_3} \cdot$$

$$\cdot \int_0^{\infty} d\Omega_1 \cdot \Omega_1^{-1-i_1-i_1-j_1-1-i_2-i_2-j_2+i_3+c+j_3+i_3+c-1} \cdot e^{-\frac{2x^2}{\Omega_1}-\frac{\Omega_1}{\Omega_0\left(1-\rho^2\right)}} =$$

$$= \sum_{i_1=0}^{\infty} \frac{A^{2i_1}}{\left(i_1!\right)^2} \cdot \frac{1}{i_1} x^{2i_1} \sum_{j_1=0}^{\infty} \frac{1}{\left(i_1+1\right)\left(j_1\right)} x^{2j_1} \cdot$$

$$\cdot \sum_{i_2=0}^{\infty} \frac{A^{2i_2}}{\left(i_2!\right)^2} \cdot \frac{1}{i_2} x^{2i_2} \sum_{j_2=0}^{\infty} \frac{1}{\left(i_2+1\right)\left(j_2\right)} x^{2j_2} \cdot$$

$$\cdot \frac{1}{\Gamma(c)\left(1-\rho^2\right)\rho^{\frac{c-1}{2}}\Omega_0^{c+1}} \cdot \sum_{i_3=0}^{\infty} \left(\frac{\rho}{\Omega_0\left(1-\rho^2\right)}\right)^{2i_3+c} \frac{1}{i_3!\,\Gamma\left(i_3+c\right)} \cdot$$

$$\cdot \frac{1}{i_3+c} \cdot \left(\frac{1}{\Omega_0\left(1-\rho^2\right)}\right)^{i_3+c} \sum_{j_3=0}^{\infty} \frac{1}{\left(i_3+c+1\right)\left(j_3\right)} \left(\frac{1}{\Omega_0\left(1-\rho^2\right)}\right)^{j_3} \cdot$$

$$\cdot \left(2x^2\Omega_0\left(1-\rho^2\right)\right)^{-1-i_1-j_1/2-i_2-j_2/2+i_3+c+j_3/2} \cdot$$

$$\cdot K_{-2-2i_1-j_1-2i_2-j_2+2i_3+2c+j_3}\left(2\sqrt{\frac{2x^2}{\Omega_0\left(1-\rho^2\right)}}\right). \qquad (9)$$

Level crossing rate of macrodiversity SC receiver output signal envelope is:

$$N_z = \int_0^{\infty} d\Omega_1 \int_0^{\Omega_1} d\Omega_2 \, N_{z_1/\Omega_1} \, p_{\Omega_1\Omega_2}\left(\Omega_1\Omega_2\right) +$$

$$= \int_0^{\infty} d\Omega_2 \int_0^{\Omega_2} d\Omega_1 \, N_{z_2/\Omega_2} \, p_{\Omega_1\Omega_2}\left(\Omega_1\Omega_2\right) =$$

$$= 2\int_0^{\infty} d\Omega_1 \int_0^{\Omega_1} d\Omega_2 \, N_{z_1/\Omega_1} \, p_{\Omega_1\Omega_2}\left(\Omega_1\Omega_2\right) =$$

$$= 2 \cdot 2 \cdot \sum_{i_1=0}^{\infty} \frac{A^{2i_1}}{\left(i_1!\right)^2} \cdot \frac{1}{i_1} 2x^{2i_1} \sum_{j_1=0}^{\infty} \frac{1}{\left(i_1+1\right)\left(j_1\right)} x^{2j_1} \cdot$$

$$\cdot f_m\sqrt{2\pi} \cdot \sum_{i_2=0}^{\infty} \frac{A^{2i_2}}{\left(i_2!\right)^2} \cdot \frac{1}{i_2} x^{2i_2+1} \cdot \frac{1}{\Gamma(c)\left(1-\rho^2\right)\rho^{\frac{c-1}{2}}\Omega_0^{c+1}} \cdot$$

$$\cdot \sum_{i_3=0}^{\infty} \left(\frac{\rho}{\Omega_0\left(1-\rho^2\right)}\right)^{2i_3+c} \frac{1}{i_3!\,\Gamma\left(i_3+c\right)} \cdot \left(\Omega_0\left(1-\rho^2\right)\right)^{i_3+c} \cdot$$

$$\cdot \frac{1}{i_3+c} \cdot \left(\frac{1}{\Omega_0\left(1-\rho^2\right)}\right)^{i_3+c} \sum_{j_2=0}^{\infty} \frac{1}{\left(i_3+c+1\right)\left(j_2\right)} \left(\frac{1}{\Omega_0\left(1-\rho^2\right)}\right)^{j_2} \cdot$$

$$\cdot \int_0^{\infty} d\Omega_1 \cdot \Omega_1^{-1-i_1-i_1-j_1-1/2-2i_2+i_1+c-1+i_1+c+j_2} \cdot e^{-\frac{2x^2}{\Omega_1}-\frac{\Omega_1}{\Omega_0\left(1-\rho^2\right)}} =$$

$$= 4 \cdot \sum_{i_1=0}^{\infty} \frac{A^{2i_1}}{\left(i_1!\right)^2} \cdot \frac{1}{i_1} x^{2i_1} \sum_{j_1=0}^{\infty} \frac{1}{\left(i_1+1\right)\left(j_1\right)} x^{2j_1} \cdot$$

$$\cdot f_m\sqrt{2\pi} \cdot \sum_{i_2=0}^{\infty} \frac{A^{2i_2}}{\left(i_2!\right)^2} \cdot \frac{1}{i_2} x^{2i_2+1} \cdot$$

$$\cdot \frac{1}{\Gamma(c)\left(1-\rho^2\right)\rho^{\frac{c-1}{2}}\Omega_0^{c+1}} \cdot$$

$$\cdot \sum_{i_3=0}^{\infty} \left(\frac{\rho}{\Omega_0\left(1-\rho^2\right)}\right)^{2i_3+c} \frac{1}{i_3!\,\Gamma\left(i_3+c\right)} \cdot \left(\Omega_0\left(1-\rho^2\right)\right)^{i_3+c} \cdot$$

$$\cdot \frac{1}{i_3+c} \cdot \left(\frac{1}{\Omega_0\left(1-\rho^2\right)}\right)^{i_3+c} \sum_{j_2=0}^{\infty} \frac{1}{\left(i_3+c+1\right)\left(j_2\right)} \left(\frac{1}{\Omega_0\left(1-\rho^2\right)}\right)^{j_2} \cdot$$

$$\cdot \left(2x^2\Omega_0\left(1-\rho^2\right)\right)^{-1/2-i_1-j_1/2-1/4-i_2+i_3+c+j_2/2} \cdot$$

$$\cdot K_{-1-2i_1-j_1-1/2-2i_2+2i_3+c+j_2}\left(2\sqrt{\frac{2x^2}{\Omega_0\left(1-\rho^2\right)}}\right) \qquad (10)$$

IV. Analysis of Numerical Results

In Fig. 3, normalized average level crossing rate of macrodiversity SC receiver output signal envelope is ploted versus SC receiver output signal envelope for several values of Rician factor, Gamma long term fading severity, parameter and correlation coefficient of Gamma shadowing.

When output signal envelope increases, the normalized average level crossing rate increases, achieves maximum, and after that decreases again.

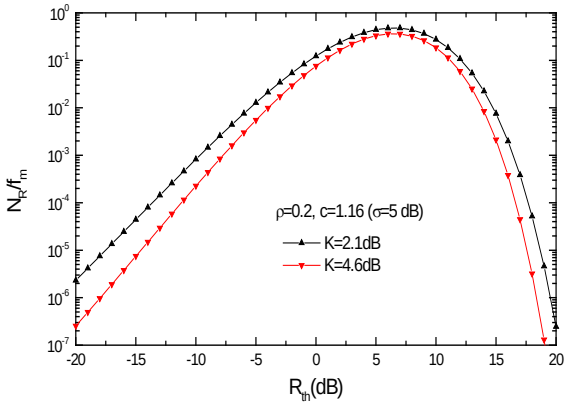


Fig. 3 Normalized average LCR versus normalized signal level for different values of Rician factor K

System performance is better for lower values of average level crossing rate. Average level crossing rate decreases and outage probability also decreases when Rician factor increases.

The Rician factor has lower values as dominant component power decreases or scattering components power increases. When Rician factor goes to infinity, Gamma shadowed Rician multipath fading channel becomes Gamma long term fading channel, and when Rician factor goes to zero, Rician multipath fading channel reduces to Gamma shadowed Rayleigh multipath fading channel. Average level crossing rate has higher values as Gamma shadowing severity parameter decreases.

The influence of Gamma parameter on average level crossing rate is higher for lower values of Rician factor. When Gamma parameter goes to infinity, Gamma shadowed Rician multipath fading channel becomes pure Rician multipath fading channel. Also, normalized level crossing rate of SC receiver output signal increases when correlation coefficient of Gamma shadow increases. When correlation coefficient goes to one, macrodiversity system becomes microdiversity system. Gamma long term fading is correlated due to both base stations are shadowed by the same obstacles. When correlation coefficient goes to one, the lowest sigbnal envelope occurs, simultaneously, on both base stations.

V. Conclusion

Macrodiversity system with macrodiversity selection combining receiver and two microdiversity SC receivers operating over shadowed multipath fading channel is considered. Received signal experiences Gamma long term fading and Rician short term fading. Macrodiversity SC receiver reduces Gamma long term effects and microdiversity SC receivers mitigate Rician short term fading effects on system performance. Microdiversity SC receiver selects the diversity branch with the highest signal envelope and macrodiversity SC receiver selects microdiversity receiver with the highest signal envelope average power at its inputs to provide service to user. Microdiversity SC receiver combine signals with multiple antennas at base stations and macrodiversity SC receiver combines signals envelope with two or more base stations geographically distributed in cell.

In this paper, probability density function and cumulative distribution function of SC receiver output signal envelope are calculated.Also, average level crossing rate at output of microdiversity SC receivers are evaluated and using these expressions, average level crossing rate of macrodiversity SC receiver output signal is calculated. These expressions rapidly converge for any values of Gamma shadowing severity parameter, correlation coefficient of shadowing and Rician factor. From obtained expressions for level crossing rate, level crossing rate of macrodiversity system in the presence of Rayleigh fading can be calculated as special case for Rician factor being zero.

In this paper, the influence of Gamma shadowing severity parameter, correlation coefficient of shadowing and Rician factor on average level crossing rate is studied. Average level crossing rate of macrodiversity SC receiver output signal envelope increases as Rician factor and Gamma shadowing severity parameter decrease. Shadowing correlation is resulting in diversity gain degradation. Average level crossing rate increases as correlation coefficient goes to one. The influence of correlation coefficient on average level crossing rate is highest when Sc receiver signal envelope has lower values.

Acknowledgment

This paper was done under projects III-44006 and TR-33035 of the Ministry of Education, Science and Technological Development of Republic of Serbia.

References

[1] P. M. Shankar, *Fading and Shadowing in Wireless Systems*, Dec 7, 2011 - Technology & Engineering, ISBN 978-1-4614-0366-1, e-ISBN 978-1-4614-0367-8, Springer, New York, Dordrecht, Heidelberg, London DOI 10.1007/978-1-4614-0367-8

[2] M. K. Simon, M. S. Alouini, *Digital Communication over Fading Channels*, New York: Wiley, 2005.

[3] D. Stefanovic, S. R. Panic, P. Spalevic "Second-order statistics of SC macrodiversity system operating over Gamma shadowed Nakagami-m fading channels," *AEU - International Journal of Electronics and Communications*, vo. 65, no. 5, pp. 413–418, 2011.

[4] A. D Cvetkovic, M. Č Stefanovic, N. M Sekulovic, D. N Milic, D. M Stefanovic, Z. J Popovic, "Second-order statistics of dual SC macrodiversity system over channels affected by Nakagami-m fading and correlated gamma shadowing", *Przeglad Elektrotechniczny*, 05/2011; 6: pp. 283-287.

[5] B. Jaksic, D. Stefanovic, M. Stefanovic, P. Spalevic , V. Milenkovic, "Level Crossing Rate of Macrodiversity System in the Presence of Multipath Fading and Shadowing", *Radioengineering*, Vol. 24, No. 1, April 2015, pp. 185-191, DOI: 10.13164/re.2015.0185

[6] V. Milenkovic, N. Sekulovic, M. Stefanovic, and M. Petrovic, "Effect of microdiversity and macrodiversity on average bit error probability in Gamma-shadowed Rician fading channels", *ETRI Journal*, Volume 32, Number 3, June 2010, 464-467.

[7] N. Sekulovic and M. Stefanovic, "Performance analysis of system with micro- and macrodiversity reception in corelated gamma shadowed Rician fading channels", *Wireless Personal Communications*, (publisher: Springer), vol. 65, no. 1, pp. 143-156, 2012, published online 12. Feb. 2011, doi: 10.1007/s11277-011-0232-8

[8] M. Bandjur, N. Sekulovic, M. Stefanovic, A. Golubovic, P. Spalevic, D. Milic, "Second-Order Statistics of System with Microdiversity and Macrodiversity Reception in Gamma-Shadowed Rician Fading Channels", *ETRI Journal*, Vol. 35, Number 4, August 2013, pp. 722-725.

[9] E. Xekalaki, J. Panaretos, and S. Psarakis, "A Predictive Model Evaluation and Selection Approach – The Correlated Gamma Ratio Distribution," *Stochastic Musings: Perspectives from the Pioneers of the Late 20th Century*, J. Panaretos, Ed., Mahwah, NJ: Lawrence Erlbaum Associates, 2003, pp. 188-202.

Dragana S. Krstić was born in Pirot, Serbia in 1966. She received the B.Sc, M.Sc and PhD. degrees in Electrical Engineering from Faculty of Electronic Engineering, University of Nis, Serbia in 1990, 1998. and 2006, respectively. She is with Faculty of Electronic Engineering, University of Nis, since 1990.

Her field of interest includes telecommunications theory, optical, wireless, mobile and satellite communication systems, etc. As author/co-author, she wrote about 200 scientific research papers, of which over 30 are printed in international journals, several in national journals, more than 100 are referred at international symposia and conferences, while over 30 are presented at national professional conferences.

Dr Krstic had several plenary and keynote lectures, panels and tutorials by invitation at international conferences. She is also the member of editorial boards of a few international journals and reviewer for many of them. Dr Krstic is a member of the technical program committees or reviewer for more than 50 conferences.

Mihajlo C. Stefanovic was born in Nis, Serbia in 1947. He received B.Sc., M.Sc. and Ph.D. degrees in Electrical Engineering from the Faculty of Electronic Engineering (Dept. of Telecommunications), University of Nis, Serbia, in 1971, 1976 and 1979, respectively.

His primary research interests are statistical communication theory, optical and satellite communications. He has written or co-authored a great number of journal publications. Dr. Stefanovic is a retired full-time professor with the Dept. of Telecommunications, Faculty of Electronic Engineering, University of Nis, Serbia.

Architecture of Asymmetric Quantum Cryptography Based on EPR

A.F.Metwaly[1], Nikos E. Mastorakis[2]

[1]Information Technology Department, Al-Zahra College for Women, Oman
[2]Technical University of Sofia, Bulgaria

Abstract

Multicast Classical transmission means that the channels and transmitted messages are both classical. This type of transmission deteriorate from many difficulties, the most important is network cryptography problems. For solving multicast classical network cryptography problems', the quantum approach has been investigated but Quantum approach requires additional resources to work in an effective way. In this paper, Generation and measuring shared entangled pair keys between the communicated peers in a multicast network is achieved by Quantum Multicast shared distribution and measurement centre " QM_{SDM} " and quantum gates. Encoding of transmitted quantum messages is handled by the basis of quantum teleportation. Teleportation or encoding at sender side will be accomplished by C_{NOT} and a $Hadamard$ gates. Decoding the teleported message is achieved by performing the correction action on received entangled pair. On the receiver side decoding will be accomplished by X and Z gates. If two members within the same multicast group need to communicate, they can by using entangled shared key pair. If two members in a different groups need to communicate, they can by complete or partial support of QM_{SDM}. By full support of QM_{SDM} the responsibility of QM_{SDM} is decoding /encoding the teleported / original transmitted quantum message between the communicated members. Optical clock synchronization is used for improving the transmission of generated entangled keys as well key update.

Keywords Quantum Key Distribution, Teleportation, Measurement, Secret Sharing

1 Introduction

The pioneering work of Bennett and Brassard [2] has been developed for the purpose of quantum cryptography. Quantum cryptography is one of the most significant prospects associated with laws of quantum mechanics in order to ensure unconditional security [3, 4, 5, 6, 10]. The quantum cryptography proves unconditional security characteristic through no cloning theory [1] as the transmitted quantum bit can't be replicated or copied but its state can be teleported. The most used quantum principles are quantum teleportation and dense coding. In quantum teleportation the quantum information can be transmitted between distant parties based on both classical communication and maximally shared quantum entanglement among the distant parties [1, 2, 3, 4]. In Dense coding the classical information can be encoded and transmitted between distant parties based on both one quantum bit and maximally shared quantum entanglement among the distant parties as each quantum bit can transmit two classical bits [1,2]. There are number of approaches and prototypes for the exploitation of quantum principles' to secure the communication between two parties and multi-parties. While these approaches used different techniques for achieving a private communication among authorized users but still most of them depend on generation of a secret random keys. At present, there're two approaches of quantum private communication. One is a hybrid of classical cryptosystem and quantum key distribution. In this approach, the employed encoding and decoding algorithms come from classical. Whilst the generated keys for message encoding and decoding which act as significant role in the cryptosystem derives from a distinguished quantum key distribution scheme. The other approach applies a completely quantum cryptosystem with natural quantum physics laws. In this approach, the encoding and decoding algorithms are quantum one and the keys for message encoding and decoding derives from a distinguished quantum key distribution scheme. The quantum communication system can be described using the same way of classical model. The messages in quantum system represented by quantum state which can be pure or mixed [3, 4, 5, 6, 7]. The most three principal components for designing a quantum communication system are cryptosystem, authentication and key management system. All included processes in these components may be classical or quantum but in any case at minimum one of these components has to apply a quantum features and laws [3, 4, 5, 6, 7]. Recently, quantum secure direct communication concept is introduced for transmitting the secured messages between the communicated participants without establishing secret keys to encode them [16, 17, 19, 20, 21, 22, 18, 23, 24, 25, 33, 34 , 35 , 36 ,37 , 38 , 39 , 40 , 41]. In [16] a ping pong protocol is introduced for directly decrypted the transmitted encoded bits between the communicated participants in every corresponding transmission without the need of QKD. In [40]

enhances the capability of ping pong protocol by adding two more unitary operations. In [19] a two-step quantum secure direct communication is proposed for transferring of quantum information by utilizing EPR pair blocks for secure the transmission. In [8] the authentication and communication process performed using GHZ states. Firstly, GHZ states are used for authentication purpose then the remaining GHZ will be used for directly transmitting the secret message. In [31] architecture of centralized multicast scheme is proposed based on hybrid model of quantum key distribution and classical symmetric encryption. The proposed scheme solved the key generation and management problem using a single entity called centralized Quantum Multicast Key Distribution Centre. In [32] a novel multiparty concurrent quantum secure direct communication based on GHZ states and dense coding is introduced. In [11] a managed quantum secure direct communication protocol based on quantum encoding and incompletely entangled states. Different quantum authentication approaches have been developed for preventing various types of attack and especially man in the middle attack [26, 27, 28, 29, 30]. In this paper, Generation and measuring shared entangled pair keys between the communicated peers in a multicast network is achieved by Quantum Multicast shared distribution and measurement centre " QM_{SDN} " and quantum gates. Encoding of transmitted quantum messages is handled by the basis of quantum teleportation. Teleportation or encoding at sender side will be accomplished by C_{NOT} and a $Hadamard$ gates. Decoding the teleported message is achieved by performing the correction action on received entangled pair. On the receiver side decoding will be accomplished by X and Z gates. If two members within the same multicast group need to communicate, they can by using entangled shared key pair. If two members in a different groups need to communicate, they can by complete or partial support of QM_{SDN}. By full support of QM_{SDN} the responsibility of QM_{SDN} is decoding /encoding the teleported / original transmitted quantum message between the communicated members.

2 Quantum State and Entanglement

The classical bit is the fundamental element of information. It is used to represent information by computers. Nevertheless of its physical realization, a classical bit has two possible states, 0 and 1. It is recognized that the quantum state is a fundamental concept in quantum mechanics. Actually, the quantum bit is the same as the quantum state. The quantum bit can be represented and measured using two states $|0>$ and $|1>$ which well known as Dirac notation [5, 7]. In classical computer all information is expressed in terms of classical bit. Classical bit can be either 0 or 1 at any time. On the other hand quantum computer uses quantum bit rather than a bit. It can be in a state of 0 or 1, also there is usage of a form of linear combinations of state called superposition state. Quantum bit can take the properties of 0 and 1 simultaneously at any one moment.

Quantum bit definition is described as follow: ***Definition:*** A quantum bit, or qubit for short, is a 2 dimensional Hilbert space H_2. An orthonormal basis of H_2 is specified by $\{|0>,|1>\}$. The state of the qubit is an associated unit length vector in H_2. If a state is equal to a basis vector then we say it is a pure state. If a state is any other linear combination of the basis vectors we say it is a mixed state, or that the state is a superposition of $|0>$ and $|1>$ [8, 9]. In general, the state of a quantum bit is described by Eq. (1) Where $|\psi>$ is a quantum state, α and β are complex numbers:

$$\Psi = \alpha\,|0> + \beta\,|1> \tag{1}$$

The quantum bit can be measured in the traditional basis equal to the probability of effect for α^2 in $|0>$ direction and the probability of effect for β^2 in $|1>$ direction [18, 20] which α and β must be constrained by Eq. (2) and Figure.1

$$\alpha^2 + \beta^2 = 1 \tag{2}$$

As well a quantum message can be represented as quantum state in a 3-dimension Hilbert space H_3 (see Eq. (3, 4))

$$|\Psi> = \alpha|0> + \beta|1> + \gamma|2> \tag{3}$$
$$|\alpha|^2 + |\beta|^2 + |\gamma|^2 = 1 \tag{4}$$

For a quantum system consists of multi-particle, the mixture system is equal to the tensor product of the physical elements of system state space. So, if we have two quantum states are denoted by Eq. (3, 4)

$$|\Psi_1> = \alpha_1|0> + \beta_1|1> \tag{5}$$

$$|\Psi_2> = \alpha_2|0> + \beta_2|1> \tag{6}$$

So the composite system can be written as

$$|\Psi> = |\Psi_1> \otimes |\Psi_2>$$
$$= \alpha_1\alpha_2|00> + \alpha_1\beta_1|01> + \beta_1\alpha_1|10> + \beta_1\beta_2|11> \tag{7}$$

If the decomposition of the multi-particle quantum system is unachievable, in this case the quantum system can be referred as entanglement state. The well-known two particles entanglement states are called Bell states. The Bell states are one of the main theories in quantum information processing which denote the entanglement concept [12, 13, 14, 18]. Bell states are certain extremely entangled quantum states of two particles denoted by EPR. As the two entangled particles will have interrelated physical characteristics even though they're disjointed by distance. Bell states are entitled in many applications but the most useful examples are quantum teleportation and dense coding [4, 5].

The four Bell states (EPR pairs) are defined by (Eq. (8))

$$|\Phi^{\pm}>=\frac{1}{\sqrt{2}}(|00>\pm|11>)$$

$$|\psi^{\pm}>=\frac{1}{\sqrt{2}}(|01>\pm|10>)$$

(8)

Bell states can be generated by utilizing the properties of both ***Hadamard*** gate and ***Controlled −NOT*** gate. The four possibilities of Bell states (EPR) according to the input bits. While the input bits are 00, 01, 10 and 11 then the generated EPR states given by (Eq. (9))

$$\begin{aligned}|\Phi^{+}>&=\frac{1}{\sqrt{2}}(|00>+|11>),\\ |\psi^{+}>&=\frac{1}{\sqrt{2}}(|01>+|10>),\\ |\Phi^{-}>&=\frac{1}{\sqrt{2}}(|00>-|11>),\\ |\psi^{-}>&=\frac{1}{\sqrt{2}}(|01>-|10>).\end{aligned}$$

(9)

As well the well-known three particles entanglement state is called ***GHZ*** given by (Eq. (10))

$$|\Psi>=\frac{1}{\sqrt{2}}(|000>\ \underset{-}{+}\ |111>) \quad (10)$$

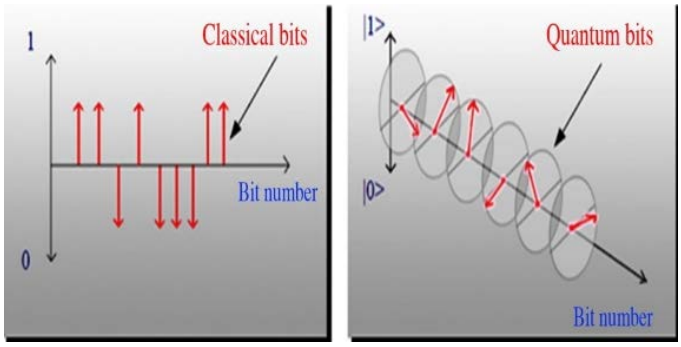


Figure 1Classical and Quantum Bits

3 Generate Shared Asymmetric Keys

This process consists of the steps required for generating and distributing shared Asymmetric keys between two members. The process begins with generating public and private keys as string of $|0>$ and $|1>$ through QM_{SDM}. Therefore, the $\mathbf{H}$ circuit selects a single public key quantum bit from the upper input and generates a single quantum bit output. The C_{NOT} circuit operates public key as control input to affect the private key which is target quantum bit. If the public key is $|0>$ then the private key output is as same as private key input. If the public key is $|1>$ then the private key output is the private key input flip-flopped as shown in Fig. 2 and given by (Eq. (11)).

$$\begin{aligned}|S_k>_{00}&=\frac{1}{\sqrt{2}}(|0>_{Q_{Kc}}\otimes|0>_{Q_{PK}}+|1>_{Q_{Kc}}\otimes|1>_{Q_{PK}})\\ |S_k>_{01}&=\frac{1}{\sqrt{2}}(|0>_{Q_{Kc}}\otimes|1>_{Q_{PK}}+|1>_{Q_{Kc}}\otimes|0>_{Q_{PK}})\\ |S_k>_{10}&=\frac{1}{\sqrt{2}}(|0>_{Q_{Kc}}\otimes|0>_{Q_{PK}}-|1>_{Q_{Kc}}\otimes|1>_{Q_{PK}})\end{aligned}$$

$$|S_k>_{11}=\frac{1}{\sqrt{2}}(|0>_{Q_{Kc}}\otimes|1>_{Q_{PK}}-|1>_{Q_{Kc}}\otimes|0>_{Q_{PK}}) \quad (11)$$

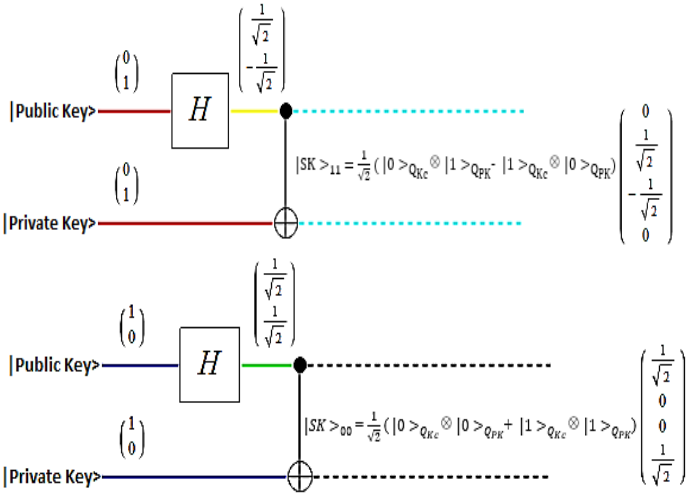


Figure 2 Generate Shared Asymmetric Keys

4 Measuring Generated Shared Asymmetric Keys

This process consists of the steps required by Member 2 for measuring received generated Asymmetric keys. Measuring is achieved by performing C_{NOT} gate and a $\mathbf{H}$ gate receptively. Result of measuring is a couple of classical bits , so Member 2 can detect which one of the four bell states is used to generate Asymmetric keys. The really essential phase for quantum teleportation and dense coding is Bell measurement. The outcome of Bell measurement is a couple of classical bits, which can be used for retrieve the original state. Bell measurement is used in Communication Process for determining which unitary operation is used to transform the original classical message so the receiver can retrieve it as shown in Fig. 3

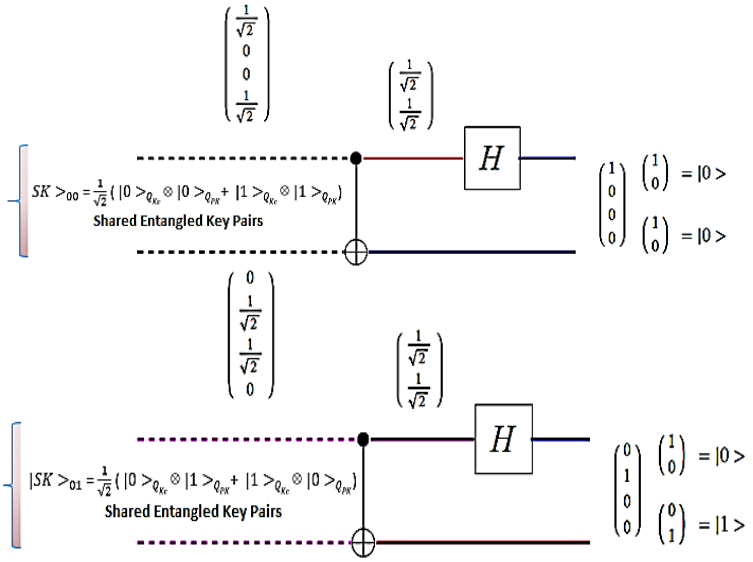


Figure 3 Measuring Generated Shared Asymmetric Keys

5 Encoding and Decoding of Transmitted Quantum Messages by Partial Support

Teleportation approaches are not restricted to two-communicator teleportation, but also generalized to several communicators quantum teleportation. One of the most used multi-communicators quantum teleportation approaches is controlled teleportation (CT). In this approach, the sender shares previous entanglement with the receiver and as a minimum one trusted center (TC). Subsequently, if the sender succeeds for teleporting the unknown quantum state to both the receiver and trusted center, afterward only one of them can create a copy of the transmitted unknown quantum state with the support of the other. As a consequence, the transmitted information is fragmented between the sender and the trusted center, so both will cooperated together for retrieving the transmitted unknown state by the sender. Meanwhile, the trusted center control the whole teleportation process the protocol is denoted as controlled teleportation (CT).

In Fig.3 shows an illustrative example of perfect teleportation as $|\psi^+>=\frac{1}{\sqrt{2}}(|00>+|11>)$ is used as a quantum channel of Asymmetric keys between sender and receiver. Currently, sender would like to transmit the unknown quantum state $|\psi>=\alpha|0>+\beta|1>$ to receiver. The unknown state will move through the teleportation circuit with a $Controlled-NOT$ gate and a $Hadamard$ gate. Sender can encode the status of one quantum message bit to member 2 on the basis of quantum teleportation. Receiver can decode the teleported message by performing the correction action on his entangled pair. In our designed circuit, Teleportation or encoding at sender side will be accomplished by C_{NOT} and $\mathbf{H}$ gates. On the receiver side decoding will be accomplished by X and Z gates. The steps required for encoding and decoding the original quantum message Ψ_μ have be illustrated below in (Eq. (12, 13)).

With the unknown state the initial state of the system is defined by (Eq. (12))

$$|\psi>_1=\alpha|0>+\beta|1>\otimes\frac{1}{\sqrt{2}}(|00>+|11>)$$
$$=\left(\alpha|0>\frac{1}{\sqrt{2}}(|00>+|11>)+\beta|1>\frac{1}{\sqrt{2}}(|00>+|11>)\right)$$

(12)

Subsequently the action of the $CNOT$ gate (using sender quantum bit as the control one and receiver quantum bit as the target one) the state becomes (Eq. (13))

$$|\psi>_2=\left(\alpha|0>\frac{1}{\sqrt{2}}(|00>+|11>)+\beta|1>\frac{1}{\sqrt{2}}(|10>+|01>)\right)$$

(13)

Since sender transmits the first quantum bit of the quantum state over the $Hadamard$ gate. So the state of overall system can be transformed as shown in (Eq. (14))

$$|\psi>_3=\begin{pmatrix}\alpha\frac{(|0>+|1>)}{\sqrt{2}}\frac{1}{\sqrt{2}}(|00>+|11>)\\+\beta\frac{(|0>-|1>)}{\sqrt{2}}\frac{1}{\sqrt{2}}(|10>+|01>)\end{pmatrix}$$
$$=$$
$$\frac{1}{2}\begin{pmatrix}|00>(\alpha|0>+\beta|1>)+|01>(\alpha|1>+\beta|0>)\\+|10>(\alpha|0>-\beta|1>)+|11>(\alpha|1>-\beta|0>)\end{pmatrix}$$

(14)

Afterward, sender computes the first two quantum bits and publish the result of his measurement through the classical channel. When receiver receives the two classical bits, he will conclude which unitary operation should be applied for restructuring the transmitted original unknown quantum state sent by sender as shown in Table.1

Table 1: Relationship between Sender Measurement and Receiver's Operation

Sender Measurement	Status of Receiver's Quantum Bit	Receiver's Operation	Status of Receiver Quantum Bit after Pauli Operation
00	$\alpha\|0>+\beta\|1>$	I	$\alpha\|0>+\beta\|1>$
01	$\alpha\|1>+\beta\|0>$	X	$\alpha\|1>+\beta\|0>$
10	$\alpha\|0>-\beta\|1>$	Z	$\alpha\|0>-\beta\|1>$
11	$\alpha\|1>-\beta\|0>$	$ZX=iY$	$\alpha\|1>-\beta\|0>$

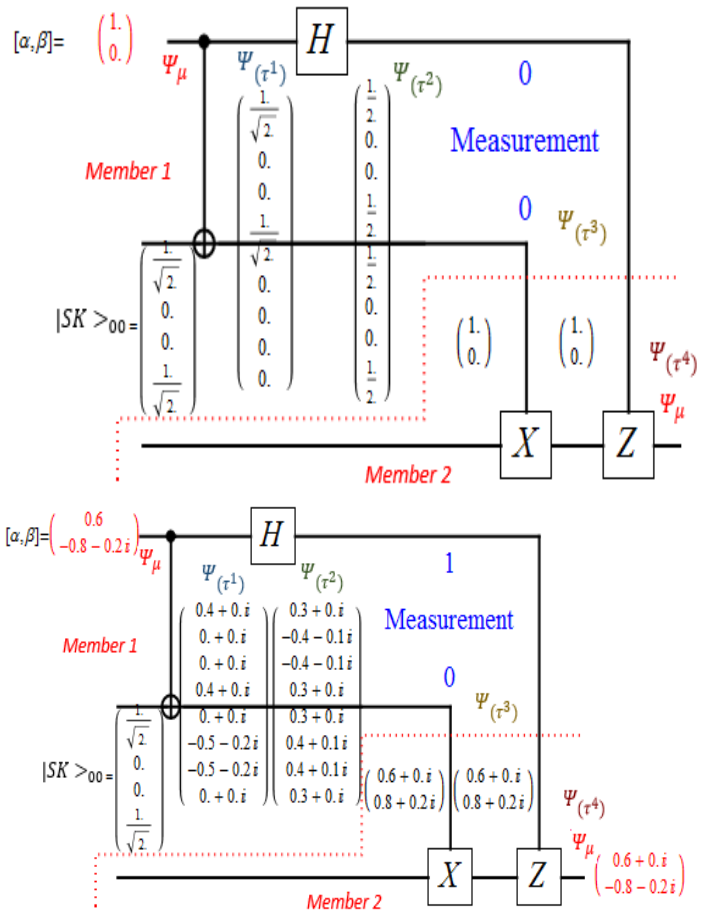


Figure 4 Encoding and Decoding of Transmitted Quantum Messages by Partial Support

6 Encoding and Decoding of Transmitted Quantum Messages by Full Support

This procedure describes the required steps for a secured communication of two members in a different groups by complete support of QM_{SDA} . The responsibility of QM_{SDA} is divided into two process. The first process is decoding of received teleported messages from member 1 which located in group 1, so now QM_{SDA} retrieves the original message. The second process, QM_{SDA} is encoding the original message and send it to member 2 which is located in group 2. Now, member 2 in group 2 retrieves the original message which was sent by member 1 in group 1 by performing the correct action. The required steps are shown in Figs. 5, 6 respectively.

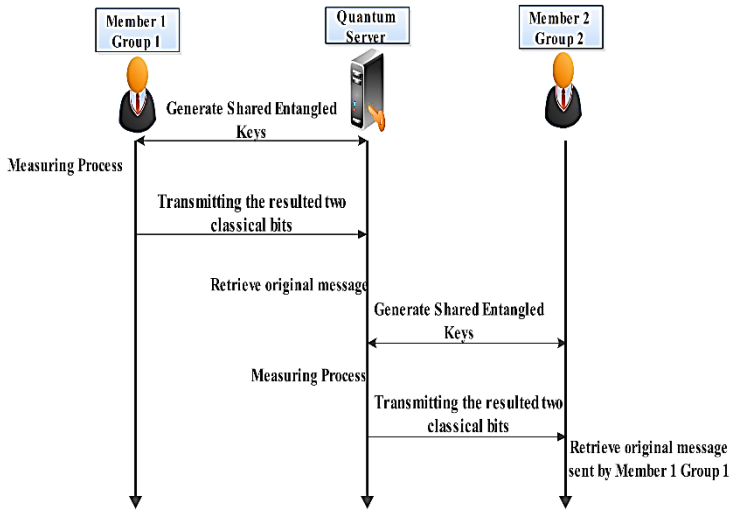


Figure 5 Encoding and Decoding of Transmitted Quantum Messages by Full Support

The protocol is then as follows:

Process 1

1) An Entangled shared key pair is generated, a separate quantum bit transmitted to member 1 and other to QM_{SDA}
2) At Member 1, measuring the Asymmetric key quantum bit and quantum message Ψ_μ by performing a C_{NOT} gate and thenceforth with a $Hadamard$ gate which resulting one of four possibilities, which can be encoded in two classical bits of information (00, 01, 10, and 11). Member 1 removes both quantum bits.
3) Member 1 transmits the resulted two classical bits to QM_{SDA} through a classical channel
4) Based on received classical bits, QM_{SDA} can perform the correct action on his entangled pair with X and Z operations. So, the result is a quantum bit identical to the message Ψ_μ which was chosen to be teleported.

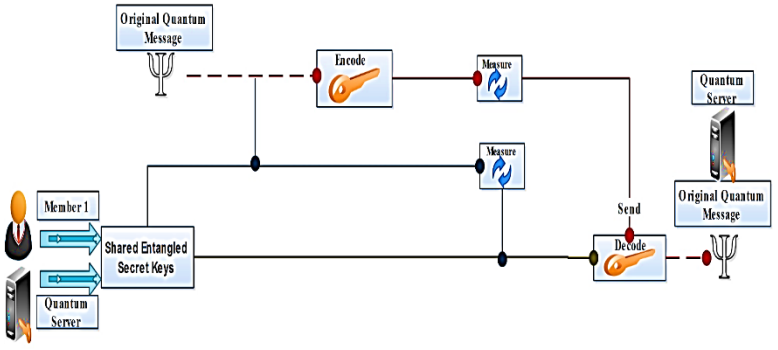


Figure 6 Full Support Process 1

ISBN: 978-1-61804-318-4

Process 2

1) The input is the retrieved quantum message Ψ_μ from process 1. An Entangled shared key pair is generated , a separate quantum bit transmitted to QM_{SDA} and other to member 2
2) At QM_{SDA}, measuring the Entangled shared key quantum bit and quantum message Ψ_μ by performing a $CNOT$gate and thenceforth with a $Hadamard$gate which resulting one of four possibilities, which can be encoded in two classical bits of information (00, 01, 10, and 11). QM_{SDA} removes both quantum bits.
3) QM_{SDA} transmits the resulted two classical bits to member 2 through a classical channel
4) Based on received classical bits, member 2 can perform the correct action on his entangled pair with X and Z operations. So, member 2 retrieves the original quantum message Ψ_μ which was chosen to be teleported by member 1

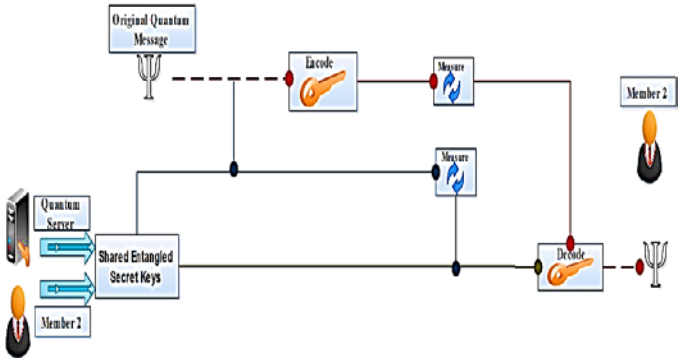


Figure 7 Full Support Process 2

Truth Table for measuring received generated Asymmetric keys. Measuring is achieved by performing CNOT gate and a $Hadamard$ gate receptively. Result of measuring is a couple of classical bits and probabilities for retrieving each state according to results are shown in Figs. 8.

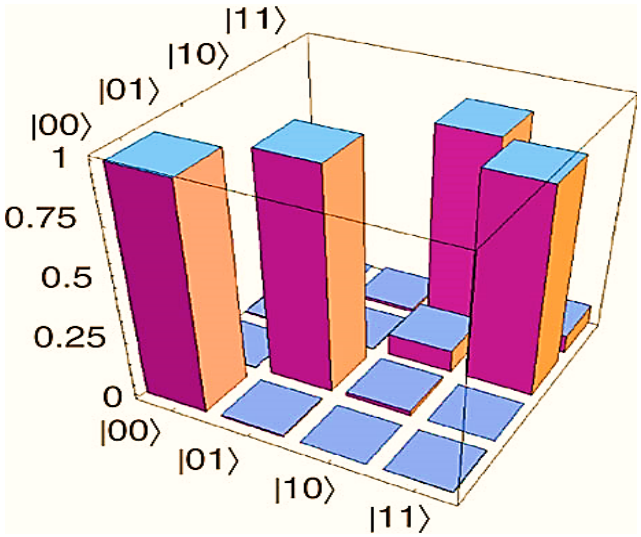


Figure 8 Truth Table for measuring received generated Asymmetric keys

The secured entangled shared key rate evaluated by applying Koashi's method and parameter approximating according to Rice and Harrington method. The formula for evaluation of secured entangled shared key rate between QM_{SDA} and communicated members is given by (Eq. (15))

$$S_R = \frac{[P_S(1 - H(P_{e1}) - P_{EFC}(P_e)H(P_e) + P_S(0)]}{t} \quad (15)$$

Where S_R is the approximate secured entangled shared key rate between QM_{SDA} and communicated members, P_S represents the estimation amount of sifted keys by a single photon form QM_{SDA} to communicated members, P_{e1} represents the estimation amount of errors which generated by a single photon , P_e represents the total number of sifted generated keys between QM_{SDA} and communicated members , P_{EFC} represents the probability of the error correction efficiency , $P_S(0)$ represents the estimation amount of sifted keys by a 0 photon pulses form QM_{SDA} to communicated members , $H(P_e)$ and $H(P_{e1})$ represent the binary entropy function and t represent the duration of established sessions between QM_{SDA} and communicated members.

The relation between generated entangled shared and purified keys rated in Kbits/s as function of the distance between communicated peers in km is illustrated in Fig.9. The relation between key rate and distance is conversely which implies as long distance enlarged the key rate is reduced.

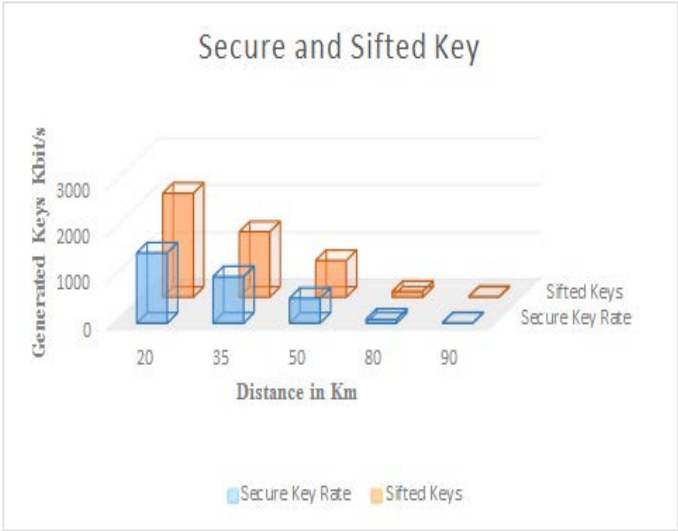


Figure 9 The relation between generated Asymmetric keys rated in Kbits/s as function of the distance between communicated peers in km

Conclusion

A proposed architecture for public quantum cryptography is investigated. The proposed architecture focus on Generation and measuring shared entangled pair keys between the communicated peers in a multicast network by Quantum

Multicast shared distribution and measurement centre " QM_{SDM} " and quantum gates. Encoding of transmitted quantum messages is handled by the basis of quantum teleportation. Teleportation or encoding at sender side will be accomplished by C_{NOT} and a $Hadamard$ gates. Decoding the teleported message is achieved by performing the correction action on received entangled pair. On the receiver side decoding will be accomplished by X and Z gates. If two members within the same multicast group need to communicate, they can by using entangled shared key pair. If two members in a different groups need to communicate, they can by complete or partial support of QM_{SDM}. By full support of QM_{SDM} the responsibility of QM_{SDM} is decoding /encoding the teleported / original transmitted quantum message between the communicated members. Optical clock synchronization is used for improving the transmission of generated entangled keys as well key update.

References

1. Wootters, W., & Zurek, W. (1982). A single quantum cannot be cloned. Nature, 299(5886), 802-803. doi:10.1038/299802a0
2. Bennett, C., & Brassard, G. (2014). Quantum cryptography: Public key distribution and coin tossing. Theoretical Computer Science, 560, 7-11. doi:10.1016/j.tcs.2014.05.025
3. Liang, H., Liu, J., Feng, S., & Chen, J. (2013). Quantum teleportation with partially entangled states via noisy channels. Quantum Inf Process, 12(8), 2671-2687. doi:10.1007/s11128-013-0555-3
4. Nielsen, M., & Chuang, I. (2000). Quantum computation and quantum information. Cambridge: Cambridge University Press.
5. Zeng, G.H. (2006) Quantum cryptology : Science Press
6. Van Assche, G. (2006). Quantum cryptography and secret-key distillation. Cambridge: Cambridge University Press.
7. Zeng, G. (2010). Quantum private communication. Beijing: Higher Education Press.
8. Barenco, A., Bennett, C., Cleve, R., DiVincenzo, D., Margolus, N., & Shor, P. et al. (1995). Elementary gates for quantum computation. Physical Review A, 52(5), 3457-3467. doi:10.1103/physreva.52.3457
9. Hirvensalo, M. (2001). Quantum computing. Berlin: Springer.
10. Sharbaf, M. S. (2009, April). Quantum cryptography: a new generation of information technology security system. In Information Technology: New Generations, 2009. ITNG'09. Sixth International Conference on (pp. 1644-1648). IEEE.
11. Jin, X. M., Ren, J. G., Yang, B., Yi, Z. H., Zhou, F., Xu, X. F., ... & Pan, J. W. (2010). Experimental free-space quantum teleportation. Nature Photonics, 4(6), 376-381.
12. Bell, J. S. (1964). On the einstein-podolsky-rosen paradox. Physics, 1(3), 195-200.
13. Aspect, A., Dalibard, J., & Roger, G. (1982). Experimental test of Bell's inequalities using time-varying analyzers. Physical review letters, 49(25), 1804.
14. Shimizu, K., & Imoto, N. (1999). Communication channels secured from eavesdropping via transmission of photonic Bell states. Physical Review A,60(1), 157.
15. Einstein, A., Podolsky, B., & Rosen, N. (1935). Can quantum-mechanical description of physical reality be considered complete?. Physical review, 47(10), 777.
16. Boström, K., & Felbinger, T. (2002). Deterministic secure direct communication using entanglement. Physical Review Letters, 89(18), 187902.
17. Deng, F. G., & Long, G. L. (2004). Secure direct communication with a quantum one-time pad. Physical Review A, 69(5), 052319.
18. Man, Z. X., Xia, Y. J., & An, N. B. (2006). Quantum secure direct communication by using GHZ states and entanglement swapping. Journal of Physics B: Atomic, Molecular and Optical Physics, 39(18), 3855.
19. Deng, F. G., Long, G. L., & Liu, X. S. (2003). Two-step quantum direct communication protocol using the Einstein-Podolsky-Rosen pair block.Physical Review A, 68(4), 042317.
20. Lucamarini, M., & Mancini, S. (2005). Secure deterministic communication without entanglement. Physical review letters, 94(14), 140501.
21. Yan, F. L., & Zhang, X. Q. (2004). A scheme for secure direct communication using EPR pairs and teleportation. The European Physical Journal B-Condensed Matter and Complex Systems, 41(1), 75-78.
22. Cai, Q. Y. (2004). The Ping-Pong protocol can be attacked without eavesdropping. arXiv preprint quant-ph/0402052.
23. Zhu, A. D., Xia, Y., Fan, Q. B., & Zhang, S. (2006). Secure direct communication based on secret transmitting order of particles. Physical Review A, 73(2), 022338.
24. Xue, P., Han, C., Yu, B., Lin, X. M., & Guo, G. C. (2004). Entanglement preparation and quantum communication with atoms in optical cavities. Physical Review A, 69(5), 052318.
25. Lee, H., Lim, J., & Yang, H. (2006). Quantum direct communication with authentication. Physical Review A, 73(4), 042305.
26. Curty, M., & Santos, D. J. (2001). Quantum authentication of classical messages. Physical Review A, 64(6), 062309.

27. Dušek, M., Haderka, O., Hendrych, M., & Myška, R. (1999). Quantum identification system. Physical Review A, 60(1), 149.
28. Zeng, G., & Zhang, W. (2000). Identity verification in quantum key distribution. Physical Review A, 61(2), 022303.
29. Ljunggren, D., Bourennane, M., & Karlsson, A. (2000). Authority-based user authentication in quantum key distribution. Physical Review A, 62(2), 022305.
30. Biham, E., Huttner, B., & Mor, T. (1996). Quantum cryptographic network based on quantum memories. Physical Review A, 54(4), 2651.
31. Metwaly, A. F., Rashad, M. Z., Omara, F. A., & Megahed, A. A. (2014). Architecture of multicast centralized key management scheme using quantum key distribution and classical symmetric encryption. The European Physical Journal Special Topics, 223(8), 1711-1728.
32. Ying, S., Qiao-Yan, W., & Fu-Chen, Z. (2008). Multiparty Quantum Chatting Scheme. Chinese Physics Letters, 25(3), 828.
33. Metwaly, A. F., & Mastorakis, N. E. (2015). Architecture of Decentralized Multicast Network Using Quantum Key Distribution and Hybrid WDM-TDM. Advances In Information Science And Computer Engineering, 504-518.
34. Metwaly, A., Rashad, M. Z., Omara, F. A., & Megahed, A. A. (2012, May). Architecture of point to multipoint QKD communication systems (QKDP2MP). In Informatics and Systems (INFOS), 2012 8th International Conference on (pp. NW-25). IEEE.
35. Farouk, A., Omara, F., Zakria, M., & Megahed, A. (2015). Secured IPsec Multicast Architecture Based on Quantum Key Distribution. In The International Conference on Electrical and Bio-medical Engineering, Clean Energy and Green Computing (pp. 38-47). The Society of Digital Information and Wireless Communication.
36. Ting, G., Feng-Li, Y., & Zhi-Xi, W. (2005). A simultaneous quantum secure direct communication scheme between the central party and other M parties.Chinese Physics Letters, 22(10), 2473.
37. Wang, C., Deng, F. G., & Long, G. L. (2005). Multi-step quantum secure direct communication using multi-particle Green–Horne–Zeilinger state. Optics communications, 253(1), 15-20.
38. Wang, J., Zhang, Q., & Tang, C. J. (2006). Quantum secure direct communication based on order rearrangement of single photons. Physics Letters A, 358(4), 256-258.
39. Qing-Yu, C., & Bai-Wen, L. (2004). Deterministic secure communication without using entanglement. Chinese Physics Letters, 21(4), 601.
40. Cai, Q. Y. (2006). Eavesdropping on the two-way quantum communication protocols with invisible photons. Physics Letters A, 351(1), 23-25.
41. Long, G. L., Deng, F. G., Wang, C., Li, X. H., Wen, K., & Wang, W. Y. (2007). Quantum secure direct communication and deterministic secure quantum communication. Frontiers of Physics in China, 2(3), 251-272.
42.

Cattle traceability: from the pasture to the slaughterhouse

Maia, A. P. M[1], Dias, E. M. [2]

***Abstract*—**Agribusiness is a productive chain in development with expressive participation in Brazil's economy. The objective of this work is to present the economic importance of the export of beef to the economy of Brazil, the progress in this sector, the growth projections for the coming years and the need to insert traceability technology into the livestock sector. Traceability technology will contribute to safer and more efficient growth, increase the chances of competitiveness in foreign markets and ensure the quality of the product in the domestic market.

***Keywords*—** agribusiness, cattle ranching, modeling, slaughter not supervised, traceability.

I. INTRODUCTION

AGRIBUSINESS contributes significantly to economic growth in Brazil, with an average of 25% of the country's total Gross Domestic Product (GDP). Of the 10 main products exported, seven are agribusiness [1].

Cattle raising is a growing sector, with great commercial importance. The market for this sector has proven more favorable every year. And for the sector to get new buyers and develop safely, we must face challenges with increasing production meat, estimated at 2% year by 2024 with the totaling 13,1milhões tons, ensure environmental sustainability and product quality (cattle) provided by the end user [1], [49].

The foreign market consumed on average 20% of the total meat production in Brazil and requires strict sanitary protocol, contributing to the surveillance. The balance of 80% of production is destined for domestic consumption, and the monitoring does not follow a pattern. Throughout the country, 75% of the slaughter of the Brazilian beef production is subject to Federal control for marketing of meat for export, 17% of slaughter is subject to State control for marketing of meat within the limits of each state and 8% of slaughter has Municipal control for marketing meat within the limits of the municipality. The lack of a unified sanitary protocol for supervising the slaughter decreases the credibility of the quality of meat offered for consumption internally [4].

The objective of this study is to present the importance of a pattern in the Brazilian supply chain of beef, in the transport process of the product (cattle) provided to producer (refrigerator), tracking all routes from the departure of the pasture until the cargo arrivals at its destination.

Included in the proposal is the automation of the logistics chain in transportation between the macro steps of supply and production in order to: (i) guarantee the origin of the product, promoting reliability; (ii) auxiliary in the process of the cargo's release and (iii) reduce the incidence of illegal refrigerators.

This paper is divided into four sections, including this introduction. The second section is an analysis of the economic importance of beef exports to Brazil. The third section presents studies about traceability and the slaughter studies not inspected in Brazil and the proposal to reduce this incidence. In the fourth section contains the conclusion and final considerations.

II. ECONOMIC IMPORTANCE OF THE EXPORT OF BEEF TO BRAZIL

Macroeconomic events in 2014 impacted the Brazilian economy in a negative way. withinflation close to the ceiling of the target set, a slowdown in industrial activity and trade balance at uncomfortable levels [23].

But among all these difficulties, agribusiness in Brazil in 2014 had reason to celebrate. Last year Brazil celebrates 100 years of beef exports. The first export took place in 1914, 200 tons of frozen meat was shipped to England, from the port of Santos through the Companhia Frigorifica e Pastoril [20].

Fig.1 - Refrigerator Decade XX

[1] Maia, A.P.M is with GAESI Gestão em Automação e TI, São Paulo, SP, 05508-900, Brazil (corresponding author) e-mail: andreapmm@hotmail.com

[2] Eduardo M. Dias is full professor of the Escola Politécnica of the Universidade de São Paulo and coordinator of GAESI - Grupo de Automação Elétrica em Sistemas Industriais, a reseach group of the Electrical Energy and Automation Department (emdias@pea.usp.br).

The export of Brazilian beef, in 2014 closed the year with $ 7.2 billion in revenue, according to data released by the Brazilian Association of Meat Exporters (Abiec - acronym in Brazil) on January 08, 2015. Compared to 2013, there was an increase of 7.7% in revenue; 3.3% in the volume exported totaling 1.56 million tons; 9% growth in exports to Hong Kong, totaling 400.5 tons and 3% for Russia, total of 314.6 tons, our main buyers account for some 50% of exports. As in Table1.

Table 1 -Results for the top 10 exporting countries [39].

Ranking	Country	Billing US$ (January to February 2014)	Volume in tons (January to February 2014)
1	Hong Kong	1.711.839.321,14	399.973,89
2	Rússia	1.314.093.693,40	314.672,41
3	European Union	928.514.318,35	127.442,31
4	Venezuela	900.806.593,06	169.599,51
5	Egypt	611.331.607,82	165.831,77
6	Chile	286.924.277,34	55.225,52
7	Iran	274.764.475,21	61.570,59
8	USA	231.357.572,53	22.214,31
9	Angola	118.374.094,78	37.442,68
10	Algéria	100.531.196,77	21.044,52

Agribusiness in Brazil contributes fundamentally to the economy and growth, accounting for 40% of exports, 37% of jobs created in the country and on average 25% of GDP (Gross Domestic Product).[23]

In 2014, Brazil grew by 0.1% of Gross Domestic Product (GDP) the sum of the wealth produced in Brazil totaled R$ 5.52 trillion, the agricultural sector contributed 0.4% in economic growth, totaling 262 billion. [31] – [39]

In the same period, agriculture accounts for the herd with approximately 210 million cattle head, which ranks Brazil as the largest commercial herd in the world, second largest beef producer, with production of 10,07 million tons, and the first in export with 2,09 million tons, maintaining its leadership since 2008. [39]

Cattle ranching continues to grow, even with the drop of 3.2% of agribusiness products in the total export, which closed the year 2014 at $ 96.5 billion [18], [26].

According to data published on December 08, 2014 by the Ministry of Agriculture (MAPA), agribusiness contributed $ 6.13 billion, for Brazilian economy, 40% of the total Brazilian exports in the period between December 2013 and November 2014.

In this period, the meat sector ranked first in terms of export value, with $ 1.43 billion. Of this total, US $ 555.98 million were to the beef [18].

Brazilian cattle raising is favorable for investments over the next nine years. According to studies conducted by the Ministry of Agriculture, Livestock and Supply (MAPA): "Agribusiness Projections: Brazil 2013/2014 to 2023/2024", released in September, 2014, the results are very favorable for the period of 2013/2024; the projection of total growth of meat production is 30.3% and outlook of the grow for beef is 22.8% of production, 15.6% for consumption and 39.7% for export [49].

The Brazilian agricultural industry is growing and investing in technologies, with a high degree of national content, facilitating the process of communication between all links in the beef production chain. Technologies aims to promote safety in the production process, the credibility of the quality and value of the product (cattle). These investments boost the Brazilian economy to enhance competition in the market with the conquest of new trade routes to export meat, opening new markets for sale of the product (cattle) and confidence for domestic consumption. In this context two projects developed with Brazilian technology stand out:

1. CTC 11002, the project developed by CEITEC (National Center for Advanced Electronic Technology), also known as “Chip do Boi”, uses integrated circuits that allow you to identify, track and authenticate individual cattle on pasture, recording and monitoring all product development (livestock) to be marketed. The device technology consists of a plastic earring used as a basis for encapsulating the chip with Radio Frequency

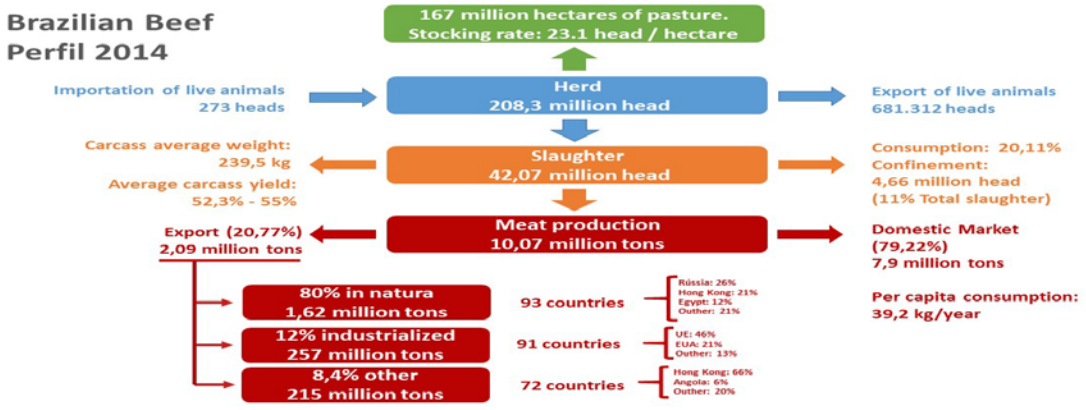


Fig.2-Brazilian Beef Perfil 2014 [39].

Identification - RFID 134KHz LF. The project objective, economically, is to reduce the cost of earring, used for cattle identification, between 30% and 35% and the production target of 70 million units per year. Contributing to traceability within the Brazilian beef supply chain. [28].

Fig.3 - Kit "Chip do Boi" [28], [29].

2. “Canal Azul”, a project developed in partnership with the Brazilian Association of Meat Exporters (Abiec); Polytechnic School, University of São Paulo (EPUSP); Institute for Software Technology (ITS); Financier of Studies and Projects (FINEP); National Center for Advanced Electronic Technology (CEITEC); and the Ministry of Agriculture, Livestock and Supply (MAPA).

   The technology used in the project contains, basically: (i) Electronic Seal: Composed of a TAG (chip + antenna) with radio frequency identification (RFID) UHF, with the objective of seals the container loaded ensures the health and integrity load; (Ii) Channel Electronic Blue (CA-e): Existing documentation only in electronic form that contains all the information required for all cargo shipment in Brazilian ports, the producer (refrigerator) records the information on the chip, the electronic seal, which follows with the load to the port and also distributes to the tax agency involved in the operation, streamlining the bureaucratic process in the analysis of the necessary export documentation. (GAESI, 2012)

   This project's economic potential is the reduction of the average time analysis to document the release of the shipment, reduce operating costs and contribute to increased competition for export.

   Process details in:(http://globotv.globo.com/tv-tem-interior-sp/nosso-campo-tv-tem/v/exportacao-de-carne/3163850/). [27],[33],[36],[38],[46]

III. Traceability for Slaughter Control Not Monitored in Brazil

Traceability is the possibility that the consumer knows the "past life" of products and can identify potential dangers to public health that the product was exposed to during its production and distribution [1].

The need to implement traceability in the beef supply chain began in 1996, when the BSE (Bovine Spongiform Encephalopathy) disease known as "mad cow disease" appeared. The need to identify and remove product from the market quickly and needs of the product (cattle) became apparent in case of risk [1].

For greater control of the product (cattle) marketed in each country, the European community by Resolution EC 1760/2000 of June 17/ 2000, provides for the implementation of a traceability program in countries that provide it with beef, through the individual product registration (cattle) for the lifting of all the animal information from birth to consumption of the final product. [1], [5], [30]

To meet this regulation, Brazil began its national traceability program in 2002 by the technical committee formed by the Ministry of Agriculture, Livestock and Supply (MAPA), National Confederation of Agriculture (CNA), Brazilian Association of Export Beef Industry (ABIEC) and the Brazilian Agricultural Research Corporation (EMBRAPA), through Normative Instruction 1 of January 09/2002 with the project of the Cattle Identification System and Certification and Buffalos (Sisbov) [1], [5].

In 2006, faced with the need to adapt the European requirements, made after a rigorous evaluation in the Brazilian traceability system, Normative Instruction 17 of July 13, 2006 was published,. the Traceability Service of Supply Chain Cattle and Buffaloes (Sisbov), revoking Normative Instruction 1 [1], [5].

The Cattle Identification System and Certification and Buffalos (Sisbov) has control and traceability of the production process in the context of rural properties. The voluntary participation of farmers is voluntary becoming mandatory only for producers who adhere to marketing in markets requiring traceability [1].

The chain of Brazilian beef supply consists of sets of interactive links, through systems: production system of raw materials, with subsystems of breeding, rearing and fattening; industrialization system with subsystems of slaughter, cutting, packaging / storage and dispatch; storage systems and port loading systems [44].

According to [42] one of the flaws in the interactive link in the meat production system, shows up at slaughter. For the amount of the refrigerator is relatively small compared to the amount of companies creating cattle. Data released in the 2006 IBGE Census of Agriculture show 2.6 million cattle ranches in the country and only 2 thousand slaughterers.

A study published in September 2014, conducted by the

Center for Advanced Studies in Applied Economics (Cepea), linked to the School of Luiz Queiroz (Esalq), the slaughter not inspected is estimated at less than 10%. These numbers are based on 2012 data on the 85% of the total slaughtered. The study did not include 100%, relative the 23.8 million tons of the Brazilian herd of the total slaughtered in 2012 and did not cover all the producing states. The Slaughter not inspected has some variations, as Table 1, according to the analyzed states.

Table 2 - Slaughter not monitored with reference to 2012 production [5].

Slaughter without Sanitary Surveillance (% of supply)	
São Paulo	5,40%
Rio de Janeiro	5,20%
Minas Gerais	15,20%
Espirito Santos	4,10%
Southeast	9,50%
Mato Grosso	5,60%
Mato Grosso do Sul	4,80%
Goiás	5,30%
Distrito Federal	6,80%
Midwest	5,70%
Paraná	5,50%
Santa Catarina	4,80%
Rio Grande do Sul	5,10%
South	5,20%
Bahia	10,40%
Northeast	10,40%
Pará	12,20%
Tocantins	9,20%

As the beef trade grows on average 2% per year, as the graph in Figure 3 shows, it is necessary to keep the slaughter index not inspected under control and close to extinction.

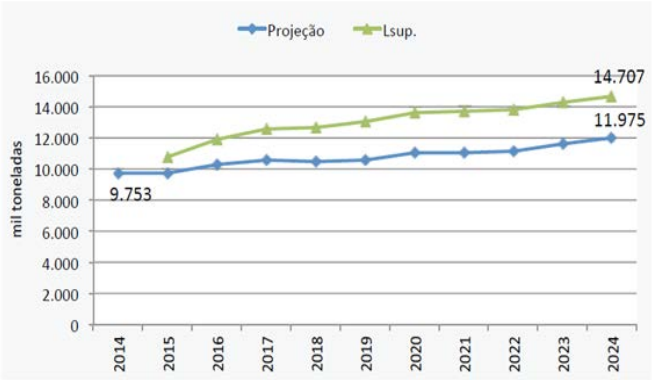


Fig.4 Meat production projection [49]

The automation of traceability within the supply chain of beef production, between the interactive link of the raw material production system and industrialization system, which is the product output (cattle) of the pasture until the accepted the cargo at destination, contributes to unify the production chain in a monitoring and certification system in order to: (i) guarantee the origin of the product, promoting reliability; (ii) to help streamline the transmitted cargo release and also (iii) minimize the incidence of illegal refrigerators.

The system of traceability in the beef supply chain, involves communication and transparency between all links in the chain and monitoring product (cattle) at any time of its production cycle. Providing greater safety in food quality, allowing greater economic benefits and confidence to the consumer.

IV. Conclusion

Invest in technology to the livestock sector is necessary for a coordinated control of meat production systems with sector's projected growth in the period 2013-2023 of 46.4% [1].

Include traceability throughout the supply chain systems of beef, with systemic nature of approach aims to avoid failures in procedures, and therefore, does not undermine the credibility of the final product.

Traceability properly inserted in the systems of beef production chain, integrating all the technologies developed, contributes to the correct flow and coordination of all production processes.

The demand for safe food health and known origin has increased considerably in the last year. Whereas every consumer must have access to safe food. The adoption of a comprehensive model of automation of traceability within the supply chain of allied beef production, among other factors, with existing technologies. Contribute to: (i) the supervision, reducing the levels of illegal slaughterhouses; (Ii) increasing the credibility of the agricultural sector and (iii) the growth controlled and insurance sector, allowing greater competition in the market for export and promoting domestic consumption.

V. References

[1] M.A.P.A (Ministry of Agriculture, Livestock and Supply). Ready to: <http://www.agricultura.gov.br/>. Consulted November 12, 2014.

[2] GTA (Transit Guide Animal). Ready to: <http://www.agricultura.gov.br/animal/mercado-interno/transito/>. Consulted November 12, 2014

[3] DIPOA (Department of Inspection of Animal Origin Products). Ready to: <http://www.agricultura.gov.br/animal/dipoa/dipoa-geral>. Consulted November 12, 2014.

[4] http://www.canaldoprodutor.com.br/area/184/Log%C3%ADstica%20e%20infraestrutura#wrapper. Consulted November 12, 2014.

[5] CEPEA - Cepea/USP estimates that uninspected cattle slaughters are less than 10% in BR

[6] SISBOV - Brazilian System of Identification and Certification of Bovine and Buffaloes. Ready to: <http://www.agricultura.gov.br/animal/rastreabilidade/sisbov>. Consulted November 12, 2014.

[7] "Carne, Osso" - Reporter Brazil, Documentary, 2011. Ready to: <http://reporterbrasil.org.br/carneosso/> - Consulted November 12, 2014.

[8] http://g1.globo.com/fantastico/noticia/2013/03/fantastico-mostra-falta-de higiene-em-abatedouros-e-abate-cruel-dos-gados.html. Consulted in November 30, 2014

[9] http://www.feedfood.com.br/abrafrigo-envia-comunicado-sobre-seu-posicionamento-frente-aos-abates-clandestinos/ .Consulted in November 30, 2014

[10] http://g1.globo.com/bahia/noticia/2013/05/apreendidas-seis-toneladas-de-carne-em-barreiras.html.Consulted in November 30, 2014

[11] .http://www.gazetadopovo.com.br/vidaecidadania/conteudo.phtml?id=1379497. Consulted in November 30, 2014

[12] http://www.canalrural.com.br/noticias/pecuaria/operacao-interdita-quatro-abatedouros-clandestinos-municipios-amazonas-27876.Consulted in November 30, 2014

[13] http://www.oparana.com.br/cidades/2013/09/ministerio-publico-fecha-o-cerco-contra-o-abate-clandestino/1152330. Consulted in November 30, 2014

[14] http://www.prefeituradepoa.sp.gov.br/novo/?p=4880. Consulted in November 30, 2014

[15] http://www.hojeemdia.com.br/noticias/economia-e-negocios/contra-clandestinos-governo-mineiro-quer-21-novos-frigorificos-1.230170. Consulted in November 30, 2014
[16] http://www.beefpoint.com.br/cadeia-produtiva/giro-do-boi/ue-exigencia-de-rastreabilidade-na-exportacao-de-carne-bovina/.Consulted in November 30, 2014
[17] http://www.beefpoint.com.br/cadeia-produtiva/giro-do-boi/ue-exigencia-de-rastreabilidade-na-exportacao-de-carne-bovina/. Consulted in November 30, 2014
[18] http://famasul.com.br/palestrascongresso/css/images/SenadoraKatiaAbreu.pdf . Consulted in November 30, 2014
[19] http://www.agricultura.gov.br/comunicacao/noticias/2014/12/exportacoes-do-agronegocio-alcancaram-uss-6-bilhoes-em-novembro. Consulted January 05, 2015
[20] http://apps.fas.usda.gov/psdonline/circulars/livestock_poultry.pdf. Consulted January 05, 2015
[21] http://stravaganzastravaganza.blogspot.com.br/2014/10/primeira-exportacao-de-carne-congelada.html. Consulted January 06, 2015
[22] http://www.abiec.com.br/img/Upl/perfil-290114-800.jpg. Consulted January 09, 2015
[23] http://www.valor.com.br/agro/3851172/exportacoes-de-carne-bovina-cresceram-77-em-2014-para-us-72-bi. Consulted January 01, 2015
[24] http://www.abiec.com.br/noticia.asp?id=1242#.VLAAtDHF_d0. Consulted January 09, 2015
[25] http://sites.beefpoint.com.br/pedrodefelicio/o-surgimento-dos-matadouros-frigorificos-no-brasil-do-inicio-do-seculo-xx. Consulted January 06, 2015
[26] http://www.abiec.com.br/news_view.asp?id=%7BCAACE975-B5D1-4337-9F3B-580E7118CB45%7D. Consulted January 06, 2015
[27] http://brasileconomico.ig.com.br/negocios/2015-01-09/exportacoes-do-agronegocio-do-brasil-recuam-32-em-2014.html. Consulted January 09, 2015
[28] http://www.abiec.com.br/news_view.asp?id=%7BF6B5E0C2-4EF1-4EBF-BA80-2484622ED8B2%7D. Consulted February, 17, 2015
[29] http://www.ceitec-sa.com/assets/documentos/produtos/Folhetos_CTC11002_bilingue_versao_02_junho2013.pdf. Consulted February, 17, 2015
[30] http://brasil.rfidjournal.com/noticias/vision?9680. Consulted February, 17, 2015
[31] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2000:204:0001:0010:PT:PDF. Consulted March 01, 2015
[32] http://www.abiec.com.br/noticia.asp?id=1232#.VS0x_fnF-So. Consulted March 01, 2015
[33] http://www.ipdeletron.org.br/wwwroot/pdf/27/gaesi.pdf. Consulted March 01, 2015
[34] http://ciencia.folhadaregiao.com.br/2014/08/nosso-insustentavel-boi-gordo.html. Consulted March 01, 2015
[35] http://jornalggn.com.br/blog/sistema-da-usp-dribla-burocracia-e-reduz-tempo-de-desembaraco-de-cargas-de-57h-para-1h30 . Consulted March 01, 2015
[36] http://globotv.globo.com/tv-tem-interior-sp/nosso-campo-tv-tem/v/exportacao-de-carne/3163850/.. Consulted March 01, 2015
[37] http://www.abreti.org.br/news196/130514b.html. Consulted March 01, 2015
[38] http://www.ceitec-sa.com/assets/documentos/produtos/folheto_CTC13001T_portugues_ingles.pdf. Consulted March 01, 2015
[39] http://www1.fazenda.gov.br/confaz/Confaz/Convenios/icms/2013/CV012_13.htm. Consulted March 01, 2015
[40] http://www.desenvolvimento.gov.br/portalmdic/arquivos/dwnl_1423144482.pdf. Consulted March 01, 2015
[41] http://www.abiec.com.br/download/Jan%20-%20Dez%20-%202014.pdf. Consulted March 01, 2015
[42] ftp://ftp.ibge.gov.br/Contas_Nacionais/Contas_Nacionais_Trimestrais/Fasciculo_Indicadores_IBGE/pib-vol-val_201404caderno.pdf. Consulted March 01, 2015
[43] Martins, F.M; Lopes, M.A. Agricultural Bulletin Federal University of Larvas - Beef Traceability in Brazil. p 1-72 August.
[44] Furquim, N.R. Safe Food an analysis of the institutional environment for beef supply in Brazil. São Paulo, 2012. 157p.
[45] Fernandez, M.L.A. Evaluation of the use of electronic tax documents in traceability loads. São Paulo, 2012. 142p.
[46] Dias, M. L. R. P. Brazilian logistics chain safe: international supply of processed beef and traceability. São Paulo, 2012. 166p
[47] ZAPIOLA, M.G. El bienestar animal y la calidad de la carne. In: Bienestar animal y calidad de la carne. Argentina: Instituto de Promoción de la Carne Vacuna Argentina– IPCVA, 2006. (Cuadernillo Técnico)
[48] Fontana, C. F. Methodology for the implementation of secure supply chain processes– São Paulo, 2009. P 175
[49] Silva, D. L. da Information system for traceability of forest products based on a service-oriented architecture– ed. rev. – São Paulo, 2012. 125p
[50] Lara, J. A. F; Soares, A.L; Lima, P.N; Ida, E; Shimokomaki, M : Traceability of beef: a requirement for food security - v. 24, n. 1, p. 143-148, jan./jun. 2003.
[51] Brazil. Ministry of Agriculture, Livestock and Supply. Agribusiness projections: Brazil 2013/2014 to 2023/2024 projections of long-term / Ministry of Agriculture, Livestock and Supply. Strategic Management Advisory. - Brasilia: MAP / ACS, 2014. 100 p

Options to Implement Bus Priority in the City of São Paulo

Luiz Cox, Alisson R. Leite, Eduardo M. Dias

***Abstract* -** This article has as objective discuss various techniques to prioritize the traffic of public buses in the road systems of medium and large cities, making comments and remarks to the specific case of the City of São Paulo. It aims to contribute to the increase of punctuality and reliability of this transport modality, in order to encourage the population of this large metropolis to migrate from individual to public transportation, bringing along uncountable benefits for the whole of its inhabitants and also to the environment.

***Keywords* -** priority, bus, buses, semaphores, controller, punctuality, service, urban mobility.

I. INTRODUCTION

Bus is one of the main categories of public transport, being accounted for about 80% of commuting by public transport in the world [1]. It is a flexible mode of transportation, reasonably cheap and easy to implement. Even in cities where rail based modes of transportation are necessary due to the high volume of trips, buses are an essential complement, mainly used to configure a feeding network for the higher capacity transportation means. By addition of extra infrastructure and some technology, bus based systems turn into BRTs ("Bus Rapid Transport"), with a transportation capacity to match those of light rail systems, in some examples reaching the lower limits of the transportation capacity of underground metro systems.

São Paulo has today one of the largest bus fleets in the world, with about 15.000 vehicles. They service daily a network of approximately 4.400 km, corresponding to about 25% of the city´s roads. São Paulo needs transportation for around 10 million passengers each workday. For this, 63% of the trips make use of public transportation [2].

Nevertheless, the continuous growth of the city´s population and consequently of the number automotive vehicles in circulation lead to the collapse of the road infrastructure, which is no more enough, particularly in the hours of peak of demand.

This causes heavy congestion, higher pollution levels, waste of natural resources and loss of quality of life. This situation affects the bus transportation system, since its vehicles have usually a lower capacity to accelerate compared to the average car, causing less regularity, quality and punctuality of service. Such situation negatively influences the migration of individual transportation to public transportation.

On the other side, the application of automation technologies in the urban environment, manly in transportation and mobility areas, make possible the better use of the road infrastructure existing in cities. It is what caracterizes ITS – Intelligent Transport Systems. Design and implementation shall always carefully consider the choice of technology, compatibility, scalability and maintainability aspects.

Automation may bring its contributions through traffic monitoring and control, via development and implementation of on-board systems or by enhancing the user information levels.

The main objective of this article is to discuss options for implementation of bus priority initiatives in São Paulo, aiming to make this transport mode more reliable, attractive and viable for its users.

II. PROBLEM

São Paulo, the most populated metropolitan region in Brazil, presented in 2012 the largest automotive vehicle fleet in country, with approximately 8,6 million units. This is equivalent to 17,3% of the whole Brazilian fleet. Between 2001 and 2012 the growth of São Paulo metropolitan area fleet was 76%, corresponding to more than 3,7 million in absolute numbers [3].

São Paulo population, as well as those of many other metropolitan regions in Brazil, suffers daily with significant congestion levels. This situation harms everyone, but particularly public bus users, system which cannot maintain the necessary regularity and comfort standards.

Although this problem worsens continuously, one must say it is not new. Already more than 40 years

Luiz Cox is with GAESI - Grupo de Automação Elétrica em Sistemas Industriais, a reseach group of the Electrical Energy and Automation Department, Escola Politécnica, Universidade de São Paulo, Av. Prof. Luciano Gualberto, trav. 3, n. 158, São Paulo/SP, Brazil, CEP 05508-970 (e-mail: luizcox@gmail.com).

Alisson R. Leite is with GAESI (e-mail: alisson_rodolfo5@yahoo.com.br).

Eduardo M. Dias is full professor of the Escola Politécnica of the Universidade de São Paulo and coordinator of GAESI (emdias@pea.usp.br).

ago the city began initiatives in order to prioritize the bus transportation in prejudice of the flow of individual vehicles. The idea was to organize a coordinate bus operation in specific corridors along the city. Soon happened the implantation of the first exclusive bus lane in the city, in the corridor Avenida Brigadeiro Luís Antônio. Today the city has 291 km of exclusive bus lanes, most of them implemented in 2013 [1].

The exclusive lanes present little or no physical segregation from the individual car lanes, exception is those of the Expresso Tiradentes.

In the following subsections we analyse some of the technology available to prioritize bus transportation that may be used in the city of São Paulo.

A. Solution involving road infrastructure

We evolve from the simple to the more complex solutions, so we have:

- Exclusive lanes: Present in Sao Paulo, they consist of a lane located to the right or to the left of the remaining lanes, they may also operate in with flow in the opposite direction. During its operation times, the use by vehicles other than buses is forbidden, including taxis. They are easy to set-up and operate in pre-determined days and times. Their main limitation is that buses still have to compete with individual cars that wish to do right or left turns. The Brazilian Traffic Code punishes the invasion of bus exclusive lanes by other vehicles with a light penalty. The value foreseen is R$ 53,20 (Arouns US$ 18,00) plus three points in the driver's record. This punishment is the same for taxis, forbidden to use these lanes.
- Bus Corridors: In Sao Paulo, bus corridors are located in the left side of some avenues, where larger bus stops are available. Taxis may use the bus corridor provided they transport passengers. Its deployment is more expensive and takes longer time than bus exclusive lanes, since it requires the construction of larger bus stops mostly in the central area dividing avenues and also buses with doors that open to the left side.

One of the main initiatives to enhance the quality of bus services is turn them into BRT systems (Bus Rapit Transit). The BRTs still do not exist in São Paulo.

There is no closed set of characteristics to precisely define a BRT service as such, but the main ones are:

1) Use of segregated lanes for the buses – there must be a physical separation between the way to be used by buses and those to be used by other vehicles to warrant the exclusivity of their use. Ideally there should be overtaking areas planed as well.
2) Availability of stations with an area reserved for passangers which already paid for the fare, so payment time does not interfere with boardoing times.
3) Boarding platforms in the same level of the bus floor, in order to minimize boarding and disembarking times.
4) A Control System. [4].

B. Solutions involving semaphoric control

Solutions based on semaphoric control aiming to prioritize bus flow may be passive or active.

Passive solutions are more appropriate where there is no centralized semaphoric control system, so that what is possible is to reprogram the semaphoric plans (times for red and green of the semaphores) in a way to consider prioritization of the existing bus flow across the junction. Such plans are variable depending on the time of the day and the day of the week. In passive solutions the definition of the plans is executed with the help of off-line computational tools or simulation tools such as the software TRANSIT, taking into account traffic demand surveys. Results achieved by the use of passive techniques are modest, so it should be used in smaller municipalities only.

On the other side active solutions make use of electronic equipment to detect the effective presence of buses and control semaphores to achieve the system operators pre-determined objectives.

It is possible to detect buses using the following technologies:

- Detector loops are wire loops installed in the superior layers of asphalt on the street, at the point where it is desirable to detect the presence of vehicles. Such loops are connected through feeder cables to electronic circuits inside the cabinets of traffic controllers installed near semaphores. The electronic circuits feed alternate current to the loops using frequencies between 10 and 200 kHz, depending on the manufacturer and model [5]. The moment vehicles flow over the loops, its inductance changes. The

ISBN: 978-1-61804-318-4

electronic circuits detect this variation and generate a pulse to the traffic controller CPU, to indicate the passage of one vehicle. Considering the amount of change in inductance and the time lapse of change, these circuits are able to deduct the type of vehicle that crosses the loop area, being able to tell if it is a bus or a private car.

- RFID Tags: These are radio transmitting devices installed inside buses. Once in the proximity of one receiving antenna, they transmit a given set of information, including vehicle identification. Receiving antennas are installed near junctions and relay the information to the traffic control system or semaphoric control [6]. There are active RFID tags, fed by an electric power source, or passive RFID tags, without need for electrical power. For bus detection application usually active devices are used, since they can be read from larger distances.
- Low Power Radios: They have similar functionality to the above described regarding RFID tags, but offer a wider variety of options for system design. They are data exchange devices based on radio-frequency, using low power in order to avoid unnecessary electromagnetic interference, which allow the traffic control system to receive information regarding the arrival of buses to crossing areas under its responsibility.
- CFTV – Video cameras capture sequences of images with a short time interval among them. After processing of these images using proper specific algorithms it is possible to identify vehicles and buses approaching junctions. This process request adjustments of image background, since conditions such as weather, luminosity and changes in the road may interfere in the adjustments to the algorithm [7-8]. In São Paulo, currently there is no connection between the public transportation cameras and those related to traffic control and enforcement.
- GPS: Global Position System, developed by the Department of Defence of the United States of America, initially for military purposes. The position is determined by passive receptors by triangulation, where signals from four satellites are used to define the position of a given point over the Earth surface. Reception of signal of three satellites makes already possible to determine latitude and longitude, while the fourth satellite add the height information and adjust the clock on the receptor side. Services offered are divided in two categories – the standard service, where time and positioning are offered to any user free of charge with accuracy of 95% and positional precision of 100m horizontally, 156m vertically and 185m in altitude, with a transfer time of 340 nanoseconds, and the precise service, available only for authorized user with accuracy above 95% and precision of 22m horizontally, 27,7m vertically and 35,4m in altitude, with transfer time less than 200 nanoseconds [9]. In equipment such as smartphones GPS is used along with other techniques to enhance higher effectiveness in position determination [10].
- AVL: Automatic Vehicle Location are embarked systems that manage the traffic variables of the vehicle being capable of inform the status of the bus and its position to a central data and to street side equipment, making it possible to monitor the vehicle in real time [11].

Data related to the arrival of buses to junctions or position of buses is relayed to the operational control centre and to the semaphoric control centre, or both. The systems can also communicate with each other to exchange relevant data.

There are various levels of automation possible, depending on the complexity available in the systems, of the specific needs of each municipality and the historic of implantation. The possibilities range from traffic control systems giving immediate unconditional priority to all arriving buses to complex systems which integrate on-line positioning and delay (or advance) information so that the semaphoric system can prioritize only delayed buses or make sure the headway is kept [12-13].

The main actions a semaphoric control system can perform to effective prioritize buses are:

1) Extension of green during vehicle approach to a given junction,
2) Anticipation of green period during vehicle approach to a given junction,
3) Change (inclusion or suppression) of a semaphoric stage.

ISBN: 978-1-61804-318-4

C. *Integrated Solutions*

Integrated solutions combine aspects linked to physical infrastructure with aspects related to semaphoric control. Some examples include the access control of vehicles to critic junctions ("metering"). Other example is the allocation of road areas for buses to wait for green nearer the junction if compared to the other vehicles.

III. PROPOSAL

The first point to be considered to choose the most adequate solution for the city of São Paulo is to define the desired impact in the bus transportation system. Such objectives should be:

1) To minimize travel time,
2) To warrant punctuality along the travel path,
3) To warrant an uniform headway for vehicles of selected lines,
4) To have other positive side-effects (to minimize emission of harmful gases, to minimize fuel consumption),

The most adequate solution for the city would be a combination of options 2) and 3) above. Punctuality brings the sensation of confort and reliability of the public transportation system and importantly influences the choice of citizens about which transport mode to use. Punctuality is particularly important when the frequency of the line is not high (interval between buses above 10 minutes). Punctuality also allows users to optimize their time, allowing them to plan arrival to the bus stops 2-3 minutes before the foreseen arrival of the vehicle.

On the other side, for high frequency services in high capacity corridors users usually arrive to the bus stops following a uniform distribution. The most important in this case is to warrant the regularity of the headways – time intervals between vehicles, so to minimize the average waiting time for the users.

To achieve these objectives it would be recommendable to implement the following steps:

1) The city must have available one real time semaphoric control system which includes a bus prioritization software module, such as the SCOOT system, which is already available and installed in some regions of the city, but needs expansion to cover all the areas where bus passenger flow is relevant.
2) To upgrade and integrate the bus operational control systems available in order to make sure the real time position of each bus is known, as well the information related to the situation of delay or advance of the said vehicle in comparison with its travel plan.
3) To implement communication ways between the operational bus control system and the semaphoric control system, directly or via a higher level system, so that the semaphoric control can act to achieve the objectives defined – to enhance punctuality and regularity of headway of São Paulo public bus system.

These systems, once integrated, or alternatively the higher level system above referenced, shall have the important role of making information about the bus system available to its users, informing routes, times and possible exception situations, as well as alternatives. Such information shall be made available to the users via Variable Message Panels and Internet.

To warrant the complete integration and interoperability of the systems, as well as scalability for future expansions, ITS systems to be installed in the city of São Paulo, must make use of open communication protocols such as UTMC or NTC-IP in its last version. The Municipal Transportation Secretary already made this decision and it is valid for the whole city and for all ITS equipment to be acquired.

IV. CONCLUSION

The city of São Paulo needs to warrant to its citizens and visitors a high quality public transportation system. Fundamental aspects for the option of the transportation mode are regularity and punctuality of service. So, we propose and recommend modernizing and integrating the operational control systems of the city buses and the existent semaphoric control system, in order to achieve these important objectives.

REFERENCES

[1] I. Gnecco Filho, "*A Prioridade para o Transporte Público"*, Coletivo – Revista Técnica SPTrans, 2012, vol.0, n.1, 2012, pp. 11-12.

[2] I. M. Whaterly, "*O Papel do Ônibus no Transporte Público de São Paulo"* , 2012, vol.0, n.1, pp. 22.

[3] Site:http://www.observatoriodasmetropoles.net/download/auto_motos2013.pdf, *Evolução da Frota de Automóveis e Motos no Brasil 2001-2012*, access date: 29/03/2015.

[4] C. L. Marte et al., "*Estudo Preliminar de Funções ITS Aplicadas na Operação de Sistemas BRT*," Série Cadernos Técnicos – ANTP, 2012, vol.8, n.0, pp. 100.

[5] L. A. Klein et al., *Traffic Detector Handbook*, 2006, vol.1, no.0, pp. 1-2.

[6] Z. Li et al., *Expert Systems with Applications, Anti-collision Algorithm Using Adaptive Hierarchical Artificial Immune System*, 2014, vol. 41, no.5, pp. 2126-2133, doi="http://dx.doi.org/10.1016/j.eswa.2013.09.011",

[7] L. C. León, R. Hirata Jr., *Car detection in sequences of images of urban environments using mixture of deformable part models*, Pattern Recognition Letters, 2013, doi="http://dx.doi.org/10.1016/j.patrec.2013.10.028".

[8] M. Bertozzi, A. Broggi, GOLD, "A parallel real-time stereo vision system for generic obstacle and lane detection," *IEEE*

Transactions on Image Processing, 1998, vol.7, No.1, pp.62-81, doi=10.1109/83.650851, ISSN=10577149.

[9] N. Ananthanarayanan, "Intelligent vehicle monitoring system using wireless communication," *Advances in Technology and Engineering (ICATE)*, 2013 International Conference on, 2013, pp. 1-5, doi=10.1109/ICAdTE.2013.6524722.

[10] Y. He, R. Martim, A. Bilgic, *Scalable low-complexity GPS and DGPS positioning using approximate QR decomposition*, 2014, vol.94, no.0, pp.445-455,doi="http://dx.doi.org/10.1016/j.sigpro.2013.07.014".

[11] N. Hounsell, B. Shrestha, "A New Approach for Co-Operative Bus Priority at Traffic Signals," *Intelligent Transportation Systems*, IEEE Transactions on, 2012, vol. 13, n. 1, pp. 6-14, doi=10.1109/TITS.2011.2172869, ISSN=1524-9050.

[12] D. Gorni, *Modelagem para Operação de Bus Rapid Transit*, Dissertação, Escola Politécnica da Universidade de São Paulo, São Paulo, 2010.

[13] F. M. Oliveira Neto, *Priorização do Transporte Coletivo por Ônibus em Sistemas Centralizado de Controle de Tráfego*, Dissertação, Universidade Federal do Ceará, Fortaleza,2004.

Methodologies and Techniques to preventive control of Dangerous Cargo Mass Notification & Advisory System

Luiz Antonio Reis, Prof. Eduardo Mario Dias, Prof. Sergio Luiz Pereira

Abstract: -The objective of the Methodologies and Techniques to preventive control of dangerous goods transport mass notification & advisory system project is to prototype a new operational system for monitoring the transportation of dangerous goods in Brazil based on regional responsibilities. This concept, based on systems used in air traffic control[1], aims to provide civil security centers with real-time knowledge of the position and contents of dangerous vehicles circulating in their area of responsibility, and, in the event of a dangerous situation, to issue warnings, alerts and crisis management information, thereby allowing intervention teams to react immediately with maximum safety.

Keywords: -Predictive control, Dangerous Goods Transport, Mass Notification Advisory System, Alert Communication Emergency

I. INTRODUCTION

The transportation of dangerous goods faces a pool of problems, mainly the fact the cargo aren´t tracked by the responsible authorities which can´t take preventive measures, it canraise the delay to identify, respond and control an accident on a given zone.

Figure 1 illustrates the nine classes that dangerous goods are classified:

Class	Symbol
Class 1 – EXPLOSIVE	
Class 2 – GASES:	
Subclass 2.1 – Flammable Gases;	
Subclass 2.2 –Non-Poisonous, nonflammable gases;	
Subclass 2.3 –Poison Gases;	
Subclass 2.3 –Corrosive Gases;	
Class 3 –Flammable liquids	
Class 4 - Solids:	
Subclass 4.1 –Flammable solids;	

ISBN: 978-1-61804-318-4

Subclass 4.2 – Spontaneous Combustion Substances;	
Subclass 4.3 –Water contact flammable gases emission.	
Class 5 - Oxidants:	
Subclass 5.1 –Oxidant Substances;	
Subclass 5.2 –Organic peroxide.	
Class 6 - Poisonous:	
Subclass 6.1 –Toxic Substance (poisonous);	
Subclass 6.2 –Infectious Substances.	

Class 7 –Radioactive Material (Category I)	
Class 7 –Radioactive Material (Category II or III)	
Class 8 - Corrosives	
Class 9 – Miscellaneous Dangerous Substance	

Figure 1 : Dangerous classes identification

The main causes of road transportation accidents are four:

a) Lack of signalization or maintenance on roads, avenues and streets;
b) Human behavior (speeding; fatigue; inexperience, lack of attention);
c) Lack of maintenance on vehicles;
d) Incompatible cargo to vehicle type (volume, weight, incorrect lashing).

The usual consequences of road transportation accidents are four:

a) Loss of lives;
b) Traffic Jam;
c) Increase of pollution of rivers, ground and air contamination;
d) Unnecessary exposure of first rescue teams without correct protection equipment;

ISBN: 978-1-61804-318-4

The transportation system is essential to economic development, it boosts the GDP - Gross Domestic Product and according to Brazilian transportation ministry 58% of cargo transportation in Brazil is made by roads, this infrastructure is one of the bottleneck of the Brazilian growth and accordingresearch of CNT (National Transport Confederation) the pathways growth 20,1% on the last ten years.

The objective of the system is to integrate the authorization of dangerous goods database with city hall and road planning database to schedule and optimize the best route.

II. Problem Formulation

Brazilian authorities must authorize all dangerous goods transport as DSV (Road System department) and CODESP (Port public security system) these authorities consult and share the information with any other areas involved.

DSV and CODESP are in compliance with the sectorial politics elaborated for two members:

a) CONIT - National Council of Transport Policy Integration
b) Transport Ministry and responsible for sectorial policy elaboration.

After elaborated, the politics are overseen by two regulatory agencies to control private companies exploration in the transport services:

a) ANTT – National Terrestrial Transport Agency
b) ANTAQ – National Aquatic Transport

A. Present Scenario

The companies need to ask authorization in order to do the dangerous good transportation.

The DSV-Road System Department analyses the request makes the emergency plan and communicates other authorities like the civil defense department, the green and environment secretary and the traffic engineering company.

The emergency plan must include three requirements:

a) The list of human resources and materials available in emergency committee;
b) The dangerous goods classification;
c) The list of companies that aren´t in compliance with dangerous transport resolution.

Once made the emergency plan the requestor company will receive the authorization to transport dangerous goods.

The factual scenario has three problems:

a) There isn´t enough surveillance;
b) The trucks aren´t tracked on real time;
c) The database of police, civil defense, road system department and traffic engineering company aren´t integrated.
d) Department and the Traffic Engineering Company aren´t integrated.

Considering those facts, if an event with a dangerous situationoccurs, the time to warning, alert and manage crisis information, calling intervention teams to react isn´t fast enough to prevent damage.

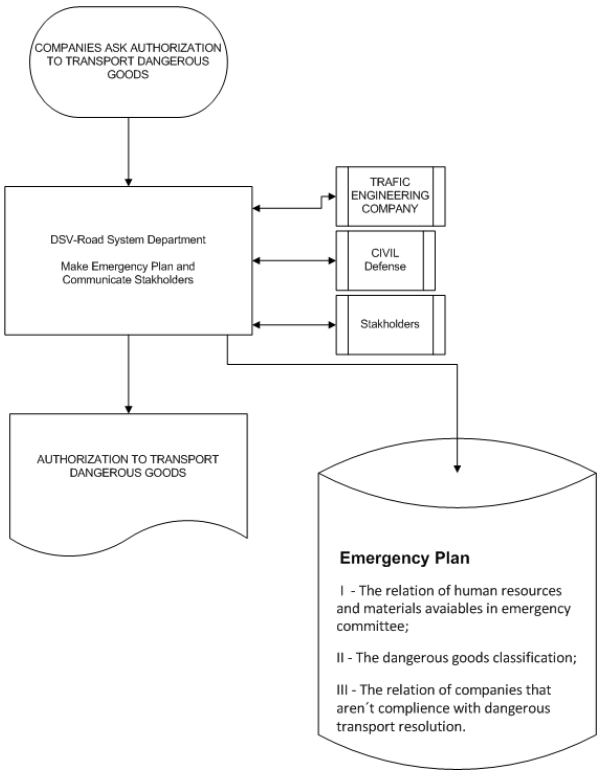


Figure 3 :Factual workflow

III Problem Solution

A. Route Planning:

In order to plan the routes, six steps are followed:

a) Post all the cities of a region in terms of coordinates
b) Launch on the region, a road network spatially well distributed and containing the major cities, as shown on figure 4.
c) Calculate distances
d) Calculate the route time
e) Calculate the route cost
f) Choose the best option

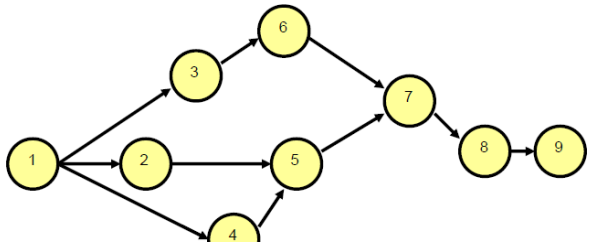


Figure 4: A road Network

B Georefences:

São Paulo city is the main city in the South America with over 20 million citizens and 17 thousand kilometers of streets and avenues. Due to a metropolitan city São Paulo has a georeferenced database planning where is possible to know the status of every work around the pathways, named GeoSampa.

The figure 5 shows GeoSampa. It is a system that integrates databases and synchronizes information. Nowadays the information can be used by the City Hall administration and there is in development new features to make information available to citizens.

Figure 5 - Geosampa screen

B. Data acquisition:

Trucks can be tracked by five main technologies:

a) RFID technology.
b) Cell Phone carriers using GSM, GPRS, WCDMA or LTE commercial networks
c) GPS – Global Positioning Satellite
d) V2I : Vehicle to Infrastructure technology, IEEE 802.11p standard.
e) Surveillance camera to catch the trucks identification

D. Integration:

The proposal of Methodologies and Techniques to preventive control of Dangerous Goods Transport Mass Notification & Advisory System project combines RFID with GSM, Cameras and GPS systems.

Integrating the georeferenced planning, authorization and track data bases, the system can communicate official information during an emergency or crisis situation that disrupts normal operation of Dangerous Cargo transport[2].

E. Route optimization:

The route optimization is analyzed between two point of view:

a) Static, the regular way to plan a route. The routes are planning considering the costs of the distance vector, the costs in our study are the roads, streets and avenues that can be measured using off-line statistics.

Dynamic with interactive system. The main advantage of this method is when some way is interrupted new on-line routes will be planned faster and with more options than a static planning. In our study, each street corners could be considered a decision point, where new routes can be traced as soon as new events happen.

F. Rolling Horizon

Another technique is Rolling Horizon where one stage is divided in *k* intervals, named projection stage or projection horizon.

In projection horizon the first *r* intervals are the head and after this the tail of the horizon.

Using the head is possible to estimate the behavior of the tail, and according the time flows the head advances into the tail and new behavior can be calculated, as shown in figure 6.

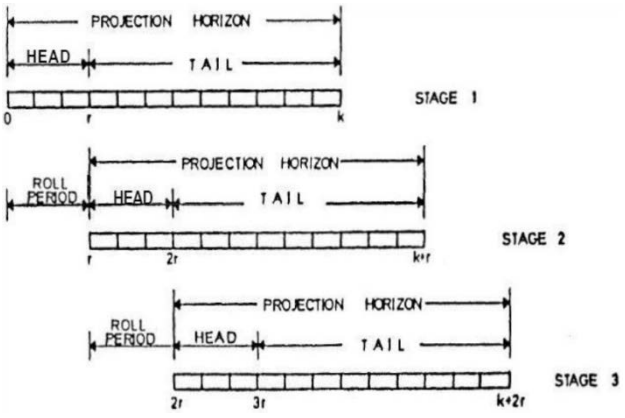


Figure 6: Illustration of the Rolling Horizon Approach [6]

ISBN: 978-1-61804-318-4

G. Human decision

The Bayes theorem is a corollary that calculates the total probability based on a conditional relationship:

$$\Pr(A|B) = \frac{\Pr(B|A)\,\Pr(A)}{\Pr(B)}$$

-Pr(A) e Pr(B) are the A and Bprobabilities

-Pr(B|A) and Pr(A|B) are the probabilities B conditioned to A and the probabilities A conditioned to B respectively.

Bayes rules show the probability to avoid two possibilities:

a) False positives- An alert without really dangerous situation.
b) False negative – A dangerous situation without alerts.

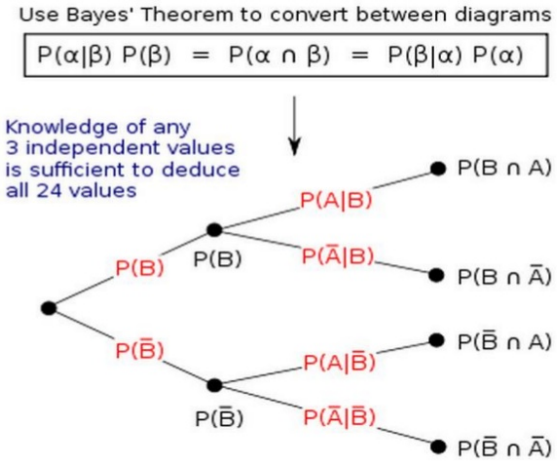


Figure 7: Bayes' Theorem

H. Problem Management

Problem Management [5] is an ITIL - Infrastructure Technology Information Library – is a technique to register know ledged causes with standard behaviors that can make procedures to take programmed actions in order to solve them.

I. Decision making

BI – Business Intelligence techniques can be used to support the best human decision.

The data can be gotten from at least seven ways:

a) Tracking systems
b) Crowd application
c) Social networks
d) Phone calls
e) Police Department
f) Municipality control

ETL – Extract Transform Load – figure 7 shows the seven steps that can be used to populate the datawarehouse:

a) Extract: Extracts data from the source system
b) Transform: Apply functions to conform data to a standard dimensional schema
c) Load: Load the data into the data mart for consumption
d) Process: Load the data from the data mart into the cube for browsing and analyses the information on cube view at OLAP – On-Line Analytical Processing

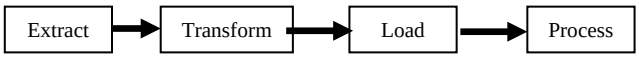


Figure 7: ETL workflow

The Knowledge management consists in three main steps:

a) Data: The data is collected from sensors and operational systems.
b) Information: The data is organized and summarized, become information.
c) Knowledge: The information is analysed, and synthetized, become knowledge to support decisions, as shown on figure 8.

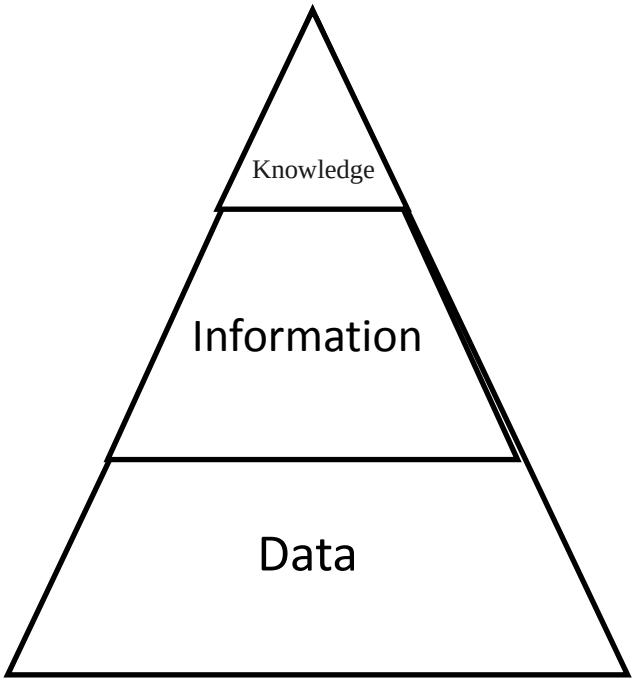


Figure 8: Knowledge management

J. Alerts

The Civil Defensedepartment can send Short Message Service (SMS) text messaging automatically by gateway to carrier system to all cellphones of any carriers, input alerts on websites, e-mail, radio, TV, social networks and crowd apps.

The system is comprised of several individual (multi-mode) resources to advise the community of an emergency in progress.

K. Nominal Situations:

A nominal situation is given when no accident happens. Based on MITRA [3] the system tracks the goods and reports nominal situations as four heartbeat signalization:

a) Current vehicle position;
b) Current vehicle speed;
c) Cargo identification;
d) Potential risks based on vehicle and areas characteristics.

L. Alert/Crisis Situations:

A crisis situation requires an action beyond the control of the driver, it can be detected if the vehicle changes the pre-planned route, reduce the speed dramatically or if the driver pushes the panic button.

In order to alert the civil population, the system uses mass notifications, traffic advisory notices, and emergency news and information systems to communicate during critical events. Civil defense also issue information about special events through their websites.

The emergency communication and advisory systems are constantly changing as technology involves on a day-to-day basis. The community should refer to this site frequently to be informed and be educated as soon as new systems are implemented and upgraded. Part of the warning system's success is the ability of the community to understand how the systems are used, their limitations and how a person should respond once the system is activated.

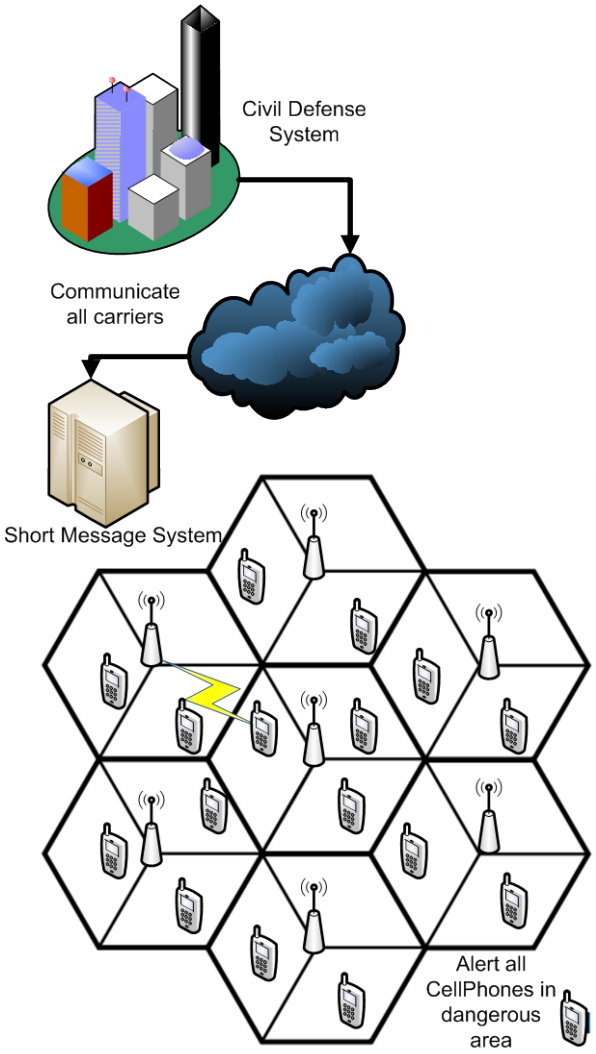


Figure 10: Cell phone Carrier alert area

The system can be improved with visual warning systems, audible warning systems (both indoor and outdoor), and e-technology systems to communicate during a crisis. It is important to understand that no single system has the capability to provide 100 percent coverage of the area. An effective and reliable notification system utilizes many redundant systems, integrated with several resources. The alerts are impacted by topography, building design and other factors in providing mass notification and communications during crisis.

The system of Communication & Advisory Resources Includes database integration among seven main authorities

a) Docks Companies,
b) Road System Department,
c) Civil Defense,
d) Ports,
e) Highway Police,
f) Cities Police,

ISBN: 978-1-61804-318-4

g) Traffic Engineering Company

The alert system includes ten ways of communication:

a) AM and FM Radio Station
b) Cable Television System
c) Local Broadcast Television System
d) Outdoor Warning Sirens located in the ports areas and the high risk areas.
e) Safety Website
f) Emergency Website
g) Crowd system, Waze app
h) Social network, twitter app, whatsup, Facebook Website.
i) Cities Main Website and Newsroom
j) All cellphone carriers (Send Short Message to all cellphones covered by the affected area)

M. Proposed Scenario

The proposal of Methodologies and Techniques to preventive control of Dangerous Goods Transport Mass Notification & Advisory System project uses RFID and GSM/LTE and GPS networks to track the vehicles.

Integrating the georeferenced planning, authorization and track databases, the system can communicate official information during an emergency or crisis situation that disrupts normal operation of Dangerous Cargo transport.

The proposed scenario integrates on the same database the main four authorities:

a) Docks Department,
b) Police Department,
c) Road System Department
d) Civil Defense Department

The problems can be reported by phone calls, website or social network applications, all of them are commercial available and can be easily implemented.

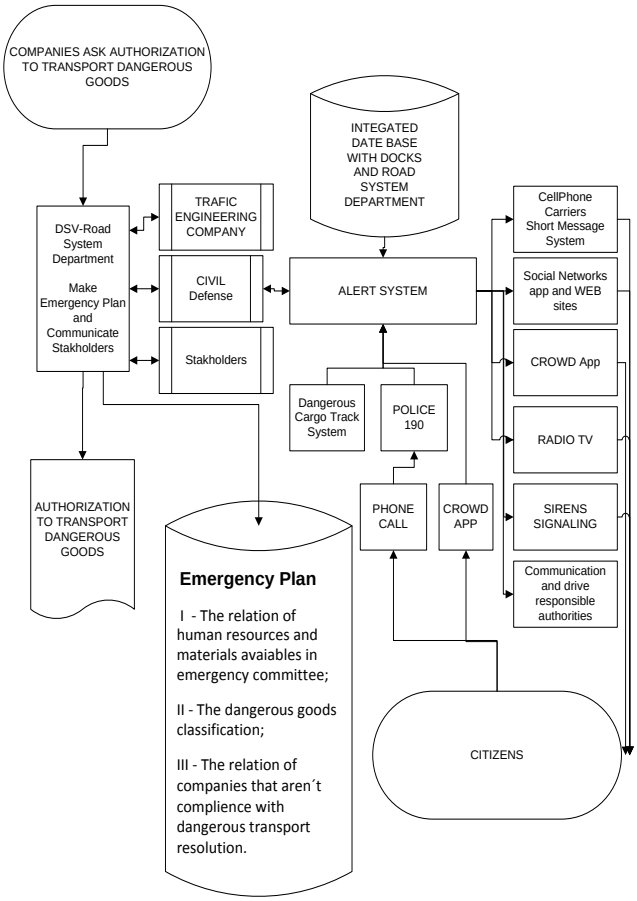


Figure 11: Proposed scenario

IV CONCLUSION

This paper addresses the possibility to integrated several resources to track vehicles, alert population and active responsible authorities during crisis deal to prevent huge damages.

The main idea is to use existing tools with high efficiency, low costs and short roadmap implementation.

REFERENCES:

[1] The application of information and communication technologies in transportGiannopoulos, G.A European Journal of Operational Research, 2004, Vol.152(2), pp.302-320
[2] The application of information and communication technologies in transportGiannopoulos, G.A European Journal of Operational Research, 2004, Vol.152(2), pp.302-320
[3] Results of the MITRA project: Monitoring and intervention for the transportation of dangerous goodsE. Planas, E. Pastor, F. Presutto, J. TixierJournal of Hazardous Materials 152 (2008) 516–526
[4]PMBOK – PMI – Project Management Institute
[5] *ITIL Foundation Essentials* from ISACA.org
[6] A Review of the Optimized Policies for Adaptive Control Strategy (OPAC) Lawrence C. Liao University of California, Berkeley

Research on the integration of automation systems involving “Transit” and “Safety” processes

Marcelo L. Fernandez, Eduardo M. Dias

***Summary* —** The transit of dangerous products in urban areas carries risks for all of society as well as the correlating environment, indicating that the transit control of these products is something that must be controlled and integrated into security systems. This article describes the proposal of this integration, with the assistance of information found in electronic fiscal documents.

***Key-Words* —** dangerous products, electronic invoices, safety, transit, vehicles.

I. The Integration Scenario

THE transportation of dangerous products, due to inherent characteristics as well as packaging and wrapping materials, carries risk for the environment, public safety and public health.

For this reason, the transportation of these products in Brazil, notably on the highway, must submit to rules and procedures established by the National Agency of Land Transport – ANTT, which controls the subject through Resolutions ANTT nº. 3665/11, complemented by the Instructions approved by ANTT Resolution nº. 420/04 and its bylaws, without modifying the specific norms of each product.

According to Decree nº 50.446, from 20 February, 2009, materials, substances or artifacts are considered dangerous products if they carry risk to human and animal health, or if the materials cause damage to the environment. In this case, the Numeric Relation of Dangerous Products is found in Chapter 3.2 of Resolution nº 420 from National Land Transport Agency (in Brazil has the following Acronym: ANTT), from 12/02/2004, and is composed of the following products: explosives, gases (flammable gases, non-flammable gases, non-toxic; toxic gases), flammable liquids, flammable solids, substances which may spontaneously combust, substances which, upon contact with water, emit flammable gases, oxidizing substances, organic peroxides, toxic substances (venomous), infecting substances, radioactive materials, corrosives, various dangerous substances.

M. L. Fernandez is with GAESI - Electrical Automation Group of Industrial Systems (Av. Prof. Luciano Gualberto travessa 3 no. 158 - São Paulo – SP – CEP 05508-970; e-mail: mlafernandez@gmail.com).

E. M. Dias is PhD in Electrical Engineering and full professor at the Polytechnic School of the University of São Paulo - USP. He is coordinator of GAESI - Electrical Automation Group of Industrial Systems and researcher at the Electrical Department of Energy and Automation. Polytechnic School, USP (emdias@pea.usp.br).

Some cities, worried about the inherent risks of these products, restrict or prohibit the circulation of vehicles which transport dangerous products in urban zones. This is the case,for example, in the city of Sao Paulo, according to a warning advertised on the Traffic Engineering Company (CET) website [1]:

> “The City of Sao Paulo, via the Department of Roadway Operating System (DSV), decided to prohibit the circulation of vehicles which transport dangerous products between 5 a.m. To 10 a.m. And 4 p.m. To 9 p.m. From Monday to Friday. (...)”.

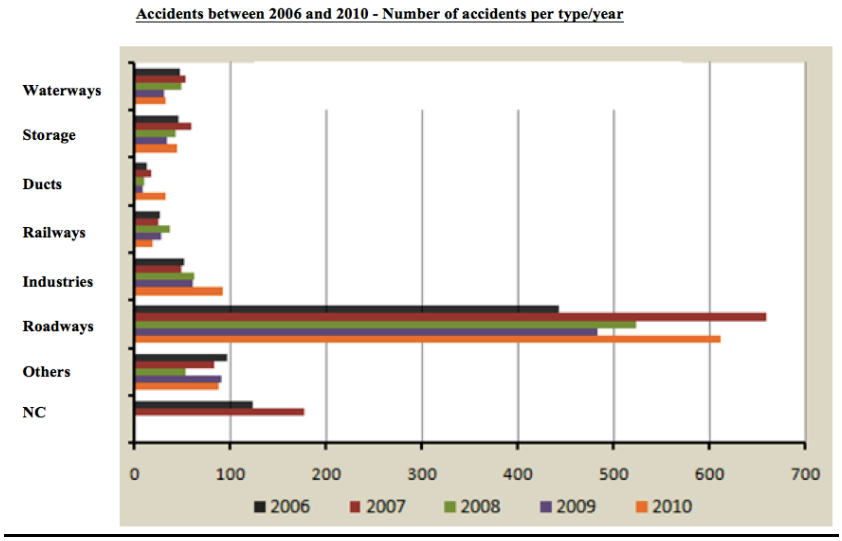
To illustrate the problematic nature involved in the circulation of these products, we present research on data regarding traffic accidents involving dangerous products in Brazil, between the years of 2006 and 2010, available in Fig. 1.

The research reflects the proportion of accidents involving roadway transportation of dangerous products in relation to other modes or industries, reaching alarming numbers. This information is available on the Ministry of Environment website [2].

For this reason, it has become necessary to develop effective controls and to integrate traffic control and urban safety, to identify dangerous cargo in order to prevent accidents and health dangers in the urban areas where these products circulate.

Therefore, the initial question becomes: how to discover that any specific vehicle is transporting cargo which is considered dangerous?

ISBN: 978-1-61804-318-4



Waterways = waterways + maritime
Others = airport + dam + commerce + port + platform + gas stations + refineries + urban streets
NC = Not Classified

Fig. 1 - Number of accidents involving dangerous products in Brazil between 2006 and 2010

A possible source for this is the information from commercial operations and cargo transportation service providers, available in electronic invoices. The Secretaries of Treasury of the 26 States and the Federal District, as well as the Brazilian Internal Revenue Service, have been working since 2005 on the implementation of fiscal documents in electronic format, in order to replace the corresponding paper documents – these are the Electronic Invoice (NF-e), to document commercial operations between companies; the Electronic Acknowledgment of Transport (CT-e), to document cargo transportation services; The Electronic Manifesto of Fiscal Documents (MDF-e), to link transportation documents (NF-e and CT-e) with the respective vehicle. These documents, as well as numbers which support their successful implementation, will be detailed below.

At this point, two important observations must be done:

a. There are several aspects involved in transport and security. This article deals with one of these aspects, the transport of dangerous products. Other common elements such as vehicle quantity, speed limits, and maximum vehicle weight permitted on each street were not analyzed at this moment;
b. The use of the electronic fiscal document is subject to fiscal confidentiality, according to articles 198 and 199 of the National Tax Code. The use of this information is restricted to the Secretary of Treasury, and its usage, as described in this article, must be preceded by juridical analysis.

II. METHODOLOGY

The thesis of this article is that the integration of transport and security can be carried out via the vehicle's license plate, the identification of the related linked fiscal document, the verification of the existence of dangerous products within the identified document and the use of security systems as detours or blocks. The methodology involves the following steps:

a. Define alternative solutions from identifying license plates. Naturally, there is not only one solution, allowing one or more alternatives to be used individually or jointly, according to the best cost-benefit;
b. The identified license plates must be transmitted to a data processing center, which has access to a database (although only partially) that details the types of vehicles associated to the plate (ex: DETRAN) and the recent electronic fiscal documents (ex: Secretary of Treasury);
c. If the presence of dangerous products is identified, the security systems must be activated, in the sense of alerting traffic teams in order to intercept and intervene the vehicle during transit, activating automatic detours, implementing gates, etc.

The following alerts and restrictions should be taken into consideration for the implementation of the proposal:

a. The transport of dangerous products between states, in which the State of Sao Paulo merely grants passage but doesn't play the role of the following within its territory: emitter, recipient, expediter, or service provider. In this case, the State of transit may be unable to provide the electronic documents and the identification of cargo transport vehicles, which have no associated electronic invoice, could be an acceptable alternative;
b. Situations in which transportation is carried out without invoice. In this case, the procedure described in the previous item is possible; the identification of cargo transport vehicles which have no associated electronic invoice, in zones where transit is not common;

ISBN: 978-1-61804-318-4

c. Situations in which the information regarding the dangerous products is not entered on the invoice. In this case, the taxpayer must be charged with previously described laws, according to infractions already described by judicial order;
d. Situations in which transport occurs by private vehicle;
e. Failures in license plate identification: for example, cases in which the plate (or its corresponding chip, when installed) cannot be identified on account of dirt.

III. Study details

Brazil has implemented, since 2005, electronic fiscal solutions. Since then, two more types of electronic documents were developed and implemented in Brazilian territory, currently aimed at transportation service providing and the identification of respective vehicles.

In the following we present the summarized description of the operation model, the existing blocks of information and the data and numbers of each electronic document. The complete details are found in the Master's dissertation of this author, entitled "Avaliação da utilização de documentos fiscais eletrônicos na rastreabilidade de cargas (Evaluation of the usage of electronic invoices in cargo traceability)" [3] as well as the WSEAS article, presented November 2013, entitled "Documentos fiscais eletrônicos e o Rastreamento de Mercadorias e veículos" (Electronic invoices and the tracing of products and vehicles) [4].

A. Electronic Invoice (NF-e)

This electronic document substituted the sales invoices between companies. The operational model for the NF-e consists of (i) the generation of an XML format file for the taxpayer, containing all the information which reflects the commercial operation to be realized; (ii) a digital signature of the taxpayer within the file; (iii) transmission of the file to treasury organizations, through the Internet and WebService technology; (iv) authorization or rejection of the NF-e by the treasury organization and the respective response notified to the taxpayer; (v) if the commercial operation is authorized, an auxiliary document is printed only for transportation of the goods (DANFE - Auxiliary Document for the NF-e) and, finally (vi) the initiation of goods transportation.

1) Summary of information provided to the taxpayer in the NF-e

The NF-e is an XML file composed of groups of information. In the following, we highlight the fields that are useful for tracking products (the complete layout is available in the taxpayers manual, at the national project site - http://www.nfe.fazenda.gov.br [5] – or at the Paulista project site– https://www.fazenda.sp.gov.br/nfe [6]):

a. Identification of Electronic Invoice: indicators that allow for the differentiation of one NF-e from another, such as serial numbers. Operational information exists, such as emission date and the date of products' release.
b. Identification of NF-e emitter: Provides registration data on emitter and emitter's address.
c. Identification of product destination: provides registration data and address.
d. Identification of pick-up location of goods;
e. Identification of delivery location.
f. NF-e Products and services of NF-e: Provide all commercialized product information. Relevant Information:
g. GTIN (Global Trade Item Number) – Barcode;
h. Description of product or service;
i. NCM Code (MERCOSUL Common Nomenclature) with 8 digits or 2 digits (generic).
j. Incidental taxes on product or service;
k. Total NF-e values;
l. NF-e Transport Information. Relevant Information:
m. Main License Plate;
n. Information on Digital Signature.

There are two pieces of information which are of fundamental importance for traceability found in the NF-e layout: the products exit date and the transporter's data (such as the vehicles license plate and/or driver details).

Both pieces of information are optional in this layout, as they can be unknown by the emittor upon emission of the electronic document. This means that it is possible to not be contained within the NF-e, upon emission.

2) Obligation of use

Currently all industrial, wholesale commerce and taxpayers which carry out trade operations abroad, interstate or with public organizations are obligated to use the Electronic Invoice. This means that industrial production and wholesale commerce are already documented via NF-e, allowing for widespread use in tracking work. If a portion of commercial operations were to be documented by paper, this work would be damaged. More than 2.5 million NF-es are emitted per day in the state of Sao Paulo alone, involving more than 620 thousand industrial and commercial establishments. Within all of Brazil, there are 8.4 million NF-e daily, emitted by over 1 million establishments.

B. Electronic Acknowledgment of Transport (CT-e)

This has the goal of altering the systematic emission of invoices referring to cargo transport.

According to the master's dissertation of this author, the operational model of the CT-e follows the same model applied to the Electronic Invoice (NF-e), previously explained, including the emission of the auxiliary paper document used to follow the transportation service (DACTE – Documento Auxiliar do CT-e).

However, there are some differences for the transport sector regarding the people involved in transport services provided. In addition to the emitter of the document and the recipient of the good, the transport sector also involves a service receiver (responsible for service contracts and payments), goods expediter (who delivers the goods to the transporter in order to provide transportation services), and the product recipient (who should receive the goods from the transporter).

Each participant of the the transportation service must emit an electronic document in the appropriate moments, according to tax legislation, following the same operation model as the NF-e, as previously seen.

1) Summary of provided information

The "Conhecimento de Transporte Eletrônico" (Electronic Acknowledgement of Transport) is an XML file made of up the following groups of information (the complete layout is available in the taxpayers manual, at the national project site- http://www.cte.fazenda.gov.br [4] – or the project website in São Paulo – https://www.fazenda.sp.gov.br/cte [8]):

a. Identification of CT-e: provide indicators, which allow the differentiation of one CT-e from another, such as serial numbers. This group provides information on services provided, such as emission date and type of transport utilized (air, roadway, waterway, railway or pipeline).
b. Identification of CT-e emitter: Provide registration data and address.
c. Information on the sender of transported goods by CT-e: Provide registration data and address.
d. NF-e Information (NF-e Access Key). Note that the CT-e also registers paper documents emitted by the sender, when this is the case.
e. Information on the product expediter: Provide registration data on the expediter (when one exists) and address.
f. Information on the product recipient: Provide registration data (when one exists) and address.
g. Information on the CT-e recipient: Provide registration data and address.
h. Values for services Provided;
i. Information related to taxes;
j. Group of CT Information- normal and substitute: provide details on transported products. Relevant Information:
    i. Predominant Product;
    ii. Information on product quantity;
    iii. Container Information– Procedural group.
    iv. Documents of previous Transport;
    v. Modal Information.

Regarding the layout of each modal, we present the most relevant portions of a rotational model, which more easily relate to tracing technology below.

k. Vehicle Data:
    i. Vehicle license plate;
    ii. Weight in Kilograms;
    iii. Type of vehicle(traction or towing);
    iv. Seals.

Finally, there is a specific group with transport data for dangerous products, with the following fields:

l. Risk Class and Subclass;
m. UN Number;
n. Packaging group;
o. Proper name for product shipping;
p. Quantity limited by vehicle.

2) Obligatory use

Currently all cargo transport in Brazil must be documented by CT-e. The State of Sao Paulo alone emits over 500 thousand CT-es per day, involving 63 thousand transportation service providers. In all Brazilian territory, there are over 120 thousand service providers.

C. Electronic Invoice Manifest (MDF-e)

This document has the function of relating the electronic invoices emitted by the shipper and the transporter, which are all the NF-es and CT-es. According to Brazilian legislation, the electronic manifest must be emitted by companies providing transport services that offer more than one mode of transportation or for other companies involved in operations with goods, who uses their own vehicles for transportation, either leased or through the hiring of third-party goods transporters, with more than one invoice.

According to the operational model, an XML file should be generated and signed digitally. The particulars of this document are the following:

a. It will contain all the information regarding the goods, driver, itinerary, value and weight of the goods and invoices;
b. As opposed to the NF-e and CT-e, the company that emits the MFD-e must finalize it at the end of the route. While any documents are pending closure, it will be impossible to authorize a new one relating to the same loading and unloading pair of the same vehicle.
c. This guarantees that the information from the beginning and end of the transportation service provided will be recognized for fiscal tax purposes. This also ensures that if, during transport, there is any alteration in the electronic documents information (vehicle, goods, documentation, driver, etc.), this must first be closed, followed by the emission of a new document with the updated configuration.

1) Summary of provided information

The Electronic Invoice Manifest is an XML file composed of the following groups of information. (The entire layout is available in the taxpayer's manual, available at the site https://mdfe-portal.sefaz.rs.gov.br/ [5]):

a. Identification of MDF-e: provides indications which allow for the differentiation of one MDF-e and another, for example, serial numbers. This group has information on services provided, such as emission dates.
b. Identification of Manifest Emitter: Provide registration data and address.
c. Information on fiscal documents linked to the Manifest: Provide invoice information which accompanies transportation and goods. Relevant information:
    i. CT-e Access key;
    ii. NF-e Access Key.
d. Totals for transported goods and tax documents. Relevant Information:

i. Total amount of CT-es related in the Manifest;
ii. Total amount of paper acknowledgment related in the MDF-e;
iii. Total amount of NF-es related in the Manifest;
iv. Total amount of paper invoices related in the MDF-e;
v. Total value of transported goods;
vi. Total brute weight of transported goods.

e. MDF-e seals.
f. Roadway layouts. Relevant Information:
i. Main vehicle license plate – obligatory
ii. Vehicle weight in kilograms– obligatory.

2) Usage obligation

The obligation for MDF-e emission will be imposed on taxpayers according to the following timetable:

a. For businesses which transport goods, interstate transport on parceled goods, with the following starting dates:
i. January 2nd, 2014, for 200 largest roadway contributors, as well as rail and air contributors;
ii. July 1st, 2014, for roadway contributors, which haven't opted for simple tax regimes, and for sea contributors;
iii. October 1st, 2014, for roadway contributors subject to simple tax regimes;
b. For contributors which supply goods via their own transportation, MDF-es must be emitted starting 03/02/2014 (if the simple tax regime was not opted for) or 01/10/2014 (otherwise).

D. Examples of technologies which identify license plates

In the following, we present types of technology which may be used for the identification of license plates and/or vehicles, which can be utilized in conjunction with one another or independently.

Regarding the exchange of information, the use of the XML standard and WebService technology may be the most recommend, if they are amply utilized by models of electronic invoices. This will depend on how the platforms which operate the traffic and security controls work.

1) OCR

OCR (Optical Character Recognition) is a non-obtrusive technology which allows for the capture of license plate images from vehicles via cameras and post-haste recognition of license plate characters.

If the license plate is identified, the information may be utilized to cross reference an electronic document database. When a vehicle with dangerous products is identified, the security system should be informed.

2) RFID Technology in the vehicle

This is a radio frequency technology linked to an object and signal-receiving equipment installed on highways or strategic points for passing vehicles. It is a technology that depends on the installation of a transponder (chip) board, or tracked object (in this case, the vehicle); which makes it an intrusive solution. However, once installed, the gathered information can be made with the vehicle in motion, preventing delays with stops on highways or roads.

The reflected response can contain pre-recorded information on the transponder, such as its ID number, chassis information, cargo or transported invoice. The reach of the antenna depends on the potency installed, allowing for up to hundreds of meters.

3) RFID Technology in the product

This technology involves the installation of radio frequency tags on objects giving them a "fingerprint", readable whenever it passes an antenna with the capability to scan it.

The solution is similar to the previously described item, with the difference that the tag will contain information on the product, fabrication date and producer. If it contains a tag, the product will be detected by passing a gate or portal with antennas that capture the signal. This has the following advantages: (i) the fact that it is a solution which allows, under certain conditions, the identification of the product from a distance, with the necessity of stopping or analyzing the cargo; (ii) the tags are read even if they are piled on top of one another.

4) Trackers

This technology has information obtained by providers of cargo tracking service companies and/or risk management. This solution offers the ability to obtain route information, vehicle stops and corresponding times, providing the ability to check any route deviations. Depending on the accuracy of the solution adopted, it is possible to have more or less number of positions per minute.

Different features can be offered, depending on the service, for example, location features (real time or periodic) and the corresponding transmission for capturing points (both should be used for GPS, satellite, cellular or other form of data communication), opening of control compartments or vehicle doors of the vehicle and even issuing commands to the vehicle.

IV. ENVISIONED ARCHITECTURE

The architecture initially involves the traffic system identifying license plates and cross-referencing this information with DETRAN databases (to identify if it is a cargo vehicle or not) and the Secretary of Treasury (to identify if it has declared dangerous cargo or not). We re-emphasize the initial question on the necessity of juridical research to develop the viability of this cross-reference.

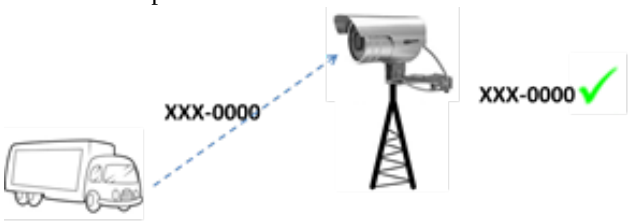
If the presence of dangerous cargo is identified, the security system will be activated through detours and roadblocks.

We will have:

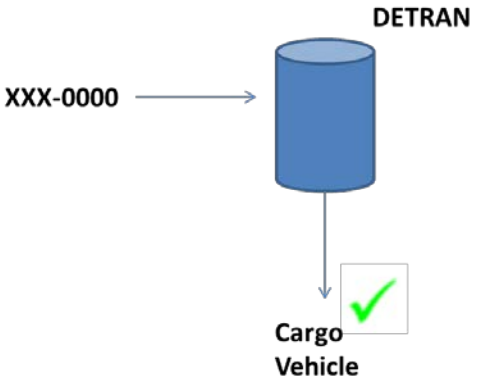
(i) The definition of alternative technology for the identification of vehicle license plates;

(ii) The installation of gates or posts with OCR cameras or antennas in strategic locations which aim to impede vehicle

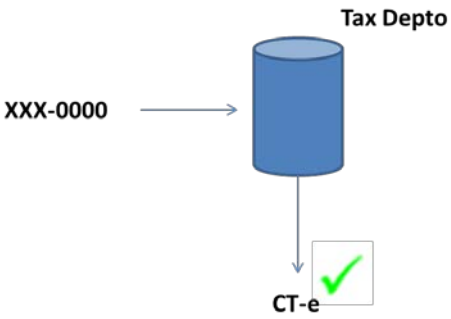
traffic including dangerous cargo, in order to record these vehicles license plates.



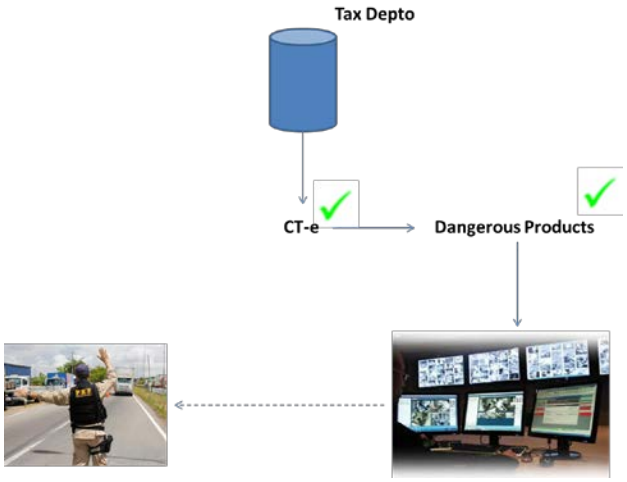
(iii) The information of the identified license plate should be cross-referenced against a database showing types of vehicles (ex: Traffic Department, called, in Brazil, by DETRAN).



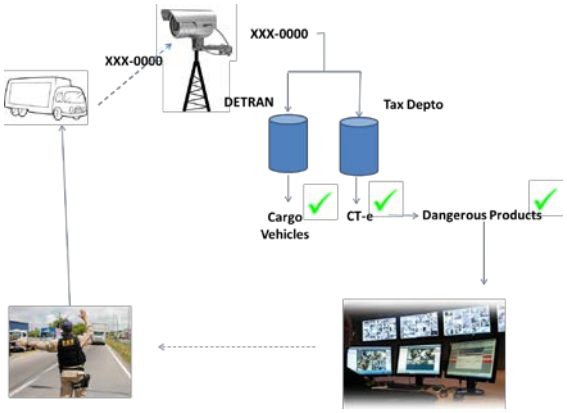
(iv) If it is a cargo vehicle, crosscheck the information with electronic invoices (MDF-e and/or CT-e), seeking to identify a CT-e whose deadline for transport would be condemned by the date upon which the license plate information is captured.



(v) Identify the CT-e, verify if the transport group for Dangerous Products has been filled out. If so, alert a center of operations, which should act to intercept the vehicle, impeding its entrance or transit in areas or hours in which dangerous products should not pass through.



In summary, we can conclude, graphically:



V. EXPECTED RESULTS

The final expected result, in the sense of reaching the initial desired result, hopes to integrate traffic monitoring systems with electronic invoices in order to identify the transit of dangerous products and act in a way which prohibits the transit of these materials in undesired locations (integration with security systems).

There are numerous available types of technology, both for traffic monitoring and for integration among databases. In the case of electronic documents, both the XML standard and WebService technology are amply utilized.

However, the result depends on the extension of the area which wishes to be monitored, from infrastructure to data communication, as well as systems which cross-reference databases.

As a benefit, we have the possibility of reducing, avoiding or controlling the transit of dangerous products in urban areas and the respective risk of accidents.

ISBN: 978-1-61804-318-4

REFERENCES

[1] Brazilian Traffic engineering Company – CET website: http://www.cetsp.com.br/consultas/transporte-de-produtos-perigosos.aspx
[2] Ministry of Environment website: http://www.mma.gov.br/seguranca-quimica/emergencias-ambientais/estatisticas-de-acidentes
[3] Fernandez, M. L. A..; Avaliação da utilização de documentos fiscais eletrônicos na rastreabilidade de cargas. Dissertação de mestrado; Escola politécnica da Universidade de São Paulo, Brasil; 2012.
[4] Marcelo Luiz Alves Fernandez, Maria Lídia R. P. Dias, Eduardo Mario Dias. Electronic tax documents and goods and vehicles track and trace. Proceedings of the 7th International Conference on Economy and Management Transformation (EMT '13).
[5] Electronic Invoice national project website. Available: www.nfe.fazenda.gov.br. Access: 10/02/2015.
[6] Electronic Invoice project website in São Paulo. Available: www.fazenda.sp.gov.br/nfe. Access: 10/02/2015.
[7] Electronic Acknowledgment of Transport national project website. Available: www.cte.fazenda.gov.br. Access: 10/02/2015.
[8] Electronic Invoice Manifest national project website. Available: https://mdfe-portal.sefaz.rs.gov.br. Access: 10/02/2015.
[9] Electronic Acknowledgment of Transport project website in São Paulo. Available: www.fazenda.sp.gov.br/cte. Access: 10/02/2015.

Obtaining automatic surveys about passenger demand in the public transportation through RFID technology

Mauricio L. Ferreira[1]; Eduardo M. Dias[2]

***Abstract* —** This paper explores the application of RFID technology to obtain data for transport network planning, with a focus on buses passengers travel behavior. The informations obtained are essential to support short-term effect actions, bringing efficiency to public transport services, and to allow medium-term planning to the public transport network. Moreover, it represents an important step to reduce the costs and the time associated with traditional research methods, widely used in the transportation sector. The proposal provides that RFID technology devices installed on buses and smart cards used for payment of the tariff will allow automatic collection of data on the passenger landings locations, making available information on the capacity of the buses and dynamic construction of matrices Origin-Destination (OD).

***Keywords*—**data collection**,** matrices Origin-Destination, public transportation, radio frequency identification (RFID), smart cards.

I. Introduction

This paper presents a proposal for the use of RFID (Radio Frequency Identification) to collect data within a wireless network. In addition, allows innovate with information about passengers travel behavior. For that, RFID devices will be installed on buses and on smart cards for the payment of the fare. As the passengers approach the doors of boarding and landing on the bus, the smart cards will be identified by the RFID devices, and the data will be collected.

The innovation possibilities offered by this technology allow measuring, with greater certainty, the use capacity of the buses, while simultaneously identifying locations of origin and destination of the passengers. The analysis of this information is crucial to get the basic knowledge needed to develop adjustment plans (short-term) and evolution at transport network system (medium and long term). Moreover, there is an expectation that the use of RFID can replace manual researchers with advantages, because it decreases costs, and significantly increases the ability to collect and record new facts uninterruptedly with greater spatial coverage.

The effects generated by the population density, economic growth and indiscriminate use of natural resources [6], leaves cities rulers under pressure to build public policies and make investments that aim to improve structures of the cities to maintain the quality of life of its citizens without prejudicing the functioning of their economic activities [43]. Among the priorities, public transport stands out [25], because the cities depend on it to democratize the mobility and the operation of production structures [44], [15] and provide environmental sustainability [22], [39].

The public transport network it therefore important because it has great economic and social role [21], and because it is essential to the quality of life of the big cities population [45]. The greater the public transport service efficiency, the greater the positive effect over the cities and the quality of life of citizens, because it collaborates to reduce congestion, pollution, energy consumption, indiscriminate use of fuels, also optimizing the use of the roads infrastructure [34]. It is central to the economy, as it allows the operation of the productive sector and the conditions for distributed consumer goods and services [8]. On the other hand, inefficient services affect tangible costs, people and cities, as they lead to an increase in fuel consumption, vehicle wear and infrastructure, cost of labor, which, added to other factors, result in higher fares, and directly influence the cost of money (inflation) [1]. In some of these cities, we can observe increased accident rates, traffic deaths and problems related to air pollution [12], [14], [42], and [2].

The importance of transport in daily life is evidenced by the consumption of time, spent in commuting between the place of residence and the activities involved (work, leisure, study, etc.) [40]. Transport also generates intangible costs,

[1]**M. L. FERREIRA** is Project Coordinator of São Paulo Transportes S.A. – SPTrans. Address: R. Boa Vista n.236, São Paulo/SP, Brazil, CEP 01014-000. MSc student of the Polytechnic School, University of São Paulo – USP. Address: Av. Prof. Luciano Gualberto, lane 3, n. 158, São Paulo / SP, Brazil, CEP 05508-970 (mauriciolima7@usp.br).

[2]**E. M. DIAS** is PhD in Electrical Engineering and professor at the Polytechnic School of the University of São Paulo - USP. He is coordinator of GAESI - Electrical Automation Group of Industrial Systems and researcher at the Electrical Department of Energy and Automation. Polytechnic School, University of São Paulo. Address: Av. Prof. Luciano Gualberto, lane 3, n. 158, São Paulo/SP, Brazil, CEP 05508-970 (emdias@pea.usp.br).

particularly those related to increased trip time, longer waiting times and discomforts. Arising from this, still affect people's views regarding shipping as a public service by interfering directly in their daily lives.

Developments in mass transit services are proposed based on public policies that seek the sustainability of cities. The rulers delegate to public agencies the managing of these developments and take on responsibility over the public transport services. On this context, they act in the planning, supervision and remuneration of the activity.

However, despite the efforts of public agencies and private suppliers, user expectation for the quality of transport services always tends to increase the level of demand on them, and always aims faster, safer and more confortable travels [18]. In fact, to attract people continuously to mass transit requires a creative attitude of public and private suppliers, to develop alternatives that are able to compete with other transport modes, mainly during the confrontation of problems that plague cities (traffic, energy consumption, pollution, accidents, etc.).

One of the main challenges that agencies face in creative planning in the transport sector is to form an expanded knowledge base about the problems, in order to obtain the best solution to eliminate them and offer efficient services. This expanded knowledge base for analysis of the information coming from the facts, arising from the data collection, is the prior stage in planning services [41]. However, obtaining more comprehensive data not always brings the quality of information as collateral [4]. On this way, planning public transport systems in many cities is still a technocratic decision, indifferent to their passengers, relegating the user's actual needs to the background. [24]

II. Short-term transport plan difficulties

To accommodate the increased demand of passengers, the agencies actions require adjust schedules, services and supply capacity [16]. Generally, the updates are performed on existing routes, adding or reducing offer [7].

As a finite system, the transport resources have limits on their ability to grow, and currently no longer leave room for increase, causing, ultimately, inefficiency in updating the schedules of services. That situation leads to a degradation on the service quality level, consequently increasing passenger dissatisfaction [28].

In this context, simple adjustments in schedules are ineffective, because the main feature of the transport system is to be interconnected and potentially integrated with other services and modes of transport, resulting in the formation of chained trips (network) [3]. This behavior is different from the discrete standard travel in connectionless service lines with transportation alternatives [16]. The planning and scheduling services to support chained trips cannot consider solely passenger quantity, neither the capacity to set the size of the offer. Rather, one should also consider the passenger displacement profiles, including determining the regions that concentrate origins and destinations of trips [31] and identification of points of articulation of the public transport network (natural local connections).

If the agencies can monitor changes in demand over the routes and adjust their ranges, the quality of service can improve and the public transport can become more attractive. Besides knowing the details of the offer (distances, capacity and time) and availability of infrastructure (streets, terminals, stations, etc.), the transport planner must have complete and systematic knowledge about the characteristics of demands for public services (number of users, origins and destinations of trips, times and places of passenger concentrations, etc.) [28].

Generally, information like this is considered in the modeling of transport plans having long-term horizons, but ignored in the short-term adjustments. However, the short-term adjustments maintain the ability of transport services and help improve public opinion on the city functioning, because it saves passengers time to produce and consume the activities that the city offers.

This is the situation that have been occurring in the city of São Paulo, Brazil. The city has a public transport network (subway, train and bus) that has been consolidated over the years. Public transport by buses is administered by the city through the São Paulo Transportation company - SPTrans, contracted to carry out the management of the 1,300 buses lines in the city. The city have 16 consortia made up of private companies and cooperativesk, responsible for 15 thousand buses operation and transport system services. These buses run 190,000 daily trips totaling 3.7 million km, carrying 9.5 million passengers [9], and [9a].

III. Intelligent Transportation Systems (ITS) Application

The Intelligent Transport Systems (ITS) uses processing and communication, sensing, navigation and technology controls applied to improvements in the management and operation of transport systems, and help to control the use of road infrastructure, security and accessibility. It reduces costs, waiting timeouts of users and negative environmental impacts [17], [32].

In general, the ITS promotes smart connection between users, vehicles and infrastructure [46]. Since 2004, São Paulo city has had benefits aggregated by intelligent transport systems that monitor public transport operations. All buses are equipped with AVL equipment and electronic ticketing (AFC - Automatic Fare Collection). Stops at exclusive lanes were interconnected with terminals by fiber-optic network, and their platforms are equipped with camera systems (CCTV), variable message panels (VMP) and monitors that display information

about the services to users. The Electronic Ticketing System, which records high usage rate of smart cards - called "Bilhete Único" - are used by 94% of the passengers [9].

The application of intelligent transport systems enabled obtaining real-time information, and its use in control and planning systems began to play an important role for public agencies and private companies supply services, allowing rapid and efficient decisions[4a]. Through automated systems, significant amount of data can be collected and treated by high-performance computing systems that help transportation engineers and planners to do their job more efficiently [23].

IV. Obtaining the Data of Passengers Demand on Public Transportation

The characteristics on travel demand can be identified through surveys, particularly those that use direct methods [27]. This type of research involves large amount of data from large samples and are recorded by interviewers (researchers, electronic equipment, or through filled forms provided by respondents), and is performed out through research, whose methods can be: census, household [33], direct approaches [38], or counts [26], widely used in the transport sector.

The agencies usually use direct methods, like field research through users interviews, which requires great efforts of planning, high expenditure of human resources, money and time [10]. For those reasons, traditional research methods often prevent the knowledge of the characteristics and the scope of the dislocations with the ideal frequency. In the city of São Paulo, O / D (Origin and Destination) household surveys are conducted with intervals of 10 years, among nearly 3.5% of the resident population [29]. In the meantime, the tabulated results have been used as reference for transport planners.

Those indirect methods of manual research, mainly the observations, are being replaced by ITS systems. More cities have been adopting electronic tracking systems installed on buses and equipment for the electronic payment [5]. These automatic systems offer advantages for the passengers and the managers, helping lower the cost and improve control, and continuously collecting data, what provides a huge amount of useful information to analyze the population demand.

The identification of locations of sources and destinations of trips and concentrations of passengers helps to solve many planning problems. This information, arranged into Matrices origin-destination (O/D), allows obtaining crucial information for the planning stages of the transport system, as they allow to know how are sought after locations in the city by the bus passengers [19]. However, despite their importance, these information are complex to be obtained [28], entailing high expenditure of resources.

After collecting the data provided by embedded electronic systems on buses and terminals, the information will be used for correct programming and dimensioning of the seats amount on buses. This, however, does not provide detailed characteristics of passenger movements; for this information would be necessary to integrate all data provided by the automatic positioning systems (AVL) with the information provided by transactions between smart cards and electronic validators (SBE). This type of data integration is already a reality in São Paulo - although the buses used in the city have a hall that accommodates multiple seats for the handicapped, the elderly and pregnant women, soon after boarding doors. The door is located on the front of the vehicle. Thus, each passenger needs to cross this space to get to the electronic validator and execute the transaction with the smart card in the equipment, which releases the lock that gives access to the hall named "paid area" on the back of the bus. At this time occurs the identification of the smart card and the coordinates (latitude/longitude), although the passenger maybe has embarked several meters before. Situations where the departure hall of bus is crowded with other passengers, or even when they have the option to stay in this area, are common. This situation produces many distorted information about the local of embarkation.

Current electronic systems, thus, generally offers advantages, mainly the higher coverage data collection, but still do not produce origin and destination reliable information, except by means of estimates and error handling models.

V. An Overview about RFID Technologies

RFID technology has been widely discussed by companies, technical community and scholars and there are extensive amount of material about its development and application [36]. It has the advantage of storing information about what is being identified and transmit this information to compatible requesters via wireless network without the need for physical contact [35], because it allows the TAG (attached in objects, products or people) to be activated and recognized at distance through the issuance of radio frequency waves [11], [30]. RFIDs have integrated circuits for modulating and demodulating radio waves by converting the reflected waves into digital information. It also integrates the technology, processing services and data storage based on miniaturized components in devices [37].

RFID technology was invented in 1948, but it was not mainstreamed for commercial applications until the 1980s. One of its first known applications was during World War II, when it was used by the British radar system to differentiate between German aircraft and their own aircraft with attached radio transponders. It became relevant within the industry sectors, trade, services and government, and has been used in a variety of applications [20] such as: control of the supply

ISBN: 978-1-61804-318-4

chain (logistics), product tracking (schedule control), authentication (quality control), access control (security), anti-theft systems (security), documentation and identification of persons (passports and hospital patients), electronic payment (smart card and smartphone).

VI. Purpose, Methodology and Initial Results

The proposal is to use RFID devices on buses and on smart cards to collect data on passenger trips. The expectation is that the use of this technology will be able to provide detailed information about passenger travel patterns and, from this, promote knowledge for planning services of public transport network [9], [9a].

Tests are being managed in the city of São Paulo with the help of companies that propose to evaluate the technology. The methodology used was organizing processes to check the operation of the technology through observations. The verification processes were divided into five stages:

1. Identify TAG with better performance;
2. Expose as feasibly (or not) the data collection on the conduct of users who keep smart cards (on personal bags, pockets of pants and shirts, in books, wallets, etc.);
3. Infrastructure testing: Program the reader equipment to identify and count shipments and passenger landings, including integration with the AVL systems for geolocation;
4. Integrate antennas (TAG) in similar plastic cards to smart cards;
5. Through data captured by devices RFID, identify the places of origin and destination of passengers boarding on buses and loading (in development).

The first step was to identify the performance of the TAG when activated by the readers. Initially, the proposal for identification of smart cards containing radio frequency technology devices had exposed some doubts about its operation. The main question was about the reach of the electromagnetic waves when reading the TAG, since the environment of buses consists of many metal parts, various physical elements and is occupied by a variable number of passengers.

A reader equipment was installed into a vehicle with two coupled antennas, and the antennas were place in the front hall of the vehicle, near the front door, inside the area where the passengers perform the shipments. Fig. 1 shows the position of the antennas, near passenger boarding door.

Fig. 1 – Antennas positioned near the boarding door.

In this first phase, four sets of TAG were provided, each with different designs and characteristics. Each set consisted of five similar TAG.

Five volunteers participated in the test development and for each one were delivered a tag of the same model. They followed the procedure below:

- Each volunteer should stand ten feet away from the bus;
- They walked to the door of bus;
- Boarded in the buses and passed between the antennas;
- Walked through the validator equipment of the vehicle;
- And finally returned to the start position.

The test examined whether the TAG was enabled and if a unique ID was transmitted to the reader. A notebook connected to the equipment monitored the TAG readings. With a similar procedure, the four sets of TAG were examined. The results are shown in Table I below:

STEP 1 - TESTS IN THE BUS - ENERGIZING TAGS						+30dBm - 2 Antennas Antenna 2 near the wind glass	TAG on volunteer hand
TAG	5 DIGIT FINAL EPC	READINGS (RECORDS)	TIME (SECONDS)	RECORDS/ SECONDS	TOTAL RECORDS	METHOD	OBSERVATION
UCODE7 AD225	22201	9	15	0,600		Enter, go to validator - return	OK
	22202	8	14	0,571			OK
	22203	7	22	0,318	28		OK
	22204	8	18	0,444			OK
	22205	7	12	0,583			OK
UCODE7 AD379	33301	6	9	0,667		Enter, go to validator - return	OK
	33302	4	7	0,571			OK
	33303	5	16	0,313	26		OK
	33304	5	15	0,333			OK
	33305	6	14	0,429			OK
Sirit - RS454	55501	10	16	0,625		Enter, go to validator - return	OK
	55502	9	13	0,692			OK
	55503	15	17	0,882	57		OK
	55504	11	15	0,733			OK
	55505	12	16	0,750			OK
UCODE7 - Preto	77701	14	13	1,077		Enter, go to validator - return	OK
	77702	12	14	0,857			OK
	77703	16	15	1,067	65		OK
	77704	10	9	1,111			OK
	77705	13	14	0,929			OK

Table I: Step 1 - Results of verification of the TAG energizing.

Source: SPtrans.

ISBN: 978-1-61804-318-4

The test showed that all TAG were energized and transmitted data to the reader device. However, the UCODE7 / PROTO TAG (Fig. 2) stood out, since it was observed a higher amount of logged events with it.

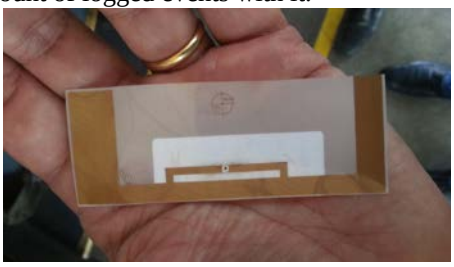


Fig. 2 - Prototype (TAG-UCODE7) used in the experiment.

The second step of the test aimed to verify, among the different models of TAG, which ones had a better performance, measured by the number of readings per second, when the TAG were hidden inside objects or clothing. At this stage, each volunteer was asked to keep the TAG in a specific location of their clothing (trouser, shirt pocket), or inside bags, wallets, books, etc., simulating the passengers standard behavior. So all the volunteers repeated the sequence of the Step 1. The results are show in Table II below:

STEP 2 - TAGS HIDDEN INSIDE DRESS OR OBJECTS (Simulating behaviour of passengers on the bus)							+30dBm - 2 Antennas Antenna 2 near the wind glass	TAG on volunteer hand
TAG	5 DIGIT FINAL EPC	TAG LOCAL IN THE BODY	READINGS (RECORDS)	TIME (SECONDS)	RECORDS/ SECONDS	TOTAL RECORDS	METHOD	OBSERVATION
UCODE7 A0235	22201	Shirt Pocket	9	15	0,600		Enter, go to validator - return	
	22202	Left front trousers pocket	6	15	0,400			
	22203	Inside wallet in the bag, cell phone in front of	6	22	0,273	28		
	22204	Behind the Badge, shirt pocket, cell	2	15	0,133			
	22205	Left back trousers pocket	5	13	0,385			from antenna 1
UCODE7 A0370	33301	Shirt Pocket	4	2	2,000		Enter, go to validator - return	Only 1 reading from antenna
	33302	Left front trousers pocket	2	1	2,000			Only getting in the bus
	33303	Inside wallet in the bag, cell phone in front of	2	15	0,133	14		
	33304	Behind the Badge, shirt pocket, cell	No records at all					
	33305	Left back trousers pocket	6	12	0,500			Only 1 reading getting in the bus
Sint - IS1654	55501	Shirt Pocket	8	16	0,500		Enter, go to validator - return	
	55502	Left front trousers pocket	7	13	0,538			
	55503	Inside wallet in the bag, cell phone in front of	14	16	0,875	39		
	55504	Behind the Badge, shirt pocket, cell	7	15	0,467			
	55505	Left back trousers pocket	3	11	0,273			
UCODE7 - Proto	77701	Shirt Pocket	11	15	0,733		Enter, go to validator - return	
	77702	Left front trousers pocket	8	13	0,615			
	77703	Inside wallet in the bag, cell phone in front of	14	16	0,875	47		Read 5 m from door out the bus
	77704	Behind the Badge, shirt pocket, cell	5	3	1,667			Read only getting in the bus
	77705	Left back trousers pocket	9	13	0,692			when leaving the bus

Table II: Step 2 - Verification of the TAG reading performance when hidden in clothing. Source: SPtrans.

The results, although satisfactory, showed that there were variations on the data caption by the reader, depending on the exposure of the TAG over the wave Radio Frequency emitted by the antennas.

Step 3 consisted in an environment test with firmware programming reader equipment. This step was important to evaluate some response characteristics of the reader equipment with the necessary requirements for this assessment.

An equipment supplier, partner in the technology development, prepared a reader with four antennas, programmed with the following requirements:

- Data selection: only first and last event records of a card are stored.
- Each logged event is associated with an exposure of longitude / latitude, showing also the time and place of the event.
- The collect of data only works while the vehicle is with the doors open.

In the data selection process, the reader storage is running only to first and last event collected from the same card. The criteria reader programming requires storage only of the initial event data collected by the antennas positioned at the boarding door, as well as the last data, by antennas positioned next the buse exit door. In this process, all intermediate events were discarded (i.e, were not stored).

For each event record, the reading equipment asks the AVL the current position provided by the GPS and associates this information with the date and the time of the event. To reduce the amount of collected events, the reader device can collect data only when at least one bus door remains opened.

To control the events, a computer program was developed with a user interface designed specifically to allow the researcher to control the operating parameters of the reader and the antennas, like opening and closing of the doors of the vehicle and the change of geographic location. In the visualizer interface are shown the new records per port (shipments); old records per port (arrivals) and the number of passengers still inside the vehicle. (Fig. 3).

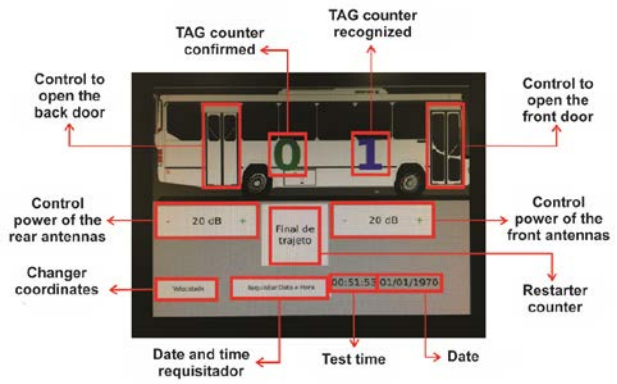


Fig. 3: Program of control of devices RFID - visualizer's interface.

The Step 4 of the experiment discussed the coexistence of radio frequency (RF) technology (that is currently used on the smart card) with RFID technology. For the development of

this stage of the experiment, the SPTrans company invited some smart card providers companies for testing the incorporation of the TAG-RFID into the smart card, so that it could be recognized by reading devices within the bus. This recognition between the reader and the smart card should enable data collect of a sequential number. In the proposal, the microchip into the TAG device (N-Bits transponder system, Read Only), received in his memory a unique sequence number that was associated with the unique number of the smart card (UID - User ID) allowing the TAG, at the time of activation, storage the number as information to the reader device (reader). Thus, all captured records would be equivalent only to the unique card recognized. The proposal is shown in Fig. 4, below:

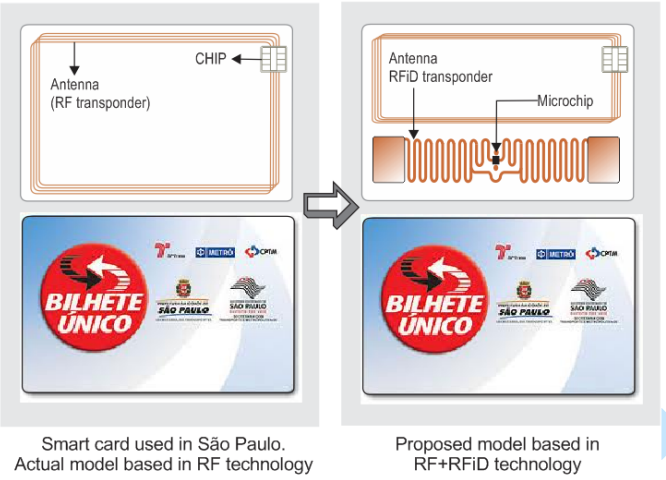


Fig. 4 - Smart card with RF technology, and example of smart card with RFID technology TAG applied.

Thirty plastic cards were produced for use in testing (Fig. 5). In all cards, in addition to the smart card technology, was also incorporated the TAG UCODE7 PROTO-type, enabled in the initial tests.

Fig. 5 – Cards with similar commercial technology used in the smart cards of São Paulo, with RFID technology incorporated devices, used in Step 4.

The fifth and final stage of the tests was predicted to obtain the locations where the RFID reader had done recognitions of TAG.

The ultimate test to find the local of shipments and passenger landings using RFID will be made soon by SPTrans. Previews tests were made in the laboratory, thanks to the possibility to simulate the ability to obtain location coordinates and operate opening and closing doors in the computer program that manages the equipment operation. In those tests, the reader equipment and two antennas were interconnected by wire to an AVL equipment mobile. Changes in localization of smart card simultaneously with the AVL equipment were recorded during the interrogations of the card while they remained at the exposure field of the antennas. Records were stored in a log file and showed the readings of TAG RFID and the position changes of the AVL equipment.

The results obtained were satisfactory; however, the SPTrans company determined that the results should be disclosed only after the completion of the final test in a bus on operation.

VII. Conclusion

This article tried to start a discussion on the possibilities of obtaining important information with the use of radio frequency technology - RFID - in the urban transportation passenger sector. It opens the prospect for future discussions on the portability of the technology in smart card ordinarily used for payment of the transport services by passengers.

The data collected by technology devices may result in crucial information for cities, through studies on passenger demand, and collaborate with the operation of its transport systems update with more precise adjustments to schedules of buses, increasing the services quality and evolving its conditions for the wellbeing of the users.

The RFID technology in smart card may allow the obtaining of information that is currently unavailable, except with great effort. That new information shall grant efficient actions in the role of supply management of passenger demand with the following benefits:

- Information about loading of passengers on the bus, between stopping points, to determine the maximum loading section. The section of maximum loading is critical in determining services supply;
- Provide information to the user on the vehicle stocking condition, as it allows to determine the capacity use in the vehicle, and, based on the results, tell users if the bus is empty or full. With this information, the user will be able to assess whether it is or not convenient to catch the approaching bus, or to wait for the next vehicle;
- Intervene and make changes online, through direct actions over the vehicle driver, in order to make it

compatible with scheduled forecasting, controlling delays and advances. The control centers can provide extra vehicles in the case of delays identification, or of concentrated passenger demand at certain points, providing better services;

- Plan the supply of places (vehicle capacity and frequency) due to the persistence of information indicating change in the use of the services profile;
- Plan the new connections, obtained by the matrices of origin and destination (O / D) depicting the movements of passengers' journey flows and connection locations of trips where passengers make the integration between services.

The possibility of using this technology in the public transport sector is based on vehicle control systems already running and in operation in several cities in the world, as well as the popularization of smart cards for paying the fare. Smart cards already commonly utilized by the users of public transport systems appear as excellent information providers.

Add radio frequency identification components on these cards do not change the current use as electronic payment of the fare method, and does not add any providence of users in its daily maintenance. However, its use allows data capture to provide indicators about time and places where users move on public transport by buses.

Among the advantages arising from the use of technology in the transport sector, especially in the activities of service management, are:

- Improve transport service for people's needs;
- Develop urban mobility plans according to the growth and functioning of cities;
- Provide important information to assist users in deciding how and when to use the public transport services.
- Reduce the cost of managing saving human resources, time and money, eliminating or reducing the development and application of manual searches;
- Making proactive management functions in the control centers on the operation of public service;
- Provide updated information to support the modeling of the transport system in the medium and long-term planning.

The use of RFID technology data related to the provision of services in the public transport system can become effective in the next years with several benefits.

The first results are promising, although this research is still in a very early stage. The initial experiments did produce positive results, showing the possibility of efficient use of RFID technology to obtain data on locations of shipments and passenger landings as well as getting the number of passengers into buses online.

REFERENCES

[1] ANTP - Redução das deseconomias urbanas com a melhoria do transporte público. Instituto de Pesquisa Econômica Aplicada (in Portuguese version) - IPEA, Associação Nacional dos Transportes Públicos - ANTP, Journal of Public Transport, year 21, 1st quarter 1999. Available: http://www.antp.org.br/_5dotSystem/download/dcmDocument/2013/01/10/057A84C9-76D1-4BEC-9837-7E0B0AEAF5CE.pdf - Last access: 10/01/2015.

[2] BATAGAN, L. - Smart Cities and Sustainability Models - Informática Econômica vol. 15, N. 15. 80 p. - March 2011. Available: http://revistaie.ase.ro/content/59/07%20-%20Batagan.pdf - Last access: 11/02/2015.

[3] BHAT, C.R.; KOPPELMAN, F.S. - Activity-based travel demand analysis: History results and future directions. In: TRANSPORTATION RESEARCH BOARD ANNUAL MEETING, 79th, Washington Proceedings. Washington: Transportation Research Board, 2000. Available: http://www.ce.utexas.edu/prof/bhat/ABSTRACTS/TSHANDBK.pdf - Last accessed 13/09/2014.

[4] BORGHETTI, L.C. - Guia de procedimentos para o desenvolvimento institucional e organizacional do sistema local de transportes urbanos, SLTU (in Portuguese version). Brasil, Brasília, DF: Empresa Brasileira de Transportes Urbanos – EBTU, 1987.

[4a] BRETTAS, L.E.S.; MELO V.A.Z.C.; DIAS, E.M. - Automation of the control of street furniture using mobile technologies, (imprinting).

[5] CUI, A. - Bus Passenger Origin-Destination Matrix Estimation Using Automated Data Collection Systems. Massachusetts Institute of Technology, p. 134, 2006.

[6] DEMOGRAPHIA - Demographia World Urban Areas - 10th Annual Edition - Revised at May 2014, 20 p. – Available: http://www.demographia.com/db-worldua.pdf – Last accessed 07/03/2014.

[7] ETTEMA, D. - Activity-based travel demand modeling. 280 f. These - Technische Universiteit Eindhoven, Eindhoven. The Netherlands, 1996. Available: http://alexandria.tue.nl/extra3/proefschrift/PRF13B/9604751.pdf - Last accessed 13/04/2014.

[8] FERRAZ, A.C.P. - A qualidade do serviço de transporte coletivo em cidades médias sob a ótica dos usuários (in Portuguese version). In: Encontro Nacional da ANPET, 2, São Paulo, 1988.

[9] FERREIRA, M.L.; MARTE, C.L.; MEDEIROS, J.E.L.; SAKURAI, C.A.; FONTANA, C.F. - RFID for Real Time Passenger Monitoring. Recent Researches in Telecommunications, Informatics, Electronics and Signal Processing. 12th International Conference on Telecommunications and Informatics (TELE-INFO '13) p. 170-175, Baltimore, MD, USA, September 2013. Available: http://www.wseas.us/e-library/conferences/2013/Baltimore/TESIMI/TESIMI-22.pdf – Last accessed 07/01/2014.

[9a] FERREIRA, M.L. et al - Real time monitoring of public transit passenger flows through Radio Frequency Identification - RFID technology embedded in fare smart cards - Latest Trends on Systems - Vol. II, page 599-605, Santorini, 2014. Available: - http://www.europment.org/library/2014/santorini/bypaper/SYSTEMS/SYSTEMS2-40.pdf - Last accessed 07/01/2014.

[10] FERREIRA, E.A. - Um método de utilização de dados de pesquisa embarque / desembarque na calibração de modelos de distribuição do tipo gravitacional (in Portuguese version). School of Engineering São Carlos, Universidade de São Paulo. São Carlos, p. 110, 1999.

[11] FINKENZELLER, K. - RFID Handbook: Fundamentals and Applications in Contactless Smart cards and Identification. ISBN 0-470-84402-7. John Wiley & Sons, 2003.

[12] FRIEDMANN, J. - The World City Hypothesis - Development and Change, Volume 17, Issue 1, January 1986 – Article first published online: 22 Oct 2008 - Available: http://onlinelibrary.wiley.com/doi/10.1111/j.1467-7660.1986.tb00231.x/abstract - Last accessed 09/03/2014.

[13] GIFFINGER, R; GUDRUN, H. - Smart Cities Ranking an Effective Instrument for the Positioning of Cities? Architecture, City and Environment – ACE, V. 12 - year IV, N. 12, 2010. Available: http://upcommons.upc.edu/revistes/bitstream/2099/8550/7/ACE_12_SA_10.pdf - Last accessed 15/03/2014.
[14] GUPTA, J. - Global Sustainable Development Governance: Institutional Challenges from a Theoretical Perspective - International Environmental Agreements: Politics, Law and Economics 2, nº 4. Kluwer Academic Publishers, 2002.
[15] HANSEN, W.G. - "How Accessibility Shapes Land Use," Journal of the American Institute of Planners, Vol. 35, Nº 2, 1959.
[16] JONES, P.; KOPPELMAN, F.; ORFUEIL, J.P. - Activity Analysis: state-of-the-art and future directions. In JONES, P. (Ed), Developments in Dynamic and Activity-Based Approaches to Travel Analysis, Aldershot: Gower Publishing, 1990.
[17] KANNINEN, B.J. - Intelligent Transportation Systems: an Economic and Environmental Policy Assessment. Transportation Research, London, v. 30A, n.1, p. a-10, 1996.
[18] KAWAMOTO, E. - Um novo enfoque do processo de escolha em transporte com tratamento baseado na psicofísica multidimensional (in Portuguese version). School of Engineering São Carlos, EESC, São Carlos, 1987.
[19] KAWAMOTO, E. - Análise de sistemas de transportes (in Portuguese version). School of Engineering of São Carlos: EESC – USP, 1994. Available: http://s6352.minhateca.com.br/File.aspx?e=at3m7z8kZiwQsQuR5QuB-yMmySh5riOqcraQ4LpvqrxYmPSjos9qLj36dpdzCY_uaNrit3Mo9mLNbYVhyrmK6OxjrW273LU5rN-HVIR3vQmTAYs985iwtElAO9x-SfDYU0xKruuHMrbDI6o4-Akkn2Ya7NOcfbVFiXuJRLOtgSZvj3B2okx1hErwPj4aKwmAl7JfADJsw6OF0cNYvzfgQQ&pv=1 - Last accessed 11/04/2014.
[20] LAHIRI, S. - RFID Sourcebook. IBM Press, 2006 - ISBN 0131851373, 97801318051375, 276 pages.
[21] LITMAN, T. - Social Inclusion as a Transport Planning Issue in Canada. Victoria, BC, Victoria Transport Policy Institute, 2003.
[22] LUCAS, Et Al - Transport, the environment and social exclusion - Joseph Rowntree Foundation, 2001.
[23] MACLEAN, S.D. and DAILEY D. J., "Wireless Internet Access to Real-Time Transit Information", Transportation Research Record 1791, 2002, pp. 92-98.
[24] MARTINS, E.R.C.; ARAGAO, J.J.G.; MIAZAKI, E.S. - Segmentação do mercado de transportes urbanos de passageiros: uma abordagem pela análise de agrupamentos (in Portuguese version). In: Congress Research and Training in Transportation, XI, Rio de Janeiro. Anais. Rio de Janeiro: ANPET, 1997, p. 887-896.
[25] MINISTÉRIO DAS CIDADES - Mobilidade Urbana e desenvolvimento urbano. First edition. November. Pólis Institute, 2005. Available: http://www.polis.org.br/uploads/922/922.pdf - Last accessed 07/07/2014.
[26] NAVICK, D.S.; FURTH, P.G. - Distance based model for estimating a bus route Origin-Destination matrix. Transportation Research Record, n. 1433, 1993.
[27] NGUYEN - Estimating origin-destination matrices from observed flows. Proceedings of the AIRO Conference. Guida, Naples, 1984.
[28] ORTÚZAR, J.D.; WILLUMSEN, L.G. - Modeling Transport, fourth Edition, John Wiley & Sons, Chichester, England, Ltd., and ISBN 978-0-470-76039-0, 2011.
[29] PESQUISA OD 2007 - Pesquisa Origem e Destino 2007 - Região Metropolitana de São Paulo - Síntese das Informações, SMT (in Portuguese version) - Secretaria dos Transportes Metropolitanos – December, 2008.
[30] POLNIAK, S. - The RFID Case Study Book - RFID application stories from around the globe – Abhisam, 2007.
[31] RAKHA, H.; PARAMAHAMSAN, H.; VAN AERDE, M. - Comparison of static maximum likelihood origin-destination formulations, 2001.
[32] RIBEIRO, J.L. D.; MOTA, E.V.O. - Desdobramento da Qualidade: modelos para serviços e para a manufatura (in Portuguese version). Porto Alegre: PPGEP, EE/UFRGS, 1996. (Technical specifications nº 5).
[33] RICHARDSON, A.J.; AMPT, E.S.; MEYBURG, A.H. - Survey methods for transport planning. Melbourne, Eucalyptus Press, 1995.
[34] RODRIGUES, M.O. - Avaliação do transporte público na cidade de São Carlos (in Portuguese version). School Engineering of São Carlos - EESC, São Carlos, 2006.
[35] SANGREMAN A. - CAMANHO T. RFID. 2010. Available: http://www.gta.ufrj.br/grad/_07_1/rfid/RFID_arquivos/_o_que_e.htm - Last accessed 26/04/2014.
[36] SANTANA, S.R.M. - RFID – Identificação Por Radiofrequência (in Portuguese version). Santos: FAAP, 2011. Available: http://www.wirelessbrasil.org - Last accessed 23/04/2014.
[37] SHAHRAM, M.; MANISH, B., (2005). RFID Field Guide: Deploying Radio Frequency Identification Systems. [S.l.]: Prentice Hall PTR, 2005.
[38] SHESKIN, I.M.; STOPHER, P.R. - Surveillance and monitoring of a bus system. Transportation Research Record 862, p. 9-15. Apud RICHARDSON, A.J.; AMPT, E.S, 1982.
[39] STEEMERS, K. - Energy and the city: density, buildings and transport - Energy and Buildings 35, Elsevier 3-14, 2003.
[40] TAGORE, M.R.; SIKDAR, P.K. - A new accessibility measure accounting mobility parameters. Paper presented at seventh World Conference on Transport Research. University of New South Wales. Sydney, Australia, 1995.
[41] TEIXEIRA, G.L. - Uso de dados censitários para identificação de zonas homogêneas para planejamento de transportes utilizando estatística espacial (in Portuguese version), 169 pages. Brasilia University – Brazil, Brasília, 2003.
[42] TOPPETA, D. - The Smart City vision: How Innovation and ICT can build smart, "live able", sustainable cities - Think Report 005/2010 - Available: http://www.thinkinnovation.org/en/portfol/the-smart-city-vision-how-innovation-and-ict-can-build-smart-liveable-sustainable-cities-2-2/ - Last accessed 11/03/2014.
[43] UITP Europe - Investing in public transport infrastructure as part of the EU Package for jobs, growth and investments. Advancing Public Transportation. - Position Paper of the international association of public transport, December 2014.
[44] VASCONCELLOS, E.A. - Transporte urbano, espaço e equidade (in Portuguese version) - Análise das políticas públicas - 3ed. Editora Annablume, 2001. Available: http://books.google.com.br/books?id=fp7HJrZZ_qMC&printsec=frontcover&hl=pt-BR&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false - Last accessed 25/03/2014.
[45] VASCONCELLOS, E.A. - Transporte e meio ambiente: conceitos e informações para análise de impactos (in Portuguese version). São Paulo: Author's edition, 2006.
[46] WASHBURN, D.; SINDHU U. - Helping CIOs Understand "Smart City" Initiatives - defining the smart city, its drivers, and the role of the CIO - Cambridge, MA: Forrester Research, Inc. 2010. Available: https://www.google.com.br/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CC0QFjAA&url=http%3A%2F%2Fc3328005.r5.cf0.rackcdn.com%2F73efa931-0fac4e28-ae778e58ebf74aa6.pdf&ei=2bxiU8i2G_KgyAGGw4DoDQ&usg=AFQjCNGVoR0aW1cvaBbSXCdaRUuWDx4T8Q&sig2=odFKd8320o84VktZmxt0Rw&bvm=bv.65788261,d.aWc&cad=rjt - Last accessed 19/03/2014.
[47] WILLUMSEN, L.G. - Simplified transport models based on traffic counts. Transport n.10, 1981.
[48] ZHAO, J. The Planning and Analysis Implications of Automated Data Collection System: Rail Transit OD Matrix Inference and Path Choice Modeling Examples. Massachusetts Institute.

Study and Implementation of Routing Protocol for Data Gathering in WSN

Madhumathy P[1], Sivakumar D[2]
[1]Research Scholar, Anna University, Chennai
[2]Professor, CSE Department, Easwari Engineering College, Chennai

Abstract – *In WSN, most existing mechanism around data gathering and optimal path selection result in collision. Collision further increases the possibility of packet drop. So the need is to eliminate collision during data aggregation. This research is an effort to come up with a reliable and an energy efficient Wireless Sensor Network (WSN) routing protocol. To find the rendezvous point for optimal transmission of data a "Splitting tree" technique is employed in tree-shaped network topology and then to determine all the subsequent position of a sink the " Biased Random Walk" model is used. In case of an event, the sink gathers the data from all the source, when they are in the sensing range of rendezvous point. Otherwise relay node is selected from its neighbor to transfer packets from the rendezvous point to sink. The proposed routing protocol simulation results proves there is significant improvement in preventing data collision and increase in the network lifetime compared with other routing protocol*

Keywords*: Relay node selection, Rendezvous point and Data aggregation, Mobile Sink*

1.INTRODUCTION

Wireless Sensor Network is a self-organized, distributed, sensing and data propagation network formed by a large number of sensor nodes. Nodes are resource constrained tiny autonomous devices. They are used to sense the environmental conditions in their immediate surroundings, process the data and communicate the processed data to the base station. Sensor nodes generate data and transmit the gathered data to a distant base station (BS) [1].WSN can be used to monitor environment, surveillance of property and collecting data of massive fields at low cost and with less manpower. Vehicles, animals or people moving around large geographic areas are attached to the sensors with robotic elements and the data is exchanged between individual sensors and infrastructure nodes to drive applications like traffic, wild life monitoring, smart homes and pollution control [2].

1.1 Current challenges in WSN

Data collection from sensors is the key issue in WSN[3]

- Reliability and robustness of transferring data is another significant challenge [3]
- With limited battery power, sensors are expected to sense for very long time hence, energy efficient data collection arises as one of the critical issues in WSN.
- WSN have limited processing and communication capabilities.
- In real applications, the deterministic lifetime of sensor node is still an issue.

1.2 Data gathering using mobile sink

A mobile sink is used for data collection from energy constrained sensor fields. It brings the sink closer to the sensors and conserves precious sensor node energy. The effectiveness of it can be determined by the total sensor energy conserved and the time consumed in gathering the sensor data from the field or from the trajectory length implemented by the mobile sinks [4]. Mobile sink in WSN optimizes the energy consumption and reduces the delay observed during data gathering. It can also reduce possibility of "routing hotspots" caused by fixed sinks due to the nearby heavy data flow. As the lifetime of the battery-operated sensor node is limited, mobile sink is deployed in a robot, vehicle or portable device to selectively activate only the sensor nodes interesting to the sink and deactivate the other nodes. This can considerably extend the lifetime of the sensor nodes to reduce unnecessary power consumption. Mobile sink technique involves controlled movement of sink towards nodes with higher energy for even distribution of energy in the entire network to avoid network partitioning [5-6].

1.3 Routing and issues on mobile sink data gathering

The sink possesses significant and easily replenishable energy reserves, it should move closer to the subset of sensor devices to collect the recorded data. The energy consumption during this process is very minimal. The sink should be within a sensor range for single hop communication and remain within the transmitter range for successful communication. This problem becomes severe while having a high density of sensors in an area. This results in inadequacy in network communication time to upload the data of nodes to the sink and if the sink moves out of transmitter range, node has to wait till the sink returns back. As a result, high delivery delay occurs. A network with a few mobile sinks calculates the gradients using proactive approach which is costly in terms of energy [7]. Mobile sink brings new challenges to densely deployed and large WSN [8].

2. LITERATURE SURVEY

.

Jae-Wan Kim et al [9] proposed IAR, an Intelligent Agent-based Routing protocol for providing efficient data delivery to mobile sink. The performance of the IAR protocol is evaluated using mathematical analysis. Results proved that the scheme effectively supports sink mobility with low overhead and an improvement over the triangular routing problem. However, retransmission will occur four times, if collision occurs and this action fails. Packet loss occurs due to the link failure between the sink and its immediate relay in this scenario.

Luo et al[10] have highlighted the difference between a mobile relay and a mobile sink. A Mobile relay collects data whenever it is closer to the sensor nodes. It transports the data to the sink through mechanical movement. As it does not use the wireless links for transmission to the sink, the latency in delivering data is significant. In contrast, a mobile sink performs various operations like distributing the load among the sensor nodes, collecting data continuously from the sensor nodes, and moving slowly and discontinuously in the data collection process.

Bi et al [11] have considered the mobile sink as moving strategy based on residual energy of the static nodes, which is used to balance the network workload and thereby prolong the life of the network.

Cheng et al [12] have proposed a query-based data collection in which mobile sink issues queries in the specific area while moving through the sensing field and the corresponding response is received through multi hop communication. The problem with such query based systems is that the mobility of the sink causes the query and response packets to take different routes. Cheng et al (2009) analyzed the prevailing query based protocols and proposed an efficient Query-Based Data Collection Scheme which consumes lesser energy and delivered packet with minimum latency. Moreover, QBDCS chooses the optimal time to send the query packet and tailored the routing mechanism for partial sensor nodes forwarding packets. The performance of the QBDCS was evaluated by comparing with a "Naïve" scheme using the simulation tool OMNeT++.

Lei & Kwon [13] propose RECPE a reliable collection protocol for aggregating data packets from the sensor nodes to the sink in a large-scale wireless sensor network. RECPE has successfully covered all the routes in the network by employing expected transmission count over forward links (ETF) method to construct a one-way collection tree, thereby reducing the effect of asymmetric link in the network. Moreover, the proposed protocol also utilized Trickle algorithm and pipeline mechanism to reduce the control information and improve the efficiency of data delivery.

3. THE PROPOSED SOLUTION

3.1 Problem Formulation

In paper [14] Mobile Sink based Reliable and Energy Efficient Data Gathering technique (MSREEDG) was proposed for data gathering in tree based network topology in WSN and compared with Biased sink mobility with adaptive stop times for low latency data collection (BSMASD) .The sink's next moving position is determined by using a Biased Random Walk model. The optimal path for data transmission estimated by rendezvous point selection method and splitting tree process. When the sensor senses the data and when it is ready for transmission, the data are encoded and transmitted to the sink. The sink nodes receive the data which is encoded from the sensors, and then it decodes the data and the resulting message is stored in local buffer. After decoding all the blocks, the original data bundle is reconstructed by the mobile sink. The increase in the packet loss can be prevented by increasing the pause time of the sink.

This process can be enhanced for multiple number of sinks and designing an efficient routing protocol for data gathering . In this paper, a relay node based routing protocol for mobile sink for data gathering for WSN is proposed.

3.2 Proposed architecture

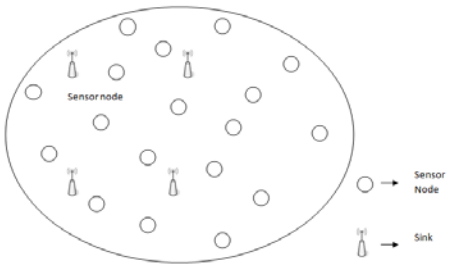


Figure1. Proposed architecture of Multiple Mobile Sink

It is assumed that the sensor nodes as well as the sink deployed in the network are aware of their own location. It is also assumed that multiple sinks move around the sensor field and the number of sinks may vary over the time. The proposed architecture is shown in Figure 1.

3.2 Data gathering routing protocol

Let S and D be the source and sink node
Let QP be the query packet
Let RePmr be rendezvous point
Let RNi be the relay node
Let RPn be the new relay path
Let RPo be the old relay path
Let RP_seq be the sequence number of relay path
Let RP_mes and R_CLR are the relay path setup and clear message with RP_seq

(1) If event occurs, initially, a rendezvous point (RePmr) is selected.
(2) Then sink has to transmit QP to RePmr once the event occurs.

The fields in the QP are shown in table-1

Table -1 Format of Query Packet

Sink ID	Hop Count	Transmitter Distance

(3) RePmr broadcasts the QP with hop count counter value as zero.
(4) When the neighboring nodes receives the QP, it rebroadcasts the packet by incrementing hop count counter as 1. Thus, as the query propagates, each sensor node Ni estimates the next new hop node towards RePmr that are in one hop communication distance.
(5) If next new hop node > 2,
Then
two nodes compares the packets arrival time (Tpa)
The next hop node with earlier Tpa is chosen
End if

Source transmits the path request message (P_REQ) to its neighbors and the neighbor node that sends the (P_REP) is chosen as next new hop node.

If D is moving in the radio range of ReP_{mr}

Then

D receives the data directly from the ReP_{mr}.

Else

D chooses RN_i from its neighbor nodes to transmit the data from ReP_{mr} to sink. (Relay node selection)

End if

Here, the node, which is nearest to D is selected as RN.

(6) If an event occurs, Ni enclosing it collectively processes the signal. One among the nodes becomes S to generate the data reports
(7) When S matches the data sent by QP, the data is forwarded to one hop distance node.
(8) If the next new hop node is failed or its battery is exhausted
Then

$$S \xrightarrow{P\_REQ} Neigh_i$$

$$Neigh_i \xrightarrow{P\_REP} S$$

S chooses the respective $Neigh_i$ as next new hop node.

End if

3.3 Relay node selection

When D does not receive the data packets for the pre-defined time interval T, it suspects that they are out of the coverage radio range. In order to prevent this action, the following steps are executed.

- ReP_{mr} transmits atleast one packet at interval of T/n period. (n is integer number which consider channel loss).
- If ReP_{mr} has no data to transmit in T/n duration, it transmits NULL packet . Thus, when D does not receive the data packet for T, it performs the following actions to select the relay node.

1) relay node request message (R_REQ) is sent to its $Neigh_i$.

$$D \xrightarrow{R\_REQ} Neigh_i$$

2) $Neigh_i$ node will reply to the sink.

$$Neigh_i \xrightarrow{R\_REP} D$$

3) sink chooses the node which is nearer to it as immediate relay node.
4) The relay path message is sent through IRN.

$$D \xrightarrow{RP\_mes} IRN_i \xrightarrow{RP\_mes} ReP_{mr}$$

Note: If D's speed is rapid, then the distance among ReP_{mr} and IRN is more. In this case, the RN_i count is increased.
5) When ReP_{mr} receives the RP_mes, data packets are transmitted in the path traversed by RP_mes. This relay path is flagged with RP_seq.
Note: When D moves away of radio range or after completing its relay path set up, there may be possibility of packet drop. This is prevented by ReP_{mr} by caching the overheard packets which are transmitted to D. The cached packets are routed to D when the ReP_{mr} receives RP_mes.
6) If D is again away from its transmission range of IRN_i, then it selects new IRN_i (as per step 3).
7) D then transmits RP_mes to the ReP_{mr} through the newly selected IRN_i in separate relay path RP_n which is flagged with RP_n_SEQ.
a) When ReP_{mr} receives relay path setup message,
If an RP_o exists for the same sink
Then

$$ReP_{mr} \xrightarrow{R\_CLR} RP_o$$

End if
If there is an old relay path for the same sink, then the ReP_{mr} transmits R_CLR message along RP_o which is flagged with RP_o_SEQ.
b)If a relay path receives a new RP_mes, then it does not remove the RP_o state. RP_o state is maintained until the path receives R_CLR for RP_o. Then a relay path is set up in the reverse path of RP_mes.

3.4 Relay Node *based Routing Protocol for Multiple Mobile Sink*

If an event occurs, the sink comes into contact with the RP and starts collecting the data. When the sink moves out of its coverage range of the RP, it selects a relay node among its neighboring nodes to transfer packets from the RP to sink. When the sink does not receive any packet for time *T*, the sink broadcasts the relay request message to all its neighbors. The neighbor nodes reply to the R_REQ by a relay message (R_REP). A R_REP includes the coordinates of the sending node. The closest node from sink is selected as the relay node and it is named immediate relay nodes . The sink transmits relay path message to RP via the selected IRN. Figure 2 shows the flow chart of the proposed

routing protocol. As soon as the RP receives the RP_mes, data packets are routed along the reverse path of the RP_mes called as relay path. Some packets may be dropped during the time between the movement of the sink out of the RP radio range and the time when the relay path is completely setup. To prevent packet loss in this interval, the RP caches the packets overheard transmitted to the sink for last *T*. These cached packets are routed to the sink when the RP receives the RP_mes. This protocol is used for increase in number of sinks. By increasing number of mobile sinks low latency can be achieved

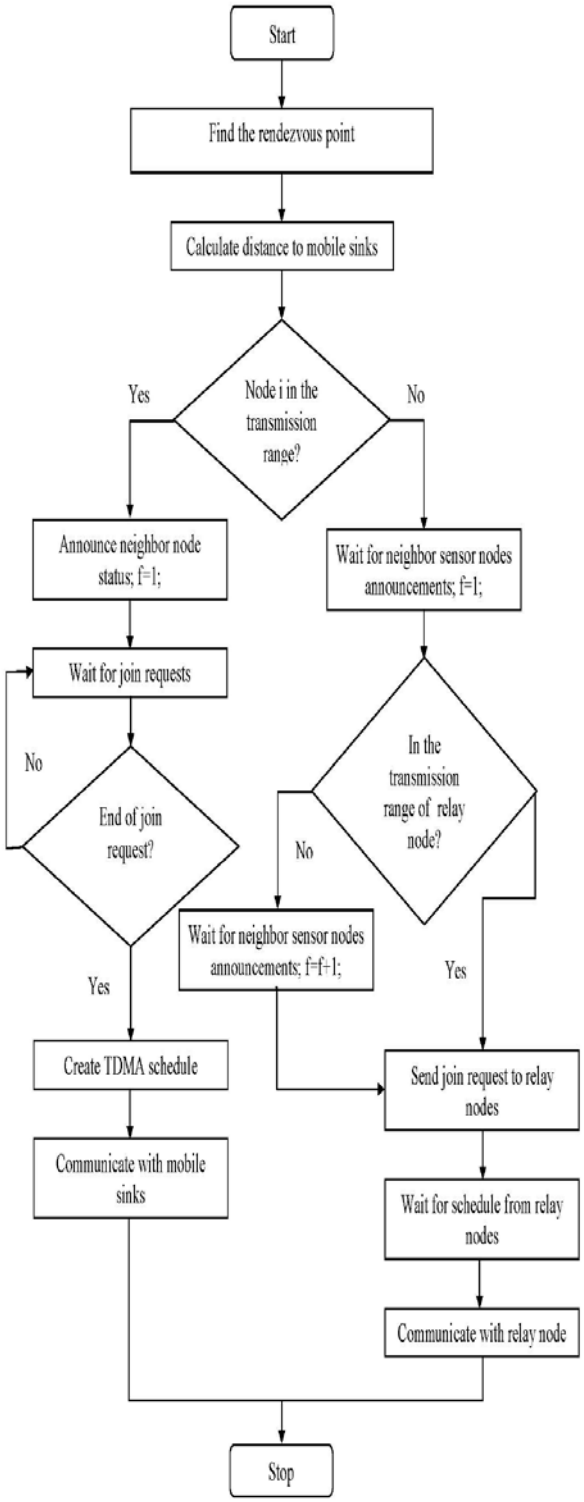


Figure 2. Flowchart of the Proposed Routing Protocol

4.SIMULATION PARAMETERS

NS2 simulation is employed to evaluate Relay node based Routing Protocol for Mobile Sink (RRPMS) which is proposed . In this case a randomly deployed sensor nodes covering the area of 600 X 600m are varied from 50,100,150,200 and 250 kbps data rate and nodes are varied from 20 to 100 nodes. The time taken for simulation is 50 sec.

4.1 *The Performance evaluation in terms of number of sinks*

This section describes the simulation results of the proposed protocol when number of sinks is increased as 1, 2 and 5.All the performance parameters were evaluated for node as well as rate.

4.2 *Simulation Results for varying Nodes*

A. Nodes Versus Delay

From Figure 3 the delay keeps reducing as and when number of sink increases. The operation with 5 sinks provides better performance than with two or single number of sinks.

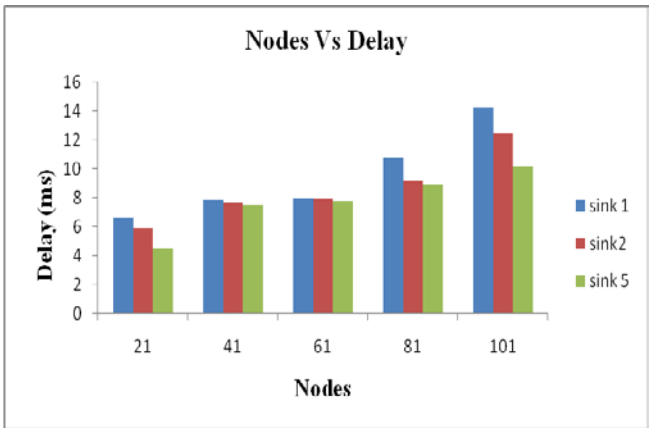


Figure 3. Nodes Vs Delay in terms of Sink

B. Nodes Versus Drop

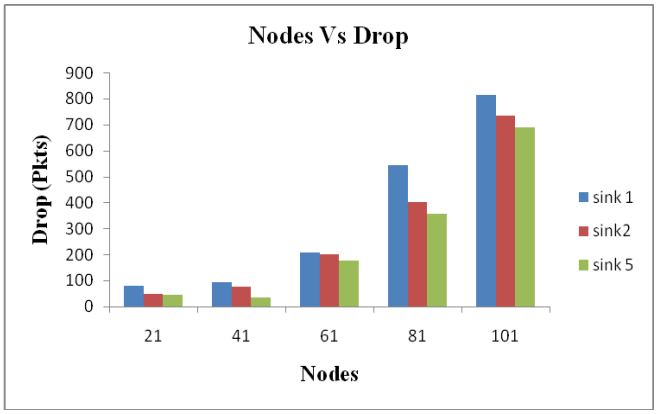


Figure 4. Nodes Vs Drop in terms of Sink

Figure 4 presents the packet drop for various node scenarios when the number of sinks is varied as 1,2 and 5. It can be seen that when number of nodes is increased, the drop decreases drastically for 5 sink scenario compared to 1 or 2 number of sinks.

C. Nodes Versus The Energy

Figure 5 presents the energy consumed by various nodes when the sink number is varied as 1,2 and 5. It can be observed that when node number is increased, energy consumption for sink 5 is higher than 1 or 2 sinks due to the increased number of beacon messages received by the sensors.

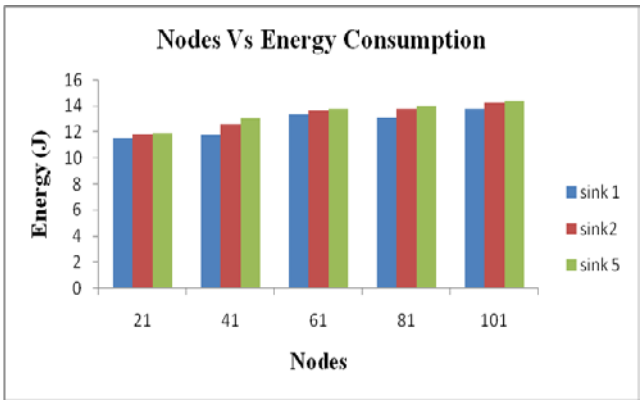


Figure 5. Nodes Vs Energy in terms of Sink

D. Nodes Versus Overhead

Figure 6 presents the overhead for various node scenarios when number of sinks are varied from 1,2 and 5. It is observed that as the number of nodes is increased, the overhead also increases for all sink scenarios. However, the overhead is very low for sink 1 operation compared with 2 or 5 sink operations.

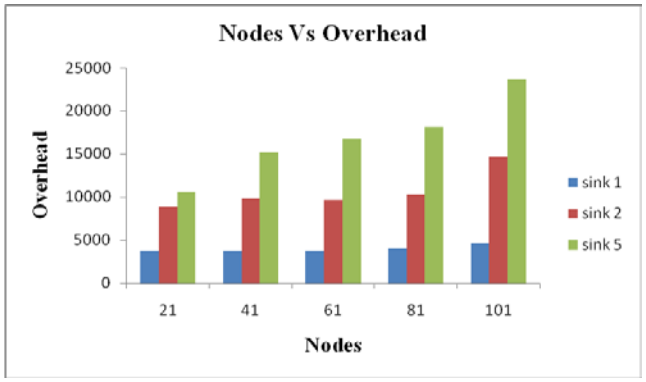


Figure 6. Nodes Vs Overhead in terms of Sink

4.3 Simulation Results for varying Data Rate

A. Rate Versus Delay

Figure 7 presents the delay Vs rate for various rate scenarios when the number of sink is varied as 1,2 and 5. It is very clear the delay is minimum for the 5 sink operation and maximum for a single sink operation.

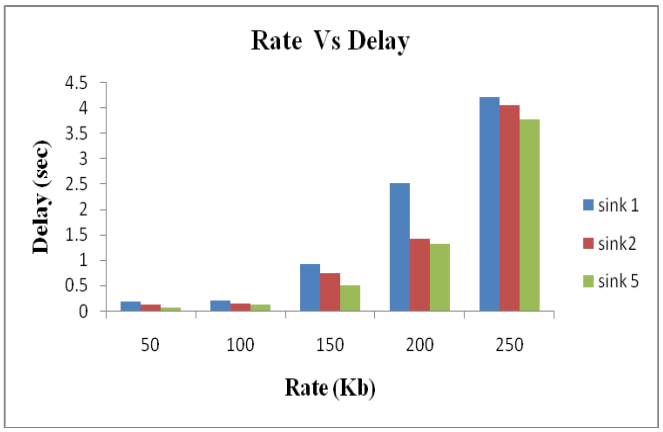


Figure 7. Rate Vs Delay in terms of Sink

B. Rate Versus Drop

Figure 8 presents the packet drop for various rate scenarios when the number of sink is varied as 1,2 and 5. It can be seen that as the number of data rate is increased, the drop increases drastically for 1 sink scenario compared to 2 or 5 number of sinks.

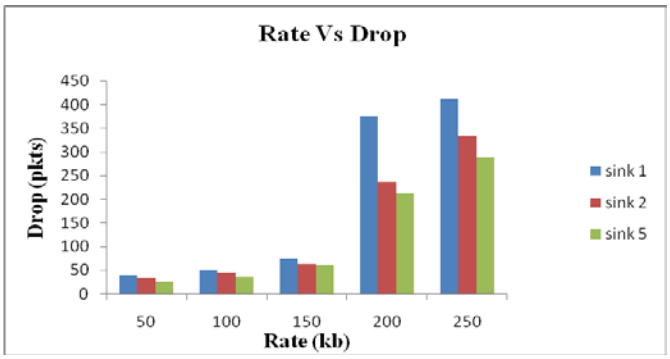


Figure 8. Data Rate Vs Drop in terms of Sink

C. Rate Versus Energy

Figure 9 presents the packet energy for various rate scenarios when the number of sink is varied as 1,2 and 5. It can be seen that as the data rate is increased, energy consumption get increased. The residual energy of sink 5 is lesser than sink 1 and sink 2 .

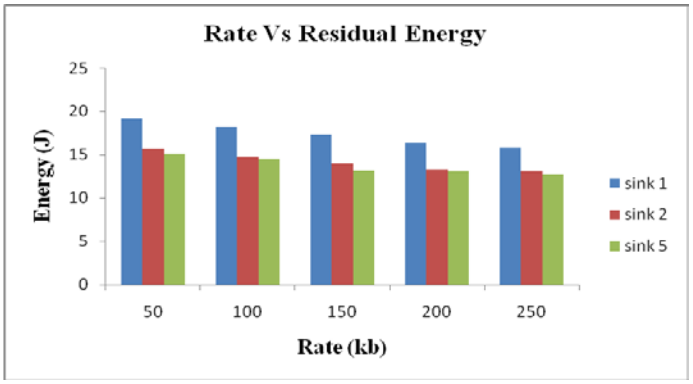


Figure 9. Rate Vs Energy in terms of Sink

D. Rate Versus Overhead

Figure 10 presents the overhead in terms of data rate for various numbers of sinks. It is observed that overhead is maximum for 5 sinks scenario when compared to other sink scenarios.

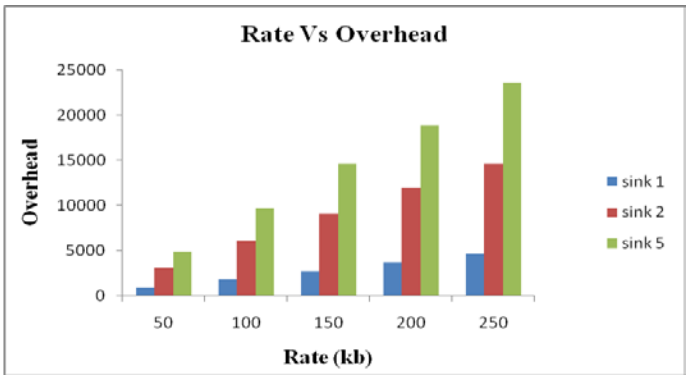


Figure 10. Rate Vs Overhead in terms of Sink

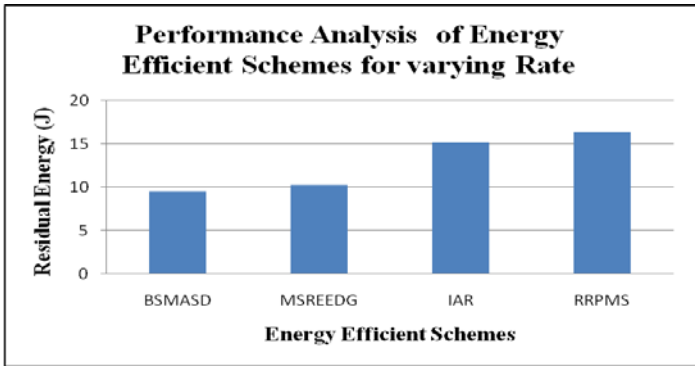


Figure 11. Performance Analysis of Energy Efficient Schemes for varying Data Rate

Figure 11 shows the cumulative remaining energy of the schemes for 200kbps data rate with 100 nodes. Residual energy of MSREEDG produces better savings in residual energy when compared with BSMASD but lesser than RRPMS.RRPMS is higher than Intelligent Agent based Routing protocol by 7%. Thus the proposed protocol increases the energy efficiency and reliability of the data.

5. CONCLUSION

The proposed routing protocol performance was compared with traditional schemes like BSMASD and IAR techniques. The parameters of comparison included packet drop, energy, delay, overhead. The simulations were carried out using NS2 simulator under various conditions of operations like varying the number of nodes and data rate. The simulation result shows that, proposed relay node based routing protocol for mobile sink is energy efficient and thus increases the lifetime of the network when compared to other routing protocols.

Conflict of Interests

"The authors declares that there is no conflict of interests regarding the publication of this paper."

ISBN: 978-1-61804-318-4

References

1. Akkaya, K, Younis, M & Bangad, M 2005, 'Sink repositioning for enhanced performance in wireless sensor networks', Computer Networks, vol. 49, no. 4, pp. 512-534.
2. Akyildiz, IF, Su, W, Sankarasubramaniam, Y & Cayirci, E 2002, 'A survey on sensor networks'. Communications magazine, IEEE, vol. 40, no.8, pp.102-114.
3. Anisi, MH, Abdullah, AH & Razak, SA 2011, 'Energy-Efficient Data Collection in Wireless Sensor Networks', Wireless Sensor Network, vol. 3, no. 10
4. Almi'ani, K, Viglas, A & Libman, L 2010, 'Energy-efficient data gathering with tour length-constrained mobile elements in wireless sensor networks', In Local Computer Networks (LCN), 2010 IEEE 35th Conference, pp. 582-589.
5. Anastasi, G, Borgia, E, Conti, M & Gregori, E 'A hybrid adaptive protocol for reliable data delivery in wsns with multipe mobile sinks', The Computer Journal, to appear (currrently available online: http://dx.doi.org/10.1093/comjnl/bxq038).
6. Khaled Almi'ani, Anastasios Viglas and Lavy Libman, *Energy-Efficient Data Gathering with Tour Length-Constrained Mobile Elements in Wireless Sensor Networks*, 35th Annual IEEE Conference on Local Computer Networks, LCN 2010, Denver, Colorado.
7. Veena Safdar, Faisal Bashir, Zara Hamid, Hammad Afzal and Jae Young Pyun, *A hybrid routing protocol for wireless sensor networks with mobile sinks*, Wireless and Pervasive Computing (ISWPC), 7th International Symposium on Dalian,2012
8. Luo, J, Panchard, J, Piorkowski, M, Grossglauser, M & Hubaux, J-P 2006, 'MobiRoute: Routing towards a Mobile Sink for Improving Lifetime in Sensor Networks', IEEE/ACM DCOSS
9. Bi, Y, Sun, L, Ma, J, Li, N, Khan, I & Chen, C 2007, 'Hums: An autonomous moving strategy in data gathering sensor networks', EURASIP Journal On Wireless Communication and Networking.
10. Chen, J, Lin, R, Li, Y & Sun, Y 2008, 'LQER: A Link Quality Estimation based Routing for Wireless Sensor Networks', Sensors, vol. 8, no. 2, pp. 1025-1038
11. Lei, JJ, Taehyun Park & Kwon 2013 'A Reliable Data Collection Protocol Based on Erasure-resilient Code in Asymmetric Wireless Sensor Networks', International Journal of Distributed Sensor Networks.
12. P.Madhumathy and D.Sivakumar, Mobile sink based reliable and energy efficient data gathering technique for WSN *Journal of Theoretical and Applied Information Technology* ,10thMarch 2014. Vol. 61 No.1

ISBN: 978-1-61804-318-4

Low Energy Adaptive Clustering Hierarchy for Three-dimensional Wireless Sensor Network

MOSTAFA BAGHOURI, ABDERRAHMANE HAJRAOUI and SAAD CHAKKOR

***Abstract*—** A great difference appears between two-dimensional (2D) and three-dimensional (3D) configurations of wireless sensor networks (WSNs). All researchers assume actually that the distribution of the nodes is done in a 2D environment. However, the WSNs are in the reality, deployed in the 3D environment. Therefore, many applications require 3D architecture. Unfortunately, the energy consumption and throughput in the 3D environment decreases considerably compared to 2D in which we can't neglect them in some applications. In this paper, we have applied the 3D architecture in LEACH protocol and we have proved by computer simulation how this 2D approximation is not reasonable since the lifetime of 3D WSN decreases by about 21% over than 2D WSN.

***Keywords*—** Wireless sensor networks, LEACH protocol, Energy-efficiency, 2D and 3D WSN, Network lifetime.

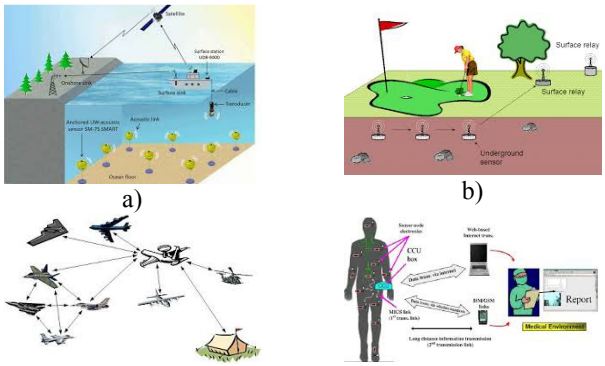
I. Introduction

In the reality the physical world we live in, is a 3D environment. Therefore, many applications, such as underwater, underground, airborne, space communications, atmospheric, forest, body or building, of WSN deployed in three-dimensional space (see Figure1). A wireless sensor network (WSN) is considered as three-dimensional (3D) when the height of deployed sensor nodes field is not negligible as compared to length and breadth of network [1]. However, with the complexity of the design and analysis of the 3D WSN, wireless sensor network in 2D plane are more studied than in 3D space.

A 3D wireless sensor network is a set wireless sensor nodes distributed in a 3D space. Each sensor node has emission to sense the events detection, such as temperature, pressure or vibration and send their measurements toward a processing center called sink [1, 2]. Due to the limitation in their battery capacity which their replacement is impossible, optimization of this unique resource has become an important issue. Node clustering is an effective technique for improving the energy efficiency and prolonging the network lifetime of a WSN [3] and has been widely studied in 2D WSNs.

LEACH [3,4] is one of the first protocols which use this technique and has been applied into the underwater environment by doing some changes [5,6,7,8]. All of these literatures considered that the nodes are distributed in tow-dimensional area.

M. BAGHOURI, is with Department of Physics, University of Abdelmalek Essaâdi, Faculty of Sciences, Tetouan, Morocco (baghouri.mostafa@gmail.com).
A. HAJRAOUI is with Department of Physics, University of Abdelmalek Essaâdi, Faculty of Sciences, Tetouan, Morocco (ad_hajraoui@hotmail.com).
S. CHAKKOR is with Department of Physics, University of Abdelmalek Essaâdi, Faculty of Sciences, Tetouan, Morocco (saadchakkor@gmail.com).



a) b) c) d)

Figure 1: Examples of three-dimensional wireless sensor networks: a) underwater, b) underground, c) airborne, and d) body

In this paper, we show that approximate the 3D field in the 2D environment deployment is not negligible when a height of network is greater.

The rest of the paper organization is done as follows: Section II summarizes the related work. Three-dimensional wireless sensor network model is provided in section III. The Simulation results are carried out in section IV. Finally we conclude our research work and give some perspectives in section V.

II. Related Work

Some works try to use the existing WSN clustering protocol for WSN in underwater environment. Reference [6] assumed UASNs are less dynamic than normal WSNs and proposed the LEACH-L, which updates its state locally, and reduced the overhead of LEACH. In [9], a clustering scheme is proposed in the context of routing scheme to extend the lifetime of UASN. Reference [10] designed a cluster structure without considering energy problem. Gu et al [11] have presented a feasible routing protocol for underground WSN in coal mine, called LEACH-mine. In the algorithm, all nodes are located in three sides of the XY projection plane and in the internal of the rectangular of the XZ projection plane. Zhou et al [12,13] have deployed a 2D WSN for coal mine, comparing to the random node deployment strategy, the strategy proposed in this work can prolong the life by two times. However, they have not considered the influence of height of the network.

Generally, in the practical applications of WSN, the sensor nodes need to be deployed and communicate in the three-dimensional area in the order to monitoring the hostile regions such as underwater, underground mine, airborne, and body environments. Therefore, to more approach to the reality situations, a 3D WSN deployment is studied detailed in this paper.

ISBN: 978-1-61804-318-4

Based on the analysis above, we find that few works on 3D deployment have been studied for WSNs. Driven by this observation; we will show by simulation that these assumptions and approximations are not reasonable in some applications of WSN.

III. Three Dimensional Wireless Sensor Network Models

A) Energy Model

This study assumes a simple model for the radio hardware where the transmitter dissipates energy for running the radio electronics to transmit and amplify the signals, and the receiver runs the radio electronics for reception of signals [7]. Multipath fading model (d^4 power loss) for large distance transmissions and the free space model (d^2power loss) for proximal transmissions are considered. Thus to transmit an $l-$bits message over a distance d, the radio expends:

$$E_{Tx}(l,d) = E_{Tx-elec}(l) + E_{Tx-amp}(l,d)$$

$$E_{Tx-elec}(l) = lE_{elec}$$

$$E_{Tx-amp}(l,d) = \begin{cases} l\epsilon_{fs}d^2, when\ d < d_0 \\ l\epsilon_{mp}d^4, when\ d \geq d_0 \end{cases}$$

Where d_o is the distance threshold for swapping amplification models, which can be calculated as $d_o = \sqrt{\frac{\epsilon_{fs}}{\epsilon_{mp}}}$

To receive an $l\ bits$ message the receiver expends:

$$E_{Rx}(l) = lE_{elec}$$

To aggregate n data signals of length l − bits, the energy consumption was calculated as:

$$E_{DA-expend}(l) = lnE_{DA}$$

B) Network Model

This section describes the network model and other basic assumptions:

1. N sensors are uniformly distributed within a square 3D rectangular field of area $A = M \times M \times M$. The Base Station is positioned at the center of the square region. The number of sensor nodes N to be deployed depends specifically on the application.
2. All nodes are deployed randomly.
3. Each sensor can sense the environment in the 3D sphere of radius r.
4. All sensors are homogeneous, i.e., they have the same capacities.
5. All the sensor nodes have a particular identifier (ID) allocated to them. Each cluster head coordinates the MAC and routing of packets within their clusters. (see Figure 2)

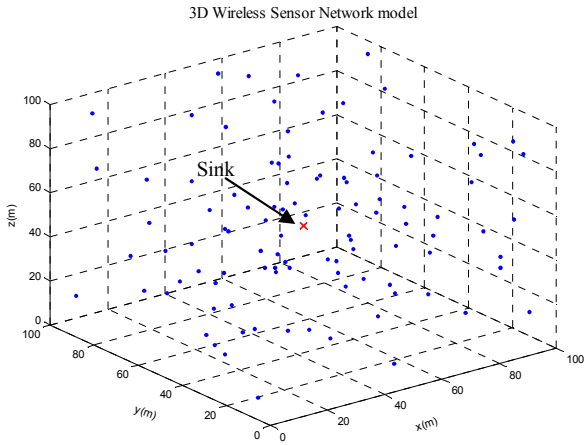


Figure 2:Three-dimensional Wireless Sensor Network model

C) Optimal number of cluster

We assume there are N nodes distributed uniformly in $M \times M \times M$ 3D region. If there are c clusters, there are on average N/c nodes per cluster. Each cluster-head dissipates energy receiving signals from the nodes and transmitting the aggregate signal to the base station. Therefore, the energy dissipated in the cluster-head node during a single frame is:

$$E_{CH} = l\frac{N}{c}E_{elec} + l\frac{N}{c}E_{DA} + l\epsilon_{mp}d_{toBS}^4$$

Where l is the number of bits in each data message, d_{toBS} is the distance from the cluster head node to the BS, and we have assumed perfect data aggregation E_{DA}.

The expression for the energy spends by a non-cluster head is given by:

$$E_{nonCH} = lE_{elec} + l\epsilon_{fs}d_{toCH}^2$$

Where d_{toCH} is the distance from the node to the cluster head.

Let E[d_{toBS}] be the Expected distance of cluster head from the base station. Assuming that the nodes are uniformly distributed, so it is calculated as follows:

$$E[d_{toBS}^2] = \int_0^{x_{max}} \int_0^{y_{max}} \int_0^{z_{max}} (x^2 + y^2 + z^2)\, f(x,y,z) dxdydz$$

Where f(x, y, z) is the probability density function of three dimensions random variable $X(x,y,z)$ which is uniform and given by:

$$f = \frac{1}{V_T} = \frac{1}{M^3}$$

If we assume that base station is the center of the network we can passing in the spherical coordinates:

$$E[d_{toBS}^2] = \int_0^{r_{max}} \int_0^{\pi} \int_0^{2\pi} r^2 f(r,\theta,\varphi) r^2 \sin\theta\, dr d\theta d\varphi$$

The area of network is aspheric with radius $r_{max} = M \times \sqrt[3]{3/4\pi}$.

If the density of sensor nodes is uniform throughout the area then becomes independent of r, θ and φ then:

$$E[d_{toBS}^2] = \frac{3}{10}\left(\frac{3}{4\pi}\right)^{\frac{2}{3}} M^2 = 0.5312M^2$$

The expected squared distance from the nodes to the cluster head (assumed to be at the center of mass of the cluster) is given by:

$$E[d_{toCH}^2] = \int_0^{r_{max}} \int_0^{\pi} \int_0^{2\pi} r^2 f(r,\theta,\varphi) r^2 \sin\theta \, dr d\theta d\varphi$$

If we assume this area is a sphere with radius $r_{max} = M \times \sqrt[3]{3/4\pi c}$ and $f(r,\theta,\varphi)$ is constant for r, θ and , (10) simplifies to:

$$E[d_{toCH}^2] = f \int_0^{M\times\sqrt[3]{3/4\pi c}} \int_0^{\pi} \int_0^{2\pi} r^3 \sin\theta \, dr d\theta d\varphi$$

If the density of nodes is uniform throughout the cluster area, then $f = c/M^3$ and

$$E[d_{toCH}^2] = \frac{3}{10} M^2 \left(\frac{3}{4\pi c}\right)^{\frac{2}{3}}$$

Therefore, the total energy dissipated in the network per round, E_{Total} , is expressed by:

$$E_{Total} = cE_{cluster}$$

Where $E_{cluster}$ is the energy dissipated in cluster which giving by:

$$E_{Cluster} = E_{CH} + \left(\frac{N}{c} - 1\right) E_{nonCH} \approx E_{CH} + \frac{N}{c} E_{nonCH}$$

This can be calculated by:

$$E_{Cluster} = l\left(\frac{N}{c} E_{elec} + \frac{N}{c} E_{DA} + \epsilon_{mp} d_{toBS}^4\right) + l\left(\frac{N}{c} E_{elec} + \frac{N}{c} \epsilon_{fs} d_{toCH}^2\right)$$

Therefore, the total energy dissipated in the network is simplified by:

$$E_{Total} = l\left(2NE_{elec} + NE_{DA} + c\epsilon_{mp} d_{toBS}^4 + N\epsilon_{fs} \frac{3}{10} M^2 \left(\frac{3}{4\pi c}\right)^{\frac{2}{3}}\right)$$

We can find the optimum number of clusters by setting the derivative of E_{Total} with respect to c to zero

$$\frac{\partial E_{Total}}{\partial c} = 0$$

$$C_{opt} = 0.2147 \times \left(N \frac{\epsilon_{fs}}{\epsilon_{mp}} \frac{M^2}{d_{toBS}^4}\right)^{\frac{3}{5}}$$

The optimal probability for becoming a cluster-head can also be computed as:

$$P_{opt} = \frac{C_{opt}}{N}$$

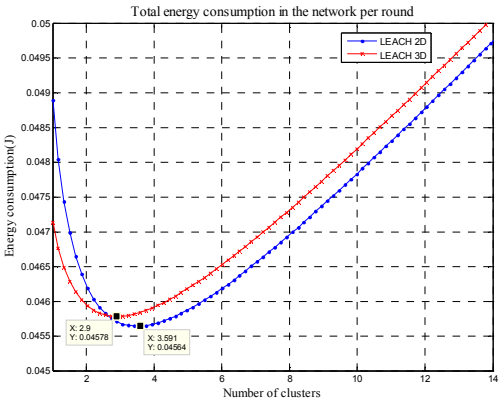


Figure 3:Variation of energy consumption for different values of clusters number *c*.

In Figure 3, we show the average energy consumption per round by each sensor node as a function of the number of clusters for two types of model, 3D and 2D WSN. Firstly, graph of the LEACH 3D model follow the same of LEACH 2D. Secondly, the graph of LEACH 3D model shows that the simulation agrees well with the analysis results. In the other hand, the 3D model consumes more energy than the 2D model which depends essentially to the no negligible value of the network height. However, this model has an optimal number of clusters less than the other model which can exploit advantageous to minimize the lifetime of the network.

IV. Simulation Results

A) *Parameter settings*

TABLE I
Energy Model Parameters

Parameter	Value
Initial Node Energy	0.5J
N	100
P	0.05
E_{elec}	50 nJ/bit
E_{DA}	5 pJ/bit
ϵ_{fs}	10 pJ/bit/m^2
ϵ_{mp}	0.0013 pJ/bit/m^4
d_{toBS}	100 m
l	500 Bytes
Rounds	2000

In this section, we study the performance of LEACH 3D protocol under different scenarios using MATLAB. We consider a model illustrate in the figure 2 with N = 100 nodes randomly and uniformly distributed in a 100m × 100m × 100m field. To compare the performance of LEACH 3D with LEACH 2D protocol, we ignore the effect caused by signal collision and interference in the wireless channel. The radio parameters used in our simulations are shown in Table1.

B) *Simulation metrics*

We define two performance metrics to evaluate both protocols as: First Node Dies (FND), or stability period and Last Node Dies (LND), or instability period. Moreover, the

ISBN: 978-1-61804-318-4

performance metrics used in the simulation study can be as follow:

- Energy consumption analysis
- Lifetime
- Throughput
- Decrease:

$$Increase = \frac{FND\ of\ LEACH\ 3D - FND\ of\ LEACH\ 2D}{FND\ of\ LEACH\ 2D} \times 100 \quad (21)$$

C) Simulation results

a. Energy consumption analysis

The performance of LEACH 3D is compared with that of the original LEACH in terms of energy and is shown in Figure 4. With the use of 3D deployment of nodes, the energy consumption of the network is decreased. This is due to the gain of the energy dissipated by height of network. From the graph it is clear that LEACH 3D decrease twice the energy savings than LEACH protocol.

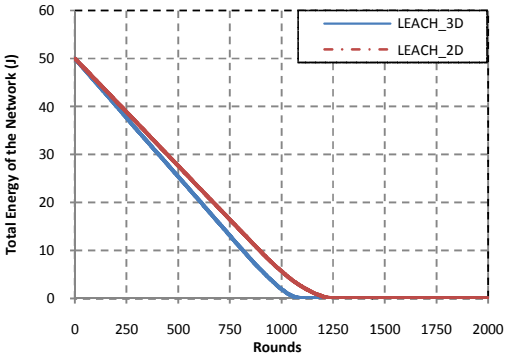


Figure 1: Energy analysis comparison of LEACH 3D and LEACH 2D.

b. Network lifetime

The number of nodes alive for each round of data transmission is observed for the LEACH 2D and 3D protocols to evaluate the lifetime of the network. Figure 5 and Figure 6 show the performance of LEACH 3D compared to LEACH 2D. It is observed that the LEACH 3D is less perform than LEACH 2D due to energy dissipation of individual node throughout the network which depend essentially on the distance between nodes and sink.

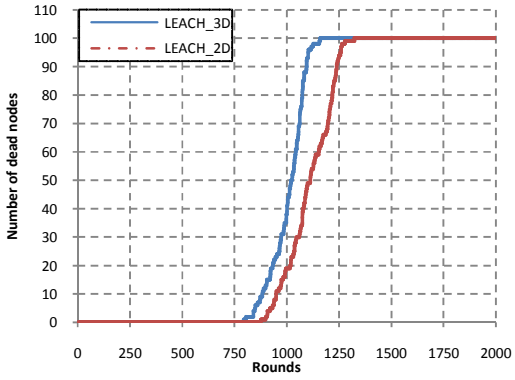


Figure 2: Number of dead nodes per round comparison of LEACH 3D and LEACH 2D.

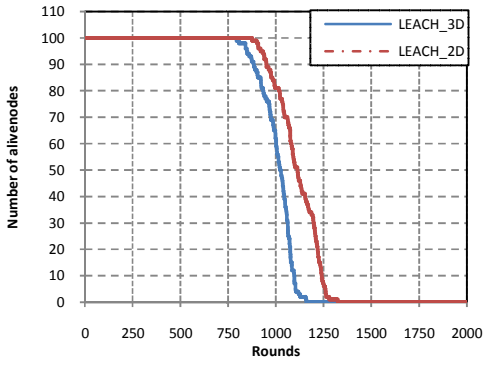


Figure 3: Number of alive nodes per round comparison of LEACH 3D and LEACH 2D.

c. Throughput

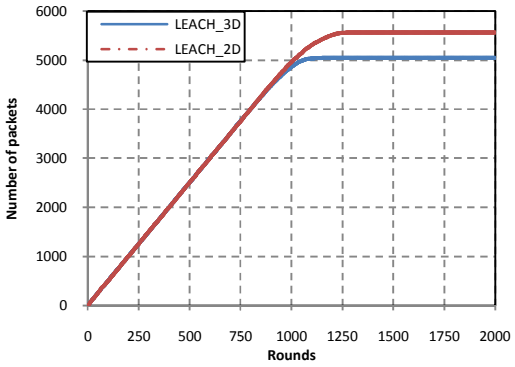


Figure 4: Performance of the protocols.

Referred to figure 7, it show clearly that LEACH 3D provide a poor throughput compared to LEACH 2D protocol, this decrease is justified by the low lifetime which give the three dimensional deployment of the nodes in the network.

d. Decrease

Generally, we can illustrate the decrease of the LEACH 3D in the Figure 8. It's noted that the throughput decreases 21% as much than LEACH 2D due to its less energy. Whereas, LEACH 2D outperforms the FND of LEACH 2D by 21% and by 28% for LND. In the other hand, LEACH 3D consumes 32% more energy than LEACH 2D.

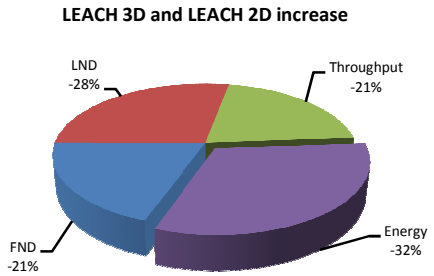


Figure 5: Decrease of LEACH 3D compared to LEACH 2D.

D) Result analysis

From our simulations, we observed that LEACH 3D consumes more energy and delivers less packets to the base

station. These results can be interpreted by the difference of distance between nodes in both situations which naturally causes by the random deployment of nodes.

V. Conclusion And Future Work

In recently, 3D wireless sensor networks have known a great prevalent due to their large applications such as underwater, space communications, atmospheric, forest or building.

The analytic of 3D WSN is more complexity than the analytic in 2D WSN. Therefore, many researches project the 3D WSN in 2D WSN. In this paper, we demonstrate by simulation, that this approximation is not reasonable if the height of network is greater than length and breadth of this network.

We strongly believe that projection of WSN in 2D environment is unjustifiable in reason that the 3D WSN is much closer to our physical word.

As future work, we will work to optimize the energy consumption of this network, since the number of cluster head in 3D WSN gives more result than 2D WSN.

References

[1] H.M. Ammari, S.K. Das, Coverage and connectivity in three-dimensional wireless sensor networks using percolation theory. IEEE Trans. Parallel Distrib. Syst. (IEEE TPDS) 20(6) (2009) Kay Romer and Friedemann Mattern. "The Design Space of Wireless Sensor Networks". IEEE Wireless Communications, 11(6):54–61, December 2004.

[2] Kay Romer and Friedemann Mattern. "The Design Space of Wireless Sensor Networks". IEEE Wireless Communications, 11(6):54–61, December 2004.

[3] Wendi R. Heinzelman, Anantha Chandrakasan, and Hari Balakrishnan, "Energy efficient communication protocol for wireless microsensor networks", IEEE International Conference on System Sciences, pp 1-10, 2000.

[4] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An application specific protocol architecture for wireless microsensor networks," IEEE Transactions on Wireless Communications, vol. 1, No. 4, pp. 660-670, October 2002

[5] Xia Li, Yongqian Wang, and Jingjin Zhou, "An energy-efficient clustering algorithm for underwater acoustic sensor networks," Control Engineering and Communication Technology (ICCECT), 2012 International Conference on, pp. 711 - 714, Dec. 2012.

[6] X.Li, S.-L.Fang, and Y.C. Zhang, "The study on clustering algorithm of the underwater acoustic sensor networks," the 14th International Conference on Mechatronics and Machine Vision in Practice (M2VIP 2007), Dec. 2007.

[7] Guangsong Yang, Mingbo Xiao, En Cheng, and Jing Zhang, " A cluster-head selection scheme for underwater acoustic sensor networks," Communications and Mobile Computing (CMC), 2010 International Conference on, vol. 3, pp. 188 - 191, Apr. 2010.

[8] Liu, G., and Wei, C., "A new multi-path routing protocol based on cluster for underwater acoustic sensor networks," International Conference on Multimedia Technology (ICMT), pp. 91-94, 2011.

[9] P. Wang, C. Li, and J. Zheng, "Distributed minimum-cost clustering protocol for underwater sensor networks (UWSNs)." IEEE International Conference on Communications (ICC 2007), June, 2007, pp. 3510-3515

[10] Salva-Garau F, Stojanovic M, "Multi-cluster protocol for ad hoc mobile underwater acoustic networks", OCEANS 2003 Proceedings ,2003-09, 1(22-26), pp.91-98.

[11] X. Gu, Y. Jin, Y. Sun and J. Yan, "Maximum lifetime routing strategies for wireless sensor networks in coal mine", in 2010 International Conference on Computer Engineering and Technology, pp. 341-344, April 2010.

[12] Z.C. Zhu, G.B. Zhou, and G.Z. Chen, "Chain-type wireless underground mine sensor networks for gas monitoring", Advanced Science Letters, vol. 4, no. 2, pp. 391-399, 2011.

[13] G.B. Zhou, Z.C. Zhu, G.Z. Chen, and N.N. Hu, "Energy-efficient chain-type wireless sensor network for gas monitoring", in International Conference on Information and Computing Science, pp. 125-128, May 2009.

Baghouri Mostafa was born in Tangier Morocco. He's a member in the Physics department, Team Communication Systems, Faculty of sciences, University of Abdelmalek Essaâdi, Tetouan Morocco, his research area is: routing and real time protocols for energy optimization in wireless sensors networks. He obtained a Master's degree in Electrical and Computer Engineering from the Faculty of Science and Technology of Tangier in Morocco in 2002. He graduated enabling teaching computer science for secondary qualifying school in 2004. In 2006, he graduated from DESA in Automatics and information processing at the same faculty. He work teacher of computer science in the high school

Abderrahmane Hajraoui is a professor of the Higher Education at University of Abdelmalek Essaâdi. He's a director thesis in the Physics department, Communication and detection Systems laboratory, Faculty of sciences, University of Abdelmalek Essaâdi, Tetouan Morocco. His research areas are: Signal processing and image, automation, automation systems, simulation systems, Antennas Antennas and radiation, microwave devices

Chakkor Saad was born in Tangier Morocco. He's a member in the Physics department, Team Communication and detection Systems, Faculty of sciences, University of Abdelmalek Essaâdi, Tetouan Morocco, and his research area is: wireless intelligent sensors and theirs applications, frequency estimation algorithms for faults detection and diagnosis system in electromecanical machines. He obtained the Master's degree in Electrical and Computer Engineering from the Faculty of Sciences and Techniques of Tangier, Morocco in 2002. He graduated enabling teaching computer science for secondary qualifying school in 2003. In 2006, he graduated from DESA in Automatics and information processing at the same faculty. He works as teacher of computer science in the high school.

An Enhanced Multidimensional Hadamard Error Correcting Code and his Application in Video-Watermarking

Andrzej Dziech[1], Jakob Wassermann[2]

***Abstract*—** this paper presents a new approach for multidimensional Hadamard Error Correcting Code and his application in Video Watermarking. The codewords of the 2D Hadamard Error Correcting Code are basic Images of this transform. The main idea of this new method is to map the 2D basis images into a collection of one-dimensional rows and apply a 1D Hadamard decoding procedure on them. After this, the image is reassembled, and the 2D decoding procedure can be applied more efficiently. With this approach, it is possible to overcome the theoretical limit of error correcting capability of n/2-1. To prove the efficiency and practicability of this new enhanced 2D Hadamard ECC, the method was applied to a video Watermarking Coding Scheme. In this case, the Watermarks are protected by this code and, therefore, robust against attacks like compression and subsampling. For this purpose, the initial video is decomposed by multi-Level Interframe Wavelet Transform. The low pass filtered part of the video stream is used for embedding the watermarks, which are protected by enhanced 2D Hadamard Error Correcting Code. Every frame of this low pass filtered sequence undergoes an 8x8 block-wise DCT before the embedding procedure is applied on selected coefficients of the spectrum by using the QIM (Quadrature Index Modulation). The experimental results show that this method seems to be very robust against strong MPEG compression and obtains low degradation of the host sequence, due to the new efficient Hadamard error correcting code**.**

***Keywords*—** Error Correcting Code, Hadamard Code, DWT, 3D Hadamard Transform, DCT, Spread Spectrum, basis images, Video-Watermarking.

I. Introduction

Many applications in telecommunication technologies are using Hadamard Error Correcting Code. Plotkin[12] was the first who discovered in 1960 error correcting capabilities of Hadamard matrices. Bose, Shrikhande[13], Peterson[14] also have made important contributions. Levenshstein[10] was the first who introduced an algorithm for constructing a Hadamard ECC.The most famous application of Hadamard Error Correcting Code was the NASA space mission in 1969 of Mariner and Voyager spacecrafts. Thanks to the powerful error correcting capability of this code it was possible to decode properly high-quality pictures of Mars, Jupiter, Saturn, and Uranus.

The Hadamard code of the length n can encode $k=log_2(n)$ messages and is denoted as (n,k) linear code. The Hamming distance is $n/2$, and it can correct $n/2$-1 errors [11]. The 2D and 3D Hadamard code are an extension of the one-dimensional case and till now does deliver any advantages.

In this paper we introduced a new type of multidimensional Hadamard ECC, we called it enhanced Hadamard Error Correcting Code. It can overcome the limit of error correcting capability of n/2-1 errors. The application of this Code in Video Watermarking gives the strong prove of its effectiveness. The reason for selecting Video Watermarking lies in strong compression ratio, which normally applied the video sequences, compression factors greater than 1:200. For example, an uncompressed HDTV video stream has a data rate of 1.2Gbit/s and for distribution reason it must be compressed to 6 Mbit/s. For embedded watermarks, it is a big challenge to survive such strong compression ratio and error correcting code plays a decisive role. This paper has followed the structure: In Chapter II we introduce the enhanced Hadamard Error Correcting Code and his error correcting capability. In the Chapter III, the authors explain the Video Watermarking Scheme and the Chapter IV presents the results and discussion.

II. Enhanced Hadamard Error Correcting Code

In this chapter, we will give in the beginning an overview about one 1D and 2D- Hadamard Code. Then we explain the enhanced version.

A. One-Dimensional Hadamard Code

To generate a Hadamard code of the length n can encode $k=log_2(n)$ messages and is denoted as (n,k) linear code. The Hamming distance is $n/2$, and it can correct $n/2$-1 errors. The code words of n bit Hadamard Code is the rows of $n*n$ Hadamard Matrix H_n.

This work was supported by European Commission FP7 Grant INDECT Project. No. FP7-218086.

[1]Andrzej Dziech, Dept. of Telecommunication, AGH University of Science and Technology al. Mickiewicza 30, 30-059 Krakow POLAND (dziech@kt.agh.edu.pl)

[2]Jakob. Wassermann, Dept. of Electronic Engineering University of Applied Sciences Technikum Wien, Hoechstaedtplatz 5, 1200 Vienna, AUSTRIA,(jakob.wassermann@technikum-wien.at)

In case of n=8 we obtain the following matrix:

$$H_8 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{pmatrix} \quad (1)$$

The code words are the rows of this Matrix H_8. The Table I shows the Code Book of the linear Code (8,3)

Message	Code Words
0 0 0	1 1 1 1 1 1 1 1
0 0 1	1 -1 1 -1 1 -1 1 -1
0 1 0	1 1 -1 -1 1 1 -1 -1
0 1 1	1 -1 -1 1 1 -1 -1 1
1 0 0	1 1 1 1 -1 -1 -1 -1
1 0 1	1 -1 1 -1 -1 1 -1 1
1 1 0	1 1 -1 -1 -1 -1 1 1
1 1 1	1 -1 -1 1 -1 1 1 -1

Table I: Code Book with n=8 and k=3

The decoding procedure of the received code word is using the Hadamard spectrum vector to determine the corresponding message. The spectrum vector d is calculated by multiplying the code vector c by the Hadamard Matrix H_8

$$d = c \cdot H_8 \quad (2)$$

When considering the message: 010 (third row), the corresponding Hadamard code is:

$$c = [1 \quad 1 \quad -1 \quad -1 \quad 1 \quad 1 \quad -1 \quad -1].$$

The decoded vector is:

$$d = [1 \quad 1 \quad -1 \quad -1 \quad 1 \quad 1 \quad -1 \quad -1] * H_8$$

$$d = [0 \quad 0 \quad 8 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]$$

The third component of the vector d has the biggest value of the spectrum; all others are zeroes, $d(3)=8$, $d(i)=0$ for i=1,..8 and $i \neq 3$. It means that the code word at the position i=3 in the codebook determines the message (010).

In the case of one error, the third component of the spectrum vector d still remains the biggest one. In the case of a corrupted code word c=[***-1*** 1 -1 -1 1 1 -1 -1] the Hadamard spectrum vector delivers:

$$d = [-1 \quad 1 \quad -1 \quad -1 \quad 1 \quad 1 \quad -1 \quad -1] * H_8$$

$$d = [-2 \quad -2 \quad 6 \quad -2 \quad -2 \quad -2 \quad -2 \quad -2]$$

The third component is still the biggest, so the message can be decoded.

In the case of two errors, it is already impossible to decode the message unambiguously. Nevertheless, it is possible to correct more than one error. If eight errors occur, our code word c=[-1- 1 1 1 -1 -1 1 1] is completely corrupted. In this case, the absolute value of the third component of the Hadamard spectrum is the biggest, and it has a negative sign. A negative sign means that the decoded code word must be inverted.

$$d = [-1 \quad -1 \quad 1 \quad 1 \quad -1 \quad -1 \quad 1 \quad 1] * H_8$$

$$d = [0 \quad 0 \quad -8 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]$$

In the case of seven errors, we have exactly the same situation as with one error, however, with one small difference: the third component has a negative sign. The following figure shows the error correcting capability of an 8 bit Hadamard code.

Fig. 1. Error Correcting Capability of 8 Bit Hadamard Code

The 8 bit Hadamard code can correct 1,7 and 8 bit errors regardless where they occur within the code words. Generally, we can say than n bit Hadamard code can correct totally n/2-1 types of errors. The number of error bits occurring ranges from $\left[1,\cdots,\frac{n}{4}-1\right]$ to $\left[\frac{3}{4}n+1,...,n\right]$

B. Two-Dimensional Hadamard Error Correcting Code

The 2D Hadamard error correcting code uses so called basis images instead of Hadamard vectors. The basis images functions are orthogonal to each other, and they can be generated from the Hadamard matrix by multiplication of columns and rows. Generally we can write

$$A_{ml} = H_n(:,l) * H_n(m,:) \quad (3)$$

In case of 4x4 Hadamard matrix

$$H_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \quad (4)$$

We can calculate the complete set of 16 such basis images.

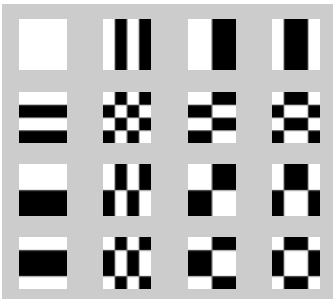


Fig. 2. Basis Images of 2D Hadamard Transform (4x4)

For instance, the pattern A_{31} is generated by Eq.(1) and has the numerical presentation

$$A_{31} = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 \end{pmatrix} \quad (3)$$

It can be visualized as

Fig. 3. Basis Image A_{31}. The "1" is interpreted as 255 (White) and "-1" as 0 (Black)

The 2D Hadamard Spectrum of such basis images, which is denoted by C, delivers a matrix where only one coefficient is differed from zero. He represents in a spectral domain the corresponding basis image. For example the Hadamard spectrum matrix of the pattern A_{31} is

$$C = H_4 * A_{31} * H_4^{-T} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 16 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (4)$$

The component C_{31}=16 and all others are zero. This fact, identification of the basis image trough it spectral coefficient, can be utilized to construct error correcting code. The code words are the pattern of basis images, and they can be decoded unambiguously by detecting the biggest absolute coefficient value inside of 2D Hadamard Spectrum according to Eq.(4).

In case that the basis image A_{31} is corrupted by some perturbation and looks likes

$$\tilde{A}_{31} = \begin{pmatrix} -1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \end{pmatrix} \quad (5)$$

It is still possible to recover the original pattern completely because the absolute value of $|C_{31}|$=10 and still stays the biggest one between the other spectral coefficients of matrix C.

$$C = H * \tilde{A}_{31} * H^{-T} = \begin{pmatrix} -2 & 2 & -2 & 2 \\ -6 & -2 & -6 & -2 \\ 10 & -2 & -6 & -2 \\ -2 & 2 & -2 & 2 \end{pmatrix} \quad (6)$$

Message	Basis Image	D.C.	Pulse Stream (code word)
0000		C_{11}	0000000000000000
0001		C_{12}	0101010101010101
0010		C_{13}	0000000011111111
0011		C_{14}	0000111111110000
0100		C_{21}	0000111100001111
0101		C_{22}	0101101001011010
0110		C_{23}	0011110000111100
0111		C_{24}	0110100110110101
1000		C_{31}	0011001100110011
1001		C_{32}	0101010110101010
1010		C_{33}	0011001111001100
1011		C_{34}	0011110011000011
1100		C_{41}	0110011001100110
1101		C_{42}	0101101010100101
1110		C_{43}	0110011010011001
1111		C_{44}	0110100110010110

Table II: 2D Code Book constructed from Basic Images

To apply the basic images as a code words, we have to map the two-dimensional structure of the pattern into one dimensional pulse stream which will be denoted by the code word. In the Table II such 2D Hadamard codebook is depicted.
The total number of errors that can be corrected is n/2-1 and correspond completely to one dimensional case. The simple enlargement from 1D to 2D doesn't bring any improvement. To overcome this limit, a new enhance Hadamard decoding procedure for 2D and 3D Hadamard code is introduced.

C. Enhance 2D Hadamard Error Correcting Code

The enhanced 2D Hadamard code makes possible to correct more errors as with the standard Hadamard method. The basic idea is to map the 2D basis images into a collection of one-dimensional rows and applying them 1D decoding procedure. After this, the image is reassembled, and the 2D decoding procedure (Eq.(4)) can be applied more efficiently. With this, the approach is possible to overcome the theoretical limits of n/2-1 errors.

To show the functionality of this method we consider the basis images A_{71} of 8x8 2D Hadamard Transform. This basis image can be derived from Eq.(1). (See Fig. 4)

$$A_{71} = H_8(:,1) \cdot H_8(7,:) = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

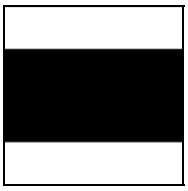


Fig. 4. Basis Image A_{71} of 2D Hadamard Transform (8x8) and its visualization. "1" is interpreted as white (255), "-1" as black (0)

This image is now corrupted by noise (Fig.5). The corresponding error matrix contains 17 errors. According to the consideration from chapter 2.1, it is not possible to recover this pattern because the number of errors exceeds the limit of n/2-1=15.

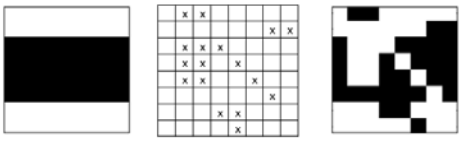


Fig. 5. Original Basis Pattern A_{71}, Error Mask, and the Corrupted Pattern

The enhanced Hadamard decoding procedure will work in this case as following
(see Fig. 6):

- The corrupted basis image (A) is separated into its rows (B).
- On each row is applied 1D Hadamard decoding procedure. Rows which contain only one error are decoded error free. Rows No. 6 and No. 8 are now without any errors (C).
- Reassemble the pattern again (D). The renewed pattern contains now fewer errors as before, namely 15.
- Apply the 2D Hadamard decoding procedure according to Eq.(4). The result will be error free pattern (E).

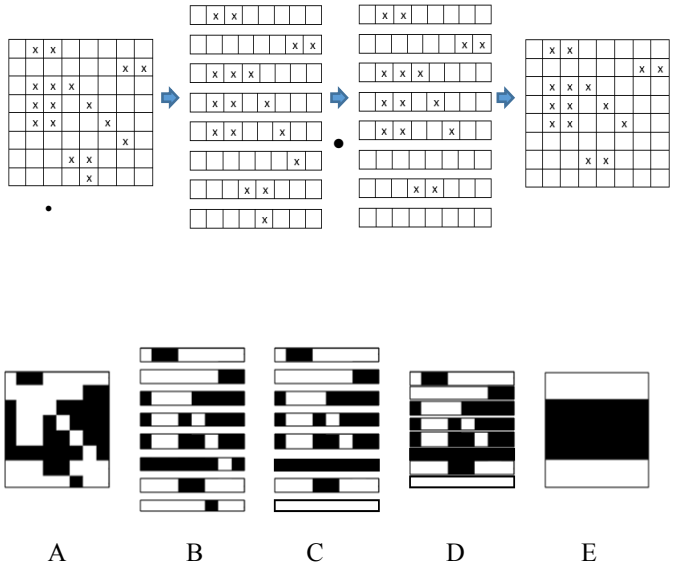


Fig. 6. Enhanced Hadamard Decoding Procedure on Error Mask and on Corrupted Basis Image

The improvement is visualized in the Fig. 7. It clearly to see that the enhanced version outperform the Standard Hadamard Code and can overcome the limit of the error capability of n/2-1 errors. In case of 8x8 basis images, the standard method can correct maximum 15 errors. The enhanced version can correct 92% of all 16 possible errors pattern inside the basis images. In case of 17 errors, 83% of all error pattern can be corrected.

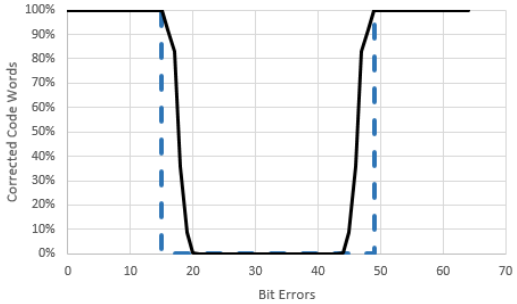


Fig. 7. Comparison of Error Correcting Capabilities of Standard Hadamard (dashed line) with 8x8 2D Enhanced Hadamard Code

ISBN: 978-1-61804-318-4

In case of 47 errors, the standard method can't correct any error but the enhanced method can correct 83% of all possible error pattern.

D. 3D Hadamard Cubes

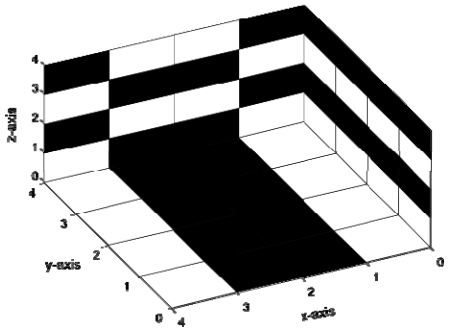
A Hadamard cube is a basis image expanded in the third dimension by multiplying the pattern with Hadamard vectors.

$$D_{mlk} = A_{ml} \cdot H_k(:) \tag{7}$$

The basis image is represented by A_{ml} and $H_k(:)$ is the k Hadamard vector. For example the pattern A_{41}

$$A_{41} = \begin{pmatrix} 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

and the Hadamard vector $H_2 = (1 \quad -1 \quad 1 \quad -1)$ generates the cube D_{412}



The decoding procedure and the corresponding error correction works similar to the correcting procedure described in Chapter II.C. Before it can be applied the cube is resolved from the front side in separate layers. On each layer, the enhanced 2D Hadamard decoding procedure is applied.

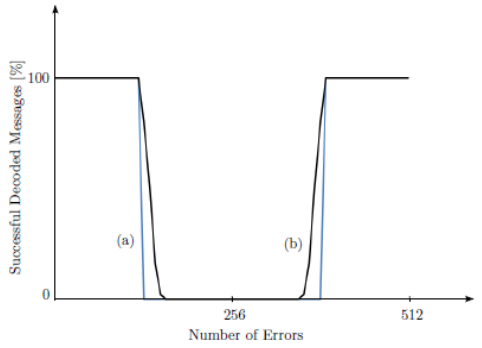


Fig. 8 Comparison of Error Correcting Capabilities of Standard Hadamard (a) with 8x8x8 3D Enhanced Hadamard Code (b).

III. Proposed Watermarking Scheme

The proposed watermarking scheme works in the raw domain and the method combines the previously described 2D Hadamard error correcting code with Discrete Wavelet Transform (DWT) of video sequences. In the Fig. 9, the whole encoding process is illustrated. The raw format of the luminance channel of the original video stream is decomposed by multi-level Interframe DWT with Haar Wavelet. This low pass filtered part of the video stream undergoes a block wise DCT Transform. From DCT spectrum, special coefficients are selected and used for embedding procedure with 3D Hadamard coded watermarks. The embedding procedure itself is realized through QIM (Quadrature Index Modulation) techniques.

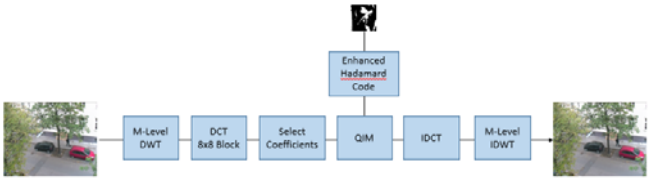


Fig. 9. Watermarking Encoding Process

The decoder procedure is depicted in Fig.10. At the beginning of the decoding procedure, the embedded video sequence undergoes the same multi-level Interframe DWT and DCT transform as on the encoder side. After the selection of the proper DCT coefficients, the inverse QIM (IQIM) is applied. It delivers the decoded code words (pulse stream). Through the help of Enhanced Hadamard error correcting code, which is described in chapter 2.3 the original watermark is extracted.

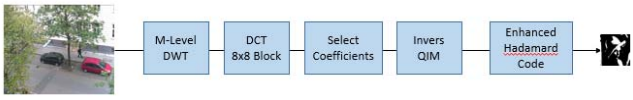


Fig. 10. Watermarking Decoding Process

A. Multi-Level DWT

As mention above a multi-Level Interframe DWT with Haar Wavelet was used to deliver a low pass filtered video. The Fig.11 illustrates the operating principle of this transform. In the first level, the two consecutive frames are averaged. In the second level, the frames from the level one are averaged and so forth. In this watermarking schemes, we used DWT levels from 12 till 16.

ISBN: 978-1-61804-318-4

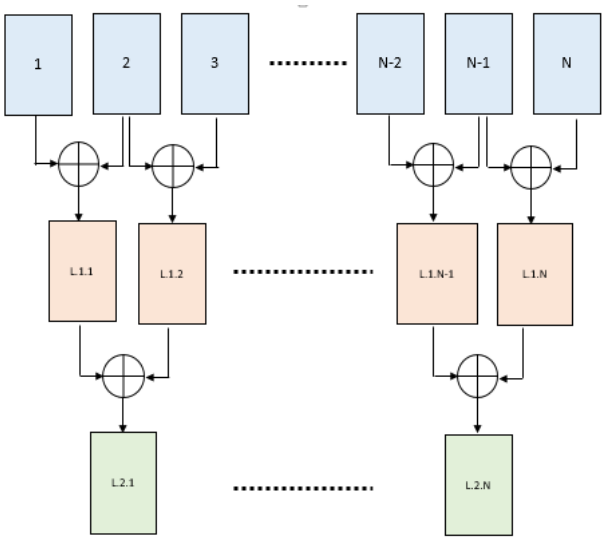


Fig. 11. Multi-Level Interframe DWT. At the first level, the two consecutive frame are averaged. At level two the consecutive frames from level one are averaged and so forth.

B. Selection of Embedding Coefficients

To realize the embedding procedure, some coefficients from the DCT spectrum of DWT filtered video sequence must be selected. The Fig.12 shows which coefficients are qualified for watermark for such process. These are mostly from the gray area.

1	9	17	25	33	41	49	57
2	10	18	26	34	42	50	58
3	11	19	27	35	43	51	59
4	12	20	28	36	44	52	60
5	13	21	29	37	45	53	61
6	14	22	30	38	46	54	62
7	15	23	31	39	47	55	63
8	16	24	32	40	48	56	64

Fig. 12. Coefficients of DCT Spectrum which fits for embedding

IV. RESULTS

The investigation was done with HDTV video sequence with the resolution of 1080*1920 and 25fps. The Video was captured with an AVCHD Camera. The watermarking processing was performed only for the luminance channel (after converting RGB into YCrCb color space) because it is more robust against distortions than any other channels. It was investigated how many embedded watermark bits survive compression attacks without to cause significant impairments. The degradation of the watermarked output video was measured with SSIM (Structural Similarity) index. SSIM is based on the human eye perception and so the expressiveness about distortion is better than in the traditional methods like PSNR (Peak Signal to Noise Ratio) or MSE (Mean Square Error) [8].

It was chosen the 3D Hadamard Code of the size of 8x8x8, which means the code word length of 512 bits. This implies the message code length of 9 bit. The DCT block size was 8x8 and from each block were selected 16 coefficients. With these, information is easy to calculate the total number of embedded watermark bits pro each frame.

$$E = \frac{H \cdot W}{B^2} \cdot \frac{M}{W} \cdot C = \frac{1920 \cdot 1080}{8^2} \cdot \frac{9}{512} \cdot 16 = 9112 \text{ Bit/Frame} \quad (7)$$

Where H is, the height and W is the width of the frame. The letter B denotes the block size of DCT transform, the letter M is the message code length, the letter W represents the code word length of the 3D Hadamard code and the letter C is the number of selected spectral coefficients.

Data Rate	Compression Ratio	DWT Levels	Delta QIM	Selected Coefficients	Embedded Bits	Extracted Watermark	SSIM Watermarking	SSIM Video
6 Mbit/s	1:200	16	11	9	9112		1	0,95
5 Mbit/s	1:240	16	11	9	9112		1	0,95
4 Mbit/s	1:300	16	11	9	9112		0,92	0,93
3 Mbit/s	1:400	16	11	9	9112		0,89	0,91
2 Mbit/s	1:600	16	11	9	9112		0,56	0,90

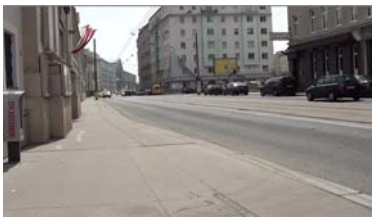
Table 2. Results

In Table 2, the results of capacity and robustness measurements are presented. The compression attacks were done by H.264 codec with different compression ratios. Because the method works in the raw domain, the original data rate is 1.2 Gbit/s. As a watermark was used a chessboard pattern of the size of 30x30 Pixel.

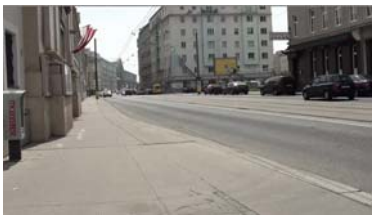
The watermarks were inserted successively into the frames. The Delta QIM gave the width of the quantization steps and was tuned to value 11. Generally the Delta value determines the noise distortion in the host video

The embedded video sequence was compressed with different compression ratios. In case of compression to 5 Mbit/s, which correspond to a compression ratio of 1:240 it is still possible to extract all watermarks error free. The quality comparison between originally compressed video and embedded and compressed shows, that there is only slight different. The SSIM index Video is in this case 98%

ISBN: 978-1-61804-318-4



a. Embedded and Compressed



b. Only Compressed

Fig. 12. a. Embbeded and Compressed with 5Mbit/s, b: Only compressed with 5 Mbit/s

V. Conclusion

A new method for robust video watermarking was introduced. It uses a new developed Error Correcting Code and watermark embedding procedure in the frequency domain. The ingesting of watermark bits are done by QIM into the DCT spectral coefficients of low pass filtered stream. This low pass filtered video is generated by Interframe multi-level DWT.

The core idea of this scheme is the usage of the new developed Enhanced 3D Hadamard Error Correcting Code, for the purpose of video watermarking. Instead of ingest watermark bits itself, Hadamard cubes are used as a code words. The code word length was selected to 512 bits, which implies the message length of 9 bit. The error correcting abilities of this Enhanced 3D Hadamard code are extremely good, especially in the case of burst errors. It is possible to correct more than 256 errors and even if the whole code word is corrupted it is possible to reconstruct it error free. This method can correct more errors as till now it was possible with standard Hadamard code.

The results are very promising. In an HDTV, a video sequence is possible to embed around 9112 bits (nearly 1, KB) watermark information per frame, which can be error free extracted after very strong compressions of the video. The compressed video has a data rate of 5Mbit/s (compression ratio 1:240). The embedded, and compressed video streams show any visual degradation.

References

[1] Hartung, F. & Kutter, M.: Multimedia Watermarking Techniques. In: *Proceeding of IEEE* 87 (1999), S. 1079-1107

[2] Ho Antony T.S., Shen Jun, Soon Hie Tan, Kot Alex C..: Digital image-in-image watermarking for copyright protection of satellite images using the fast Hadamard transform. In: *Geoscience and Remote Sensing Symposium* 6 (2002), S. 3311 – 3313

[3] Chen, B., W. G., Wornell: Quantization Index Modulation for Digital Watermarking and Information Embedding of Multimedia. In: *Journal of VLSI Signal Processing* 27 (2001), S. 7-33

[4] I.J.Cox, J. Killian, F. L. & vol.6 pp., T. S.: Secure Spread Spectrum Watermarking for Multimedia. In: *IEEE Transactions on Image Processing* 6 (1997), S. 1673-1687

[5] Lin, S. D., S. C.-F.: A Robust DCT-Based Watermarking for Copyright Protection. In: *IEEE Transactions on Consumer Electronics* 46 (2000), S. 415

[6] Pik-Wah Chan, M. R. L.: A DWT-Based Digital Video Watermarking Scheme with Error Correcting Code. In: *Information and Communications SecurityLecture Notes in Computer Science* 2836 (2003), S. 202-213

[7] Swanson, M.D., K. M. T. A.: Multimedia Data-Embedding and Watermarking Technologies. In: *Proceeding of IEEE* 86 (1998), S. 1064-1087

[8] Zhao, D., C. G. L. W.: A Chaos-Based Robust Wavelet-Domain Watermarking Algorithm. In: *Chaos, Solutions and Fractals* 22 (2004), S. 792

[9] Combinatoric in Space, The Mariner 9 Telemetry System, http://www.math.cuden.edu/~wcherowi/courses/m6409/mariner9talk.pdf

[10] Levenshtein, V.I.: Application of Hadamard matrices to a problem in coding. *Problems of Cybernatics, 5*, 166-184, 1964

[11] MacWilliams, F.J., Sloane, N.J.A.: The theory of error-correcting codes. *North Holland, New York, 1977*

[12] Plotkin, M.*: Binary Code with specified minimum distance, IRE Transactions, IT-6:445-450, 1960*

[13] Bose, R.C., Shrikhande, S.S.*: A note on the result in the theory of code construction. Inf. And Control 2, 183-194 (1965)*

[14] Peterson, W.W.: Error correcting codes. The M.I.T. Press, Massachusetts Institute of Technology & J.Wiley & Sons, New York 1961

Open Communication Supports Innovation

Stachová Katarína, Hudáková Monika, Stacho, Zdenko

Abstract— Systematic creation and usage of human potential is a precondition of building and developing of strengths and competitive advantages of organizations. A company has a great potential, when it is able to apply the knowledge and experience of its employees and achieved "innovation" results of its work, whether they have been recorded or not; and when it is able to absorb the latest knowledge from external environment, use all available resources and means with optimal efficiency and sustain their optimal structure, which indicates that open communication in a company is the key factor affecting its ability to succeed in competitive environment and it is also a predictor of open communication towards external partners. In our contribution, we focused on open and broad communication, since it has the most significant impact on formal as well as informal labour relations which have a fundamental role in team creation as well as information sharing and knowledge continuity. In a questionnaire research we focused on finding out whether and in what extent organisations operating in Slovakia focus on communication. The paper also provides a simple instrument suitable to identify bottlenecks in this sphere.

***Keywords*—**Communication, communication methods, open and broad communication, research

I. Introduction

COMMUNICATION influences the innovation process in each organisation in a great extent. Many studies implied that a great amount of problems occurred during the innovation process is a result of ineffective or unreliable communication, particularly among different functional departments of the organisation involved in the innovation process [1], [2]. It is therefore important to put emphasis on the development of open and quality communication. It is possible to use new technologies like "groupware" (group software), intranet or social knowledge networks for this purpose [3], [4], [5]. Teamwork activities and such aspects as communication and support for innovation can lead to group learning and making performance [6].

Communication is understood as transmission and receiving of notifications between two or more subjects [7] - [9]. Notification is any subject of communication, arbitrary verbal and non-verbal facts and events having a signal nature [7], [10], [11]. Communication is predominantly a relationship, since at least two subjects are necessary in order to communicate [7], [12].

Appropriate communication should fulfil two basic tasks, particularly [7]:

- to exchange the greatest possible amount of notified content between communicating entities in a certain time unit
- effectiveness of communication,
- to transmit notifications between communicating entities with regard to the existence of disturbance (disturbing elements) with the lowest possible losses – reliability of communication.

Contrary to interpersonal communication, taking place between two and more people, organisational communication deals with exchange and transmission of information within the whole organization [12], [13]. It concerns a great number of people and a great and varied amount of communication patterns and connections often occurs in it. Communication within an organisation takes place at two basic levels with regard to the number of people and amount of information it relates to, particularly formal communication, informal communication [12], [14].

If formal communication did not reflect management organisational structure it could happen that information necessary to solve a problem reaches a wrong recipient, arrives late or gets lost altogether [8], [15], [16].

In relation to management organisational structure and adaptation to formal communication network, we can talk about vertical, horizontal, diagonal organisational communication.

Vertical communication takes place in line with organisational structure. In interpersonal communication, it means that it takes place *between superior and subordinate employees* downwards as well as upward [12].

Top-down communication advances from a superior towards a subordinate, from a higher management level towards a lower one. It has a great meaning in assigning of tasks and specification of duties. It can be verbal or written, e.g. orders, personal conversations, telephone conversations, meetings, letters, circular letters, handbooks, guidelines, etc. Information flowing in this direction has a key meaning for subordinates in effective performance of their working

This work was supported in part by the project VEGA 1/0890/14 Stochastic Modeling of Decision-making Processes in Motivating Human Potential and in part by the project *IGP VŠEMVS 6/2015:* "The Analysis of Internal and External Environment of Small and Medium Enterprises under Conditions of the Slovak Republic"

Katarína Stachová is with Department of Management, School of Economics and Management in Public Administration in Bratislava, Furdekova 16, 85104 Bratislava, Slovak Republic (katarina.stachova@vsemvs.sk).

Monika Hudáková is with Department of Small and Medium Entrepreneurship, School of Economics and Management in Public Administration in Bratislava, Furdekova 16, 85104 Bratislava, Slovak Republic (monika.hudakova@vsemvs.sk).

Zdenko Stacho is with Department of Management, School of Economics and Management in Public Administration in Bratislava, Furdekova 16, 85104 Bratislava, Slovak Republic (zdenko.stacho@vsemvs.sk).

activities [8].

Bottom-up communication advances from subordinates towards superiors. Information flows through a formal channel, normally created according to management organisational structure. Information advances from the lowest, respectively a lower organisational level towards a higher, up to the highest level. This kind of communication represents, respectively contains verbal or written information and news regarding predominantly activities of the given organisational unit (however it does not always have to apply). Managers should realise that bottom-up communication is their valuable source of information from the position of subordinates. Managers can find out how tasks, activities, problems, etc. are perceived by their employees. This information can significantly influence their view of reality and help them decide. It at the same time prevents many problems, if subordinates show trust and provide necessary information already at the beginning of a problem, e.g. they do not try to hide insufficiencies [12].

Horizontal communication takes place between employees (groups of employees) at the same organisational level. It thus represents a communication of equal co-workers from the viewpoint of subordination and superiority. Exchange of information at the level of executive employees, who provide e.g. information on defects regarding the previous operation of a production line, or foremen in a processing plant or managers of individual sections, divisions etc. can be an example [8].

Diagonal communication is a diagonal information flow between employees at different organisational levels without a direct organisational relationship. Mainly verbal and written forms are used [12].

Although vertical communication is decisive for organisations, neither horizontal nor diagonal flows are negligible. On the contrary, they can have a great impact on their effectiveness. It is mainly important where character of work requires a common participation of several employees in a task solution, i.e. when cooperation is necessary. It is also used where information flow needs to be fastened, better understanding of a notification needs to be achieved, and common efforts need to be coordinated. The more independent work of organization's departments is, the more urgent and effective horizontal and diagonal communications can be. Horizontal and diagonal information is less "filtered" than information in vertical communication. It is an advantage that such obtained information is probably more complete than other kinds and enables to interpret its meaning according to the own concept [12].

Organisational informal communication works in organisations through informal communication channels. It is a direct result of the behaviour of people as such. There are sympathies and antipathies, friendships etc. at each workplace. These contacts do not reflect formal management organisational structure. Transmission of different information from different sources at various places in the organisation takes place in informal communication [8].

Many organisations do not approve of informal communication, as it restricts the extent of control over information flow and can at the same time significantly misrepresent information (or it can be made up). This informal information can at the same time contradict the formal one. Development of excessive informal communication can lead to an increase in non-productive time, a decrease in time for working duties fulfilment [17].

However, adequate informal communication can have a positive impact on organisation, as it can reveal channels which are covered but necessary [18]. Informal communication is in many cases fast and effective, and at the same time it fulfils the need of people to communicate. Manager should be able to use these advantages of informal communication and look for ways how to ensure a higher preciseness of information flowing through it [8], [19], [20].

The aforementioned shows that open effective communication is generally very important for success of innovations in organisations, however its impact is neither negligible in the need of involving the highest possible number of employees in the innovation process [14], [21], [22].

II. Materials and Methods Used in the Research

In order to determine a suitable research sample, two stratification criteria were set out. The first criterion was a minimum number of employees in the organisation, which was determined at 50 employees. The given stratification criterion excluded micro and small enterprises from the research on the one hand, however, on the other hand, the justness and need to focus on a formal system of human resources management in companies with more than 50 employees were observed and especially declared by means of this criterion. The second stratification criterion was a region of organization's operation, while the structural composition of the research sample was based on the data of the Statistical Office of the Slovak Republic).

According to the Statistical Office of the Slovak Republic the number of companies with a number of employees 50 and more was 3,261. The regional structure of companies with more than 50 employees is shown in Table 1.

Table 1 Regional structure of companies with more than 50 employees

Region	**Whole Slovakia**	**Western Slovakia**	**Central Slovakia**	**Eastern Slovakia**
Districts	All districts	Bratislava, Trnava, Trenčín, Nitra	Banská Bystrica, Žilina	Košice, Prešov
Number of companies	3,261	2,005	644	612

Source: data processed according to the Statistical Office of the Slovak Republic

Determining an optimal research sample of the given basic group of companies, Confidence Level of the research was set

at 95 %, and Confidence Interval of the research was set at H = +/- 0.10. On the grounds of the given criteria an additional, respectively relevant research sample for individual regions of Slovakia was set in the analysed years. It is shown in Table 2.

Table 2 Size of the research sample for individual regions of Slovakia

Region	**Western Slovakia**	**Central Slovakia**	**Eastern Slovakia**
Districts	Bratislava, Trnava, Trenčín, Nitra	Banská Bystrica, Žilina	Košice, Prešov
Number of companies over	2,005	644	612
Size of the research sample	92	84	83

Source: Own processing

Approximately 500 organisations were included in the research, however due to a great extent and the form of data collection only approximately 65 % of questionnaires used to be returned comprehensively completed. Subsequently, 259 organisations, corresponding to the optimal research sample determined on the grounds of stratification criteria, were selected from these organisations.

Key methods used in the conducted research include logical methods, adopting the principles of logic and logical thinking. Particularly the methods of analysis, synthesis, deduction and comparison were applied from this group of methods. Mathematical and statistical methods were also applied in the paper. From software products available on the market, a text editor, a spreadsheet and statistical software were used in the research work, particularly including SPSS 15.0 statistical software for Windows®.

III. Analysis of the Open and Broad Communication in All Directions

Within focusing of our research on the course and way of communication in organisations, we were finding out whether organisations have established a functional system of communication, whether horizontal or vertical, and whether they provide their employees the feeling of safety to such extent that they submit comments.

First of all, we focused on finding out the overall informedness of individual employee categories about organisation´s formal strategic information. Answers to question: ”To what extent do you use the following methods to communicate key information to your employees?” demonstrated that most preferred way of organisations to communicate key information to employees is verbal form, either at team meetings and meetings in 72 % or directly to a particular employee in 67 %. Written method is the second most preferred - in the form of either whole organisational electronic communication or directly to a particular employee. Least used method to communicate key information to employees is communication through a link element whether through an employee representative or a union authority. Some organisations stated that they also use notice boards and company magazine for the purpose of such communication.

Table 3 Extent of the usage of methods to communicate key information to your employees

Extent of the usage of methods to communicate key information	**%**
Through an employee representative or a union authority	23
In writing, directly to employees	67
Electronic communication	51
Verbally, directly to employees	71
Team meetings, meetings	72

Source: Own research

We subsequently focused on vertical bottom-up communication, mainly concerning provision of a safe room for employees to express their opinions about working issues as well as whole organisation. Like in vertical top-down communication, it is possible to use several ways, e.g.: employees can express their opinions to organisational management through a direct superior, trade unions, regular meetings and assemblies, boxes designed for it, surveys among employees, directly to a superior, etc. The research implies that methods mostly used by employees in interviewed organisations to communicate information to management are verbal communication methods, through a direct superior in 69 %. Other methods are represented in significantly lower amounts (Table 4).

Table 4 Extent of the usage of methods to communicate information by employees to management

Do you use the following methods to communicate information by employees to management?	**%**
Through a direct superior	69
Through representatives	26
Through a personnel employee	33
Through a survey	20
Anonymous box	13

Source: Own research

Whether an organisation provides its employees the feeling of safety in expressing their opinions and comments is mostly expressed by the way of their submission itself. In organisations we analysed, comments are most often submitted directly to a superior or at departmental meetings, where room is created for this purpose, or directly at managerial meetings. Unfortunately, the research showed that in 8 % of organisations, employees either do not submit comments at all, because they are afraid, or 10 % organisations have so called comment box used by employees to submit comments.

Table 5 Most often used forms of submitting comments in organisations

Most often used forms of submitting comments in organisations	**%**
are not submitted	10

openly told to a superior	43
room at departmental meetings	30
room at managerial meetings	15
anonymous box	9

Source: Own research

IV. Analysis of the Level of Communication in a Company

Questions focused on analysing the sphere "Open and broad communication in all directions" along with score evaluation of individual answers are provided in Table 6.

Table 6 Questions analysing the sphere of communication in a company, interconnected with score evaluation

Questions and answer choices	**Score**
Do company managers attend meetings focused on the quality of communication in the company?	
Yes, yearly	10
Yes, once during the time of their work at a managerial position	5
No	0
Do they copy channels through which formal communication flows, and corporate organisational structure?	
Yes	10
No	0
Are channels intended for formal communication strictly followed?	
Yes, formal communication only flows through designated channels and directions	10
There is an effort to communicate through these channels	5
No	0
How is vertical communication supported?	
It is supported by managers and managing employees in maximum possible extent	10
Managers and managing employees of the company only support it in a limited extent	5
It is minimal and is not supported by management	0
How is informal communication in the company affected by management?	
It is supported by managers and managing employees of the company, and it is led so that no misunderstandings caused by distorted information occur	10
It is tolerated by managers and managing employees of the company, however its content is not purposefully corrected	5
It is suppressed and limited in maximum extent by managers and managing employees of the company	0

Source: Author´s

Table 7 The level of communication in a company on the grounds of a sum of the scores of individual questions

Feature of an innovative industrial enterprise	**Your result**	**Your level**
Open and broad communication in all directions	50 – 40	A
	39 – 15	B
	0 - 14	C

Source: Author´s

Within the sphere "Open and broad communication in all directions", companies were divided into the following three groups:

A. Both formal and informal communications work perfectly in the company. Formal communication flows through channels which copy corporate organisational structure and are fully used to transfer all necessary information in both directions, i.e. from top management to performing employees and vice versa. Company managers use vertical communication at assigning tasks or specifying duties and they choose different ways of communication according to importance, necessary speed or extent of information. Both written and verbal communications are used, e.g. orders, personal conversations, telephone conversations, meetings, letters, circular letters, manuals, guidelines, etc. On the grounds of such quality formal communication, employees and teams can effectively fulfil their working activities and tasks. Company managers constantly support vertical communication, which is how they find out how innovations, tasks, activities, problems etc. are perceived by employees. Besides, formal channels also enable a quality transfer of information vertically between employees at the same level within a corporate organisational structure. This communication enables an effective transfer of information between performing employees, who provide each other information on the previous operation of a production line, or between individual foremen or managers of individual divisions, sections, etc. A diagonal transfer of information between employees at different organisational levels, between whom no direct organisational relationship exists is also at a high level, due to which information gets very quickly from a source to a target, undistorted. Informal communication is supported to a certain extent by managers and managing employees, especially in the sense that they try to specify information transferred by such communication so that no misunderstandings caused by distorted information occur.

B. Both formal and informal communications work in the company. Formal communication flows through channels which copy corporate organisational structure and are fully used to transfer all necessary information from top management to performing employees; however communication from performing employees to management is deficient. Company managers use different ways of vertical communication, i.e. at assigning tasks or specifying duties according to importance, necessary speed or extent of information. Both written and verbal communications are used, e.g. orders, personal conversations, telephone conversations, meetings, letters, circular letters, manuals, guidelines, etc. On the grounds of such quality formal communication, employees and teams can effectively fulfil their working activities and tasks. The fact that company managers only support vertical communication in a limited extent results in the fact that they do not have information on how innovations, tasks, activities, problems, etc. are perceived by employees, due

to which a certain tension is created in the company. Formal communication channels also enable a transfer of information vertically between employees at the same level within a corporate organisational structure; however it is not fully used, as employees are not supported in it, which leads to a weak feeling of companionship of employees to their company. A diagonal transfer of information between employees at different organisational levels, between whom no direct organisational relationship exists is rather limited; information often gets from a source to a target through a mediator, which causes a certain deceleration or distortion of information. Informal communication is also supported by managers and managing employees, however its content is neither monitored nor purposefully corrected, i.e. information thus flowing is frequently distorted and misleading.

C. Formal communication largely works in the company, however informal communication is suppressed. Formal communication flows through channels not precisely copying the corporate organisational culture, i.e. it sometimes happens that information necessary to solve a problem is not received by a correct receiver, it arrives late or gets lost, due to which employees and teams cannot effectively fulfil their working activities and tasks. Communication from performing employees to management is minimal and is not supported by management; anonymous box is frequently used as the only way of such communication. Employees are afraid to submit comments or express their opinions, which results in the fact that information on how innovations, tasks, activities, problems, etc. are perceived by employees remain unnoticed, due to which considerable tension occurs in the company. A vertical transfer of information between employees at the same level within a corporate organisational structure is not formally supported, i.e. information always needs to flow from top management downwards and back. A diagonal transfer of information between employees at different organisational levels, between whom no direct organisational relationship exists is not formally supported. Informal communication is suppressed and limited in the greatest possible extent by company managers and employees. It is reflected in avoiding discussion and any expression of own opinion. Employees know that if they want to retain their jobs, or advance in their careers, it is necessary to silently agree with everything their company does.

To reveal bottlenecks in the sphere of "Open and broad communication in all directions", Table 8 was created, from which it can be particularly specified which part of open and broad communication in all directions needs to be focused on in order to achieve a higher level in this sphere.

Table 8 Reveal bottlenecks in the sphere of communication in a company

Number of question / answer	1	2	3	4	5
Excellent	a	a	a	a	a
Average	b		b	b	b
Bad	c	b	c	c	c

Source: Author´s

The proposed methods enable a complex communication process evaluation in a company in a short time interval, while the results thus obtained by the analyzing employees help define an actual level of a company at a given time. Individual levels are described in the paper. Based on this analysis, companies are able to identify the bottlenecks preventing them from innovation potential development. The proposed methods help illustrate what policy and philosophy are actually applied and potentially enabled by management. Management thus obtains useful information on both practical results and problematic aspects of their current procedures and activities related to innovative company creation.

V. Conclusion

Key precondition for organisations intending to ensure sustainable development is a continuous development of its human potential, representing the ability of organisation to generate new ideas subsequently reflected in innovations, while ensuring the most important part – implementation of these innovations itself. Systematic creation and usage of human potential is a precondition of building and developing of strengths and competitive advantages of organisations. If an organisation wants to be "innovative" it should have several characteristic features. In our contribution, we focused on open and broad communication, since it has the most significant impact on formal as well as informal labour relations which have a fundamental role in team creation as well as information sharing and knowledge continuity.

In a questionnaire research we conducted at School of Economics and Management in Public Administration in Bratislava, we focused on finding out whether and in what extent organisations operating in Slovakia focus on communication. The analysis implied that organisations prefer a direct verbal form in top-down as well as bottom-up communication however employees did not submit comments because they were afraid to do so, or they only discussed them at the horizontal level in 10 % of interviewed organisations, which implies a great insufficiency in communication and thus in labour relations. With regard to the found facts, we proposed a simple method comprising three steps (Analysis of present level of communication in an organisation, Definition of the level of communication in an organisation, Specification of bottlenecks in the sphere of communication in an organisation), based on which organisations will be able to analyse their present level of communication as well as reveal their bottlenecks in this sphere. Practical justification of the given part of the research is mainly seen in analysing of the attitude of organisations operating in Slovakia to directing and advancing in the sphere of human resources management. On the grounds of our presentation of obtained results,

organisational managements have a possibility to compare their own present states in the given sphere with states declared by interviewed organisations and on its basis, to consider possibilities of its enhancement. Creation of the method is also considered as a benefit. We also consider as necessary to continue in this research in order to be able to enhance, modify, streamline and develop individual procedures on the grounds of new information obtained from interviewed organisations.

ACKNOWLEDGMENT

The article is related to VEGA 1/0890/14 Stochastic Modeling of Decision-making Processes in Motivating Human Potential and Grant Agencies of VSEMvs project, No 6/2015: "The Analysis of Internal and External Environment of Small and Medium Enterprises under Conditions of the Slovak Republic"

REFERENCES

[1] L. Cannavacciuolo, G. Capaldo, P. Rippa, "Innovation processes in moderately innovative countries," *The competencies of knowledge brokers Source of the Document International Journal of Innovation and Sustainable Development*, vol. 9 no. 1, 2015, pp. 63-82.

[2] A. Pilková, J. Papula, J. Volná, M. Holienka, "The influence of intellectual capital on firm performance among Slovak SMEs", *International Conference on Intellectual Capital, Knowledge Management and Organisational Learning. ACPI.* 2013 pp. 329-338.

[3] Z. Stacho, R. Stašiak-Betlejewska, "Approach of organisations operating in slovakia to employee performance evaluation," *Economic Annals-XXI*, vol. 5-6. 2013 pp. 83-87.

[4] Z. Stacho, H. Urbancová, K. Stachová, "Organisational arrangement of human resources management in organisations operating in Slovakia and Czech Republic," *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis.* vol. 61, no. 7, 2013, pp. 2787-2799

[5] B. Arogyaswamy, A. Nowak, "Innovation in information and communications technology in Poland: Prospects and strategies Business Transformation through Innovation and Knowledge Management," An Academic Perspective - Proceedings of the *14th International Business Information Management Association Conference, IBIMA,* 2010. Istanbul, Turkey 23-24 June 2010 pp. 1271-1286

[6] A. Ceschi, K. Dorofeeva, R. Sartori, "Studying teamwork and team climate by using a business simulation: How communication and innovation can improve group learning and decision-making performance," *European Journal of Training and Development* vol. 38 no. 3, 2014. pp. 211-230.

[7] I. Poláková, "Trainee program," *Modern management* (*Moderní řízení)*, Prague: Ekonomia, 2007. pp. 69-71.

[8] F. Horňák, D. Cagáňová, M. Čambál, "Development of Managerial Creativity," *Advanced Materials Research. Vol. 482-484* 3rd International Conference on Manufacturing Science and Engineering. 2012. pp. 996-999.

[9] K. Gubíniová, G. Pajtinková-Bartáková, "Customer Experience Management as a New Sourceof Competitive Advantage for Companies" The Proceedings of the *5th International Scientific Conference on Trade, International Business and Tourism „Application of Knowledge in Process of Business Dynamization in Central Europe"* Bratislava: Ekonóm, 2014. pp. 162-168

[10] J. Papula, J. Volná, "A Descriptive Analysis of Intellectual Capital Concept Implementation within Slovak Companies", *Driving the Economy through Innovation and Entrepreneurship: Emerging Agenda for Technology Management.* Springer, India, 2012, pp. 443-451

[11] I. Dudová, "Social Protection as the Key Pillar of Social Systems in the European Union," *Economic Annals-XXI* (*Ekonomičnij časopis XXI)*, no 7-8. 2014 pp. 36-39.

[12] M. Sedlák, *Manažment.* Bratislava: Iura Edition, 2009. 434 p.

[13] L. Oblak, L. Zadnik-Stirn, "The possibility of solving economic and environmental protection problems in wood industry companies by the application of the method of fuzzy goal programming." *Ekológia* Bratislava, vol. 19 no.4, 2000. pp. 409-419

[14] R. Kampf, M. Hitka, M. Potkány, "Medziročné diferencie motivácie zamestnancov výrobných podnikov Slovenska," *Journal Communication (Časopis Komunikácie)* vol. 4, 2014, pp. 98-102.

[15] R. Bačík, R. Štefko, J. Gburová, "Marketing pricing strategy as part of competitive advantage retailers," *Journal of Applied Economic Sciences.* vol. IX, no 4 (30), 2014., pp. 602-607.

[16] A. Remišová, Z. Búciová, "Measuring corporate social responsibility towards employees," *Journal for East European Management Studies*, vol. 17, no. 3, 2012, pp. 273-29.

[17] H. Urbancová, "Results of analysis of organisational culture in organisations in the Czech Republic and Slovak Republic" *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis.* vol. 60, no. 7. 2012, pp. 433-440.

[18] Y. Awazu, P. Baloh, K.C. Desouza, C.H. Wecht, J. Kim, S. Jha, "Information - Communication technologies open up innovation," *Research Technology Management, vol.* 52, no. 1, 2009. pp. 51-58.

[19] M. Hudáková, "Komunikácia verejného a súkromného sektora v krízových situáciách SR," *Ekonomika poľnohospodárstva*. Vol.8, no 3, 2008. pp. 55-59.

[20] J. Šujanová, P. Gabriš, M. Ličko, P. Pavlenda, R. Stašiak-Betlejewska, "Aspects of Knowledge Management in Slovak Industrial Enterprises," *Proceedings of the 13th European Conference on Knowledge Management.* vol. 60, no. 4, 2012, pp. 1135-1144.

[21] J. Závadský, M. Hitka, M. Potkány, "Changes of employee motivation of Slovak enterprises due to global economic crisis" *E+M Economics and Management* vol. 1, 2015. pp. 57-66

[22] M. Hudáková, "Komunikácia v krízových situáciách" *Regióny - vidiek - životné prostredie - I. časť* Nitra: SPU, 2006. pp. 155-159.

Performance Evaluation of the DNP3 Protocol for Smart Grid Applications over IEEE 802.3/802.11 Networks and Heterogeneous Traffic

Alcides Ortega, Ailton A. Shinoda, Christiane M. Schweitzer, Fabrizio Granelli, Aleciana V. Ortega, and Fabiola Bonvecchio

***Abstract*—**The Distributed Network Protocol Version 3.0 (DNP3) is a communications protocol used between components of a power grid system. DNP3 is designed to operate in network scenarios with high node density and can be used for both wired and wireless communications. Recently, with the evolution of power grid systems towards the smart grid, the use of DNP3 has been proposed for smart grid applications. Unfortunately, there exist few studies that evaluate the performance of DNP3 by means of computer-based simulations based on open-source software. Therefore, this paper proposes an evaluation of the performance of DNP3 over a mixed (wired/wireless) network between IEEE 802.3 and IEEE 802.11b, encapsulated over TCP/IP, using the Network Simulator Version 2.35 (NS-2). The purpose of the paper is to investigate the feasibility of using DNP3 over a network carrying heterogeneous traffic, such as monitoring and teleprotection, by measuring the delay required to complete a messaging operation. The simulation results show that DNP3 represents a feasible communications protocol for smart grid applications, which achieves delays shorter than 12 ms for teleprotection and below 90 ms for monitoring.

***Keywords*—**EEE 802.3, IEEE 802.11b, IEEE 1815 DNP3, NS-2, Simulator, Smart Grid.

I. Introduction

THE smart grid is a new concept that combines enhanced power grid systems and advanced techniques of Information and Communication Technologies (ICT) to monitor and man- age the transport of electricity from distribution substations of utilities to end consumers in real time [1].

Communications in the smart grid architecture are fundamental as they involve many actions of control and monitoring, for the proper operation of the electric systems [2]. Such communications are performed through a telecommunications infrastructure that can mainly be composed of wired and wireless communications systems [3].

The choice of wired and/or wireless communications by the utilities is based on various factors, such as geographic characteristics of the area of deployment and deployment and operational costs [4]. For instance, the uses of wireless communications in the smart grid system provide a lower cost and speed of deployment and portability when compared to wired communications. In this sense, wireless communications have been applied to sensors, IEDs (Intelligent Electronic Derive) and RTUs (Remote Terminal Unit) [5], [6].

A communications protocol that has recently been considered for smart grid applications and can be used for both wired and wireless communications is the IEEE 1815 DNP3 [7]. This is an open protocol for control and monitoring of the electric components of a power grid system. The DNP3 protocol is non-proprietary and standardized and optimized for low communication overhead, being one of the most widely used in the world, especially in North America, Latin America, Asia and Australia. The reasons for such wide adoption of DNP3 are due to its high security, interoperability, and adaptation to different applications, such as electric power substation automation, water plants treatment, and sewage, oil and gas system [8], [12].

The performance of DNP3 for smart grid applications has been evaluated in several studies. For instance, a real-setting testbed for research was developed and described in [9] to measure the messages delivery delay in a mixed communication of DNP3 over Transport Control Protocol/Internet Protocol (TCP/IP). Experimental results showed that DNP3 is in line with the requirements of the relay protection. Also, the authors claimed that DNP3 over TCP/IP is a suitable solution for real-time monitoring applications. However, it shows performance limitations to support time-critical applications.

Since experimental research for the smart grid requires a deep knowledge of hardware programming and a strong investment in hardware equipment to set up a real-setting scenario, most of works on the performance evaluation of DNP3 have been based on computer-based simulations. Unfortunately, these works employ licensed software, which does not allow new implementations or add new functions by non-proprietary partners in order to perform a simulation of DNP3 in a more realistic scenario of smart grid. Since licensed software is expensive and only proprietary partners can accomplish the update of new features, open-source software has been widely used for many years to conduct research in

This work had the financial support of the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, 2014/06022-9) and Universidade Estadual de Mato Grosso do Sul (UEMS).

Alcides Ortega, Ailton A. Shinoda, Christiane M. Schweitzer, Aleciana V. Ortega is with UNESP, Campus Ilha Solteira, SP Brasil (e-mail: ortega.tex@gmail.com, shinoda@dee.feis.unesp.br, chris@mat.feis.unesp.br, aleciana@gmail.com).

Fabrizio Granelli is with DISI, University of Trento, Italy (e-mail: granelli@disi.unitn.it).

Fabiola Bonvecchio is with University of Leoben, Austria (e-mail: fabiola.bonvecchio@stud.unileoben.ac.at).

both academia and research communities.

The NS-2 is a widely used open-source simulation framework that has been implemented and improved with the cooperation of many partners during many years. NS-2 allows abstraction of all communication protocols and their performance evaluation for different network topologies and configuration of various network traffic types.

In our previous work presented in [10], we evaluated the performance of DNP3 in an IEEE 802.3 Ethernet network encapsulated over TCP/IP by means of computer-based simulations using NS-2. Since NS-2 is based on open-source software, it allowed us to implement a new framework (or patch in the NS-2 terminology) in order to test the main functions of DNP3 in several scenarios. The functions that can be evaluated through simulations in NS-2 are: message request from a master station of outstations, transmission and retransmission of unsolicited messages sent from outstations, transmission and retransmission of unsolicited messages with read function, transmission and retransmission of unsolicited messages with different function of reading and transmission and retransmission restart command of outstations.

In this paper, we extend the work presented in [11] by evaluating the performance of DNP3 in a smart grid scenario composed of both wired and wireless communications systems using the IEEE 802.3 and IEEE 802.11b Standards, respectively, with. The simulated scenario considered for the work presented in this paper integrates heterogeneous traffic for monitoring and teleprotection applications, containing DNP3 packets and TCP packets, to verify analyze the performance as the number of retransmissions for packet loss and latency threshold, in which it was found that used in a wired network the DNP3 works well for monitoring applications, up to the limit of use of 85% of the bandwidth of a network and above this percentage without quality of service is unfeasible because the packet lost, re-transmissions and latency are significantly increased.

A. Simulation Model

A new scenario was considered to perform the Simulation of a heterogeneous network using wired networks and wireless, through the results of this Simulation will be possible to analyze the performance of the operation of the DNP3 in a scenario involving a control center and two substations of power electric distribution.

The configuration of this scenario are the following: the first substation is located to 22,8km of the control center, while the second is located to 21,3km and in both there is a base station that connects to the control center via the wired technology. In the first substation there are five outstations being that of them four use the wireless communication with the base station and one with wired, in the second substation there are four outstations being that three of which use wireless communication with the station base and one with wired.

In the first simulation was held the transmission of data packets of DNP3 only and in the second part of the simulation to evaluate both the performance of DNP3 in a mixed data traffic, with the results will be possible to analyze the possibility of use on the same network and packet transmission DNP3 with other packet and use the network more efficiently.

This work is organized into sections. Section II presents the DNP3 Protocol. The simulations and their results in throughput, delay, are described in section III. Finally, section IV presents the conclusions.

II. DNP3 Protocol

The DNP3 is an open communication protocol and optimized developed for the SCADA (Supervisory Control and Data Acquisition) that is used to communicate among power grid equipment, with two classes of devices defined. Central stations (Masters) are usually devices with some processing power and data storage. Slave stations (Outstation) are devices located in field (transmission lines, transformer substations) to collect data from sensors and send the central station [13].

DNP3 protocol was initially designed to operate in traditional power grid and through the serial links. Recently, with the migration from traditional power grid to power grid systems, the reuse of existing communication protocols is widely regarded as a solution to cost-efficient and compatible with previous versions, and how the DNP3 does not specify its own network and lower layers, the DNP3 protocol over TCP/IP has been proposed as a communication protocol for smart grid, because is designed to be as robust as possible in regards to detecting and recovering from error [12]. DNP3 can be used in all kinds of communication be it wireless or wired [7].

A. Protocol Topology

The DNP3 standard defines four types of architecture:

Point to point: it is the simplest architecture, wherein a Master station communicates with a single remote station (Outstation).

Multipoint: a central station communicates with several slave stations. The communications for data requests are performed between the central station and each slave station in sequential mode. Each slave station monitors the messages from the central station and only responds when the destination address of the received message matches its address.

Hierarchical: it consists of a master that communicates with a slave station that in turn becomes a central station of another station. In this case, the second slave station gets its name from the central substation or Submaster.

Data concentrator: this topology can handle different protocols where the DNP3 may be running in a central station or in a central substation. The data concentrator collects information from various IEDs to be transmitted to the central station [7].

III. Results and Discussions

In this section we introduce the test scenario used in this paper and the setup of our simulations and discussions of the results.

Simulation Environment: Simulations are carried out in the

open-source network simulator NS-2 version 2.35 compiled into Fedora Core Linux operating system.

Traffic model: the DNP3 traffic generator with 292 bytes of information encapsulated in TCP/IP protocol, are added to the 20 bytes of the TCP header and 20 bytes of IP header totaling a data packet of 332 bytes, sent every 2 s of slave stations (outstations) to the master station and a data packet with confirmation of receipt sent to the master station to the outstation to the size of 40 bytes, the function of the DNP3 protocol used was the sending unsolicited messages, this function and used in DNP3 for when there is any change of state in any monitored equipment is automatically sent to the control center, are also used generator traffic TCP with packet size of 1040 bytes and 40 bytes are confirmation that the TCP ACK. The TCP traffic was used for simulating the utilization of FTP (File Transfer Protocol) that are used for file transfer rate between computers.

Simulation Scenarios: In this job we present the communication between the substations of distribution of electric power to the control center. The substations are located generally in the entrance of cities and the control center within the town, the Fig. 1 shows the scenario with two substations for the distribution of electric power and a control center, the first substation are five node, four node are IEDs (outstation) and one node representing a data server, three of the four outstation communicate with a base station (BS) via wireless and one via wired, communication data server and wireless too.

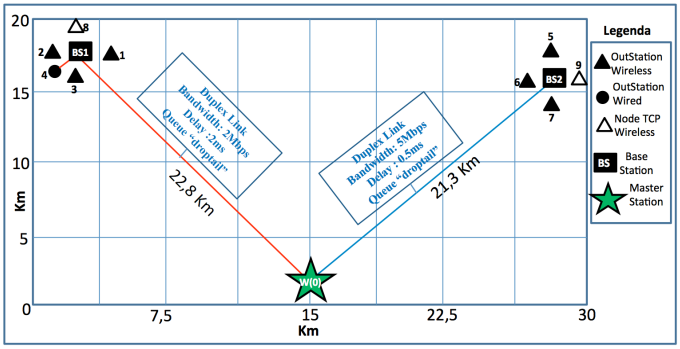


Fig. 1 Simulation scenario.

Communication with the BS and control center located 22,8 km and through wired, while the second substation is located 21,3 km from central control, this substation there is a BS connected via wireless with three outstations and a data server. Our goal was to make the time measurements of latency and amount of data packets retransmitted the DNP3 protocol in a mixed network consisting of 802.3 and 802.11b network from outstations to the master station at the control center in the second experiment we used the same scenario but with the traffic of heterogeneous data packets containing TCP and DNP3 data in order to check the latency time and the amount of lost packets.

Table I presents the configuration and composition of the simulated network topology, as well as the distance and the type of network interface connection used between the node and the BS, type of queue and packet traffic. The central station is connected to the gateway W(0), the communication of the nodes to the gateway are performed through the BS. In this Table the outstation are the IEDs and the nodes are the data server, the topology of the network and connections.

On NS-2 all transport protocols are defined through an agent. In Fig. 1, using the parameters of Table I, have seven agents defined in outstations: TCP/SimpleTcp, two agents TCP/Newreno defined on node8 and node9 and two TCPSink defined in node W(0).

TABLE I- ASSUMED NETWORK PARAMETERS BETWEEN NODE AND BS.

Node	Network Interface	BS	Distance	Traffic	Queue
Outstation1	Wireless	1	300 m	DNP3	Droptail
Outstation2	Wireless	1	200 m	DNP3	Droptail
Outstation3	Wireless	1	200 m	DNP3	Droptail
Outstation4	Wired	1	10 m	DNP3	Droptail
Outstation5	Wireless	2	300 m	DNP3	Droptail
Outstation6	Wireless	2	200 m	DNP3	Droptail
Outstation7	Wireless	2	100 m	DNP3	Droptail
Node8	Wireless	1	300 m	FTP	Droptail
Node9	Wireless	2	300 m	FTP	Droptail

In this simulation was used the proactive DSDV (Destination Sequenced Distance Vector) Protocol maintains a table of possible routes for packet traffic throughout the network, all nodes employ wireless DSDV Protocol. In the Table II presents the configuration assumed between the connection of the BS at node w(0), as the type of traffic, queue, distance, latency, packet traffic and bandwidth.

TABLE II- ASSUMED NETWORK PARAMETERS BETWEEN BS AND GW.

BS	Duplex Link	Latency	Distance	Traffic	Queue	Gateway
1	2 Mb	2 ms	22,8 Km	DNP3 FTP	Droptail	W(0)
2	5 Mb	0,5 ms	21,3 Km	DNP3 FTP	Droptail	W(0)

Table III presents the main parameters of the considered network model in simulation of heterogeneous packet transmission, in the range of the gateway W(0) to the BS, and BS until wireless nodes.

Simulation of TCP application: TCP agent attached to the node W(0) and another agent of reception (Sink) linked to wireless Node8 and Node9. In addition, the associated application between the node W(0) and the wireless node Node8 and Node9 is the FTP. Agent Sink receives the TCP packets and generates recognition packages ACKs.

The advantage of TCP for applications is the versatility and robustness of this protocol, making it suitable to global networks, since it verifies that the data is sent correctly, in the proper sequence and no mistakes over the network, ensuring the transmission of packets.

One of the most well known uses of TCP protocol is cyber applications, such as SSH (Secure Shell), FTP, HTTP (Hyper Text Transfer Protocol), among others.

TABLE III - MODELED NETWORK PARAMETERS.

Parameter	Value
Channel	WirelessChannel
Propagation	TwoRayGround
Network interface	Wireless Phy
MAC layer	802_11
Queue type	*DropTail*
Link layer	LL
Model antenna	OmniAntenna
Maximum packet in the queue	60
Number of wireless nodes	8
Base Station	2
Number of wired node	2
Routing protocol	DSDV
Coverage area (XxY)	20km x 30km
Frequency	2.4 Ghz
Transmission rate	11Mb
Packet size	292, 1040 bytes
Type of application	DNP3, TCP
Event	Unsolicited messages

Simulation of DNP3 application: In Fig. 2 are illustrated the node and the agents (TCP/SimpleTcp) and application (TCPAppmod) linked in each node to perform the simulation of DNP3 in NS-2.

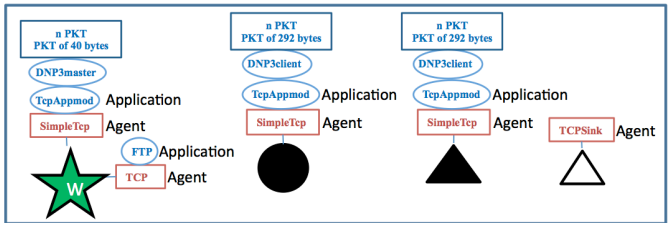


Fig. 2 Agents and applications running on each node type in DNP3 in NS-2.

The DNP3master receives the DNP3 protocol packets and after acknowledgment the TCPAppmod generates the DNP3 recognition packages with the size of 40 bytes. The advantage of this patch with the TCPAppmod for applications is the versatility and safety, making it suitable for global networks, since it verifies that the data is sent correctly and in the proper sequence with a specified time for receipt of the confirmation packages.

Table IV presents the time requirements to delay delivery of a message according to each type of application on the smart grid system.

TABLE IV – REQUIREMENTS OF MESSAGE DELIVERY DELAY [14].

Type	Delivery Delay	Application
Protection	3 ~ 16 ms	Trip, Closing, Reclosing.
Real-time monitoring	16 ~ 100 ms	State reporting
Low speed	≥ 100 ms	File transferring

Simulation result of throughput of the transmission of the DNP3 packet: Fig. 3 shows the flow of packets transmitted from Outstation1 to Outstation7 for the Master station. With the network parameters of Table I to Table III applied to the scenario of Fig. 1, the DNP3 traffic is generated every 2 s with a packet size of 292 bytes encapsulated in TCP/IP protocol, totaling 332 bytes of the packet transmitted, the yield expected average is about 166 bytes/sec (0,166 Kbytes/s).

The graph shown in Fig. 3 in which the Y axis shows the throughput in Kbytes/s and the X axis simulation time in seconds', we note that there is no variation of throughput of packets with the same trafficking in heterogeneous packets, keeping the same throughput rate of (0,166 Kbytes/s), all the outstation of the station master, was generated a thousand packet were in each outstations, and the master station received a total of 7000 packets, the throughput is constant in simulation.

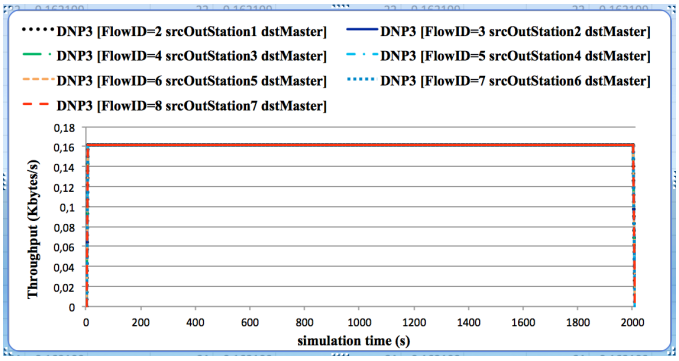


Fig. 3. Average flow of the seven statistical Outstation.

Fig. 4 illustrates the DNP3 delivery traffic delay of outstation to the master station by the BS1. The greater distance between the outstations and the master station is illustrated in Fig. 1 and specified in the Table I, the outstation which is farthest from the BS1 is the outstation1 with 300m and with the lowest distance is the outstation4 with 10m, noting that the outstation1 uses wireless technology and the outstation4 wired.

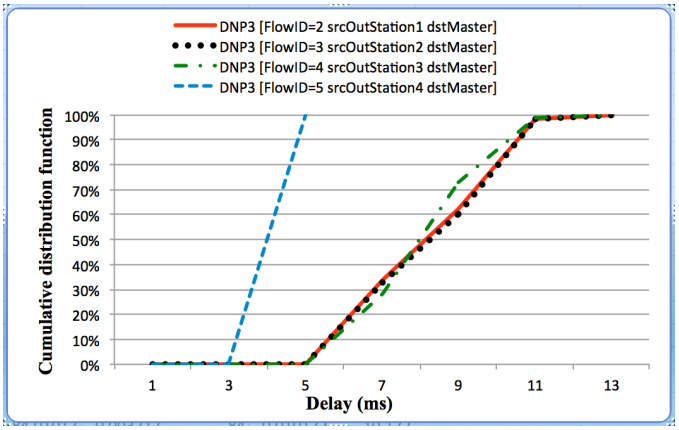


Fig. 4 Delay of Outstations from BS1 without TCP traffic.

The expected result was that the delay time of the outstation4 would be the lesser of the other outstation, because besides stay nearest and wired, and the graph show exactly this expected result. Analyzing the results of the graphics with the normal traffic of DNP3 without the FTP traffic, the results demonstrate that there was a perceptible difference between the outstation with wired and wireless, but between wireless outstations there was not perceptible differences in delay time, even with one difference in distances between them.

The results obtained through the cumulative distribution function (CDF) in a sample of 1000 packets, demonstrated through the graph curves that the trend is that packets sent by outstations with wired network and those who are closest to

BS1 and not using one heterogeneous network has the lowest delay time. The graphs illustrate the results exactly than 90% of the total sent packets from the outstation4 has the lower time delay longer than 5 ms.

While the other outstations using wireless networks to communicate with the BS1 and then wired up to the master station has a longer delay time in relation to wired outstation in graphic can be observed that 90% of the total packets sent from outstation1 to outstation3 has the time delay greater than 5 ms and less than or equal to 11 ms.

Fig. 5 illustrates the delay in delivery of the DNP3 traffic of outstation to the Master station via the BS2. The distance between the outstation's and the Master station, the greater distance is between the outstation5 until the BS2, nearest the BS2 is the outstation7 with 100 m.

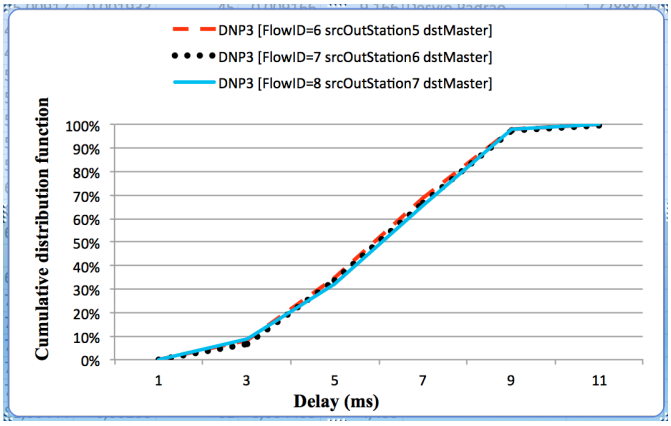


Fig. 5 Delay of Outstations from BS2 without TCP traffic.

The results obtained through the CDF in a sample of 1000 packets demonstrate through the curves that the trend is that packets sent by outstations are more alike and with few differences, in General 90% of packets sent has a delay less than 9 ms.

Fig. 6 shows the comparison of the average statistical throughput or instantaneous flow (measured at each 1 s) of packets transmitted from node wired W(0) to BS. This, in turn, makes the routing of packets to the Node8 and Node9 of the TCP, the simulation for both began the 2 s and finished in 2004 s.

The graph presents the result of simulations of TCP traffic, the black line identified in the graph and the result of the flow of TCP traffic from Node8 to the BS1 and BS1 until the Gateway W(0) with a transfer rate of 540 Kb. It should be noted that the average flow of the TCP traffic from BS1 was 240 Kbytes/s and no changes TCP traffic already identified with the color blue.

A result of the TCP traffic throughput from Node9 to the BS2 and the BS2 until the Gateway W(0), also with a transfer rate of 512 Kb. It should be noted that the average throughput of the TCP traffic from sub BS2 was 275 Kbytes/s and with few fluctuations during the simulation due to discard packets.

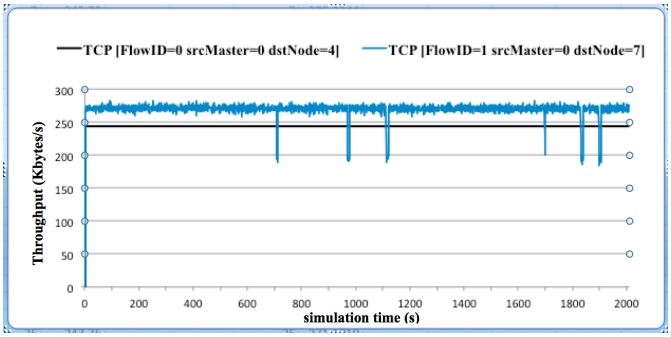


Fig. 6 Average throughput statistic of TCP traffic.

The results showed that using the TCP transport protocol and DSDV routing protocol, the packages were led with an average and instantaneous throughput almost constant for the wireless network, and node8 with a few variations along the node9 simulation.

Fig. 7 illustrates the delay in delivery of the DNP3 traffic of outstation to the Master via the BS1 with TCP traffic with a 512 Kb rate. Note that the outstation4 who obtained 90% of packages delivered less than 4 ms with TCP traffic crossed the 4 ms to 77 ms.

The others which obtained the time of delay in the delivery of packages with 90% less than or equal to 10 ms passed for 91 ms. with the result of simulations with TCP traffic, the delay time only serves to real-time monitoring as well as for data acquisition from sensors that are not employed to teleprotection, because in this case the maximum delay time allowed is up to 16 ms.

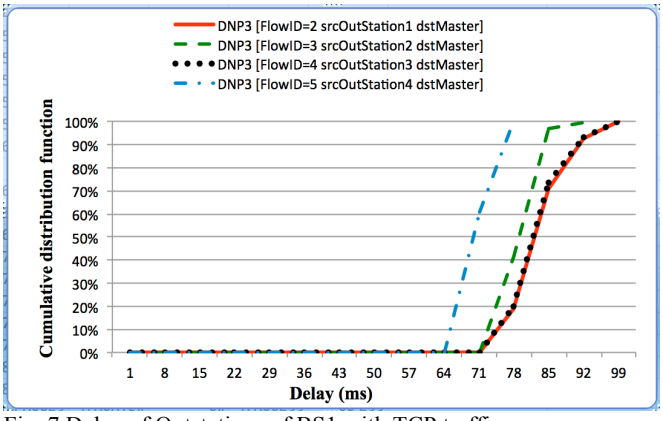


Fig. 7 Delay of Outstations of BS1 with TCP traffic.

Fig. 8 illustrates the delay in delivery of the DNP3 traffic of outstation to the Master station by the BS2 with the introduction of the TCP traffic.

Note that the results, obtained previously with 90% of packages delivered and confirmed with a shorter 9 ms, passed to 75 ms in/from the outstation8 that is the closest to the BS2, others passed 84 ms. The results of the simulations demonstrate that with the TCP traffic is possible DNP3 communication for real-time monitoring.

ISBN: 978-1-61804-318-4

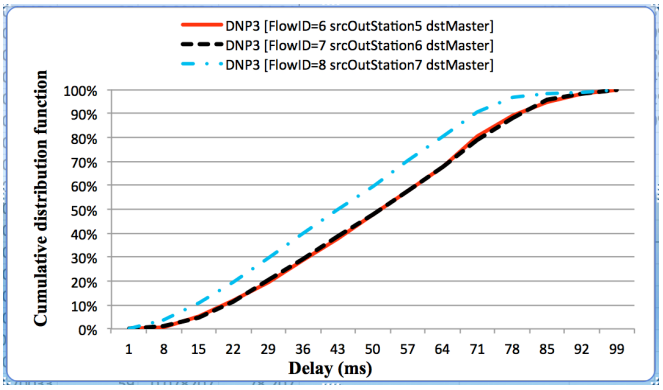


Fig. 8 Delay of Outstations of BS2 with TCP traffic.

Fig. 9 illustrates the delay in delivery of TCP traffic packages identified in red color of the node8 of substation1 resulting in a line with the time delay of 78 ms, the color identified with the blue Node9 of BS2 in form resulted in the delay time sine of 85 ms.

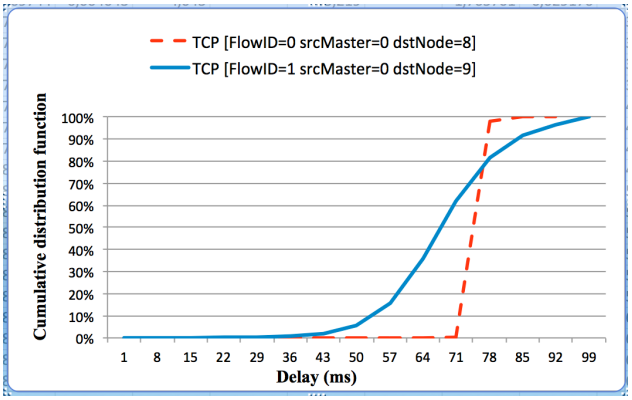


Fig. 9 Delay of node8 of substation1 and node9 of substation2.

IV. Conclusions

This work aimed to perform simulations in a scenario involving two substations and one control center. It is through the simulation scenario in NS-2 analyzing the performance and viability of the use of the DNP3 under TCP/IP in a mixed network with normal traffic and with heterogeneous traffic, using the FTP data traffic through the TCP transport protocol.

The results of the simulations showed that when the DNP3 used in a mixed network between IEEE 802.3 and IEEE 802.11b over TCP/IP even long distance are feasible for tele protection. The results of the delay was under 16 ms, the best case was obtained with the wired node reaching the delay time below 5ms, already with the wireless node was also obtained good results with time delay under 11 ms.

With heterogeneous traffic is infeasible to use for tele-protection, because the delay times were above 16 ms, but can be used for real-time monitoring and data acquisition of smart meters because the times obtained were below 100 ms around 91ms in case higher than forecast.

As future work is to be performed a simulation with mixed networks with heterogeneous traffic with QoS (quality of service), also using the WiMax network.

Acknowledgment

This study was supported by grants of Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, 2014/06022-9), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (bolsista da CAPES - Proc. nº 99999.002669/2014-05) and Universidade Estadual de Mato Grosso do Sul (UEMS).

References

[1] R. Hassan and G. Radman, "Survey on smart grid," in *IEEE Southeast-Con 2010 (SoutheastCon), Proceedings of the*, March 2010, pp. 210–213.

[2] X. Fang, S. Misra, G. Xue, and D. Yang, "Smart grid — the new and improved power grid: A survey," *Communications Surveys Tutorials, IEEE*, vol. 14, no. 4, pp. 944–980, Fourth 2012.

[3] Zheng, D. Gao, and L. Lin, "Smart meters in smart grid: An over view," in *Green Technologies Conference, 2013 IEEE*, April 2013, pp. 57–64.

[4] X. Lu, W. Wang, and J. Ma, "An Empirical Study of Communication Infrastructures Towards the Smart Grid: Design, Implementation, and Evaluation," *Smart Grid, IEEE Transactions on*, vol. 4, no. 1, pp. 170-183, March 2013.

[5] Parikh, M. Kanabar, and T. Sidhu, "Opportunities and challenges of wireless communication technologies for smart grid applications," in *Power and Energy Society General Meeting, 2010 IEEE*, July 2010, pp. 1–7.

[6] H. Kanchev, D. Lu, F. Colas, V. Lazarov, and B. Francois, "Energy management and operational planning of a microgrid with a PV-Based Active Generator for Smart Grid Applications," *Industrial Electronics, IEEE Transactions on*, vol. 58, no. 10, pp. 4583–4592, Oct 2011.

[7] G. Clarke, D. Reynders, and E. Wright, *Practical Modern SCADA Protocols: DNP3, 60870.5 and Related Systems*, 1st ed., ser. Engineering: instrumentation & control. Elsevier, 2004, no. pp. 66-164.

[8] IEEE, "IEEE draft standard for exchanging information between networks implementing IEC 61850 and IEEE std 1815 (distributed network protocol - DNP3)," *IEEE P1815.1/D4.00, June 2012*, pp. 1-283, June 2012.

[9] R.Lasseter,J.Eto,B.Schenkman,J.Stevens,H.Vollkommer,D.Klapp, E. Linton, H. Hurtado, and J. Roy, "CERTS Microgrid Laboratory Test Bed," *Power Delivery, IEEE Transactions on*, vol. 26, no. 1, pp. 325-332, Jan 2011.

[10] A.Ortega and A.Akira Shinoda, "Simulationin NS-2 of DNP3 protocol encapsulated over TCP/IP in smart grid applications," in *Innovative Smart Grid Technologies Latin America (ISGT LA), 2013 IEEE PES Conference On*, April 2013, pp. 1–8.

[11] A. Ortega, C. Schweitzer, and A. Akira Shinoda, "Performance analysis of smart grid communication protocol DNP3 over TCP/IP in a heterogeneous traffic environment," in *Communications and Computing (COLCOM), 2013 IEEE Colombian Conference on*, May 2013, pp. 1–6.

[12] S. Bagaria, S. Prabhakar, and Z. Saquib, "Flexi-DNP3: Flexible distributed network protocol version 3 (DNP3) for SCADA security," in *Recent Trends in Information Systems (ReTIS), 2011 International Conference on*, Dec 2011, pp. 293–296.

[13] IEEE, "IEEE standard for electric power systems communications-distributed network protocol (DNP3)," *IEEE Std 1815-2012*, pp. 1–821, Oct 2012.

[14] X. Lu, Z. Lu, W. Wang, and J. Ma, "On Network Performance Evaluation toward the Smart Grid: A Case Study of DNP3 over TCP/IP," in *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, Dec 2011, pp. 1–6.

Autoregressive Model of Channel Transfer Function for UWB Link inside a Passenger Car

Aniruddha Chandra, Pavel Kukolev, Tomáš Mikulášek, Aleš Prokeš
Department of Radio Electronics, Brno University of Technology, Brno 61600, Czech Republic.
E-Mail: aniruddha.chandra@ieee.org

***Abstract*—This article reports results of channel frequency transfer function measurements inside the passenger compartment of a four seated sedan car. The examined frequency range spans from 3 GHz to 11 GHz and covers the ultra wide band, a promising candidate for future automotive wireless standards. It is found that the complex transfer function may be decomposed into two terms, the first one being a real valued long term trend that characterizes frequency dependency with a power law, and the second term forms a complex correlative discrete series which may be represented via an autoregressive model. The proposed model is validated by comparing the simulated transfer function with the measured data. Simulated values for the coherence bandwidth, power delay profile, and the root mean square delay spread are also in good agreement with the experimental values.**

***Index Terms*—AR model, UWB, transfer function, frequency dependency, intra-vehicle.**

I. Introduction

A. Background

In the recent past, a number of ultra wide band (UWB) link measurements in passenger cars were being carried out [1]. Due to its large dynamic range, a vector network analyzer (VNA), is often preferred for such measurements. The two requirements for VNA based setups: transmitter (Tx) and the receiver (Rx) antennas should be within cable length, and the channel should be static, are also satisfied for in-car sounding experiments. A VNA provides channel transfer function (CTF) in the frequency domain, and proper characterization of the measured CTF is crucial for analysis of experimental data.

B. Contributions

- We propose an autoregressive (AR) process for channel frequency transfer function modelling of UWB links in car. To the best of our knowledge, this has not been attempted so far.
- We demonstrate that the AR process may be applied after removing the long term frequency trend from the transfer function. The method is different from earlier work on characterizing frequency dependency of intra-vehicular wireless channels, such as [2], where only simple models of large scale frequency variation was reported.

C. Organization

The next section provides description of the experimental setup and discusses the transfer function modelling in detail. The simulated transfer functions, following the model developed in Section 2, are validated against the measured data in Section 3. Finally, Section 4 concludes the paper and provides some directions for extending the work.

II. Measurement and Modelling

A. Experimental Setup

The measurements were performed inside a four seater passenger car, Skoda Octavia, parked in the basement garage. A 4 port VNA (Agilent E5071C) swept the entire UWB frequency range from 3 GHz to 11 GHz, with a frequency step of 10 MHz, creating a $N = (11-3)\times10^9/(10\times10^6)+1 = 801$ point data set for each run. Port 1 and port 4 of the VNA were connected to the Tx and the Rx antennas, respectively, and the forward transmission coefficient, s_{41}, approximates the CTF, $H(f)$. A pair of identical conical monopole antennas were used as Tx and Rx, which were connected to the VNA through phase stable coaxial cables. The gain of a conical monopole antenna in the frequency range of interest (3-11 GHz) is almost constant [3]. Thus it is possible to analyze the measured CTF without considering antenna effects. The experimental setup and antenna positions inside the passenger compartment are shown in Fig. 1.

B. Long Term Variations

The magnitude of CTF has a overall downward slope with respect to frequency and the first step of our modelling involves separating this long term variation or trend, i.e. we express the complex CTF as

$$H(f) = \tilde{H}(f) \cdot |H(f)|_{\text{trend}} \tag{1}$$

where $\tilde{H}(f)$ denotes the complex CTF after de-trending. The well known free space path loss formula suggests that the CTF is inversely proportional to frequency [4]. For real life wideband propagation a simple power law

$$|H(f)|_{\text{trend}} = K \cdot f^{-m} \tag{2}$$

is generally used [5]. There also exists another exponential model [6]

$$20 \cdot \log_{10} |H(f)|_{\text{trend}} = K' \cdot \exp(-m'f) \tag{3}$$

In Fig. 2 we depict a typical measured CTF and the corresponding trends with least mean square error fitting. The root mean square error for both the trends are comparable for the whole set of experimental data. However, we considered the power law given by (2) over the exponential one in (3) as

ISBN: 978-1-61804-318-4

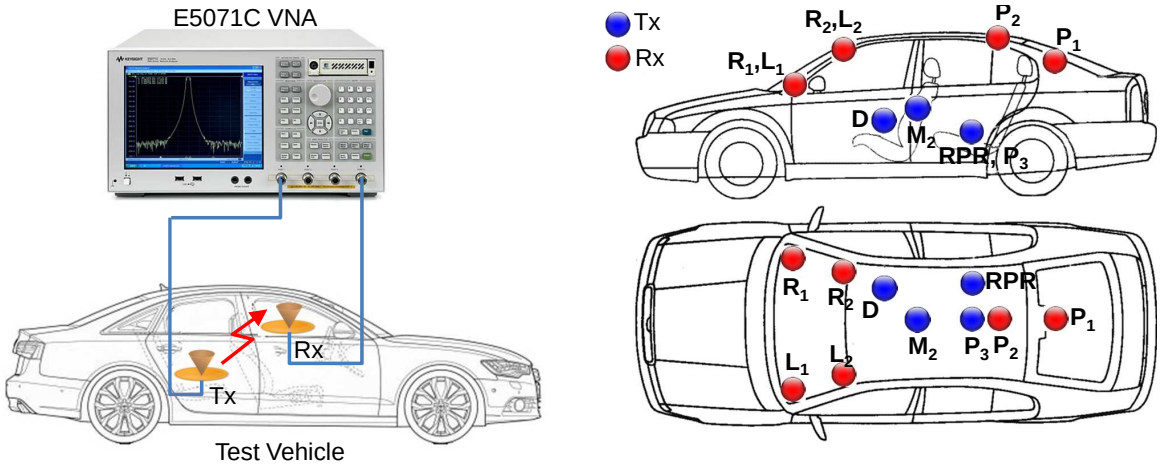


Fig. 1. Measurement setup (left) and antenna placement inside car (right).
Tx legends - D: driver, RPR: rear passenger on right, P_3: middle of backseat, M_2: midpoint between two front seats.
Rx legends - L_1: left dashboard, R_1: right dashboard, L_2: left windshield, R_2: right windshield, P_1 and P_2: positions at rear part of the ceiling.

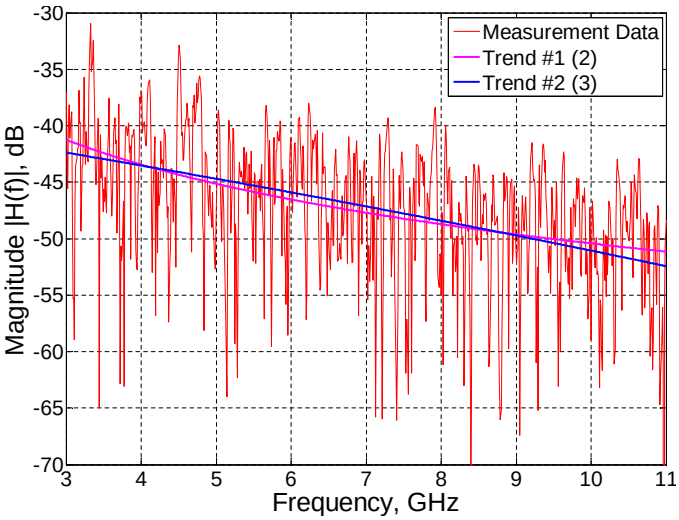


Fig. 2. CTF and estimated trends (Tx position: P_3, Rx position: L_2).

TABLE I
PARAMETER VALUES OF THE FREQUENCY TREND FOR DIFFERENT TX AND RX ANTENNA POSITIONS (MARKINGS ARE AS PER FIGURE 1).

Tx position	Rx position	Tx-Rx gap [m]	K [dB]	m	LoS/ nLoS
D	R_2	0.56	95.9481	0.6988	LoS
P_3	P_2	0.60	106.0515	0.7583	LoS
RPR	P_2	0.70	200.0803	1.2355	LoS
M_2	L_2	0.73	55.9457	0.4913	LoS
M_2	R_2	0.76	80.6947	0.6145	LoS
P_3	P_1	0.76	192.7777	1.1910	LoS
RPR	P_1	0.84	115.0652	0.8015	LoS
D	R_1	0.85	135.7254	0.9257	LoS
M_2	P_2	0.87	150.9870	0.9788	LoS
D	L_2	0.97	198.5332	1.2165	LoS
D	L_1	1.16	174.1665	1.1257	LoS
D	P_2	1.23	141.4975	0.9498	nLoS
P_3	L_2	1.23	125.0928	0.8776	nLoS
RPR	R_2	1.25	135.1506	0.9250	nLoS
P_3	R_2	1.28	178.1539	1.1457	nLoS
RPR	L_2	1.44	204.2820	1.2859	nLoS
D	P_1	1.48	62.0210	0.5556	nLoS
RPR	R_1	1.57	152.7682	1.0283	nLoS
P_3	L_1	1.62	148.1906	1.0062	nLoS
P_3	R_1	1.65	147.4560	1.0030	nLoS
RPR	L_1	1.74	157.8843	1.0637	nLoS

the former one is currently adopted in major UWB standards. The parameters for different experiments are listed in Table I.

For indoor UWB propagation, a value range of $0.8 < m < 1.4$ was reported earlier [7]. The m values in Table I roughly follows the limits. Our results are also consistent with previous measurements inside car compartment [8] where a $1/f^2$ decay was observed in the power spectra. The experiments conducted at other parts (e.g. under the chasis) of the vehicle with less favourable propagation modes results in a higher m [9].

C. AR Model for Short Term Variations

After finding out the long term trends, we proceed with the characterization of the normalized CTF, namely, $\tilde{H}(f)$. The variations of $\tilde{H}(f)$ resembles a correlated series, and for such a series with low peaks and deep fades, an AR model is preferred over moving average (MA) or hybrid ARMA models.

AR model for wideband indoor radio propagation was first presented in [10], and later applied to UWB channel modelling in [11] for indoor scenarios and in [12] for underground mines. The normalized CTF under a q order AR process assumption may be mathematically expressed as

$$\tilde{H}(f_n) - a_1\tilde{H}(f_{n-1}) - a_2\tilde{H}(f_{n-2}) \cdots - a_q\tilde{H}(f_{n-q}) = \xi_n \quad (4)$$

where, f_n; $n = 1, 2, \cdots N$, is the nth discrete frequency in the CTF vector, a_k; $k = 1, 2, \cdots q$, are the complex AR process coefficients, and ξ_n is the nth sample of a complex Gaussian process with variance σ_ξ^2. A z-transform, $\tilde{H}(z) = \sum_n \tilde{H}(f_n)z^{-n}$, allows us to view the CTF as the output of a linear filter with transfer function, $\mathcal{G}(z) = \tilde{H}(z)/\xi(z)$, driven

ISBN: 978-1-61804-318-4

by white Gaussian noise [10], i.e.

$$\mathcal{G}(z) = \frac{1}{1 - \sum_{k=1}^{q} a_k z^{-k}} = \prod_{k=1}^{q} \frac{1}{1 - p_k z^{-k}} \tag{5}$$

The poles of the filter, p_k; $k = 1, 2, \cdots q$, account for individual multipath clusters. Although a second order AR process was sufficient for indoor [11] and underground mines [12], we used a fifth order $(q = 5)$ process as the car compartment exhibits multiple overlapped clusters. The poles (p_k) and the noise variance (σ_ξ^2) are found by solving the Yule-Walker equations in MATLAB. For finding initial conditions, we used a built in function `filtic()` with the first q entries of $\tilde{H}(f_n)$ and the AR model coefficients as arguments.

III. Simulation Results and Discussions

A. Channel Transfer Function and Coherence BW

The CTF, $H(f)$, is obtained through combining the long term frequency dependence using parameters from Table I with the simulated short term AR model based variations. The measured and simulated transfer functions for one particular Tx-Rx pair is shown in Fig. 3.

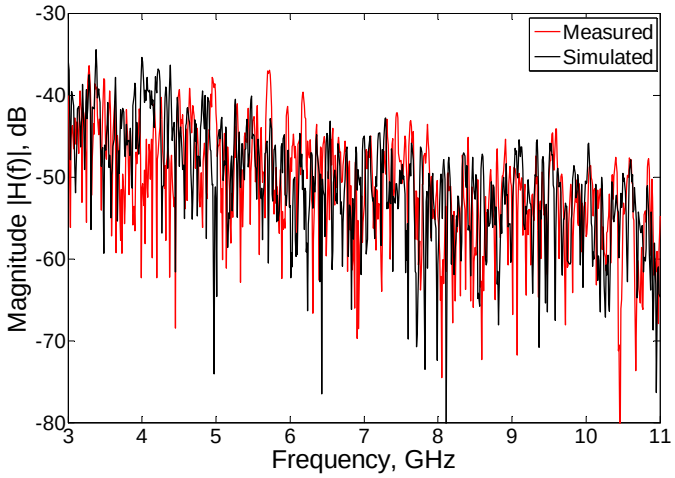


Fig. 3. Measured and simulated CTF (Tx position: RPR, Rx position: L_1).

The frequency autocorrelation function (ACF), $R(\Delta f)$, may be found from the channel transfer function as [13]

$$R(\Delta f) = \int_{-\infty}^{\infty} H(f) H^*(f + \Delta f)\, \mathrm{d}f \tag{6}$$

which provides a measure of the frequency selectivity. The range between DC or zero frequency, where normalized ACF attains its peak value of unity, and the frequency where ACF falls to 50% of or 3 dB lower than its peak value, is defined as the coherence bandwidth (BW), B_C. From Fig. 4, it can be seen that the measured and simulated transfer functions manifest almost similar B_C values.

A channel is considered *flat* in the coherence BW interval, i.e. if two different frequencies are separated by more than B_C, the channel exhibits uncorrelated fading at these two frquencies. There is a more direct method available for calculation

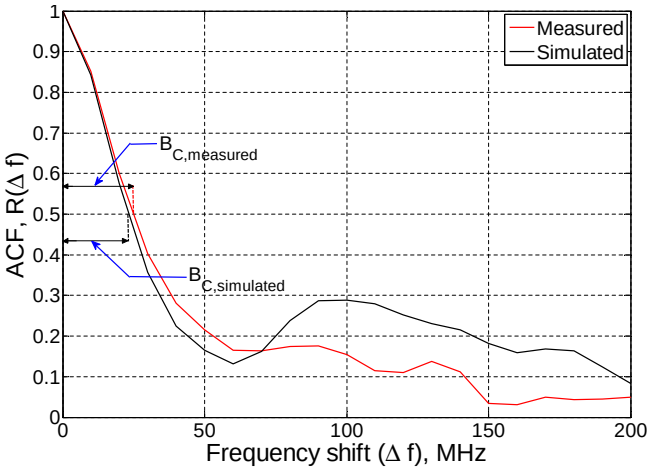


Fig. 4. Comparison of frequency ACF (Tx position: P_3, Rx position: R_2).

of coherence BW [14]. However, we computed B_C via the classical approach as the BW spans over only few samples for the current frequency step size (10 MHz), and there might be large approximation errors involved in the direct method.

B. PDP and RMS Delay Spread

Next, the complex channel impulse responses (CIRs), $h(t) = \mathcal{F}^{-1} H(f)$, are obtained through inverse fast Fourier transform (IFFT) method. Power delay profile (PDP) is closely related with the CIR, $\mathrm{PDP}(t) = |h(t)|^2$, and describes the variation of the average received power in dBm (when the transmit power of VNA is set to 0 dBm) as a function of delay time [15]. Fig. 5 shows the comparison of the measured PDP with the simulated PDP, and one can find that there is a close match.

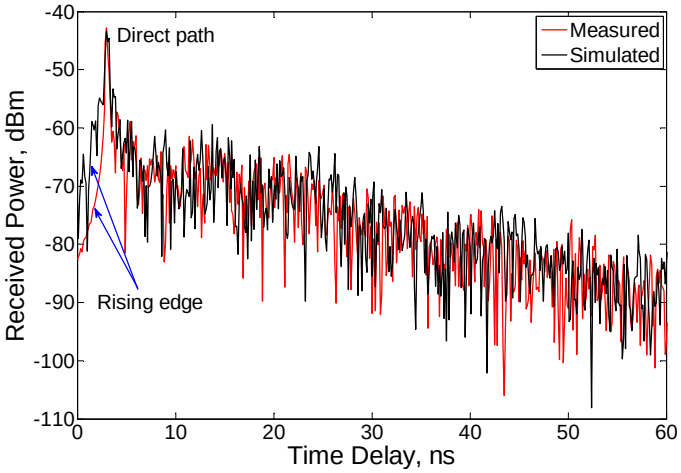


Fig. 5. Measured and simulated PDPs (Tx position: P_3, Rx position: P_1).

A quantitative comparison between the measured PDP and the simulated PDP can be performed by noting the similarity of the root mean square (RMS) delay spreads obtained for both the delay profiles. RMS delay spread is the second central

ISBN: 978-1-61804-318-4

moment of the PDP

$$\tau_{\mathrm{rms}} = \sqrt{\int_0^{\tau_{\max}} (t-\bar{\tau})^2 \cdot \mathrm{PDP}_n(t)\ \mathrm{d}t} \tag{7}$$

where $\tau_{\max}$ denotes the maximum excess delay, $\mathrm{PDP}_n(t) = \mathrm{PDP}(t)/\int_0^{\tau_{\max}} \mathrm{PDP}(t)\mathrm{d}t$ is normalized PDP, and $\bar{\tau} = \int_0^{\tau_{\max}} t \cdot \mathrm{PDP}_n(t)\mathrm{d}t$ is the mean excess delay.

For calculating the RMS delays, the rising edge of the PDP is cut off and the time origin is shifted to the time index that corresponds to the peak. This time shifting helps in rendering the delays as excess delays relative to the peak or first arriving path which has a zero delay. Although the rising edge may be suppressed with appropriate windowing (Hamming, Hann, Blackman etc.) during the IFFT post-processing, the process is avoided here. Further, only those multipath components (MPCs) having a delay less than $\tau_{\max} = 60$ ns are considered. This step ensures that the truncated PDP does not hit the noise floor. According to the Agilent E5071C VNA data sheet, the noise floor is -120 dBm/Hz. Hence, for a 10Hz IF bandwidth, it is good enough to consider MPCs upto -110 dBm. Finally, the PDPs are normalized so that the peak occurs at 0 dB. The measured RMS delay values are between 5 to 10 ns, and are consistent with time domain measurements of intra-vehicle UWB links [16].

The comparison of RMS delay spreads reveals that the simulated PDP matches closely with the measured PDP as the percentage of error

$$\%\ \mathrm{error} = \frac{\tau_{\mathrm{rms,simulated}} - \tau_{\mathrm{rms,measured}}}{\tau_{\mathrm{rms,measured}}} * 100 \tag{8}$$

is typically 10%, with values ranging between 2% to 30%. It is also interesting to note that the error is always positive, i.e. the simulated PDP slightly overestimates τ_{rms}.

IV. Conclusions and Future Work

The key finding of the paper is, the transfer function of a intra-vehicle UWB channel can be modelled with an AR process after removing the frequency dependent trend. Simulated transfer functions using the AR process parameters and the trend parameters exhibit close match with the measured values. The similarity of coherence BW, PDP, and RMS delay spreads further validates the model.

The next step is to study the dependence of the parameters for the AR process and the long term trends on frequency and Tx-Rx separation. This would lead to development of a comprehensive model for intra-car communication which can predict transfer function for arbitrary Tx/ Rx placement with a certain degree of accuracy. It would be also interesting to investigate whether this model can be extended to intra-car millimeter wave links.

Acknowledgment

This work was supported by the SoMoPro II programme, Project No. 3SGA5720 *Localization via UWB*, co-financed by the People Programme (Marie Curie action) of the Seventh Framework Programme (FP7) of EU according to the REA Grant Agreement No. 291782 and by the South-Moravian Region. The research is further co-financed by the Czech Science Foundation, Project No. 13-38735S *Research into wireless channels for intra-vehicle communication and positioning*, and by Czech Ministry of Education in frame of National Sustainability Program under grant LO1401. For research, infrastructure of the SIX Center was used. The generous support from Skoda a.s. Mlada Boleslav are also gratefully acknowledged.

References

[1] I. G. Zuazola, J. M. Elmirghani, and J. C. Batchelor, "High-speed ultra-wide band in-car wireless channel measurements," *IET communications*, vol. 3, no. 7, pp. 1115–23, Jul. 2009.

[2] M. Schack, J. Jemai, R. Piesiewicz, R. Geise, I. Schmidt, and T. Kürner, "Measurements and analysis of an in-car UWB channel," in *IEEE Vehicular Technology Conference (VTC 2008-Spring)*, Singapore, May 2008, pp. 459–63.

[3] J. S. McLean, R. Sutton, A. Medina, H. Foltz, and J. Li, "The experimental characterization of uwb antennas via frequency-domain measurements," *IEEE Antennas and Propagation Magazine*, vol. 49, no. 6, pp. 192–202, Dec. 2007.

[4] W. Q. Malik, D. J. Edwards, and C. J. Stevens, "Frequency-dependent pathloss in the ultrawideband indoor channel," in *IEEE International Conference on Communications (ICC)*, vol. 12, Istanbul, Turkey, Jun. 2006, pp. 5546–51.

[5] J. Kunisch and J. Pamp, "Measurement results and modeling aspects for the UWB radio channel," in *IEEE Conference on Ultra Wideband Systems and Technologies (UWBST)*, vol. 1, Baltimore, MD, USA, May 2002, pp. 19–23.

[6] A. Álvarez, G. Valera, M. Lobeira, R. P. Torres, and J. L. C. García, "New channel impulse response model for UWB indoor system simulations," in *IEEE Vehicular Technology Conference (VTC 2003-Spring)*, vol. 1, Jeju, Korea, Apr. 2003, pp. 1–5.

[7] C. C. Chong, Y. E. Kim, S. K. Yong, and S. S. Lee, "Statistical characterization of the UWB propagation channel in indoor residential environment," *Wireless Communications and Mobile Computing*, vol. 5, no. 5, pp. 503–12, Aug. 2005.

[8] T. Kobayashi, "Measurements and characterization of ultra wideband propagation channels in a passenger-car compartment," in *IEEE International Symposium on Spread Spectrum Techniques and Applications*, Manaus, Amazon, Brazil, Aug. 2006, pp. 228–32.

[9] C. U. Bas and S. C. Ergen, "Ultra-wideband channel model for intra-vehicular wireless sensor networks beneath the chassis: from statistical model to simulations," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 1, pp. 14–25, Jan. 2013.

[10] S. J. Howard and K. Pahlavan, "Autoregressive modeling of wide-band indoor radio propagation," *IEEE Transactions on Communications*, vol. 40, no. 9, pp. 1540–52, Sep. 1992.

[11] S. S. Ghassemzadeh, R. Jana, C. W. Rice, W. Turin, and V. Tarokh, "Measurement and modeling of an ultra-wide bandwidth indoor channel," *IEEE Transactions on Communications*, vol. 52, no. 10, pp. 1786–96, Oct. 2004.

[12] A. Chehri and P. Fortier, "Frequency domain analysis of UWB channel propagation in underground mines," in *IEEE Vehicular Technology Conference (VTC 2006-Fall)*, Montreal, Quebec, Canada, Sep. 2006, pp. 1–5.

[13] I. Cuiñas and M. G. Sánchez, "Measuring, modeling, and characterizing of indoor radio channel at 5.8 GHz," *IEEE Transactions on Vehicular Technology*, vol. 50, no. 2, pp. 526–35, Mar. 2001.

[14] M. Ghaddar, L. Talbi, and G. Y. Delisle, "Coherence bandwidth measurement in indoor broadband propagation channel at unlicensed 60 GHz band," *Electronics Letters*, vol. 48, no. 13, pp. 795–97, Jun. 2012.

[15] G. J. M. Janssen, P. A. Stigter, and R. Prasad, "Wideband indoor channel measurements and BER analysis of frequency selective multipath channels at 2.4, 4.75, and 11.5 GHz," *IEEE Transactions on Communications*, vol. 44, no. 10, pp. 1272–88, Oct. 1996.

[16] A. Chandra, J. Blumenstein, T. Mikulásek, J. Vychodil, M. Pospísil, R. Maršálek, A. Prokeš, T. Zemen, and C. Mecklenbräuker, "CLEAN algorithms for intra-vehicular time-domain UWB channel sounding," in *International Conference on Pervasive and Embedded Computing and Communication Systems (PECCS)*, Angers, France, Feb. 2015, pp. 224–29.

Machine Learning and the Detection of Anomalies in Wikipedia

Mentor Hamiti, Arsim Susuri and Agni Dika

***Abstract*—**This work analyses the current trend in applying machine learning in detection of anomalies, with the specific aim of analyzing anomalies in Wikipedia articles. Ever since it was created, in 2001, Wikipedia has grown with immense speed, enabling anyone the ability to edit articles, thus, establishing itself as one of the largest information sources on the Internet. Having become this popular, Wikipedia has become the source of an ever-increasing number of articles, created, modified and enhanced by different editors and, inadvertently, susceptible to various acts of vandalisms. This article aims to provide an overview of the initial research and developments in the field of machine learning applications in detecting anomalies in Wikipedia and future trends.

***Keywords*—**machine learning, Wikipedia, anomalies, vandalism, detection of anomalies.

I. Introduction

Ever since its inception, in 2001, Wikipedia has continuously grown to become one the largest information source on the Internet. One of its unique features is that it offers the ability to anyone to edit the articles. This popularity, in itself, means that, a number of articles can be read, edited, and enhanced by different editors and, inevitably, be subject to acts of vandalisms through illegitimate editing.

Vandalism means any type of editing which damages the reputation of an article or a user in Wikipedia. A list of typical vandalisms along with their chances of appearance, as shown in Fig. 1, was created as a result of empirical studies done by Priedhorsky et al. [1]. Typical examples include massive deletions, spam, partial deletions, offences and misinformation.

In order to deal with vandalism, Wikipedia relies on the following users:

- Wikipedia its users' ability and willingness to find (accidentally or deliberately) damaged articles
- Wikipedia administrators and
- Wikipedia users with additional privileges

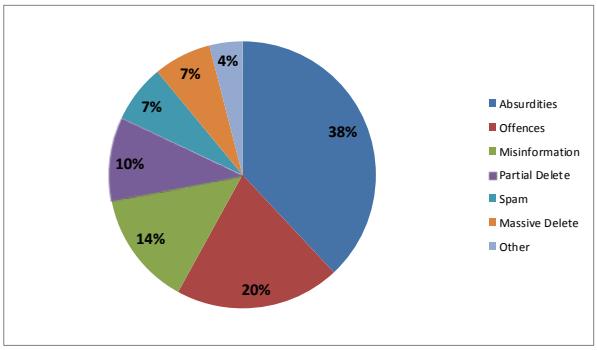


Fig. 1. Categories of vandalism based on empirical approach [1]

These users use special tools (e.g. Vandal Fighters) to monitor recent changes and modifications that enable retrieval of bad expressions or which are implemented by blacklisted users.

Wikipedia was subject to different statistical analysis from various authors. Viégas et al. [2] uses visualization tools to analyze the history of Wikipedia articles. When it comes to vandalism, authors were able to identify (manually) massive deletions as a jump in the history flow of a particular article page.

Since late 2006, some bots (computer programs designed to detect and revert vandalism), have appeared on Wikipedia. These tools are built on the primitive included in the Vandal Fighters. These use lists of common phrases, and consult databases containing blocked users or IP addresses in order to separate legitimate editing from vandalism.

One drawback of these approaches is emphasized that these world use static list of obscenities and grammatical rules which are difficult to maintain and easily "fooled". These detect only 30% of vandalisms committed.

Consequently, there is a need to improve the detection of this kind. One of the possible improvements is the application of machine learning.

The prior success implemented in interference detection, spam filtering for email, etc., is a good indicator for the opportunity that the machine learning shows in improvements in detecting anomalies in Wikipedia.

ISBN: 978-1-61804-318-4

II. Wikipedia Vandalism Detection

To define the vandalism detection task, we have to define some key concepts of MediaWiki (the wiki engine used by Wikipedia).

An article is composed of a sequence of revisions, commonly referred to as the article history. A revision is the state of an article at a given time in its history and is composed of the textual content and metadata describing the transition from the previous revision [3].

Revision metadata contains, among others, the user who performed the edit, a comment explaining the changes, a timestamp, etc. An edit is a tuple of two consecutive revisions and should be interpreted as the transition from a given revision to the next one. Wikipedia vandalism detection is a one-class classification task.

The goal is, given any edit, determine whether it is destructive or not. Through machine learning, anomalous contributions (edits) can be detected by inspecting Wikipedia edits. An edit $e = (r_-, r_+)$ is defined as a set of two consecutive revisions of an article which contains the original revision (r_-) and the new revision (r_+) once the changes have been submitted.

A revision r is a version of a Wikipedia article that, besides the article markup text, includes additional data (meta data) about the latest editing, such as the editor's user identification, his/her comment on the nature of the changes made, and a timestamp at which he/she edited the article.

Evaluating a vandalism detection system requires a corpus of pre-classified edits. Four different corpora have been reported in the literature:

1. Webis-WVC-07 - The Webis Wikipedia Vandalism Corpus 2007 (Webis-WVC-07) was the first Wikipedia vandalism corpus and consists of 940 human-annotated edits of which 301 are labelled as vandalism. It was compiled in 2007 and was first used by Potthast et al. [4]. English Wikipedia was the sole source for all edits.
2. PAN-WVC-10 - The PAN Wikipedia Vandalism Corpus 2010 (PAN-WVC-10), compiled in 2010 via Amazon's Mechanical Turk comprises 32439 edits from 28468 English Wikipedia articles of which 2394 have been annotated as vandalism.
   The dataset was created by 753 human annotators by casting 193022 votes, so that each edit has been annotated at least three times, whereas edits that were difficult to be annotated received more than three votes (Potthast [5]). The PAN-WVC-10 was first used in the 1st International Competition on Wikipedia Vandalism Detection (Potthast et al. [6]).
3. PAN-WVC-11 - The PAN Wikipedia Vandalism Corpus 2011 (PAN-WVC-11) from 2011 is an extension of the PAN-WVC-10. It was used in the 2nd International Competition on Wikipedia Vandalism Detection (Potthast and Holfeld [7]) and is the first multilingual vandalism detection corpus. The corpus comprises 29949 Wikipedia edits in total (9985 English edits with 1144 vandalism, 9990 German edits with 589 vandalism, and 9974 Spanish edits with 1081 vandalism annotations).
4. Wikipedia History Dump Wikipedia records all revisions of all articles and all other Wikipedia pages and releases them as XML or SQL dump files.

A. Wikipedia Bots

The vandalism problem on Wikipedia is probably as old as the encyclopedia itself. Kittur et al. [8] observe that the total number of vandalism edits is increasing over time. Although they report the total vandalism proportion to remain at the same level, increasing vandalism is a serious objective in the online encyclopedia.

To tackle this problem, the Wikipedia community resorts to manually protecting articles from being edited in case they are heavily vandalized.

Additionally, since 2006, vandalism detection bots are used, which automatically patrol for vandalism edits and partially revert them. Most often these bots use simple heuristic rules, word blacklists, and lists of blocked user IPs to identify vandalism edits (e.g. VoABot II or ClueBot).

The ClueBot NG bot which replaces ClueBot, uses machine learning approaches. It tries to enhance the heuristics-based techniques, which were difficult to maintain and easy to bypass. The bot uses a pre-classified edit dataset annotated by Wikipedia users to train an Artificial Neural Network.

AVBOT [9] is a bot created to automatically search for any vandalisms edits in Spanish articles of Wikipedia. So far, it has reverted more than 200,000 vandalism edits [10].

B. Approaches based on Machine Learning

Since 2008 Wikipedia vandalism detection based on machine learning approaches has become a field of increasing research interest. In Table 1 existing vandalism detection approaches from the literature are shown.

Potthast et al. [4] contributed the first machine learning vandalism detection approach using textual features as well as basic meta data features with a logistic regression classifier. Smets et al. [11] used a Naive Bayes classifier on a bag of words edit representation and were the first to use compression models to detect Wikipedia vandalism. Itakura and Clarke [12] used Dynamic Markov Compression to detect vandalism edits on Wikipedia.

Mola Velasco [13] extended the approach of Potthast et al. [4] by adding some additional textual features and multiple wordlist-based features. He was the winner of the 1st International Competition on Wikipedia Vandalism Detection (Potthast et al. [6]).

ISBN: 978-1-61804-318-4

TABLE I

VANDALISM DETECTION CLASSIFICATION OBTAINED FROM VARIOUS AUTHORS

Authors	Balanced Data	Classifier	Precision	Recall	PR-AUC	Corpora
Smets et al. [11]	x	Probablistic Sequence Modeling	0.3209	0.9171	-	Simplewiki
Smets et al. [11]	x	Naive Bayes	0.4181	0.5667	-	Simplewiki
Tran and Christen [20]	√	Gradient Tree Boosting	0.870	0.870	-	Historical Dump
Potthast et al. [4]	x	Logistic Regression	0.830	0.870	-	Webis-WVC-07
Velasco [3]	x	Random Forest	0.860	0.570	0.660	PAN-WVC-10
Adler et al. [15]	x	ADTree	0.370	0.770	0.490	PAN-WVC-10
Adler et al. [17]	x	Random Forest	-	-	0.820	PAN-WVC-10
West and Lee [18]	x	ADTree	0.370	0.770	0.490	PAN-WVC-10
Harpalani et al. [19]	x	LogitBoost	0.606	0.608	0.671	PAN-WVC-10
West and Lee [18]	x	ADTRee	-	-	0.820	PAN-WVC-11

West et al. [14] were among the first to present a vandalism detection approach solely based on spatial and temporal meta data, without the need to inspect article or revision texts.

Adler et al. [15], in a similar fashion, built a vandalism detection system on top of their WikiTrust reputation system (Adler and De Alfaro [16]). Adler et al. [17] combined natural language, spatial, temporal and reputation features used in their aforementioned works (Adler et al. [15], Mola Velasco [13], West et al. [14]). Besides Adler et al. [17], West and Lee [18] were the first to introduce ex post facto data as features, for whose calculation also future revisions have to be considered.

Their resulting multilingual vandalism detection system was the winner at the 2nd International Competition on Wikipedia Vandalism Detection (Potthast and Holfeld [7]).

Harpalani et al. [19] stated vandalism edits to share unique linguistic properties. Thus, they based their vandalism detection system on a stylometric analysis of vandalism edits by probabilistic context-free grammar models. They showed that this approach outperforms features based on shallow patterns, which match syntactic structures and text tokens.
Supporting the current trend of creating cross language vandalism classifiers, Tran and Christen [20] evaluated multiple classifiers based on a set of language independent features that were compiled from the hourly article view counts and Wikipedia's complete edit history.

C. Features of Anomalies

The literature provides an ever-growing set of features that are employed to model anomalous edits. After the first contributions to the Wikipedia vandalism detection task, most authors used a subset of existing features and added some new ones to their approaches.

Tables 2 provide an overview of textual data anomalies features that were used so far in the literature.

For the sake of simplicity, we use the following abbreviations to distinguish various authors: A17 (Adler et al. [17]), G14[1], J22 (Javanmardi et al. [22]), M3 (Mola Velasco [3]), P4 (Potthast et al. [4]), W18 (West and Lee [18]), and Wa21 (Wang and McKeown [21]).

Textual features are calculated by analyzing the new revision's markup text or rather both revisions' markup texts of an edit. Meta data features are compiled from the revision's meta data or are calculated by analyzing additional Wikipedia data, such as history dumps or article dumps.

While Mola Velasco [3] used three feature categories by considering textual, meta data and language features, his language features (wordlist-based features) could be categorized as textual features. Javanmardi et al. [22] split their features into four categories, namely textual, meta data, user and language model. The user category comprises user-related meta data features.

[1] https://github.com/webis-de/wikipedia-vandalism-detection

TABLE II

SOME TEXTUAL FEATURES USED BY VARIOUS AUTHORS, DESCRIBING ANOMALOUS EDITS IN WIKIPEDIA

Category	Feature	A17	G14	J22	M3	P4	W18	Wa21
Frequency	All words	√	√	√	√			
	Average term	√	√		√	√		
	Bad words	√	√	√	√			
	Biased words	√	√	√	√			
	Emoticons		√					
	Good/markup words	√	√	√	√			
	Sex words	√	√	√	√			
	Vulgarism		√	√	√	√	√	√
	Web slang							√
Impact	All words	√	√	√	√			
	Bad words	√	√	√	√			
	Emoticons		√					
	Good/markup words	√	√	√	√			
	Sex words	√	√	√	√			
	Vulgarism		√	√	√	√	√	
Ratio	Alphanumeric	√	√	√			√	
	Non-alphanumeric		√	√	√			
	Size	√	√	√	√	√		√
	Upper to all	√	√	√	√	√	√	
	Upper to lower	√	√		√			
Other	Blanking		√	√				
	Character diversity		√		√			
	Characters added or removed		√				√	
	Compressibility	√	√	√	√	√		
	Context relation					√		
	Digit ratio	√	√	√	√			
	External links added		√	√				
	Inserted wiki markup						√	
	Inserted words			√				
	Internal links added		√	√				
	Longest char sequence	√	√	√	√	√	√	
	Longest word	√	√	√	√	√	√	
	Punctuation misuse							√
	Removed words			√				
	Replacement similarity		√			√		
	Size increment	√	√	√	√		√	

ISBN: 978-1-61804-318-4

III. Conclusion

This brief review shows the overall progress in applying machine learning in detecting anomalies in Wikipedia. The problem of vandalism has grown over the years, along with the growth of popularity of Wikipedia. Applying machine learning as a tendency to automate detection of vandalisms is a great opportunity for maintaining and improving the credibility of Wikipedia, without compromising the ability of Various Wikipeida users to enhance articles online through editing.

Having in mind that in order to properly implement machine learning in detection of anomnalies there is a requirement to properly characterize anomalies. This is why implementation of specific features of anomalies in creating machine learning based anomaly detectors.

Based on our research, we can also conclude that, apart from English, German, French and Spanish, little or no progress is made in other language sections of Wikipedia, thus providing excellent grounds for future research. Furthermore, development of new language– independent methods to enhance detection of anomalies could improve the effect of machine learning approach on the credibility of Wikipedia.

References

[1] R. Priedhorsky, J. Chen, S. T. K. Lam, K. Panciera, L. Terveen, and J. Riedl. "Creating, destroying, and restoring value in Wikipedia," in proceedings of the international ACM conference on supporting GroupWork (GROUP), Sanibel Island, FL, pp 259-268, 2007.

[2] F. B. Viégas, M. Wattenberg, and K. Dave, "Studying cooperation and conflict between authors with history flow visualizations," in proceedings of the ACM Conference on human factors in computing systems (CHI), Vienna, Austria, pp 575-582, 2004.

[3] Santiago M. Mola-Velasco, "Wikipedia Vandalism Detection," - WWW 2011, Hyderabad, India. 2011.

[4] Martin Potthast, Benno Stein, and Robert Gerling, "Automatic vandalism detection in wkipedia," in advances in information retrieval, pp 663-668. Springer Berlin Heidelberg, 2008.

[5] Martin Potthast, "Crowdsourcing a wikipedia vandalism corpus," proceeding of the 33rd international ACM SIGIR conference on research and development in information retrieval - SIGIR '10, p 789, 2010.

[6] Martin Potthast, Benno Stein, and Teresa Holfeld, "Overview of the 1st International Competition on Wikipedia Vandalism Detection," in Martin Braschler, Donna Harman, and Emanuele Pianta, editors, Working Notes Papers of the CLEF 2010 Evaluation Labs, September 2010.

[7] Martin Potthast and Teresa Holfeld "Overview of the 2nd International Competition on Wikipedia Vandalism Detection", in Vivien Petras, Pamela Forner, and Paul D. Clough, editors, Notebook Papers of CLEF 11 Labs and Workshops, September 2011.

[8] Aniket Kittur, Bongwon Suh, Bryan A. Pendleton, and Ed H. Chi, "He says, she Says: conflict and coordination in Wikipedia;" in ACM Conference on Human Factors in Computing Systems, pp 453-462, 2007.

[9] Emilio-José Rodriguez-Posada, "AVBOT: Detecting and fixing Vandalism in Wikipedia" in proceedings of the CEPIS UPGRADE, vol. XII, issue no. 3, 2011.

[10] http://es.wikipedia.org/wiki/Especial:Contribuciones/AVBOT.

[11] Koen Smets, Bart Goethals, and Brigitte Verdonk, "Automatic vandalism detection in wikipedia: Towards a machine learning approach," in WikiAI '08: Proceedings of the AAAI Workshop on Wikipedia and Artificial Intelligence, 2008.

[12] Kelly Y. Itakura and Charles L. a. Clarke, "Using dynamic markov compression to detect vandalism in the Wikipedia," Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval - SIGIR '09, p 822, 2009.

[13] Mola Velasco Santiago Moisés Mola Velasco, "Wikipedia vandalism detection through machine learning: Feature review and new proposals - lab report for pan at clef 2010," in CLEF (Notebook Papers/LABs/Workshops), 2010.

[14] Andrew G. West, Sampath Kannan, and Insup Lee, "Detecting wikipedia vandalism via spatio-temporal analysis of revision metadata," in Proceedings of the Third European Workshop on System Security, EUROSEC '10, pp 22-28, New York, NY, USA, 2010.

[15] B. Thomas Adler, Luca De Alfaro, and Ian Pye, "Detecting wikipedia vandalism using wikitrust," notebook papers of CLEF, 2010.

[16] B. Thomas Adler and Luca De Alfaro, "A content-driven reputation system for the wikipedia," proceedings of the 16th international conference on World Wide Web WWW 07, 7(Generic):261, 2007.

[17] B. Thomas Adler, Luca De Alfaro, Santiago M. Mola-Velasco, Paolo Rosso, and Andrew G. West, "Wikipedia vandalism detection: Combining natural language, metadata, and reputation features," in proceedings of the 12th International Conference on Computational Linguistics and Intelligent Text Processing - Volume Part II, pp 277-288, Berlin, Heidelberg, 2011.

[18] Andrew G. West and Insup Lee, "Multilingual vandalism detection using language-independent & ex post facto evidence - notebook for pan at clef 2011. In CLEF (Notebook Papers/Labs/Workshop), 2011.

[19] Manoj Harpalani, Michael Hart, S Signh, Rob Johnson, and Yejin Choi, "Language of Vandalism: Improving Wikipedia Vandalism Detection via Stylometric Analysis," ACL (Short Papers), (2009):pp 83-88, 2011.

[20] Khoi-Nguyen Tran and Peter Christen, "Cross Language Prediction of Vandalism on Wikipedia Using Article Views and Revisions," Advances in Knowledge Discovery and Data Mining, pp 268-279, 2013.

[21] William Yang Wang and Kathleen R. McKeown, "Got You!: Automatic vandalism detection in Wikipedia with web-based shallow syntactic-semantic modelling," in Proceedings of the 23rd International Conference on Computational Linguistics, COLING '10, pp 1146-1154, Stroudsburg, PA, USA, 2010.

[22] Sara Javanmardi, David W. McDonald, and Cristina V. Lopes, "Vandalism detection in wikipedia: A high-performing, feature-rich model and its reduction through lasso," in Proceedings of the 7th International Symposium on Wikis and Open Collaboration, WikiSym '11, pp 82-90, New York, NY, USA, 2011.

The National Vehicle Identification System in Brazil as a tool for mobility improvement

Eduardo M. Dias, Jilmar A. Tatto, Dariusz A. Swiatek

GAESI, Poli/USP, Departamento de Engenharia de Energia e Automação Elétricas

University of São Paulo

Brasil

***Summary* —** This paper presents the adoption, benefits and effects of a National Vehicle Identification System in Brazil in terms of its application in the areas of transport and logistics. The System, created on 2006, is still in implementation phase, but its pilot program has already shown promising results. In order to understand and verify the requirements of the National Vehicle Identification System (SINIAV) were analyzed the foundations established by the legislature and the subsequent regulation by the National Traffic Council. Also, given its appliance by SINIAV, radio frequency technology was analyzed in terms of safety and expected results. The importance of an open source technology for this system was also scrutinized. The results of the system were verified regarding its application in the areas of logistics and transportation, using as indicators the prevention, monitoring and suppression of vehicle theft and robbery. SINIAV is expected to be deployed without any proprietary protocols and to allow data sharing between federal entities, which will benefit to public safety and mobility, thus becoming a qualitative leap for development of smart cities in Brazil.

***Key words*—** radio frequency identification, RFID, SINIAV, identification of vehicles.

I. INTRODUCTION

Systems for Automatic Vehicle Identification (AVI) have been used around the world for more than a decade. Structures that support access, control and identification in general are applied in many different uses all over the world. The installations can be spotted at airports, seaports, bus terminals, bus-stops, taxi stands, centers of the cities, industrial and residential estates. Such systems are convenient to use due to solutions such as proximity or radio frequency access control allow easy registration of passing vehicles without need to delay their flow.

Another feature provided by RFID controlled AVI is facilitating security through automation of vehicle recognition and identification. Furthermore congestion, queues and miscommunication can cause delay and have a negative impact on business and the wellbeing of people.

Cities and business currently are investing heavily in infrastructure and facilities to increase security and mobility. Maintaining mobility is a challenge itself for city administration and business management. Meeting that challenge while conforming to new emission standards which become a vivid need of modern societies and thanks to use of control systems such as AVI is becoming feasible. Vehicle Identification Systems can help ensuring the speedy flow of traffic while contributing to a green environment [1]. Automatic remote detection prevents cars from standing still with the motor running. It prevents congestion and ensures swift and fuel efficient transportation of vehicles and goods [2].

Finally Automatic Vehicle Identification is much more than just vehicle control. Not only cars, but all kinds of vehicles can use such systems. Vehicles can be tracked as part of logistics or production processes. A lot of taxi queues are managed using AVI systems: through the identification of taxis the queue is managed and vehicles are allowed access to the passenger pickup area in the right order. Weighing bridges for trucks are equipped with AVI to automatically identify the vehicle and connect it to the weighing information before allowing the truck to proceed to the next step in the process [3]. All kinds of vehicles in mining industrial estates are tagged to keep track of them during the production process in the mines while ensuring high levels of safety. Trains are equipped with AVI to identify them entering and leaving train stations and keep track of them while performing their duty [4]. Even a lot of bus terminals are managed using AVI equipment.

In the future AVI will be actively used in fleet management to track vehicle location, hard stops, rapid acceleration, and sudden turns using sophisticated analysis of the data in order to implement new policies (e.g., no right/left turns) that result in cost savings [5, 6, 7].

E. M. Dias, Eduardo M. Dias is PhD in Electrical Engineering and Full Professor at the Polytechnic School of University of São Paulo (EPUSP), Av. Prof. Luciano Gualberto, trav. 3, n. 158, São Paulo/SP, Brazil, CEP 05508-970 and coordinator at GAESI - Electrical Automation Group of Industrial Systems, a research group of the Electrical Energy and Automation Department (emdias@pea.usp.br) of EPUSP.

J. A. Tatto is a MSc student at EPUSP and Municipal Secretary for Transport of São Paulo/SP, Brasil (Telefone para contato: 55 11 3396-8000; fax: 55 11 3396-8000; e-mail: jilmar.tatto@uol.com.br).

Dariusz A. Swiatek has PhD in Human Geography and Urban Planning, is associated professor at the Institute of Geography and Spatial Organization, Polish Academy of Sciences – Warsaw, Poland (on sabbatical) and member of research team of GAESI - the Polytechnic School of University of São Paulo, Brazil (daswiatek@googlemail.com).

ISBN: 978-1-61804-318-4

***

Brazil has a fleet of more than 86 million vehicles, of which about 2.5 million are trucks. There are registered more than 7,5 million vehicles in São Paulo. Rio de Janeiro has about 2,4 million vehicles, while the fleet of Belo Horizonte is estimated for about 1,6 million and the Curitiba is now close to 1,5 million [8]. The territorial extension of the country together with the number of vehicles generate security and control problems, as well as complicates supervision of traffic and cargo transport on highways and urban roads. The theft charges [9], especially of electronic products, has been increasing, and in some places becoming serious issue [10]. Moreover, major cities suffer from congestion and lack of precise interventions to promote improvements of mobility and thus diminish the environmental, economic and social costs [11] e.g. time lost in traffic [12].

At a time when concepts of smart [13] and sustainable cities [14] are established, more efficiently integrating citizens to the management process, it is necessary to promote such integration with the use of technology [15] such as the Intelligent Transport Systems (ITS) which are already a reality in Brazil promoting advances in mobility.

The National System for Automatic Vehicle Identification (SINIAV) presents itself as an solution the existing problems in the areas of security and mobility, increasing the use of technology in vehicles and enforcement activities, ensuring at the same time, a preserving fundamental right for citizens and improving public management. This paper aims at analyzing the role of SINIAV as nationwide ITS, highlighting its objectives, the technology linked to its implementation and expected results, as well as proposing new use of information in urban mobility context, the information to be produced by the system when effectively in operation.

II. INTELLIGENT TRANSPORT SYSTEMS (ITS)

ITS are set of advanced applications which, without embodying technology as such, aim to provide innovative services related to different modes of transport and traffic management. According to the Brazilian National Public Transport Agency, ITS "mean multimodal control centers and operations, advanced traffic signal systems, monitoring systems and remote surveillance (cameras, sensors, probes, software), parking management, management of traffic incidents, emergency response, electronic payment, dynamic pricing and user information in real time "[16].

Initially, ITS technologies were designed to solve specific problems. The problem of speeding on public roads can be circumvented by deploying radars for surveillance that will allow capturing the image of the vehicles, and applications enabling panelizing the drivers. Closed circuit cameras which monitor vehicle flow on the roads of cities and enable quick intervention of agents where accidents occur. The traffic lights operating with fixed or real-time time schedule. In the latter case, they are sensitive to the amount of vehicles circulating and tailor its commands to the direction of increased flow through data collected by the controller reprograms for the central computer which sends the traffic signal time new rules to the controller. The collection of bus fares of public transport has benefited from the deployment of tickets and electronic turnstiles, reducing considerably the system of financial losses and facilitating passenger embarkation process. Buses are also equipped with cameras and automatic vehicle location systems (AVL) in order to enable the monitoring of passengers, travelling time and speed of the vehicle. Moreover ITS is used in number of other ways in context of transport and transit.

However, it is worth noting, that the intelligence of such systems should not be restricted to isolated solutions, which do not communicate with each other. The intelligence of these systems, at present, is in its design that ensures its own evolution, flexibility and integration with other systems. It can be achieved in number of ways, one of them accepted both in academia and public administration area, are open protocols that enable the connectivity, interoperability, ease of component replacement, the expansion of competition and cost reduction. The National System for Automatic Vehicle Identification, which will be described below, can be classified among the ITS which meets abovementioned standards.

III. THE NATIONAL SYSTEM FOR AUTOMATIC VEHICLE IDENTIFICATION (SINIAV)

The development of a system for the identification of vehicles, on national scale, was approved by the Brazilian Legislature for more than a decade. Supplementary Law No. 121 of February 9, 2006, created the National System for the Prevention, Control and Suppression of Vehicle and Cargo Theft and Robbery, aim to modernize and technologically adapt equipment and procedures related to prevention activities, surveillance and repression of theft and robbery of vehicles and cargo [17].

The federal entities, with the exception of municipalities, could then establish action plans to combat theft and robbery of vehicles and cargo throughout the national territory. The National Traffic Council (CONTRAN) has competence to establish the anti-theft devices and mandatory vehicle identification signs.

Thus, to identify vehicles, CONTRAN issued Resolution No. 212, on November 13, 2006 [18], later repealed by Resolution No. 412 on August 9, 2012 [19], establishing the SINIAV, based on the radio frequency identification technology (RFID), which should provide the transit executive entities with modern and interoperable tools for planning, monitoring and management of traffic and vehicle fleet.

IV. RFID TECHNOLOGY

The RFID technology to be used in SINIAV was began to be employed, in the 1930s, by the military in radars to identify objects. In the 1970s, it began to be used in the form tags for animal identification. In the field of mobility, the pioneer in the use of RFID was Norway, implementing an electronic toll collection system. This system has been replicated in various american cities up to the Rio-Niterói Bridge in 1996 [20].

Since that time number of other uses were given to RFID technology in the world. Noteworthy is the use of technology by large supermarket chains, retailers and industries to monitor the flow of their products [21]. Logistics, where the technology has been of great service to the traceability of supplies and products, either within or between industries and companies [22], in educational institutions [23] and cargo circulation by road [24]. Technological innovation increasingly sought by companies

ISBN: 978-1-61804-318-4

that use tags, chips or Quick Response Codes - QR Code [25]. Agro business is also adopting RFID to control the chain of plant products [26], as well as other such as meet [27].

RFID is based on emission and collection of electromagnetic waves. For this process chips with data storage capacity are read by an external device and wireless technology are used. The reading may be restricted, but information can be also recorded, being of great use in automation processes.

The similarity of this technology to the barcode facilitates understanding how the RFID operates. In both cases there are readers and devices containing information, but while in the case of barcode the reading is made by an optical reader, the RFID reading is by radio signals via antenna or transponder [20].

SINIAV system consists of electronic identification device called "electronic board" installed in the vehicle reading antennas, processing plants and computer systems.

A. *Electronic board*

The electronic board used in SINIAV must be isolated and possess a unique and unalterable number and series for each vehicle, including information about unique serial number; vehicle license plate number; vehicle category; type of vehicle; foreign fleet vehicle.

The minimum storage capacity of the board must be 1024 bits and provide information necessary to operate the system in accordance with the memory allocation map defined in Annex II of Resolution No. 412.

B. *Antenna*

The Resolution No. 412 defines the antenna as the aggregate device software and firmware, responsible and able to read and write information on the electronic board.
The antenna must enable integrated communications operation, and allow the reading of the electronic board installed in vehicles that are at any speed within the range of 0 to 160 km/h.

The safety of data transfer between the electronic board and antenna reader must ensured by the use of encryption keys recognized by the National Traffic Department (DENATRAN), with the consent of CONTRAN.

C. *Processing centrals*

As for the definition and characterization of processing centrals of SINIAV the Resolution No. 412 does not define it. Just as the previous resolution on the matter, there was an indication of the central between system components without having any explicit rule for its implementation and operation.

D. *Computer systems*

The early deployment of SINIAV in the States and the Federal District presupposes the existence of a reading equipment, registration and active supervision and its connection to a computer system for recording data of the electronic boards, connected to RENAVAM system.

Systems, information antennas and servers that are interconnected, must have security system that maintains the integrity of its specifications and content.

E. *Protocols*

In the Resolution No. 412, the CONTRAN determined that the protocol used for communication between the electronic board and the antennas cannot be proprietary.

The Brazilian Technical Standard specified in Annexes of the Resolution should be followed or, in absence of that, the International Technical Standard similar or equivalent in order to ensure the interoperability of the system throughout the national territory.

V. HOMOLOGATION OF EQUIPMENT

The devices of which the SINIAV system will consist of must be approved by the highest traffic executive of the country in accordance with the technical characteristics specified in Annex II of Resolution No. 412 and in particular Ordinance of the highest traffic executive of the country, with the consent of CONTRAN.

Even during the validity of Resolution No. 212, the DENATRAN published Ordinance 570, on 27 June, 2011, which establishes rules and defines the minimum requirements for certification and homologation of the products used to build the National Automatic Vehicle Identification System - SINIAV [28] .

VI. IMPLEMENTATION AND OPERATIONALIZATION OF SINIAV

Resolution No. 412 imposes on the bodies or entities of National Traffic System (NTS), the responsibility for the implementation and operation of SINIAV within the limits of the powers assigned to them. DENATRAN, the highest traffic executive entity in Brazil, will be the developing, deploying and operating the computerized system and national database, which integrate the computer systems and local databases, as well as determining the specification of technical requirements of the system, storage and transmission of information, the frequency of databases updates and transmission of information, especially regarding the security.

At the state and Federal District level, the organs or transit executives entities should perform the installation of the electronic board in vehicles and their registration in the national database of SINIAV and RENAVAM.

Municipalities, organs and SNT members entities, will have obligation to integrate their systems to the national database of SINIAV, directly or through agreements with other agencies or entities SNT members. However, public bodies and entities which are not part of the SNT, within their authority, can be incorporated to SINIAV, through an agreement with agency or public entity member of the SNT.

Resolution No. 412 allows private companies whose line of work is set by the highest executive transit entity of Brazil and which express interest in SINIAV, integrate to the system through organ or member of the SNT entity but does not allow full access to information.

Private companies will have access only to vehicles owned by them or whose owners have authorized such access.

According to Resolution No. 131 on 9 December, 2012, of CONTRAN the implementation process of the SINIAV, should have started on 1 January 2013 and be completed until 30 June, 2015.

VII. DATA BASE AND CONFIDENTIALITY OF INFORMATION

ISBN: 978-1-61804-318-4

The SINIAV system contains data of passing vehicle passing and exceptionally other data.

The record of the vehicle passing next to the antenna will be sent simultaneously to local and national databases and should not contain or store information which makes it possible to identify the vehicle owner.

As the exception in the database will be registered vehicles of undocumented circulation or with some kind of restriction. Only the systems provided by the highest traffic executive of the country may feed the database with the exceptions, this action will be forbidden any other means.

The exclusion of the record from database will be made only by the proper organ. The registration of monocoque, engine and plate number will be made by RENAVAM in the exception database, can only be made when demonstrating the need, effectiveness and security according to DENATRAN criteria.

Resolution No. 412 was expressed the confidentiality of information obtained through the SINIAV system, which must follow the terms of the Constitution and the laws on the matter. This information is for use of public bodies and entities that comprise, for the purposes and powers conferred on them.

VIII. SINIAV PILOT PROJECT - DETRAN-RIO DE JANEIRO AND SEGULL COMPANY

The pilot project performed by the DETRAN RJ and Segull, brought together representatives of national bodies and entities as well as the representation of the State of Rio de Janeiro. The project ended on 30 April 2011.

The work proceeded in accordance with the legal provisions of SINIAV and their technical requirements. The following components were used:

- control center installed in DETRAN-RJ – which simulated actions of DENATRAN and DETRAN;
- emission station - where electronic boards of SINIAV were recorded;
- verification station (checkpoint)- installed at the exit of emission station;
- supervision station (checkpoint) - installed on existing fixed unit;
- movable supervision unit- with antenna reader, equipment of optical character recognition (OCR), movable computer and communication via global system for mobile units (GSM) with central control and IP camera for closed circuit television (CCTV);
- 120 electronic boards provided by the Wernher Von Braun Laboratory;
- 112 vehicles.

After performing all tests and procedures, assessment of each step was made. The conclusion on the system performance was satisfactory what opened way for future deployment of the system.

CONCLUSION

Villages, towns, cities and metropolitan areas are subject of theft of vehicles and cargo, movement of vehicles with cloned plates, the occurrence of accidents, traffic of dangerous goods and congestion. The problems are most severe in metropolitan regions where the flow is higher, but also demand a response in the peripheral municipalities where safety of the citizens is ensured.

The National System for the Prevention, Control and Suppression of Vehicle and Cargo Theft and Robbery was proposed as a security measure and enabled the design of SINIAV for the identification of vehicles with the use of the RFID technology, in all Brazilian states.

The use of RFID is occurs in supplies and products control industry but also in the case RJ State experience presented above, which also shows the safety for traffic control and tracking vehicles and cargo, being a viable technology, safe and easy to apply in the immense field of Intelligent Transportation Systems.

As the relevant legislation enables municipalities to integrate data based on national SINIAV, either directly or through agreements with other agencies or SNT members, allow access, within legal limits to information for further studies and use for the planning of urban mobility, help in design and creation of other technologies and applications that facilitate monitoring of traffic and circulation of vehicles within cities.

The implementation of centers of urban mobility and public safety integrating the monitoring of events, the processing of traffic and transportation information and facilitating the rapid activation of the respective operational areas becomes possible and viable, especially in large urban centers, with the SINIAV implementation. Especially as the availability of information in real time, the use of open and standardized protocols for communication are required to the functioning of smart and sustainable cities.

REFERENCES

[1] Levinson, et al, 2004, “Weighting Waiting: Evaluating the Perception of In-Vehicle Travel Time Under Moving and Stopped Conditions”, Transportation Research Record: Journal of the Transportation Research Board, No. 1898, pp. 61-68.

[2] Fanke H and Dangelmaier W, 2004, “A web-based multi-agent system for transportation management to protect our natural environment”, Cybernetics and Systems, vol. 35:7, pp. 627-638.

[3] Böse F, Piotrowski J and Scholz-Reiter B., 2008, “Autonomously controlled storage management in vehicle logistics–applications of RFID and mobile computing systems”, International Journal of RF Technologies: Research and Applications, pp. 1-20.

[4] Jinghui Q, Bo S and Qidi Y, 2006, “Study on RFID Antenna for Railway Vehicle Identification”, 6th International Conference on ITS Telecommunications, pp. 237-240.

[5] “Internet of things – from Research and Innovation to Market Development”, ed. O. Vermesan and P. Friess, 2014, River Publishers Series in Communication, 143 p.

[6] Oneyama, H., Oguchi, T. and Kuwahara, M., 2001, “Estimation Model of Vehicle Emission Considering Variation of Running Speed”, Journal of the Eastern Asia Society for Transportation Studies , Vol.4, No.5, pp.105-117.

[7] Lu, M. et al, 2005, “Perspective of Mitigating Shock Waves by Temporary In-Vehicle Dynamic Speed Control”, Paper Submitted to the 84th Annual Meeting of the Transportation Research Board, TRB 2005 Annual Meeting CD-ROM, Washington, D.C.

[8] DENATRAN, (2014. Abr 04). “Frota de Veículos” [Online]. Disponível em : http://www.denatran.gov.br/frota2014.htm

[9] FETCESP, (2014, Abr. 04). “Levantamento estatístico”, [Online] Disponível em: http://www.fetcesp.net/estatisticas-de-seguranca.php

[10] G1, (2014, Dez. 30) “Número de roubos de carga aumenta 94,8% no RJ, segundo dados do ISP”, [Online]. Disponível em: http://g1.globo.com/rio-de-janeiro/noticia/2014/12/numero-de-roubos-de-carga-aumenta-948-no-rj-segundo-dados-do-isp.html

[11] E. A. Vasconcellos, “Congestionamento no trânsito e financiamento da mobilidade – avaliação dos estudos no Brasil e das perspectivas metodológicas”, Revista dos Transportes Públicos – ANTP, Ano 36 pp. 7-27, 1º quadrimestre, 2014.

[12] A. C. Moraes, “Congestionamento urbano: custos sociais”, Revista dos Transportes Públicos – ANTP, Ano 36, pp. 41-48, 3º quadrimestre, 2013.
[13] A. Camargo, “Depoimento”, Cadernos FGV Projetos, Ano 9, nº 24, pp. 8-11, Ago, 2014.
[14] C. Leite, “Cidades sustentáveis. Cidades inteligentes”,
[15] P. Junqueira, “Por dentro do Centro de Operações da Prefeitura do Rio de Janeiro”, Cadernos FGV Projetos, Ano 9, nº 24, pp. 76-84, Ago, 2014.
[16] ANTP, “Sistemas Inteligentes de Transportes”, Série Cadernos Técnicos, v. 8, p. 12. Disponível em: http://www.antp.org.br/_5dotSystem/download/dcmDocument/2013/03/18/9AB9A3EB-97DC-4711-9751-162AD361D7F0.pdf
[17] BRASIL, Lei Complementar n. 121, de 9 de fevereiro de 2006. Cria o Sistema Nacional de Prevenção, Fiscalização e Repressão ao Furto e Roubo de Veículos e Cargas e dá outras providências. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/lcp/lcp121.htm
[18] BRASIL. Resolução CONTRAN n. 212, de 13 de novembro de 2006. Dispõe sobre a implantação do Sistema Nacional de Identificação Automática de Veículos – SINIAV em todo o território nacional. Disponível em: http://www.denatran.gov.br/download/Resolucoes/RESOLUCAO_212.rtf .
[19] BRASIL. Resolução CONTRAN n. 412, de 9 de agosto de 2012. Dispõe sobre a implantação do Sistema Nacional de Identificação Automática de Veículos – SINIAV em todo o território nacional. Disponível em: http://www.denatran.gov.br/download/Resolucoes/(Resolução%20412.2012).pdf
[20] A. C. A. Ritto, “Tecnologias para gestão do trânsito de veículos: a solução Seagull para o SINIAV”, Rio de Janeiro: Claudio Ventura Comunicação, 2012.
[21] M. C. Pedroso, R. Zwicker, C. A. Souza, “Adoção de RFID no Brasil: um estudo exploratório”, Revista de Administração Mackenzie, v. 10, n. 1, pp. 12-35, jan./fev. 2009.
[22] C.C. Navarro, A. P. C. Grillo, R. S. Lima, “Análise e proposição de melhorias no processo logístico em uma multinacional de tecnologia da informação e automação”, Encontro Nacional de Engenharia de Produção, Rio de Janeiro, 2008.
[23] D. H. Barbosa, “Desenvolvimento de um modelo de referência para a aplicação da tecnologia RFID na logística de ambulatórios de ensino”, 2012, 201 pp., Tese (Doutorado) – Escola de Engenharia de São Carlos, São Carlos, 2012. Disponível em: http://www.teses.usp.br/teses/disponiveis/18/18156/tde-11072014-101719/pt-br.php
[24] V. Nassar, M. L. H. Vieira, “A aplicação do RFID na logística: um estudo de caso do Sistema de Infraestrutura e Monitoramento de Cargas do Estado de Santa Catarina”. ***Gest. Prod***., São Carlos, v. 21, n. 3, pp. 520-51, 2014. Disponível em: http://dx.doi.org/10.1590/0104-530X966.
[25] M. C. M. O. Nemoto, “Inovação tecnológica: um estudo exploratório de adoção do RFID (Identificação por Radiofrequência) e redes de inovação internacional”, 2009, 126 pp., Tese (Doutorado) – Escola Politécnica da Universidade de São Paulo, São Paulo, 2009. Disponível em: http://www.teses.usp.br/teses/disponiveis/12/12139/tde-18122009-105036/pt-br.php
[26] R. Candido, “Modelagem de processo ‘supply chain’ informado usando tecnologia RFID: estudo de caso para a cadeia do agronegócio”, 2013, 281 pp., Tese (Doutorado) – Escola Politécnica da Universidade de São Paulo, São Paulo, 2013. Disponível em: http://www.teses.usp.br/teses/disponiveis/3/3135/tde-26062014-105310/pt-br.php
[27] M. L. R. P. Dias, “Cadeia logística segura brasileira: suprimento internacional de carne bovina industrializada e rastreabilidade”, 2012, 187 pp., Dissertação (Mestrado) – Escola Politécnica da Universidade de São Paulo, São Paulo, 2012. Disponível em: http://www.teses.usp.br/teses/disponiveis/3/3143/tde-19112012-112942/pt-br.php
[28] BRASIL. Portaria DENATRAN n. 570, de 27 de junho de 2011. Estabelece regras e define os requisitos mínimos para a certificação e homologação de produtos do Sistema Nacional de Identificação Automática de Veículos - SINIAV. Disponível em: http://www.denatran.gov.br/download/Portarias/2011/PORTARIA_DENATRAN_570_11.pdf

ISBN: 978-1-61804-318-4

Combining KNN and Decision Tree Algorithms to Improve Intrusion Detection System Performance

Kazem Fathi1, Sayyed Majid Mazinani2
Islamic Azad University1, Sari Branch, Iran
Imam Reza International University2, Mashhad, Iran

Email: { kzm_fathi@yahoo.com, smajidmazinani@hotmail.com}

ABSTRACT — Two types of algorithms are realized which have been used within the supervised of model of intrusion detection systems. These algorithms are either of type eager or lazy as far as their performance is concerned. At the learning phase, the lazy algorithms are fairly simple, however, the eager algorithms are highly effective. On the other hand the classification phase is in at most contrast with learning phase. The aim of this research is, taking the advantages of both lazy and eager algorithms to achieve a hybrid algorithm. This approach necessitates employing an eager algorithm of Decision Tree, on the training set, which has led to the creation of a set of Decisions. This set of Decisions is applied on the training set, which results in having a set of binary vectors. In order to enhance the training set these binary vectors were added as new attributes. After that with the lazy algorithm of nearest neighbors, we have classified the samples. The outcome of test results from existing algorithms has been compared with our proposed algorithm. The results show that the proposed algorithm outperforms where the volume of samples are high. The performance of the hybrid algorithm is also remarkable within platforms, with limited or very high processing resources.

Index Terms — Intrusion **Detection System, Machine Learning, Classification.**

I. INTRODUCTION TO IDS

The growth of attacks on networks continues to increase. Attacks range from relatively benign ping sweeps to sophisticated techniques exploiting security vulnerabilities [1].

Intrusion detection is the task of detecting and responding to computer misuses [2]. In other words, an IDS is typically a computer system, which monitors activity to identify malicious activities [3].

II. MACHINE LEARNING

Machine learning is an answer for how to construct computer programs that automatically learn with experience [4]. Machine learning algorithms automatically extract knowledge from machine readable information [5]. Two types of machine learning are supervised learning and unsupervised learning. Supervised learning uses labeled training data, while unsupervised learning uses unlabeled training data. Supervised algorithms can classify samples into specific classes [6].

Eager learning is a form of supervised learning, at this form there is a learning module, classification module and a model. Eager learning algorithms invest most of their effort in the learning phase. Usually these algorithms construct a representation of the target function from training instances. Classification of new instances is usually done by rules that uses the model [7].

Opposite of eager learning there is also lazy learning as a different form of supervised learning. In different contexts, memory-based learning algorithms are named lazy [8].

In lazy learning during the learning phase, all input samples are stored and no other attempt will be made [9]. The search for the optimal hypothesis takes place during the classification phase [10].

III. SOME MACHINE LEARNING ALGORITHMS

Decision Tree is a representation of how to make a decision according to a particular attribute set [6]. Any given Decision Tree is completely deterministic, [11]. The decision tree nodes have some attributes, with the

ISBN: 978-1-61804-318-4

branches involving alternative values of those attributes. The leaves represent the classes. To make a decision by a decision tree, select a set of values for an attribute and start at the root of the tree. Decision Tree is a form of supervised learning.

The naive Bayes model is a heavily simplified Bayesian probability model [6]. In simple terms, a naive Bayes classifier assumes that the presence or absence of a particular feature is unrelated to the presence or absence of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 3" in diameter. A naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of the presence or absence of the other features.

For some types of probability models, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood; in other words, one can work with the naive Bayes model without believing in Bayesian probability or using any Bayesian methods. Rule induction is a form of eager learning. In this form during the learning phase, the rules can be induced from the training samples. The goal of rule induction is generally to induce a set of rules from data that captures all generalizable knowledge within that data, and that is as small as possible at the same time [12]. The rules that are extracted during the learning phase can easily be applied during the classification phase when new unseen test data are classified.

There are several instance-based learning algorithms. One of the best known is k-Nearest Neighbor (k-NN) [13]. The learning phase of k-NN is simply storing training samples. During the classification phase, k-NN use of a similarity-based search strategy to determine an optimal hypothesis function. New instances will be compared to the stored instances and will be classified the same class label as the k most similar stored instances [7].

IV. Proposed Hybryd algorithmes

After eager and lazy algorithms, hybrids are considered. Used hybrids are a combination of the k-NN classifier and Decision Tree. Goal of constructing hybrids is to create relationships between memory-based learning and eager learning. Combining eager and lazy learners will produce machine learners that put effort in both the learning and classification phase. The hybrid will use the hypothesis as induced by Decision Tree, and the one created during memory-based learning [7].

We combined k-NN with rule-induction. Although rules seams fully different from instances that are being used in K-NN, there is a relation between them. Decisions represent a subset of training instances that match with specific conditions. Therefore, k-NN classification can be applied to Decisions [12].

In proposed Replacement Based Decisions (RBD) to create hybrid, Decisions are induced from the training set using Decision Tree algorithm. Then the decisions are transformed into vectors, representing the binary decision-features. The vectors that are produced would seem as new training set [12]. When using RBD, the binary rule-features replace the original features in the instances. From the k-NN view, RBD attempts to reduce the sensitivity of K-NN to noise and irrelevant features [7].

After RBD we propose a second type of hybrid, named Adding Based Replacement (ABD). When using this hybrid, the initial features of the instance are not replaced by the binary decision-features. These binary decision-features are added to initial features. Therefore, this new algorithm is a k-NN classifier with extra added features. The rule-features is not noise filtering step, but we expect that the rule-features can help to repair K-NNs sensitivity to noise. Setting more weights on more important features in k-NN helps to achieve more accuracy [7].

V. ATTRIBUTE SELECTION

In general there are some reasons to use feature subset selection for machine learning algorithms. First of all, it can cause more performance. Secondly, it provides a reduced hypothesis search space, and third, attribute selection reduces the storages [5].

The first feature selection method that we used, is information gain. In order to define information gain precisely, a measure commonly used in information theory, called entropy [9].

The heuristic method, by which CFS measures “goodness” of feature subsets, takes into account the usefulness of individual features for predicting the class. CFS evaluates and hence ranks feature subsets rather than others [13].

VI. DATASET DETAILS

The data used are from the University of California, Irvine Knowledge Discovery and Data Mining (UCI

KDD) website [14]. The data files give us the information to create and train the algorithms. Neither [14] nor [15], the main references in this research for this dataset, mention what is attacking. The testing data for the 10 percent of data set contains 311029 examples. These examples contain 60593 normal items and 250436 attack items. Therefore, this data is most likely atypical because it contains more attacks than normal data. The training dataset contains 494020 items. There are 97277 normal items and 396743 attack items.

VII. EVALUATION METHODS

To evaluate the algorithms Weka is used. Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

To evaluate the performance of algorithms, different metrics were used. Metrics are calculated using the confusion matrix, which shows the predicted and actual classifications.

TABLE I. CONFUSION MATRIX

	Predicted Negative	Predicted Positive
Actual Negative	a	b
Actual Positive	c	d

False positive and false negative are two types of errors that can be made. The number in "b" in the confusion matrix corresponds with the type I errors. The number in "c" equals the number of type II errors made. Both false positives and false negatives have to be reduced [16], accuracy has to be as close to 1 as possible. The accuracy is the proportion of the total number of predictions that were correct, and is measured by: $Accuracy = \frac{a+d}{a+b+c+d}$. Precision is the proportion of the predicted positive cases that were correct; $\Pr ecision = \frac{d}{d+b}$. Another metric to evaluate algorithms is recall, it is the proportion of alerts that were correctly identified; $\mathrm{Re}call = \frac{d}{d+c}$. Because the equation mentioned above for accuracy may not be an adequate performance measure, the F-Measure was used. The sparseness of the positive examples could cause the average classification accuracy of the testing set to be unreliable [17]. For this reason, the weighted harmonic mean of precision and recall, with given β, can be calculated as: $F = \frac{(\beta^2+1)\times precision \times recall}{\beta^2 \times precision + recall}$

IIX. RESULTS

The results of the experiments are summarized in Table II; each cell is mean of 7 experiments.

TABLE II. EXPERIMENTS SUMMARIZED RESULTS

		Precision	Recall	F-Measure
RIPPER	Full	0.985	1	0.992
	CFS	0.977	0.992	0.984
	IG-10	0.994	0.999	0.996
	IG-6	0.990	0.998	0.994
	IG-2	0.990	0.993	0.991
KNN	Full	0.983	0.998	0.99
	CFS	0.973	0.979	0.976
	IG-10	0.984	0.997	0.99
	IG-6	0.994	0.999	0.996
	IG-2	0.984	0.990	0.987
Naive Bayes	Full	0.965	0.998	0.981
	CFS	0.976	0.928	0.951
	IG-10	0.976	0.988	0.982
	IG-6	0.976	0.989	0.982
	IG-2	0.972	0.937	0.954
Decision Tree	Full	0.970	0.993	0.982
	CFS	0.971	0.993	0.982
	IG-10	0.994	0.999	0.996
	IG-6	0.971	0.993	0.982
	IG-2	0.972	0.937	0.954
RBD	Full	0.986	0.998	0.992
	CFS	0.975	0.998	0.986
	IG-10	0.997	0.999	0.998
	IG-6	0.997	0.999	0.998

ISBN: 978-1-61804-318-4

	IG-2	0.993	0.996	0.994
ABD	Full	0.987	0.994	0.99
	CFS	0.978	0.999	0.988
	IG-10	0.991	0.997	0.994
	IG-6	0.996	0.998	0.997
	IG-2	0.989	0.986	0.996

It can be concluded that the results indicates the RBD, ABD and Rule Induction algorithms have far better performance in comparison with other algorithms. In order to find out which of these three algorithms are superior to others, their F-measure parameters are compared as is depicted in Figure 1.

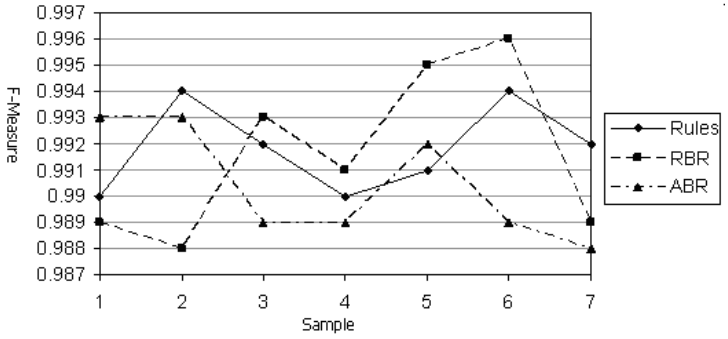


Fig 1. F-measure of 3 algorithms

For the next stage of analysis the three parameters of F-measure, Precision and Recall are compared for the RBD, ABD and Rule Induction algorithms. This comparison is done under the condition where few attributes are available. For example when the attributes are filtered with Infogian-2 method, Figures 2, 3, 4".

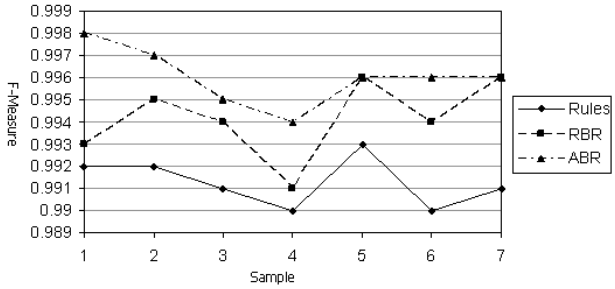


Fig 2. F-measure of 3 algorithms - Info-2

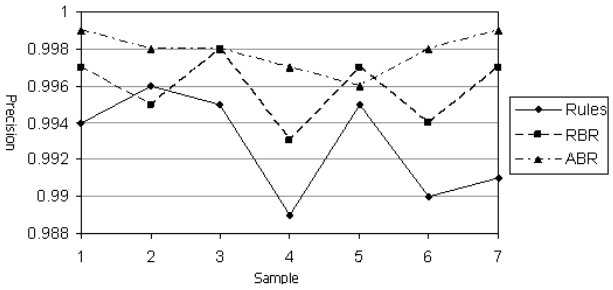


Fig 3. Precision of 3 algorithms - Info-2

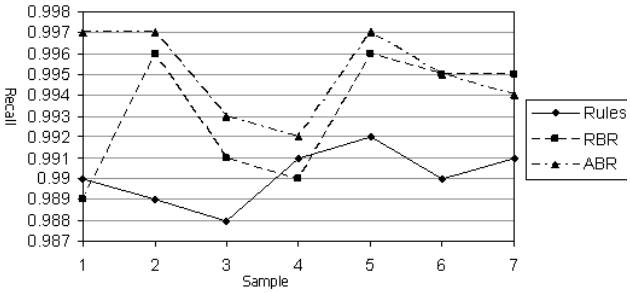


Fig 4. Recall of 3 algorithms - Info-2

Figure 5, shows the F-measure for RBD, ABD and Rule Induction algorithms where all attributes have been filtered with CFS method.

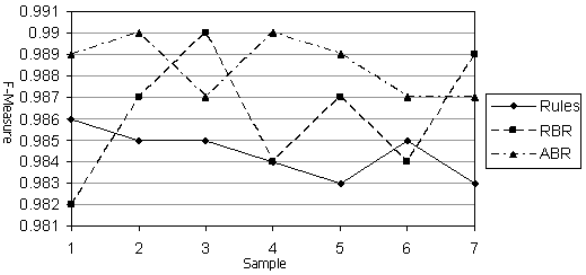


Fig 5. F-measure of 3 algorithms - CFS

Eventually it can be concluded that the Rule Induction algorithm in most instances has far superior performance under the condition where there are high volume attributes and highly susceptible to any attacks. Under circumstances where the volume of training set is not very big and the processing capability is not very high (in common environments), the Rule Induction algorithm is suggested. However when there are limitations on any or some parameters, either we are forced to select some of more effective attributes and have a high volume of training set and high processing capability is available, the ABD algorithm is preferable. Finally cause to the ABD is sensitive to process power, when there is limitation of the processing power, RBD algorithm performs better than the ABD algorithm.

REFERENCES

[1] Jackson, T., Levine, J., Grizzard, J., and Owen, H. (2004). An investigation of a compromised host on a honeynet being used to increase the security of a large enterprise network. In Proceedings of the 2004 IEEE Workshop on Information Assurance and Security. IEEE.

[2] J.X. Huang, J. Miao, Ben He, "High performance query expansion using adaptive co –training", Information Processing & Management 49 (2) (2013) 441–453

[3] Y. Li u, X. Yu, J.X. Huang. A." An, Combining integrated sampling with SVM ensembles for learning from imbalanced datasets", Information Processing &Management 47 (4) (2011) 617–631.

[4] S.X. Wu, W. Banzhaf, "The use of computational intelligence in intrusion detection systems: a review", Applied Soft Computing 10 (2010) 1–35.

[5] Yixue Wang, A Sort of Multi-Agent Cooperation Distributed Based Intrusion Detection System, Modem computer,2008

[6] S. J. Russell and P. Norvig, Artificial Intelligence: A Modern Approach (International Edition). Pearson US Imports & PHIPEs, November 2002.

[7] Hendrickx, I. (2005). Local Classification and Global Estimation. Koninklijke drukkerij Broese & Peereboom.

[8] van den Bosch, A. and Daelemans, W. (1998). Do not forget: full memory in memory-based learning of ord pronunciation. In NeMLaP3/CoNLL98, pages 195–204.

[9] Mitchell, T. (1997b). Machine Learning. McGraw-Hill International Editions.

[10] Daelemans, W., van den Bosch, A., and Zavrel, J. (1999). Forgetting exceptions is harmful in language learning. In Machine Learning, volume 34, pages 11–41.

[11] D. Fisher, L. Xu, J. Carnes, Y. Reich, S. Fenves, J. Chen, R. Shiavi, G. Biswas, and J. Weinberg, "Applying ai clustering to engineering tasks," pp. 51–60, Dec 1993.

[12] van den Bosch, A. (2004). Feature transformation through rule induction, a case study with the k-nn classifier. In Proceedings on the workshop on advances in Inductive rule learning ath the ECML/PKDD 2004, pages 1–15.

[13] Daelemans, W., Zavrel, J., van der Sloot, K., and van den Bosch, A. (2005). TiMBL: Tilburg Memory-Based Learning-Reference.

[14] S. Hettich and S. D. Bay, "Kdd cup 1999 data. http://kdd.ics.uci.edu//databases/kddcup99/kddcup99.html.irvine, ca: University of california, department of information and computer science."

[15] P.A.C.P. Stolfo J, Fan W. Lee W, "Cost-based modeling for fraud and intrusion detection: results from the jam project," in DARPA Information Survivability Conference and Exposition, vol. 2 of DISCEX, pp. 130–144, 2000.

[16] Gowadia, V., Farkas, C., and Valtorta, M. (2005). Paid: A probabilistic agent-based intrusion detection system. In Computers & Security, volume 24, pages 529–545.

[17] Kubat, M. and Matwin, S. (1997). Addressing the curse of imbalanced training sets: one-sided selection. In Proceedings of the 14th International Conference on Machine Learning, pages 179–186. Morgan Kaufmann.

An Inter-banking Cryptographic Algorithm to Tackle Rogue Trading

*Clara Aglaya Corzo P
University of Manchester
Department of Computer Science
Email: ccorzope@banrep.gov.co

Ning Zhang
University of Manchester
Department of Computer Science
Email: nzhang@man.ac.uk

Fabio Augusto Corzo S
Retired professor
Email: fcorzos@yahoo.com

***Abstract*—The misuse and or abuse of privileges by a banking trader can cost a bank millions. Here we present a framework for tackling inter-banking rogue trading fraud. The proposed workflow is more dynamic and complete, since it contemplates any two parties interacting on the basis of financial transactions recorded by their users in related but distinct automated financial systems. An auditing solution that involves intra-system, inter-system, intra-organizational and inter-organizational data.**

I. Introduction

Information and Communications Technology (ICT) has changed dramatically during the past twenty years. Network velocity has improved and better trading platforms have emerged for banking traders.Therefore, interconnections and also interactions have increased. Traders have more counterparts to choose from because they can involve more banks from more countries. The number of financial transactions entered by traders and also the aggregate amount of money negotiated by them has grown. Unfortunately, better trading facilities also mean highly dependent financial transactions are fabricated worldwide between banks introducing vulnerable cross related banking relationships which can create a domino effect if one bank fails to settle [12] [6].

One mistake or error from a trader can be very expensive for a financial institution. Particularly, the misuse and/or abuse of privileges by a trader has cost banking institutions millions [20] [19]. During the past two decades, there have been a couple of fraud cases by traders that caused imminent bank losses. Examples are Nick Leeson's case at Barings Bank losing £827 million, Yashuo Hamanaka 's case at Sumitomo Corporation losing $ 2.6 bn, John Rusnak 's case at Allied Irish Bank losing £355m, Jerome Kerviel 's case at Société Générale losing $5bn [18], and Toshihide Iguchi 's case at Daiwa Bank losing $1.1bn. They as traders (authorised users) of financial institutions managed to conceal great money failures. The above described problem, the misuse and or abuse of privileges by traders to perform unauthorised trading is commonly known as rogue trading. The problem requires the reinforcement of controls and the introduction of security systems to detect and prevent rogue trading [16] [14] [11] [13] [5].

Here we present a framework for tackling inter-banking rogue trading fraud. Timestamps are fundamental building blocks for the proposed solution. Therefore, section 2 provides background on this topic. Section 3 explains the strategy used to design the algorithm of the audit log file. Mainly, the solution is based on the idea of having an inter-system and an intra-system oriented audit log files. The above based on the fact that one workflow results from directly or indirectly related financial transactions, and that all transactions have some parameters that remain constant throughout the flow. The framework led us to a naive workflow oriented audit design architecture, which is presented in section 4. The novel solution provides means to early detect the concealment of unauthorised trading or transactions by authorised users. Section 5 summarises the achievements of this work.

II. Timestamps

A timestamp is a record or mark of a specific date and time. Timestamps can serve various functions within a company. For example, in some companies all the postal letters received are timestamped on reception. The timestamp, in this case, may be used to measure the efficiency of each department in processing the enquiries on the letters. Typically, timestamps are used as evidence that something has happened or was created on or before a certain date and time.

In the banking sector, the moment in time at which a financial transaction was recorded can determine the conditions of the transaction. For instance, a deal made on a trading floor to invest in a market instrument at 10 a.m. may be 10 times cheaper than making the same investment five minutes later. If evidence is required to prove the time at which a financial transaction was recorded then the timestamp is of great value. Time-stamping constitutes a mandatory process in most of the banking related activities.

Electronic timestamps should be integrity protected. That is, it should be very difficult to change the time and date of an existing timestamp. Timestamping schemes are of special interest to our proposed solution because of this characteristic, which is essential for a secure auditing solution. A timestamp may be vulnerable if it can either be back-dated or forward-dated. Assume two documents (d) are time-stamped, one at time t and the other one a little later at time $t+1$. Thus, a message is said to be back-dated if by looking at two timestamped documents d_t and d_{t+1} one could infer that d_{t+1} was timestamped before d_t when it was actually constructed after d_t. A message is said to be forward-dated if by looking

ISBN: 978-1-61804-318-4

at two timestamped documents d_t and d_{t+1} one could infer that d_t was time-stamped after d_{t+1} when it was actually constructed before d_{t+1}. It should be very difficult to change an already emitted timestamp.

Cryptographic primitives are fundamental in the construction of trustworthy timestamps because they help provide integrity protection. There are two well known timestamping schemes involving cryptography: the simple timestamping scheme and linking timestamping schemes.

A. Simple Timestamping Scheme

In the simple timestamping scheme [17], each timestamp does not depend on data from any previously generated time-stamps. An entity that requires a time-stamp token [1] sends a TSA[2] a request for a time-stamp with the hash value of a document $h(d)$. The TSA will generate a timestamp token of the form

$$TS = h(d)||t||ID||S \tag{1}$$

where t is the current time, ID is the identity of the TSA, and S is the TSA's signature of all items. This token is then sent back to the requester and stored with the document. This process is similar to having a reliable person manually timestamping documents one after the other. The simple timestamping scheme process relies completely on the party that is time-stamping the documents i.e. TSA.

The problem with this scheme is that an altered token, generated by a malicious TSA, cannot be detected easily. Linking timestamping schemes were developed to address this problem.

B. Linking Timestamping Schemes

Linking timestamping schemes can be classified as either linear linking schemes or aggregation schemes. In contrast to the simple timestamping scheme, in a linking scheme the timestamped data is dependent on previously timestamped data.

1) Linear Linking Schemes: A linear linking scheme (LLS), was first introduced by Haber and Stornetta [2]. In this scheme, all time- stamped documents are linked together in a chain using collision resistant hash functions. Each new timestamp depends on the previously issued timestamp, as shown in Figure 1.

More precisely, a time-stamp (TS) on a i^{th} submitted document d is of the form

$$TS_i = [i||t_i||Id_{R_{t_i}}||h(d)_{t_i}||L\_b_{t_i}||S_{t_i}] \tag{2}$$

where t_i is the current time, $Id_{R_{t_i}}$ is the requester's identifier, $h(d)_{t_i}$ is the hash value generated from d, and S_{t_i} the TSA's signature. The $L\_b_{t_i}$ expression is a recursive equation,

[1]Time-Stamp token: a data object that binds a representation of data to a particular time and date, establishing evidence that it existed before that time.

[2]TSA: A Time Stamping Authority (TSA) is a party that offers a mechanism that provide evidence that data existed from a specific date and time. A TSA is usually regarded as a trusted third party (TTP).

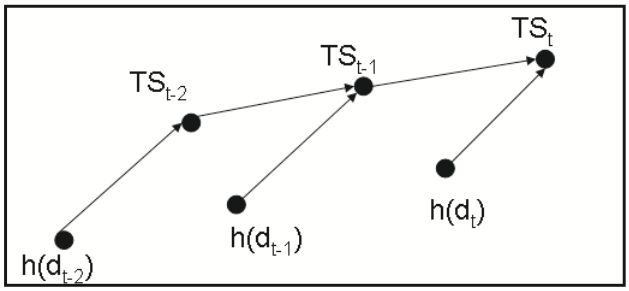


Fig. 1: Graphical representation of the linear linking scheme presented by Haber and Stornetta in [2]

known as the linking bits, which links each time-stamp to the previously issued time-stamp. The recursive equation can be represented as

$$L\_b_{t_i} = t_{i-1}||Id_{R_{t_{i-1}}}||h(d)_{t_{i-1}}||L\_b_{t_{i-1}} \tag{3}$$

If a user's timestamped document $h(d)_{t_i}$ is challenged, then the challenger first checks the signature S_{t_i}. Then, in order to check that TS_i has not colluded with TSA, the challenger can call $Id_{R_{t_{i+1}}}$ and ask to produce a timestamp using $L\_b_{t_i}$. The challenger can call other users and authenticate further the linking information. This scheme no longer relies completely and solely on the TSA signature. As each timestamp is also dependent on previously issued time-stamps, the TSA cannot easily forward-date nor back-date a TS because it would require bits from other previously issued timestamps (of other documents and probably of other users). These bits would have to be constructed in collaboration with other users of the TSA. The problem with this scheme is the amount of time that the verification process may require. In the worst case it may be necessary to compute all previously issued time-stamps in order to verify a given scenario timestamp. Aggregation timestamping schemes were introduced to overcome this problem.

2) Aggregation Schemes: Aggregation schemes are very similar to the linear linking scheme, except that an aggregation method is used instead of the recursive equation to link the documents. That is, in an aggregation scheme, each new timestamp is created using the data from two or more documents as well as data from the previously issued timestamp. There are two well known aggregation schemes; one was proposed by Haber and Stornetta [2], and the other was proposed by Benaloh and de Mare.

In the approach by Haber and Stornetta [2], a TS required for a i^{th} submitted document d is

$$TS_i = [i||t_i||Id_{R_{t_i}}||h(d)_{t_i}||L\_b_{t_i}||S_{t_i}] \tag{4}$$

where t_i is the current time, $Id_{R_{t_i}}$ is the requester's identifier and $L\_b_{t_i}$ is a recursive equation that links each time-stamp to x previously issued time-stamps.

The recursive equation can be represented as

$$L\_b_{t_i} = A||B||...||C \tag{5}$$

Where

$$A = t_{i-1}||Id_{R_{t_{i-1}}}||h(d)_{i-1}||h(L\_b_{t_{(i-1)}}) \quad (6)$$

$$B = t_{i-2}||Id_{R_{t_{i-2}}}||h(d)_{i-2}||h(L\_b_{t_{(i-2)}}) \quad (7)$$

$$C = t_{i-x}||Id_{R_{t_{i-x}}}||h(d)_{i-x}||h(L\_b_{t_{(i-x)}}) \quad (8)$$

where x is the number of previously issued time-stamps to which the new issued timestamp token will be linked.

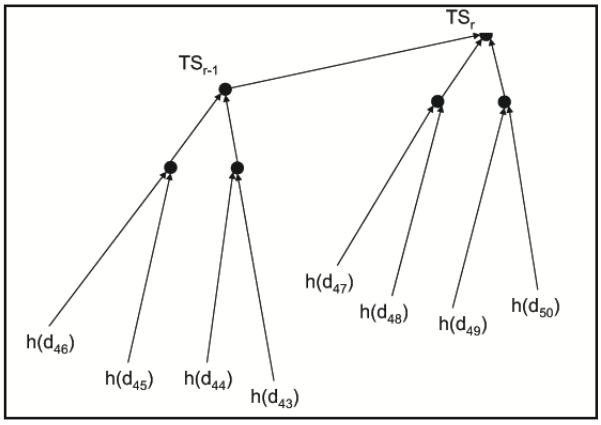


Fig. 2: Graphical representation of aggregation scheme presented by Haber and Stornetta in [7]

An improvement to this approach was presented by Haber and Stornetta in [7], in which each timestamping procedure was divided into rounds. Each TS_r for round r is a cumulative hash of TS_{r-1} for round $r-1$ and of all the documents submitted to the TSS during round r (See Figure 2).

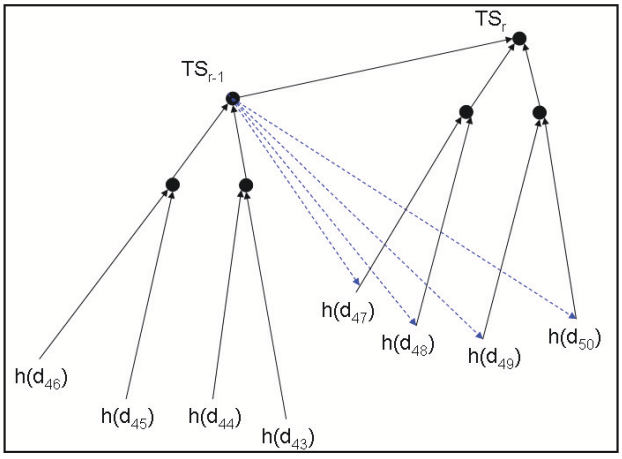


Fig. 3: Graphical representation of aggregation scheme presented by Benaloh and de Mare in [4]

The other well known aggregation scheme is that introduced by Benaloh-de-Mare [4]. This aggregation scheme is very similar to the one presented by Haber and Stornetta. The difference is that in this proposal, each time-stamped document d of each round r is hashed together with the previously issued time-stamp TS_{r-1} (see Figure 3). A binary tree is built after the end of each r round on both of the above aggregation schemes.

Aggregation schemes are also known as tree linking schemes. Another example, and well known aggregation scheme, is the tree-like linking data structure introduced by Ralph Merkle [1]. The Merkle's data structure is known as a Merkle Hash Tree (MHT) and it has the form of a binary tree. Leaves of an MHT contain the hash value of data or blocks of data. Each parent in the tree contains a hash value of the concatenation of the data of its two children (see Figure 4). MHTs were developed for generating a cryptographic digital signature which did not require asymmetric encryption.

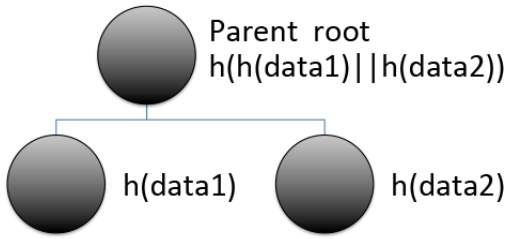


Fig. 4: Graphical representation of a simple Merkle Hash Tree (MHT)

The advantage of this structure for authenticating is that an entity that needs to authenticate one data from a MHT requires only those hash values starting from the leaf and progressing to the root. This reduces the amount of data required for the authenticating procedure from y, where y is the number of data to be authenticated from a tree, to $log_2 y$. One further advantage of this scheme is that in a MHT it is impossible to add new leaves once the root has been computed. This can help to detect data that are added or deleted illegitimately by authorised users.

Coming back to aggregation schemes in general, the problem with them is that they only provide relative temporal authentication (RTA) to some degree. In other words, the scheme does not provide a technique to be able to prove that a document in one same round was time-stamped after or before other document of the same round. The problem exists because one time-stamp is shared for each group of document.

However, cryptographic aggregate schemes are very useful and they have been applied to solve a variety of problems. For example, they have been used to extend the lifetime of certifications [3], they have been used to provide collective signatures [4], they have also been used to create certificate revocation lists for digital certificates [8] [10]. They are, how-

ever, mainly known for their use as time-stamping schemes [9] [15].

III. The Proposed Solution Strategy

Our solution begins with an observation. Some parameters of a financial transaction remain constant during the e-trading workflow activity. In other words, some values in the set of transactions related to one deal - during the settlement process of an e-trading deal do not change. See the following related financial transactions:

1) An inter-banking deal of Bank A entered by Ann via the Financial Exchange system (FE). Bank A wants to buy Bank B a £5000 market instrument MI_1. In order to exemplify, we name this transaction TN1.

TABLE I: FE Transaction Detail

Transaction 1 (TN1)			
Byr	Sys	MI Desc	Val MI
Bank A	FE	MI_1	£5000

2) Similarly, an inter-banking deal of Bank B is entered by Fraser via the Financial Exchange system (FE).BAnk B wants to sell Bank A a £5000 market instrument MI_1. In order to exemplify, we name this transaction TN2. TN1 and TN2 match and therefore a business is settled.

TABLE II: FE Transaction Detail

Transaction 1 (TN2)			
Slr	Sys	MI Desc	Val MI
Bank B	FE	MI_1	£5000

3) TN1 and TN2 match and therefore business is settled.
4) TN1 triggers a payment transaction of Bank A which will be made by Bob via the ACH system. A payment of £5000 in order to acquire MI_1.

TABLE III: ACH Transaction Detail

ACH Transaction 3 (TN3)				
Byr *	Slr	Sys	MI Desc	Val MI
Bank A	Bank B	ACH	MI_1	£5000

5) TN2 triggers a market instrument transfer of Bank B made by Sue via the CSD system. The transfer of a £5000 market instrument MI_1.

TABLE IV: CSD Transaction Detail

CSD Transaction 4 (TN4)				
Byr	Slr *	Sys	MI Desc	Val MI
Bank A	Bank B	CSD	MI_1	£5000

See that whereas the initial deal settled between Bank A and Bank B triggers financial transactions inside both banks. That is, the FE deal made by Ann in the name of Bank A triggers a payment which should be made at Bank A by Bob. Similarly, the FE deal made by Fraser in the name of Bank B triggers a transfer which is entered by Sue from the seller bank.

TABLE V: Transaction Set Detail Matrix

Transactions					
Trans	Byr	Slr	Sys	MI Desc	Val MI
TN1	Bank A	Bank B	FE	MI_1	£5000
TN2	Bank A	Bank B	FE	MI_1	£5000
TN3	Bank A	Bank B	ACH	MI_1	£5000
TN4	Bank A	Bank B	CSD	MI_1	£5000

By looking at all the above presented financial transactions in one same matrix it can be seen that some data remain constant throughout the workflow.

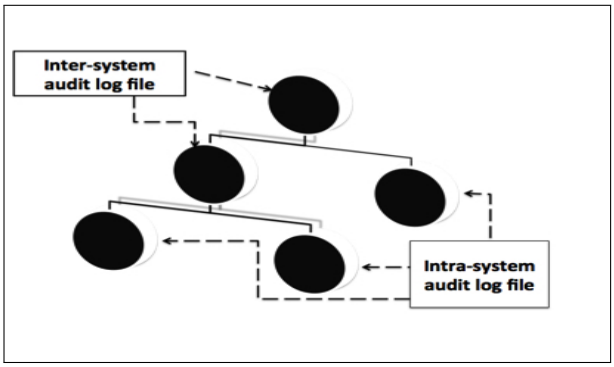
The crosscheck of those values which remain constant (i.e. £5000, MI_1,- in Bank A and Bank B -) can help detect unauthorised financial transactions entered by mistake or on purpose by any authorised user within the workflow activity. What someone buys someone sells. When someone is paying something bought it is also true that someone will have to transfer what has been sold. Based on the above observation our strategy is to define a workflow oriented audit trail based on an interleaving cryptographic algorithm. This audit trail contains data records from different sources. It contains data from financial transactions carried out by users from the same bank and from other counter-part banks. The transactions are related to the same deal and are generated from different automated financial systems. This is a dynamic workflow which is based on information coming from each financial transaction -from each bank-.

IV. The Audit Log Architecture Design

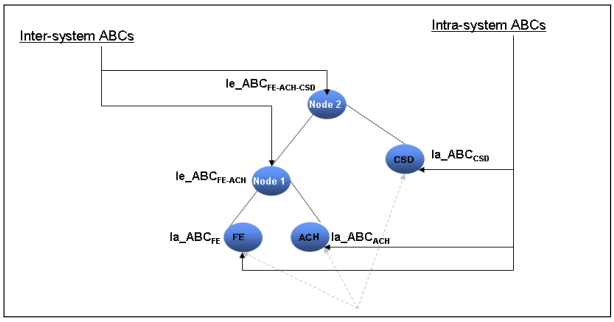
We name this audit log the Automated Banking Certificate (ABC). One complete ABC has the form of a binary tree (see Figure 5a). The leaves of the tree are called Intra-system ABCs. Each Intra-system ABC is the audit log file of a financial transaction performed in one automated financial system. The parents in the tree are called Inter-system ABCs. An Inter-system ABC is used to bring together the audit trail of two or more related Intra-system ABCs from the same bank. For example, in Figure 5b, Ia_ABC_{FE}, Ia_ABC_{ACH} and Ia_ABC_{CSD} are intra-system ABCs (they are leaves in the tree) whereas Ie_ABC_{FE-ACH} is an inter-system ABC (they are parents in the tree).

One complete ABC data structure combines two linking schemes; the linear linking scheme and an aggregation scheme. Each Intra-system ABC is created at an automated financial system using the linear linking scheme. That is, an Intra-system ABC contains linking data to the previously issued Intra-system ABC which in turn also contains linking data to the previously issued intra-system ABC (this is repeated recursively). For example, on Figure 5c one can see that $Ia\_ABC_2[FE]$ contains linking data to $Ia\_ABC_1[FE]$. Similarly, $Ia\_ABC_3[FE]$ has linking data to $Ia\_ABC_2[FE]$. A data dependent chain is therefore created with other Intra-system ABCs. The inter-system ABC will contain data from two or more constructed linking scheme data chains using the Merkle Hash Tree (MHT) aggregation scheme.

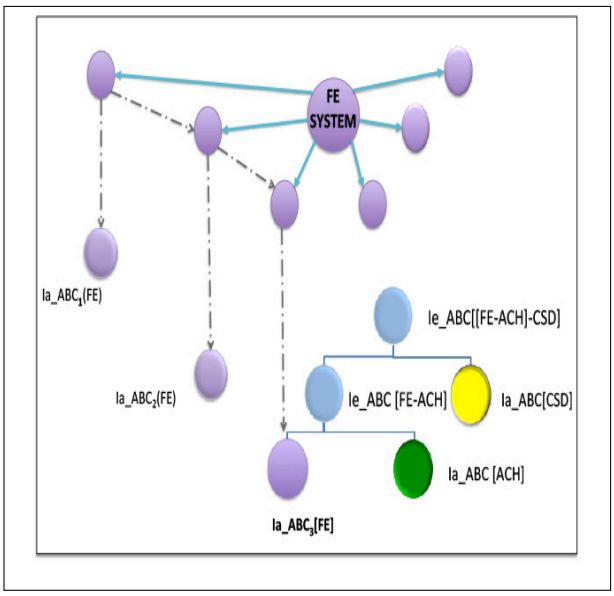
ISBN: 978-1-61804-318-4



(a) General design



(b) Design of one complete audit log file



(c) Complete audit log file showing FE system in context

Fig. 5: Graphical representation of the Automated Banking Certificate

A. The Workflow Oriented Audit Log File

An Intra-system ABC is created by a trusted party located at each automated financial system. Such a trusted party is implemented using a software agent. The trusted party at FE is called the FE_SP_Agent. The trusted party at ACH is called the ACH_SP_Agent. The trusted party at CSD is called the CSD_SP_Agent. Each trusted party has the ability to retrieve data from all the financial transactions successfully recorded at the automated financial service at which it is located.

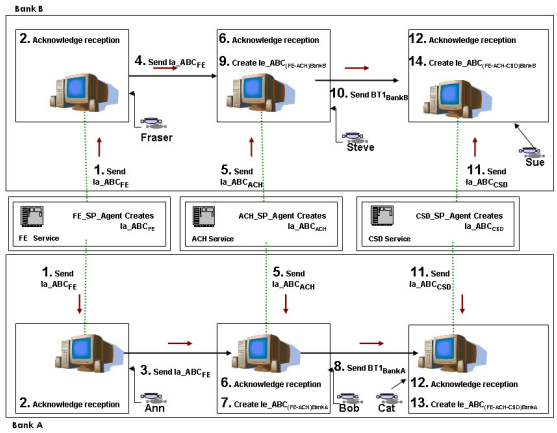


Fig. 6: The Intra-system Automated Banking Certificate (Intra-system ABC) construction process

An Intra-system ABC is created after each financial transaction is successfully recorded at an automated financial service. This occurs during e-trading workflow activity. In our scenario, the creation process is as follows (see each task addressed by a corresponding number on Figure 6):

- Ann, at Bank A, logs onto the financial service FE. Then, she performs an inter-banking e-trading deal to buy a market instrument.
- Fraser, at Bank B, logs onto the financial service FE. Then, he performs an inter-banking e-trading deal to sell a market instrument. These two deals (i.e. Ann's deal and Fraser's deal) should match. Otherwise, the FE system will not accept the deal.
- The FE_SP_Agent detects transactional activity. It retrieves and saves the transactional data into its database. Then, the FE_SP_Agent issues an Intra-system ABC, denoted as Ia_ABC_{FE}. Ia_ABC_{FE} contains the hash values of transactional items and values that will be cross-verified at a later time. The Ia_ABC$_{FE}$ is signed by the FE_SP_Agent.
- 1. FE_SP_Agent sends Ia_ABC$_{FE}$ to both of the FE_User_Agents.
  Ia_ABC$_{FE}$ is sent to FE_User_Agent$_{Ann}$ and to FE_User_Agent$_{Fraser}$.
- 2. On reception, FE_User_Agent$_{Ann}$ and FE_User_Agent$_{Fraser}$ will acknowledge each received Intra-system ABC by signing it with its respective private key. This will be done after checking that the signature of the Ia_ABC$_{FE}$ is authentic.
- 3. FE_User_Agent$_{Ann}$ sends Ia_ABC$_{FE}$ to ACH_User_Agent$_{Bob}$.
- 4. FE_User_Agent$_{Fraser}$ sends Ia_ABC$_{FE}$ to ACH_User_Agent$_{Steve}$
- Bob logs onto the financial service ACH. Then, he performs an inter-banking payment to settle the agreed

ISBN: 978-1-61804-318-4

e-trading deal.

- The ACH_SP_Agent detects the transactional activity. The ACH_SP _Agent retrieves and saves the transactional data into its database. Then, it generates an Intra-system ABC (denoted as Ia_ABC$_{ACH}$) which contains the hash values of transactional items and values of the financial transaction. These will be cross-verified at a later time. The Ia_ABC$_{ACH}$ is signed by the ACH_SP_Agent.
- 5. The ACH_SP_Agent sends the Ia_ABC$_{ACH}$ to a trusted party in Steve's PC and to a trusted party in Bob's PC. ACH_SP_Agent sends Ia_ABC$_{ACH}$ to ACH_User_Agent$_{Bob}$ and to FE_User Agent$_{Steve}$.
- 6. On reception, ACH_User_Agent$_{Bob}$ and FE_User_Agent$_{Steve}$ will acknowledge each received Intra-system ABC by signing it with its respective private keys. This will be done after checking that the signature of the Ia_ABC$_{ACH}$ is authentic.
- 7. ACH_User_Agent$_{Bob}$ compares Ia_ABC$_{FE}$ with Ia_ABC$_{ACH}$. If Ia_ABC$_{ACH}$ is proved to contain the data of a settlement for a financial transaction contained in Ia_ABC$_{FE}$, then they belong to the same transaction set. If they belong to the same transaction set then ACH_User_Agent$_{Bob}$ will create an Inter-system ABC. The new Inter-system ABC contains data that links Ia_ABC$_{FE}$ and Ia_ABC$_{ACH}$. The new Inter-system ABC is denoted as Ie_ABC$_{(FE-ACH)_{BankA}}$.
- 8.ACH_User_Agent$_{Bob}$ sends Ie_ABC$_{(FE-ACH)_{BankA}}$ to CSD_User_Agent$_{Cat}$ together with the two Intra-system ABCs (i.e. Ia_ABC$_{FE}$ and Ia_ABC$_{ACH}$). The sent items are referred as $BT1_{BankA}$.
- 7.Similarly, when two intra-system ABCs arrive ACH_User_Agent$_{Steve}$ compares Ia_ABC$_{FE}$ with Ia_ABC$_{ACH}$. If they belong to the same transaction set then ACH_User_Agent$_{Steve}$ creates an Inter-system ABC. They belong to the same transaction set if they are directly or indirectly related between them. The new Inter-system ABC contains data that binds itself to Ia_ABC$_{FE}$ and to Ia_ABC$_{ACH}$. The new Inter-system ABC is denoted as Ie_ABC$_{(FE-ACH)_{BankB}}$.
- 8. ACH_User_Agent$_{Steve}$ sends Ie_ABC$_{(FE-ACH)_{BankB}}$ to CSD_User_Agent$_{Sue}$ together with the two Intra-system ABCs (i.e. with Ia_ABC$_{FE}$ and Ia_ABC$_{ACH}$). The sent items are referred to as $BT1_{BankB}$.
- 9. Sue logs onto the financial service CSD. Then, she performs an inter-banking transfer of the market instrument to the new owner, i.e. Bank A.
- 10. The CSD_SP_Agent detects the transactional activity. It retrieves and saves the transactional data into its database. Then, the CSD_SP_Agent generates an Intra-system ABC (denoted as Ia_ABC$_{CSD}$), which contains the hash values of transactional items and values of the financial transaction. These will be cross-verified at a later time. The Ia_ABC$_{CSD}$ is signed by the CSD_SP_Agent.
- 11. The CSD_SP_Agent sends the Ia_ABC$_{CSD}$ to a trusted party in Cat's PC and to a trusted party in Sue's PC. CSD_SP_Agent sends Ia_ABC$_{CSD}$ to CSD_User_Agent$_{Cat}$ and to CSD_User _Agent$_{Sue}$.
- 12. On reception CSD_User_Agent$_{Cat}$ and CSD_User_Agent$_{Sue}$ will acknowledge each received Intra-system ABC. This will be done after checking that the signature on Ia_ABC$_{CSD}$ is authentic.
- 13. CSD_User_Agent$_{Cat}$ compares Ia_ABC$_{ACH}$ in $BT1_{BankA}$ with Ia_ABC$_{CSD}$. If they belong to the same transaction set then CSD_User_Agent$_{Cat}$ creates an Inter-system ABC. This Inter-system ABC contains data that binds the ABC to $BT1_{BankA}$ and to Ia_ABC$_{CSD}$. The new Inter-system ABC is denoted as Ie_ABC$_{(FE-ACH-CSD)_{BankA}}$.
- 13. CSD_User_Agent$_{Sue}$ compares Ia_ABC$_{ACH}$ in $BT1_{BankB}$ with Ia_ABC$_{CSD}$. If they belong to the same transaction set then CSD_User_Agent$_{Sue}$ creates an Inter-system ABC. The new Inter-system ABC contains data that binds the ABC to $BT1_{BankB}$ and to Ia_ABC$_{CSD}$. The new Inter-system ABC is denoted as Ie_ABC$_{(FE-ACH-CSD)_{BankB}}$.

From the above description it can be seen that an Ia_ABC is created for each settled financial transaction.

An Intra-system ABC (Ia_ABC$_q$) data structure can be represented as shown in equation 9

$$Ia\_ABC_q = X||E_{K_{r_T}}(X) \tag{9}$$

$$X = h(X1)||h(X2) \tag{10}$$

$$X1 = q||t_q||h(l\_b_{q-1}) \tag{11}$$

$$X2 = nv||rv||mii||bi||si \tag{12}$$

In these equations, (h) stands for a hash function, (q) stands for a sequential consecutive number given to each issued Intra-system ABC. Each sequential number contains an identifier associated to the financial service where it was created, (t_q) is the current date and time, nv is the nominal value of a market instrument, rv is the real value of a market instrument, mii is the market instrument identifier, bi is the buyer's identification, and si is the seller's identification. The $(l\_b_{(q)})$ represents the linking bits to the previously issued Intra-system ABCs. The linking bits are calculated by a recursive equation that contains data from the previously issued Intra-system ABC. See the recursive equation in 13.

$$l\_b_{(q)} = ((nv||rv||mii||bi||si)_{q-1}||h(l\_b_{q-1})) \tag{13}$$

The linking bits integrity protect the data in the audit trail. This means that if an Intra-system ABC is changed then the data in the changed Intra-system ABC will no longer be linked to the Intra-system ABC issued previous to its creation. Thus, changes will be detected by checking data integrity.

An Inter-system ABC is created at the banks side on each PC using an aggregation sheme. An inter-system ABC is the

hash value of its two children. The two children can be either two Intra-system ABCs or one Intra-system ABC and one Inter-system ABC.

An Inter-system ABC contains linking data to two Intra-system ABCs or to one Intra-system ABC and one Inter-System ABC.

The resulting data structure, after all the settlements are completed, should contain a complete transaction set, namely the complete ABC.

The Inter-system ABCs are constructed using the Merkle Hash Tree aggregation linking scheme. Thus, each Inter-system ABCs contains the hash value of its two children, i.e. two Intra-system ABCs or an Intra-system ABC and an Inter-system ABCs. The Ie_ABC of an FE Intra-system ABC and an ACH Intra-system ABC is the hash value that results from the signature of an FE Intra-system ABC (noted as Ia_ABC$_{ACH}$_S) and the signature from an ACH Intra-system ABC (noted as Ia_ABC$_{ACH}$_S). Equation 9 shows that the Inter-system ABC is the hash value of the signatures of two Intra-system ABCs; one is from the FE transaction and the other is from the ACH transaction.

$$A = Ia\_ABC_{FE_S} \tag{14}$$

$$B = Ia\_ABC_{ACH_S} \tag{15}$$

$$Ie\_ABC_{(FE-ACH)_q} = h(A||B) \tag{16}$$

where,

$$D = h(l\_b_{(q-1)})_{FE} \tag{17}$$

$$E = h(nv||rv||mii||bi||si)_{FE} \tag{18}$$

$$Ia\_ABC_{FE_S} = E_{K_{r_{FE\_SP\_Agent}}} h(q||dt||D||E) \tag{19}$$

$$F = h(l\_b_{(q-1)})) \tag{20}$$

$$G = h(nv||rv||mii||bi||si) \tag{21}$$

$$Z = h(q||dt||F||G) \tag{22}$$

$$Ia\_ABC_{ACH}\_S = E_{K_{r_{ACH\_SP\_Agent}}}(Z_{ACH} \tag{23}$$

The Inter-system ABC (Ie_ABC$_{FE-ACH-CSD}$) resulting from one Ie_ABC$_{FE-ACH}$ and Ia_ABC_{CSD} can be represented using equation 26

$$v = Ia\_ABC_{CSD}\_S \tag{24}$$

$$w = Ie\_ABC_{FE-ACH} \tag{25}$$

$$Ie\_ABC_{FE-ACH-CSD} = h(w||v) \tag{26}$$

where,

$$B = h(l\_b_{(q-1)}) \tag{27}$$

$$C = h(nv||rv||mii||bi||si)_{CSD} \tag{28}$$

$$Ia\_ABC_{CSD}\_S = E_{KrCSDSPAgent}(q||dt||B||C) \tag{29}$$

In these equations, (q) stands for a sequential consecutive number given to each Intra-system ABC issued, (dt) is the current date and time, $(l\_b_{(q)})$ are the linking bits, nv is the nominal value of a market instrument, rv is the real value of a market instrument, mii is the market instrument identifier, bi is the buyer's identification, and si is the seller's identification. The linking bits are calculated by a recursive equation that contains data from the previously issued Intra-system ABCs. That is,

$$l\_b_{(q)} = (nv||rv||mii||bi||si)_{q-1}||h(l\_b_{q-1}) \tag{30}$$

Inter-system ABCs are constructed only if an Intra-system ABC is proved to belong to a transaction set, i.e. triggered by one same deal. This authentication process is called Transaction Authentication Service, TAS.

Acknowledgements: The research was supported by the University of Manchester. In the case of the first author, she wants to inform that the opinions expressed here are personal and do not necessarily represent those of the Banco de la Repblica or those of its Board of Directors.

V. Conclusion

We have designed and presented the Workflow Oriented Auditing Architecture. The proposed workflow is more dynamic and complete, since it contemplates any two parties interacting on the basis of financial transactions recorded by their users in related but distinct automated financial systems. In the new definition these interactions can be described in one unique audit trail. This concept expands the current ideas of audit trails by adapting them to actual e-trading workflow activity.

In this audit workflow, external tasks cannot be isolated from directly related internal tasks of financial institutions, which is important since isolating the external tasks may lead to an inability to monitor the settlement of e-trading

deals. Based upon this finding a workflow oriented security service was designed. This security service can detect multiple financial transactions that belong to one single transaction set, that is, a group of transactions that are triggered by an initial transaction. This was based on the observation that all authorised transactions were either triggered by or triggered other transactions in the workflow.

This new security service uses data items that remain constant for the complete settlement of the e-trading deal. These items are used to identify which financial transactions belong to the same transaction set. The new framework led us to design a workflow oriented audit trail, namely the Automated Banking Certificate (ABCs).

An ABC is a workflow-oriented integrity protected data structure. It provides an intra-system, inter-system, inter-organizational and intra-organizational audit trail. The data structure of the ABC is designed to record a complete transaction set in each complete ABC. Two linking schemes are used in the design. The resulting structure achieves an interleaving property that can be used to trace an e-trading workflow activity. A new dynamic auditing solution was designed.

REFERENCES

[1] Ralph Merkle, A Certified Digital Signature, Advances in Cryptology CRYPTO 89 , Lecture Notes in Computer Science volume 435 pages 218-238, Springer-Verlag, 1988.

[2] Stuart Haber and Scott Stornetta, How to Time-Stamp a Digital Document, Advances in Cryptology – CRYPTO 90, Lecture Notes in Computer Science , Springer –Verlag, Berlin–Heidelberg, 1991, V537 pp. 437–455.

[3] Dave Bayer , Stuart Haber and Wake Scott Stornetta, Sequences II - Improving the Efficiency and Reliability of Digital Time-Stamping, ISBN 978-1-4613-9325-2 (print version) 978-1-4613-9323-8 (online version), Springer–Verlag, Berlin–Heidelberg, 1993, pp. 329–334.

[4] Josh Benaloh and Michael de Mare, One-Way Accumulators : A Decentralized Alternative to Digital Signatures, Advances in Cryptology EUROCRYPT 93, Lecture Notes in Computer Science , Springer-Verlag, 1994, V 795 pp. 274–285 .

[5] A Mounji, B Le Charlier and D Zampunieris, Distributed Audit trail Analysis, Proceedings of the Symposium on Network and Distributed System Security, IEEE, 1995, ISBN 0-8186-7027-4, pp. 102–112.

[6] Lisa Meulbroek and Carolyn Hart, The Effect of Illegal Insider Trading on Takeover Premia, European Finance Review , 1997,V.1 pp. 51–80.

[7] Stuart Haber and W Scott Stornetta, Secure Names for Bit-Strings, Proceedings of the 4th ACM Conference on Computer and Communications Security - CCS '97, Communications of the ACM , ISBN 0-89791-912-2, 1997,pp. 28–35.

[8] Paul C Kocher, On Certificate Revocation and Validation, Proceedings of the Second International Conference on Financial Cryptography - FC '98, Lecture Notes in Computer Science , Springer-Verlag, isbn 3-540-64951-4, 1998, pp. 172–177.

[9] Ahto Buldas, Peter Laud, Helger Lipmaa and Jan Willemson, Time-Stamping with Binary Linking Schemes, Proceedings of the 18th Annual International Cryptology Conference on Advances in Cryptology - CRYPTO '98, Lecture Notes in Computer Science pages , Springer-Verlag, isbn 3-540-64892-5, 1998, V.1462 pp. 486–501.

[10] Moni Naor and Kobbi Nissim, Certificate Revocation and Certificate Update, IEEE Journal on selected areas in Communications, 2000, V.18 I.4 pp. 561–570.

[11] Laurence H Meyer, Speech, Mr Meyer looks at the prospects for strengthening risk management derivatives, BIS Review 17, 2000, found at $http://www.bis.org/review/r000228c.pdf$ on February 2015.

[12] Helen Allen, John Hawkins and Setsuya Sato, Electronic trading and its implications for financial systems, BIS papers No. 7, 2001, found at $http://www.bis.org/publ/bppdf/bispap07d.pdf$ on February 2015.

[13] Brian F Cooper and Hector Garcia-Molina, Peer-to-Peer Data Trading to Preserve Information, ACM Transactions Communications on Information Systems Pages , April 2002 V.20 N.2 pp. 133–170.

[14] Manish Agrawal, Hermant Padmanabhan, Lokesh Pandey, H. R. Rao,Shambhu Upadhyaya , A Conceptual Approach to Information Security in Financial Account Aggregation, 6th International Conference on Electronic Commerce- ICEC '04, ACM ISBN 1-58113-930-6, pp. 619–626.

[15] Alexis Bonnecaze, A Distributed Time Stamping Scheme, Proceedings of the Conference on Signal Image Tech - SITIS 05, 2005.

[16] Ian Molloy, Chen Cheng, Pankaj Rohatgi, Trading in Risk: Using Markets to Improve Access Control, Proceedings of the Fifteenth New Security Paradigms Workshop -NSPW '08, 2008, found at $http://www.nspw.org/papers/2008/nspw2008-molloy.pdf$ on February 2015.

[17] International Organization for Standarization and International Electrotechnical Comission ISO/IEC - Standard 18014-3, Information Technology-Security Techniques - Time-Stamping Services, ISO, 2009.

[18] J Lim, Ng Kah Hwa, Computer Fraud And Ethics : The Société Générales Trading Fraud, International Conference on Computer and Management - CAMAN 2011, IEEE, ISBN 978-1-4244-9282-4, 2011, pp. 1–3.

[19] James B. Stewart and Peter Eavis, Revisiting the Lehman Brother Bailout That Never Was, The New York Times, September 29 2014.

[20] Jill Treanor Foreign exchange fines: banks handed £2bn in penalties for market Rigging, The Guardian, Wednesday 12 November 2014.

Urban growth and LULC change from 1975 to 2015 through RS/GIS in Samara, Russia

M.S. Boori, A. Kupriyanov, V.A. Soifer, and K. Choudhary

***Abstract*—** The main focus of this study is to know the changes in urban accumulation, population, land use and its correlation with the population, migration and urbanization led problems related with water and environmental degradation. This study illustrates the spatio-temporal dynamics of urban growth and land use changes in Samara city, Russia from 1975 to 2015. Landsat satellite imageries of five different time periods from 1975 to 2015 were used to know the changes. Supervised classification methodology has been employed using maximum likelihood technique in ArcGIS 10.1 Software. By applying classification methods to the satellite images four main types of land use were extracted: water, built-up, forest and grassland. Then, the area coverage for all the land use types at different points in time were measured and coupled with population data. The results demonstrate that, over the entire study period, population was increased from 1146 thousand people to 1244 thousand from 1975 to 1990 but later on first reduce and then increase again, now 1173 thousand population. Built-up area is also change according to population. The present study revealed an increase in built-up by 37.01% from 1975 to 1995, than reduce - 88.83% till 2005 and an increase by 39.16% from 2005 to 2015, along with the increase in population, migration from rural areas owing to the economic growth and technological advantages associated with urbanization.

***Keywords*—** Urban growth, land use/cover; remote sensing; change detection analysis and GIS

I. Introduction

The official foundation date of Samara is 1586. That time small fortress was built at the confluence of the Volga and Samara rivers. It was protecting the eastern borders of the Russian state from nomads. After building the quay, Samara settlement became the economic and diplomatic center of Russia. In 1780, the town became the capital of Simbirsk region. The economy of Samara was growing quickly at the end of the 19th and beginning of the 20th centuries (bread trading and milling business). The population of Samara at the beginning of the 20th century was about 100,000. It was large trade and industrial center of the Volga region of Russia [1]. During the World War II, it was chosen to be the USSR capital in case of Moscow fall. Here defense industry was developing fast after the World War II. Soon the city became so called "closed city" of the USSR. The spaceship of Yury Gagarin (first man in space) "Vostok" was built here. Now Samara is Russian large industrial and cultural center with multinational population and dramatic history. Samara is a large industrial center of the whole Volga river region. The city is among top Ten Russian Cities by industry volume. There are over 150 large and medium industrial plants in the city. About 25% of all bearings and 70% of all cables produced in Russia are made in Samara. It is producing various outer space vehicles and machinery, aircraft, power stations, refinery, cranes. Samara food industry is known for its chocolate, vodka "Rodnik" and "Zhiguli" beer. Samara is one of the largest transportation junctures in Russia; it is crossed by the shortest ways from central and Western Europe to Siberia, Middle Asia and Kazakhstan [1].

Urban sprawl is defined as an inefficient urban development often linked to sparse building density over rural areas [2, 3]. Only 3 present earth surface covered by urban area [4, 5] but due to urbanization, population growth, economic development and unplanned development are the main cause of environmental and social problems in modern cities. Urban areas are faced with distinctive, or 'systemic', issues arising from their unique social, environmental and economic characteristics [6]. Some glitches such as health risks including air pollution, occupational hazards, traffic injury, risks caused by dietary and social changes [7] as well as destruction of vegetation, agricultural lands, population of underground and surface water sources and climate change [8] are associated with urban expansion. These parameters are decreasing the quality of life in urban and rural societies. In developing cities, information about unplanned settlements is often unavailable. It is critically important to properly characterize urban expansion before developing a comprehensive understanding of urbanization processes [9, 10). The unplanned and uncontrolled rapid growth has resulted in serious negative effects on the urban dwellers and their environment [11]. As all over the globe cities are growing very quickly so it is necessary to protect natural resources with urban growth [12]. More than ever, it is imperative that urban planning focus on evidentiary models and valid spatial data.

Earlier studies show that urbanization happens because people move into urban areas to seek economic opportunities and to improve their standard of living. People in rural area have to depend on changeable environmental conditions and

This work is financially supported by the Russian Scientific Foundation (RSF), grant no. 14-31-00014 "Establishment of a Laboratory of Advanced Technology for Earth Remote Sensing".

M. S. Boori is with the Samara State Aerospace University, 34 Moskovskoye shosse, Samara - 443086, Russia (corresponding author to provide phone: 9874329875; e-mail: msboori@gmail.com).

A. Kupriyanov is with the Samara State Aerospace University, 34 Moskovskoye shosse, Samara - 443086, Russia (e-mail: akupr@smr.ru).

V. A. Soifer is with the Samara State Aerospace University, 34 Moskovskoye shosse, Samara - 443086, Russia (e-mail: soifer@ssau.ru).

in times of drought, flood or pestilence, survival becomes extremely problematic. This is very different in urban where all the facilities are well build to make human life more comfortable and the main attraction of urban is easy access to wealth [13, 14]. Usually land uses and urban growth in remote sensing involves the analysis of two registered, aerial or satellite multi- spectral bands from the same geographical area obtained at two different times. Such an analysis aims at identifying changes that have occurred in the same geographical area between the two times considered [15]. Satellite remote sensing is a potentially powerful means of monitoring land-use change at high temporal resolution and lower costs than those associated with the use of traditional methods [16]. Remote sensing data is very useful because of its synoptic view, repetitive coverage and real time data acquisition [17]. The digital data in the form of satellite imageries, therefore, enable to accurately compute various land cover/land use categories and help in maintaining the spatial data infrastructure which is very essential for monitoring urban expansion and land use studies [17. 18]. Land use/cover changes is a widespread and accelerating process, mainly driven by natural phenomena and anthropogenic activities, which in turn drive changes that would impact natural ecosystem [19, 20]. Understanding landscape patterns, changes and interactions between human activities and natural phenomenon are essential for proper land management and decision improvement.

To know the spatial patterns of Samara city urban growth over in a timeframe, city must be systematically mapped, monitored, and accurately assessed using satellite images with conventional ground truth verification data. This type of analysis work provides a scenario of where growth is occurring and helps to identify the environmental and natural resources threatened by such development and suggest the likely future directions and patterns of growth. The current study has three specific objectives: (1) investigate the growth pattern of Samara city during 1975 – 2015 by using remote sensing and GIS; (2) analyze the temporal and spatial characteristics of urban expansion in Samara from 1975 to 2015 and (3) detect and evaluate the land use and land cover change due to urbanization between 1975 to 2015; (4) analyze the main factors governing urbanization and land use and land cover change; (5) evaluate current local environmental and natural resource protection and development policies.

II. Study Area

Samara region is situated in the South-East of the Eastern European Plain in the middle flow of the greatest European river, the Volga, which separates the region in two parts of different size, Privolzhye and Zavolzhye. Study area (fig. 1.) Samara known from 1935 to 1991 as Kuybyshev, is the sixth largest city in Russia and the administrative center of Samara Oblast. Geographical coordinates are 53°12´10´´N, 50°08´27´´E (fig. 1). The region occupies an area of 53.6 square kilometers (0.31% of the territory of Russia) and forms a part of the Volga Federal District. It is situated in its southern part. The Volga acts as the city's western boundary; across the river are the Zhiguli Mountains, after which the local beer (*Zhigulyovskoye*) is named. The northern boundary is formed by the Sokolyi Hills and by the steppes in the south and east. The region stretches form 335 km from the North to the South and for 315 km from the West to the East. The land within the city boundaries covers 46,597 hectares (115,140 acres). Population: 1,164,685 (2010 Census); 1,157,880 (2002 census); 1,254,460 (1989 Census). The metropolitan area of Samara-Tolyatti-Syzran within Samara Oblast contains a population of over three million. Formerly a closed city, Samara is now a large and important social, political, economic, industrial, and cultural center in European Russia. It has a continental climate characterized by hot summers and cold winters. In this research work we use 25km^2 radius from the city center of Samara.

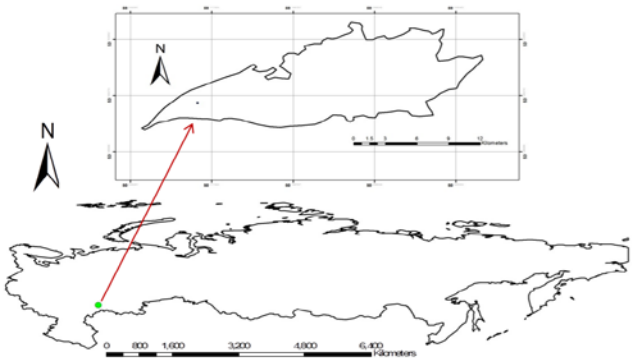


Fig. 1 The position of Samara in Mainland Russia.

III. Materal and Methods

A. Data

Landsat-TM images represent valuable and continuous records of the earth's surface during the last 4 decades (USGS, 2014). Moreover, the entire Landsat archive is now available free-of-charge to the scientific public, which represents a wealth of information for identifying and monitoring changes in manmade and physical environments [21, 16]. Several studies acknowledged the importance of pre-processing (i.e., data selection, co-registration, radiometric calibration and normalization) in performing accurate and reliable change detection analysis [16]. A selection of multi-sensor, multi-resolution, and multi-temporal images was used for this study [22, 18]. The specific satellite images used were Landsat MSS (Multi-Spectral Scanner) for 1975, Landsat TM (Thematic Mapper) for 1985-1995, Landsat ETM+ (Enhanced Thematic Mapper plus) for 2005 and 2015, an image captured by a different type of sensor. According to [23, 24], the time interval between images for the investigation of Land Use/Cover changes Levels I and II [25] should be between 5 and 10 years and the spatial resolution should be 10m or larger, so that the selected images and sensors comply with these criteria. Another reason for selecting these images was their availability and cloud cover. All satellite and supporting data used for this study are identified in Table 1.

TABLE I. Data used in this study.

Data	**Pass & Row**	**Year**	**Spatial resolution (m)**

Satellite images			
Landsat MSS	183/023	June 1975	79
Landsat TM	169/023	Sept. 1985	30
Landsat TM	169/023	May 1995	30
Landsat ETM+	169/023	Aug. 2005	15-30
Landsat ETM+	169/023	March 2015	15-30
Supporting data			
Topographic map		2000	1:25000
Field data/GPS		2015	10

B. Image Preprocessing

Digital image processing was manipulated by the ArcGIS software. The scenes were selected to be geometrically corrected, calibrated and removed from their dropouts. These data were stratified into 'zones', where land cover types within a zone have similar spectral properties. Other image enhancement techniques like histogram equalization are also performed on each image for improving the quality of the image. Some additional supporting data were used in this study. Digital topographical maps, 1:50,000 scale, were used for image georeferencing for the land use/cover map and for increased accuracy of the overall assessment. The images obtained as standard products were geometrically and radiometrically corrected but, because of the different standards and references used by the various image- supplying agencies, all images were georeferenced again at the pre-processing stage. At this stage, 20 points were selected as GCPs (Ground Control Point) for all images. Data sources used for the GCP selection were: digital topographic maps, GPS (Global Positioning System) acquisitions. Then, all five images were geometrically corrected up to orthorectified level. The data of ground truth were adapted for each single classifier produced by its spectral signatures for producing series of classification maps. Using ArcMap, we made a composite raster data of TM and ETM+ using Arctoolbox data management tools. Landsat images are composed of eight different bands, each representing a different portion of the electromagnetic spectrum. By combining all these bands, composite raster data are obtained. Table 2 shows all bands of MSS, TM and ETM+, which was used for band combination.

TABLE II. Band width of used data in this study.

Enhanced Thematic Mapper Plus (ETM+)	**Thematic Mapper (TM)**	**Multispectral Scanner (MSS)**
Band 1 Visible (0.45 - 0.52 µm) 30 m	Band 1 Visible (0.45 - 0.52 µm) 30 m	Band 4 Visible green (0.5 to 0.6 µm)
Band 2 Visible (0.52 - 0.60 µm) 30 m	Band 2 Visible (0.52 - 0.60 µm) 30 m	Band 5 Visible red (0.6 to 0.7 µm)
Band 3 Visible (0.63 - 0.69 µm) 30 m	Band 3 Visible (0.63 - 0.69 µm) 30 m	Band 6 Near-Infrared (0.7 to 0.8 µm)
Band 4 Near-Infrared (0.77 - 0.90 µm) 30 m	Band 4 Near-Infrared (0.76 - 0.90 µm) 30 m	Band 7 Near-Infrared (0.8 to 1.1 µm)
Band 5 Near-Infrared (1.55 - 1.75 µm) 30 m	Band 5 Near-Infrared (1.55 - 1.75 µm) 30 m	
Band 6 Thermal (10.40 - 12.50 µm) 60 m Low Gain / High Gain	Band 6 Thermal (10.40 - 12.50 µm) 120 m	
Band 7 Mid-Infrared (2.08 - 2.35 µm) 30 m	Band 7 Mid-Infrared (2.08 - 2.35 µm) 30 m	
Band 8 Panchromatic (PAN) (0.52 - 0.90 µm) 15 m		

C. Classification of Images

After preprocessing, first use unsupervised classification and get maximum possible classes on the basis of grave levels. Then used supervised classification method with maximum likelihood algorithm in ArcGIS 10.1 Software. Maximum likelihood algorithm (MLC) is one of the most popular supervised classification methods used with remote sensing image data. This method is based on the probability that a pixel belongs to a particular class. The basic theory assumes that these probabilities are equal for all classes and that the input bands have normal distributions. However, this method needs long time of computation, relies heavily on a normal distribution of the data in each input band and tends to over-classify signatures with relatively large values in the covariance matrix. It requires the least computational time among other supervised methods, however, the pixels that should not be unclassified become classified and it does not consider class variability. Ground verification was done for doubtful areas. Based on the ground trothing, the misclassified areas were corrected using recode option in ArcGIS. The error matrix and Kappa methods were used to assess the mapping accuracy. Four land use/cover types are identified in the study area viz., (i) Forest (ii) Grassland (iii) Built-up (iv) Water body (table 3).

TABLE III. Description of Land Use/Cover classes.

Land use class	**Description**
Built-up	Residential, commercial & services, industrial, transportation & roads, mixed

Forest	Pine, coniferous trees, citrus orchards,
Grassland	Grass belt, agriculture, parks, trees, brain land
Water bodies	River, permanent open water, lakes, ponds

D. Land use/cover change detection and analysis

For performing land use/cover change detection, a post-classification detection method was employed. A pixel-based comparison was used to produce change information on pixel basis and thus, interpret the changes more efficiently taking the advantage of ‘‘-from, -to’’ information. Classified image pairs of two different decade data were compared using cross-tabulation in order to determine qualitative and quantitative aspects of the changes for the period of 1975 to 2015. A change matrix [26, 27] was produced with the help of ArcGIS software. Quantitative areal data of the overall land use/cover changes as well as gains and losses in each category between 1975 and 2015 were then compiled.

Observations of the Earth from space provide objective information of human activities and utilization of the landscape. The classified images provide all the information to understand the land use and land cover of the study area. Change detection analyses describe and quantify differences between images of the same scene at different times. The classified images of the five dates can be used to calculate the area of different land covers and observe the changes that are taking place in the span of data. This analysis is very much helpful to identify various changes occurring in different classes of land use like increase in urban built-up area or decrease in vegetation land and so on [28].

E. Annual urban growth rate

We use following formula to know the intensity of urban expansion called annual urban growth rate (AGR):

$$\mathbf{AGR} = \frac{UA_{n+i} - UA_i}{nTA_{n+i}} \mathbf{X\ 100\%}$$

where TA$n+i$ is the total land area of the target unit to be calculated at the time point of $i+n$; UA$n+i$ and UAi the urban area or built-up area in the target unit at time $i+n$ and i, respectively and n is the interval of the calculating period (in years). Generally, the target calculating unit is set to the administrative district so as to link with administration or economic statistics. In this research, we preferred the geographical gridding unit since the administrative borders have been changed so frequently in this city. The maps were therefore gridded as 1 km×1 km units and the annual urban growth rates of each unit were then calculated. Lastly the grid-based annual urban growth rates were clustered by using natural break method and mapped to evaluate the spatial features of the ‘expansion’.

IV. Results and Discussion

This work provides a methodological framework by integrating RS-GIS, metric analysis and spatial analysis to facilitate the assessment of urbanization or urban growth and changing land use patterns. Remote sensing and GIS helped monitor urbanization process and assess the status of urban agglomeration. The temporal changes facilitate the investigation and characterization of impacts on land use/cover and surrounding environment from settlement sprawl associated with accelerating urbanization. In this study, time series data used are Landsat MSS, TM and ETM+ from 1975 to 2015. First unsupervised and later on supervised classification is done on satellite image series to analyze morphological growth. By comparing the area in square kilometer, the percentage increase in urban growth can be measured. Final maps produced are shown in Fig. 2–3. During this study, it was found that there is an increase in settlement by 6.48% (127.37 km^2) from 1975 to 2015.

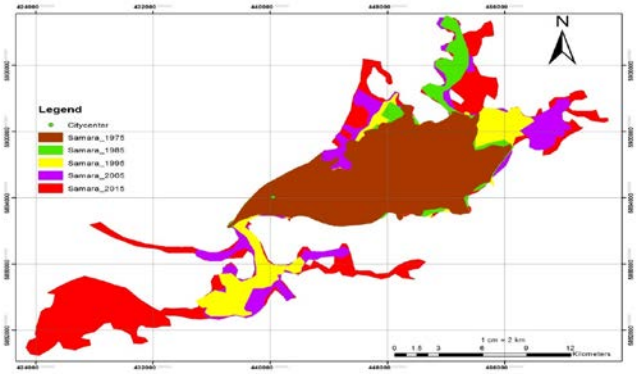
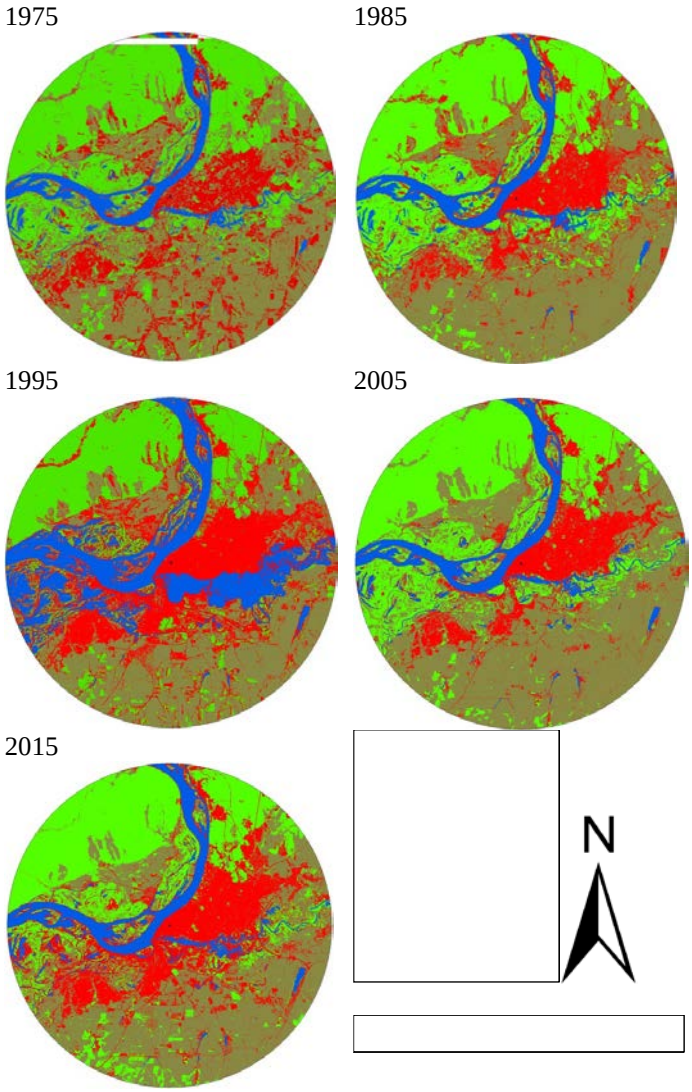


Fig. 2 Samara city growth in 1975, 1985, 1995, 2005 and 2015 map.

A. Land use/land cover images

The results obtained through the analysis of multi-temporal satellite imageries were diagrammatically illustrated in Figs. 2–3 and data are registered in Tables 4.and Fig. 2 depicts total city growth status in different years. Fig. 3 depicts land use/cover change in different land use categories. Table 4 shows the land use for different purposes in Samara. This gives an idea to the planners about urban sprawl in Samara, a greater perception of problems, the available options to rectify and develop a better plan. In future analysis, a highly detailed structural analysis of the large-scale and heterogeneous inner structures of urban morphology using satellite data with higher geometric resolution (e.g., Ikonos or Quickbird) is expected to augment information for planning purposes [29]. Digital analysis techniques can be used for identification and classification of all land cover classes from other classes in an efficient manner. If large area is to be estimated, it is more effective and accurate by this technique. A brief account of these results is discussed in the following paragraphs.

ISBN: 978-1-61804-318-4



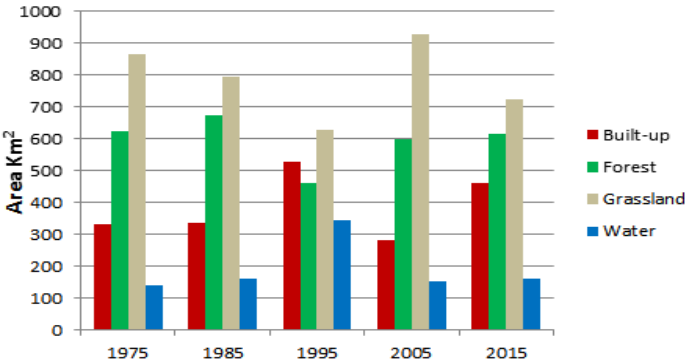


Fig. 3 The land use and land cover map of Samara in 1975, 1985, 1995, 2005 and 2015.

B. *Land use/land cover Status*

Accuracy assessment of the land use/cover classification results obtained showed an overall accuracy is more than 90% for all images. These data reveal that in 2015, about 31.38% (616.14 km^2) area of Samara block was under forest, 36.85% (723.49 km^2) under grassland, 8.30% (162.91 km^2) under water body and 460.87% (23.47km^2) under built-up land. During 1975 the area under these land categories was found about 31.81% (624.56 km^2) under forest, 44.03% (864.50 km^2) under grassland, 7.17% (140.85 km^2) under water body and 16.99% (333.50 km^2) under built-up land (table 4). First unban area was increase till 1995 then reduce but later on again increase due to increased population. Initially forest area was decreased and later on increase due to governmental protection. Grassland class cover highest area in the study area, it was increase and highest in 2005 but in last radiuses. Water class is stable with small variation (fig. 3).

TABLE IV. Land use/cover area in kilometer square

Class	1975	%	1985	%	1995	%	2005	%	2015	%
Built-up	333.50	16.99	336.48	17.14	529.48	26.97	280.40	14.28	460.87	23.47
Forest	624.56	31.81	673.94	34.32	462.69	23.57	600.57	30.59	616.14	31.38
Grassland	864.50	44.03	792.86	40.38	628.54	32.01	928.97	47.31	723.49	36.85
Water	140.85	7.17	160.13	8.16	342.70	17.45	153.47	7.82	162.91	8.30
Total	1963.41	100.00	1963.41	100.00	1963.41	100.00	1963.41	100.00	1963.41	100.00

C. *Land use/land cover Change*

Table 5 shows land use land cover change matrix from 1975 to 2015. Data registered in Table 5 and Figs. 4 reveal that both positive and negative changes occurred in the land use/cover pattern of the Samara block. During the last four decades the grassland in the study area has decrease from 864.50 km^2 in 1975 to 723.49 km^2 in 2015 which accounts for -19.49% of the total study area. The forest has slightly decreased from 624.56 km^2 in 1975 to 616.14 km^2 in 2015 which accounts for -1.36%. The built-up area has increased from 333.50 km^2 in 1975 to 460.87 km^2 in 2015 which accounts for 27.63%. The water body has been increased from 140.85 km^2 in 1975 to 162.91 km^2 in 2015. This increase in water body accounts for 13.54%. To understand land encroachment for different land categories during the last four decades, a change detection matrix (table 5) was prepared which reveals that:

Cross tabulation is a means to determine quantities of conversions from a particular land cover to another land cover category at a later date. The change matrices based on post classification comparison were obtained and are shown in tables 5 and fig 4. Built up area covered 333.5 km^2 in 1975 and 336.59 km^2 in 1985, while the grassland covered an area of 792km^2 in 1985 and 629.68 km^2 in 1995. 383.83km^2 of the forest area which was forest in 1995 was still forest cover in 2005. From 2005 to 2015 149.10 km^2 grassland and 60.50km^2 forest convert in built-up. During the same period, 115.29km^2 grassland had been converted to forest area (table 5).

Table 5. Land use/cover change matrix from 1975 to 2015.

ISBN: 978-1-61804-318-4

2005-2015

CLASS	BUILT_UP	FOREST	GRASSLANI	WATER	Total
Built-up	245.14	5.00	18.41	10.32	278.87
Forest	60.50	496.84	42.42	0.32	600.08
Grassland	149.10	115.29	662.95	4.16	931.49
Water	5.56	0.40	0.00	146.78	152.74
Total	460.30	617.53	723.77	161.58	1963.19
Change rate %	39.41	2.82	-28.69	5.47	

1995-2005

CLASS	BUILT_UP	FOREST	GRASSLANI	WATER	Total
Built-up	216.68	88.07	224.33	1.16	530.25
Forest	4.36	383.83	74.63	0.04	462.86
Grassland	26.17	51.46	551.93	0.20	629.76
Water	31.49	76.83	80.71	151.34	340.37
Total	278.71	600.19	931.60	152.74	1963.24
Change rate %	-90.25	22.88	32.40	-122.84	

1985-1995

CLASS	BUILT_UP	FOREST	GRASSLANI	WATER	Total
Built-up	227.05	15.53	52.54	42.26	337.37
Forest	114.36	385.95	62.90	111.00	674.22
Grassland	187.79	61.34	514.24	29.13	792.51
Water	1.04	0.04	0.00	157.98	159.06
Total	530.25	462.86	629.68	340.37	1963.16
Change rate %	36.37	-45.66	-25.85	53.26	

1975-1985

CLASS	BUILT_UP	FOREST	GRASSLANI	WATER	Total
Built-up	151.21	20.51	157.05	4.28	333.05
Forest	31.93	526.99	60.97	4.66	624.55
Grassland	145.54	120.89	573.31	24.87	864.60
Water	7.92	4.89	0.92	127.14	140.87
Total	336.59	673.28	792.25	160.95	1963.07
Change rate %	1.05	7.23	-9.13	12.47	

Figure 4 show that in 1975 to 1985, there are not any big changes. From 1985 to 1995 there is an around 40% change, forest and grassland have negative change but built-up area was increase around 36.37%. In the year of 1995 to 2005, there is dramatic negative change in built-up area. It`s show migration of population in another places. But during this period of time forest and grassland have positive change, which show less human interferation in the area. Final in 2005 to 2015, built-up area again increase (39.41%) and that`s why grassland was reduce (-28.69%).

CLASS	BUILT_UP	FOREST	GRASSLANI	WATER
Changes 75-85	1.05	7.23	-9.13	12.47
Changes 85-95	36.37	-45.66	-25.85	53.26
Changes 95-05	-90.25	22.88	32.40	-122.84
Changes 05-15	39.41	2.82	-28.69	5.47

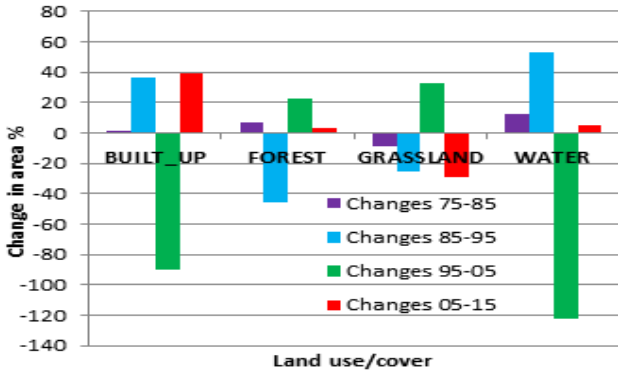


Fig. 4 Diagrammatic illustration of land use/cover change in percent during the last four decades (1995–2015) in the Samara block.

D. Temporal Properties of the Urban Expansion

The urban area of Samara city expanded from 333.50 km^2 in 1975 to 460.87 km^2 in 2015 at annual average rate of 0.69 km^2/year (table 6).

TABLE VI. Growth rate of Samara area.

Year	Area SqKm	Growth Rate	Annual GR	Population (Thousands)
1975	333.5			1146
1980	334.99			1221
1985	336.48	0.89	0.08	1241
1990	432.98			1244
1995	529.48	36.44	3.64	1208
2000	404.94			1173
2005	280.4	-88.83	-8.88	1140
2010	370.63			1164
2015	460.87	39.16	3.91	1173

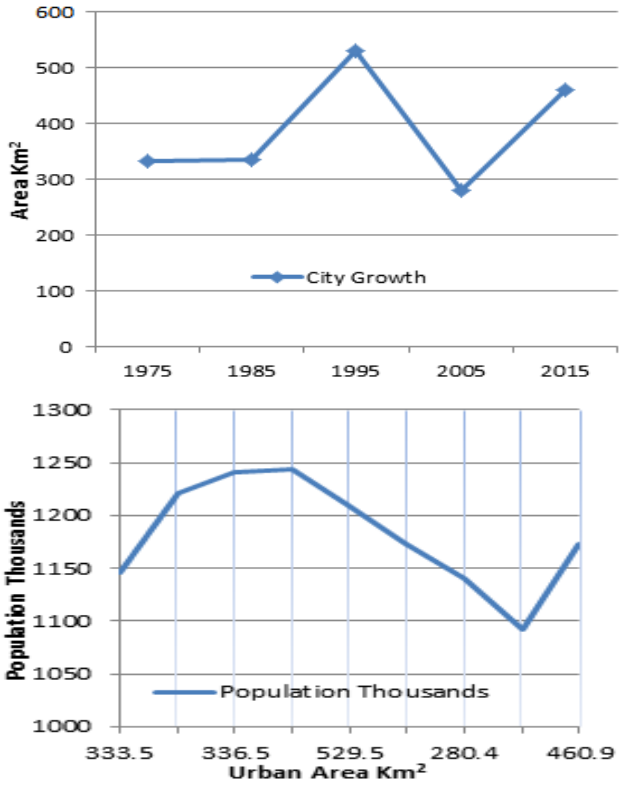


Fig. 5 Temporal change of urban area with population.

Last four decade Samara city experienced with high, low, positive and negative urban growth speed (fig.5). Satellite data correlate with historical maps or statistical data. From 1975 to 1985 there is small urban growth but from 1985 to 1995 a very high urban growth was fiend with 3.64 annual growth rate (table 6). Later on from 1995 to 2005, it was reducing dramatically around -8.88% per year. In last city was again increase with very high speed from 2005 to 2015 with 3.91 annual growth rates.

V. Driving Factors for Urban Growth and Land Use Change

Urban expansion and subsequent landscape changes are governed by geographical and socio-economic factors, such as population growth, policy and economic development. In most cases, urban expansion and associated land use/cover changes resulted from a combination of following factors such as:

ISBN: 978-1-61804-318-4

Industrial Profile: Samara Region has highly developed industry and a diversified economy structure. Industry accounts for about 40% of the gross regional product. It includes production and processing and energy sectors. The development of the region's economy is based on high-tech processing industries with high added value: automobile manufacture, air and spacecraft manufacture, which account for up to 35% of the total volume of shipped production of processing industries; enterprises with high degree of processing: chemical and metallurgical. The region manufactures 30% of new passenger cars made in Russia, 31% of polymer materials for floor, wall and ceiling coatings, 23% of anhydrous ammonia, 16% of sanitary products made from ceramics, 13% of ceramic floor tile, 7.7% of automobile gasoline and 9% of diesel fuel, 8.5% of plastics in primary forms, 7.3% of beer, 5.0% of confectionery products and 4% of mineral fertilizers. Mining of minerals accounts for approx. 17% of industrial production. About 99% of them are fuel and energy raw materials. Production and distribution of energy resources makes up about 11% of regional economy [1]. That`s why industry is the main cause of urban growth and land use change.

Development of Agricultural Complex: The agricultural complex of Samara Region is one of the leading sectors in regional economics, having its strategic importance both in provision of food safety and in maintaining socioeconomic stability in the region. It is a diversified production and economy system of over 500 collective agricultural companies, 2.5 thousand farmer-ships, 267.2 thousands of private plots and about 1000 companies of food and processing and servicing industry. It accounts for 5-7% of the cost of the gross regional product and about 3% of the capital assets. Rural areas are inhabited by 631.6 thousand people, or 19.7% of the population of the Samara Region. Agricultural complex employs about 92 thousand people (over 6% of the regional workforce). Total land area in the Samara region is more than 4 million hectares, of which 3.8 million hectares are agricultural land (more than 7% of agricultural land in the Volga Federal District), including about 2.9 million hectares of arable land. The main agricultural productions are growing cereals, oilseeds and forage crops, potatoes, vegetables, fruits and berries, milk and meat production. Regional agribusiness produces 2% of agricultural output of the Russian Federation and 7% - of the Volga Federal District. In 2013 agriculture in the Samara Region showed high growth rates in the main indicators among the Russian regions [1]. The volume of gross agricultural production in all categories of farms in 2013 was estimated at 69.5 billion rubles, gross agricultural production index in comparison to the level of 2012 is estimated at 108.4 % (106.2 % countrywide). So requirement and production of agriculture is second leading cause of land use change and its effect of urban growth.

Transport and Communication: In 2010, the Concept of development of the regional transport and logistics system of the Samara region for 2011 – 2015 was approved. Construction of modern transport and logistics infrastructure at the junction of the main transport routes West – East and North – South in the Samara Region will allow to process export-import, domestic and international cargo flows on the basis of interaction of four transportation modes and to ensure entry into the system of handling the cargo flows of international transport corridors and the cargo flow in the direction China – Europe. In order to ensure the coherence and consistency of decision-making regarding the development of regional transport and logistics system, the Coordinating Council on the development of transport logistics cluster of the Samara region was formed under the Samara Region Government. Three major Russian gas pipelines cross the Samara region: Chelyabinsk – Petrovsk, Urengoy – Petrovsk, Urengoy – Novopskovsk, as well as oil and product pipelines included in the systems of OJSC "Transnefteprodukt" and JSC "Transneft", with the total length of over 5000 kilometers. Infrastructure of the communications industry is one of the most important resources of social and economic development as well as urban growth and effect on land use change [1].

International Trade and Foreign Investments: During the recent years, a significant number of large commercial investment projects were implemented in the region, including those involving foreign companies. Over 450 enterprises with the foreign capital participation are already operating; the largest of them are listed below:

- The Russian-American enterprise "GM-AvtoVAZ" – production of cars
- the Russian-American enterprise "PES / SCC" – production of wire harnesses for cars
- the Russian-Cypriot enterprise CJSC "Acom" – production of batteries
- the Russian-German enterprise "Henkel Plastic Automotive components" – manufacturing of plastic products
- the Russian-American enterprise "Samara Optical Cable Company" – production of cables
- the Russian-Chinese-Cypriot enterprise "Tomet" – production of fertilizers
- the Russian-American enterprise LLC "Combine of ceramic structures" – manufacturing of ceramic products
- the Russian-French enterprise "Tarkett" – production of flooring
- the Russian-French enterprise "Danone-Volga" – production of yogurt
- OJSC "Confectionery Association "Russia" – production of confectionery
- Branch of the Russian-British enterprise "Coca-Cola Inchcape HBO-BBC Eurasia" and the Russian-American company "Pepsi International Bottlers (Samara)" – production of soft drinks.

The foreign companies are also active in the financial services sector. The offices of Raiffeisenbank, Citibank, Societe Generale Vostok Group and Barclays Group are operating on the territory of the region. In May 2007 an office of the European Bank for Reconstruction and Development was opened in Samara. These investments are major cause of attractions of people from surrounding and other parts of country for employment and in last its cause of urban growth and land use/cover change [1].

VI. Conclusion

This research work examined the urban growth of Samara city, which is the most important historical, cultural, industrial and commercial city of Russia. Satellite data and census data were used to monitoring the dynamic phenomena of urbanization with the help of remote sensing and GIS technology. Samara land expansion is based on Samara and Volga River and social factors such as population growth, migration and economic development. Despite the popular belief that Samara gardens and vegetation cover were destroyed and converted to built-up areas, this study demonstrated that development occurred mainly in available open spaces in the city and remaining lands between the buildings. Conversion of vegetation and orchards to built-up area, however, has been a more recent phenomenon. The study reveals that the major land use in the study area is vegetation (forest and grassland). The area under vegetation has decreased by 7.66% (149.43 km^2) due to afforestation work during 1975 to 2015. The second major category of land in the study area is built-up area which was increased by 6.48% (127.37 km^2) due to conversion in forest and grassland. Thus, the present study illustrates that remote sensing and GIS are important technologies for temporal analysis and quantification of spatial phenomena which is otherwise not possible to attempt through conventional mapping techniques. Change detection is made possible by these technologies in less time, at low cost and with better accuracy.

Acknowledgment

This work is financially supported by the Russian Scientific Foundation (RSF), grant no. 14-31-00014 “Establishment of a Laboratory of Advanced Technology for Earth Remote Sensing”.

References

[1] Ministry of Economic Development, Investments and Trade of the Samara region (2015) www.economy.samregion.ru

[2] L. Altieri, D. Cocchi, P. Giovanna, M. Scott, M. Ventrucci, Urban sprawl scatterplots for Urban Morphological Zones data. Ecol. Indic. 36, 315–323, (2014).

[3] J. Xiao, Y. Shen, J. Ge, R. Tateishi, C. Tang, Y. Liang, Z. Huang, Evaluating urban expansion and land use change in Shijiazhuang, China, by using GIS and remote sensing. Landscape and Urban Planning 75, 69–80, (2006).

[4] M. Herold, N. Goldstein, K.C. Clarke, The spatio-temporal form of urban growth: Measurement, analysis and modeling. Remote Sensing of Environment, 86(3), 286–302, (2003)

[5] X. Liu, R.G.Jr. Lathrop, Urban change detection based on an artificial neural network. International Journal of Remote Sensing, 23, 2513–2518, (2002)

[6] S.E. Gill, J.F. Handley, E.A. Roland, S. Pauleit, N. Theuray, S.J. Lindley Characterizing the urban environment of UK cities and towns: a template for landscape planning. Landscape Urban Plann. 87 (3), 210–222, (2008)

[7] Li Xin-Hu, Liu Ji-Lai, V. Gibson, Y.G. Zhu, Urban sustainability and human health in China, East Asia and Southeast Asia. Environ. Sustainability 4, 436–442, (2012)

[8] N.B. Grimm, J.M. Grove, S.T.A. Pickett, C.L. Redman, Integrated approach to long-term studies of urban ecological systems. Bioscience, 50(7), 571–584, (2000)

[9] X. Xinliang, X. Min, Quantifying spatiotemporal patterns of urban expansion in China using remote sensing data. Cities 35, 104–113, (2013)

[10] M.S. Boori, V. Vozenilek, K. Choudhary, Land use/cover disturbances due to tourism in Jeseniky Mountain, Czech Republic: A remote sensing and GIS based approach. The Egyptian Journal of Remote Sensing and Space Sciences, *18*(1), 17–26, (2015) *Doi:10.1016/j.ejrs.2014.12.002*

[11] J. Chadchan, R. Shankar, An analysis of urban growth trends in the post-economic reforms period in India. Int. J. Sustainable Built Environ. 1, 36–49, (2012)

[12] A. Latif, S.M. Sabet, Urban sprawl pattern recognition using remote sensing and GIS, case study Shiraz City, Iran. In Proceedings of urban remote sensing joint event Shanghai, China. 20–22 May, 2009,

[13] M.B.C. Soh, Crime and urbanization: revisited Malaysian case. Procedia Social Behav. Sci. 42, 291–299, (2012)

[14] M.S. Boori, V. Vozenilek, K. Choudhary, Land Use / Cover Change and Vulnerability Evaluation in Olomuc, Czech Republic, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., II-8, 77-82, (2015) *Doi:10.5194/isprsannals-II-8-77-2014*

[15] J. Radke, S. Andra, O. Al-Kofani, B. Roysan, Image change detection algorithms: a systematic survey. IEEE Trans. Image Process. 14 (3), 291–307, (2005)

[16] M.E.Bastawesy, Hydrological Scenarios of the Renaissance Dam in Ethiopia and Its Hydro-Environmental Impact on the Nile Downstream. J. Hydro. Engin., 2014, http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0001112.

[17] M.S. Boori, R.R. Ferraro, Global Land Cover classification based on microwave polarization and gradient ratio (MPGR). Geo-informatics for Intelligent Transportation, Springer International Publishing Switzerland. 71, 17-37, (2015) *Doi:10.1007/978-3-319-11463-7-2*

[18] M.S. Boori, V. Vozenilek, J. Burian, Land-cover disturbances due to tourism in Czech Republic. Advances in Intelligent Systems and Computing, Springer International Publishing Switzerland. 303, 63-72, (2014) Doi:*10.1007/978-3-319-08156-4-7*

[19] S. Mukherjee, Land use maps for conservation of ecosystems. Geog. Rev. India 3, 23–28, (1987)

[20] A. Ruiz-Luna, C.A. Berlanga-Robles, Land use, land cover changes and costal lagoon surface reduction associated with urban growth in northwest Mexico. Land. Ecol. 18, 159–171, (2003)

[21] M.G. Turner, C.L. Ruscher, Change in landscape patterns in Georgia. USA Land. Ecol. 1 (4), 251–421, (2004)

[22] G. Chander, B.L. Markham, D.L. Helder, Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. Rem. Sen. Envi., 113 (5), 893–903, (2009)

[23] P. Gamba, F. Dell’Acqua, B.V. Dasarathy, Urban remote sensing using multiple data sets: Past, present, and future. Information Fusion, 6, 319–326, (2005)

[24] J.R. Jensen, Remote sensing of the environment: An earth resource perspective (2nd ed.). Upper Saddle River, NJ: Pearson/Prentice Hall. 2007,

[25] M.S. Boori, V. Vozenilek, Land-cover disturbances due to tourism in Jeseniky mountain region: A remote sensing and GIS based approach. *SPIE Remote Sensing*. 2014, Vol. 9245, 92450T: 01-11. *Doi:10.1117/12.2065112*

[26] J.R. Anderson, E. Hardey, J. Roach, R.E. Witmer, A land use and land cover classification system for use with remote sensor data. US geological survey professional paper, Washington, DC, 964, pp. 28, (1976)

[27] Q. Weng, A remote sensing-GIS evaluation of urban expansion and its impact on surface temperature in the Zhujiang Delta, southern China. Inter. J. Rem. Sens., 22 (10), 1999–2014, (2001)

[28] M.S. Boori, V.E. Amaro, Land use change detection for environmental management:using multi-temporal, satellite data in Apodi Valley of northeastern Brazil. *Applied GIS*, 6(2): 1-15, (2010)

[29] M.S. Boori, V.E. Amaro, A remote sensing and GIS based approach for climate change and adaptation due to sea-level rise and hazards in Apodi-Mossoro estuary, Northeast Brazil. *International Journal of Plant, Animal and Environmental Sciences*, 1(1): 14–25, (2011)

[30] H. Taubenbock, M. Wegmann, A. Roth, H. Mehl, S. Dech, Urbanization in India – Spatiotemporal analysis using remote sensing data. Comput. Environ. Urban Syst. 33, 179–188, (2009)

First A. Author Prof. Dr. Mukesh Singh Boori is Senior Scientist in Samara State Aerospace University, Russia (03/2015 - Present) and Assistant Professor in JECRC University India (01/2013 - Present). He was involve in European Union Project as well as Visiting Assistant Professor in Palacky University Olomouc, Czech Republic (04/2013 – 06/2015), Ruhr University Bochum Germany (09/2014 – 12/2014) and University of Leicester, UK (Honorary Fellow 2014) funded by European Union. He was Scientist in Satellite Climate Studies Branch (NOAA/NASA), selected by National Research Council (NRC), Central Govt. of USA, Washington DC, USA. At the same time he completed his Postdoc from University of Maryland, USA (10/2012). He has done PhD (*EIA & Management of Natural Resources*) from Federal University – RN (UFRN), Natal –RN Brazil (08/2011), funded by Brazil-Italy Govt. fellowship. He has done Predoc (*Earth & Environmental Science*) from Katholieke University Leuven, Belgium (08/2008), selected by Ministry of Human Resource Development (MHRD) New Delhi, India and funded by Govt. of Belgium. He has done MSc (*Remote sensing & GIS*) from MDS University Ajmer (2004) and BSc (*Bio-group*) from University of Rajasthan, Jaipur, India (2002). In early career, he was scientist in JSAC/ISRO (2006-2007) and before that Lecturer (PG) at MDS University Ajmer (2005-2007Sessions). He received international awards/fellowships from UK, USA, Brazil, Italy, Indonesia, Belgium, Czech Republic, Germany, Russia and India. He known Six Language and visit four Continents for Awards, Meetings, Trainings, Field Trips and Conferences. He is an active Organizing Committee Member in Earth & Space Science Conferences, Co-Chaired a session and gave Conference Opening Ceremony Speech as Reynold Speaker (08/2012) at Chicago, USA. He is editor and member of more than 10 International Scientific Societies/Journals/Committees, related to Earth & Space Science, which include organize conferences. His prime research interest is “EIA and Management of Natural Resources through Remote Sensing & GIS Technology”. He has more than 50 International Publications including Books as a first author on Vulnerability, Risk Assessment and Climate Change.

ISBN: 978-1-61804-318-4

NoSQL: Robust and Efficient Data Management on Deduplication Process by using a mobile application

Hemn B.Abdalla, Jinzhao Lin, Guoquan Li

***Abstract*—**Present Information technology has an enormous responsibility to deliver Efficient Data Management in all kinds of Industries and Online Service-oriented companies. Each and every day spends enormous amount of money for online data storage and maintenance, nobody has the time to spend work hours reviewing the data or entering it into new software and then hiring someone to maintain it. In this paper, it is focused on Deduplication of data storage that is a more economical process to keep unique and speed data access. The Deduplication will manage the various content, customer data technique to avoid Deduplication while processing of data storage then fast data access and secure data storage. It is used clustering method in this paper so that rapid retrieval of data from the database we are applying the Mapping technique for categories and linking data to access information. For data processing, we use MongoDB (NOSQL) its robust database, for apply various new method and algorithm in complex data (large scale data).

Keywords-*MongoDB; Deduplication; DataSecurity; Clustering; Mapping.*

I. INTRODUCTION

IN the past couple years; we saw the number of mobile users day by day was increasing and growth in all the places in the world. We have experienced a significant shift in the way we access the internet today with mobiles that become the primary access point for internet usage. Before usually people's developed a mobile application with saved data or any files on (MySQL & SQL Server). NoSQL systems provide data partitioning and replication as built-in features [3], [4], [5]. So today we are creating a new theory of collecting data on MongoDB by using mobile application.

MongoDB is Document database that is call as NOSQL Query then it provides the high performance that makes a read and writes fast. In that storage system, we apply the Deduplication method it present to find the duplication record as well as content, the focus of Deduplication process allows only original data, so finally we get high data Security.

Today rapidly improve the complexity of massive amounts businesses have the space to store stacks of candidate resumes and applications. Nobody has the time to spend work hours reviewing the data or entering it into new software and then hiring someone to maintain it. Many people are not appropriate for your open positions; their information can be stored as a passive person in our database. Once they were loaded into our RESUME Database, they can be retrieved at a later date. We provide a secure professional system that allows us to select search criteria such as job applied for using date range, specific skills, and location. We can immediately provide lists of all your person's contact details, and their resumes (data) from our database. We can make these available to your business representatives for your critical reporting and assessment or data needs.

This research work is supported by University Innovation Team Construction Plan of Chongqing, the National Science Foundation of China under Grant No. 61301124.

II. METHODOLOGY

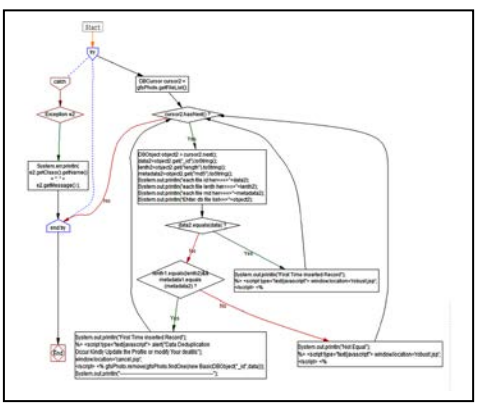
A. *Data Deduplication Process*

Deduplication is one of the Latest technologies in the current environment because of its ability to reduce the value of costs. But it comes in many flavors and organizations need to understand each one of them if they are to choose the one that is the best. Deduplication method can be applied to data or content of the data in earliest storage, support storage, cloud storage or data in flight for reproducing, such as Local area network and Wide area network transfers. So eventually, it offers the below benefits.

Deduplication system chunks data stream firstly. It tries to find duplicate chunks from the already stored chunks, and only stores new chunks to disk. [6] The Deduplication system performs all Deduplication, provides all quality, and captures interactions between source data and result data.

Algorithm1: Cryptographic hash value from the data chunks/blocks: Hash functions accelerate table or database lookup by detecting duplicate records in a large file, a Cryptographic hash function allows one to verify easily that some input data maps to a given hash value.

In this algorithm which provides the file information like chunks information. In this paper, we implement the when data user can store the data mean we will find two major concepts (file. Chunk) (file. File). Now apply the algorithm to find the duplication occur or not that finds the file chunk size as well as Metadata, ((it adjusts the dimension target table so that those sets of duplicates already in the target table reduced to single records).



ISBN: 978-1-61804-318-4

Figure 1. Flow Chart code implementation for Deduplication process

B. *Clustering*

Algorithm2: Hierarchical K-Meansclustering is the task of grouping a set of objects its general term so that in our paper apply the Hierarchical K-Means algorithm, in this algorithm provide group of data like cluster the data from MongoDB that converts into single group or group object, MongoDB is ready to push the general data into cluster data, if there is any data object, our database that data will store cluster data (Group data), and repeat this step to perform found data point and store MongoDB for example in our project consider the resume information in the resume data has different domain like Java, JSP and PHP based on that field data object (data) will cluster and Store (domain is call as data point).

Hierarchical k-means clustering the algorithm divides the dataset recursively into clusters. Clustering analysis is an important technique. [7] The k-means algorithm is used by setting k to two to split the dataset into two subsets. Then, the two subsets are divided again into two subsets by setting k to two. The recursion terminated when the dataset was split into single data points, or a stop criterion is reached. [1]

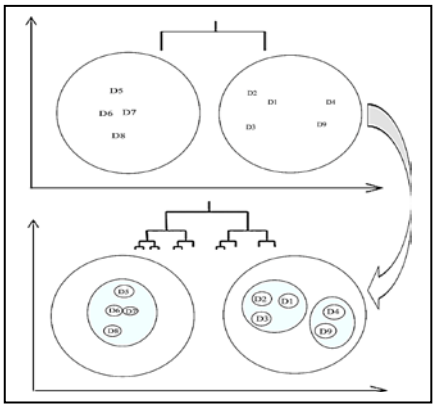


Figure 2. Hierarchical k-means clustering

Hierarchical k-means has O (n) run time process. Such a run task is possible because both the k-means algorithm and all operations concerning trees are possible in O(n). Traversing a tree is always done via depth- or breadth- first search.

Code Implementation for Clustering Process

```
GridFS gfsPhoto = new GridFS(db, i);
GridFSInputFile gfsFile = gfsPhoto.createFile(imageFile);
gfsFile.setFilename(dbFileName);
gfsFile.put("_id",data);
gfsFile.put("title",n);
gfsFile.setContentType("application/msword");
gfsFile.save();
```

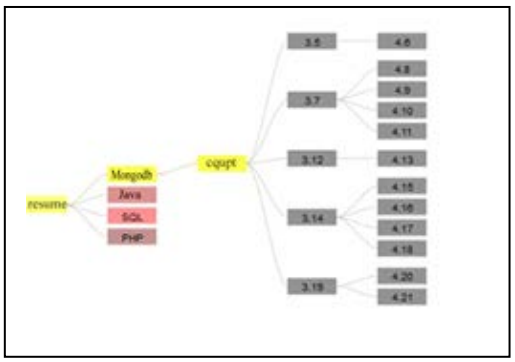


Figure 3. Shown Mapping our system chart

C. *Mapping*

In computing and warehousing management, storage mapping is the process of creating data element or value mappings between two return data models. Data mapping is used as a first step for a broad range of data integration tasks including Data transformation or data mediation between a data source and a destination. [2]

1) Identification of data relationships as part of the data lineage analysis.
2) The discovery of hidden sensitive data such as the last four digits social security number hidden in another user id as part of a data masking or de-identification project.
3) Consolidation of multiple databases into a single database and identifying redundant columns of data for consolidation or elimination.

D. *Equations*

Dijkstra's original variant found the shortest path to two nodes, but a more common difference fixes a single node as the "source" node and finds shortest paths from the source to all other nodes in the graph, producing a shortest path tree.

Algorithm3: In this algorithm mainly provide on Mapping of two data object so apply Dijkstra's we retrieve the data and find the record fast access is use full for large-scale database for example we consider the resume data apply the mapping point or node point Current location, role key Skill, domain that point are used find the database result in this outcome are shown by table format its very useful for fast retrieval and find the accurate data.

Math Algorithm functions: Map reduction

Input: A reduced Map G(V, E,f, R) where is R is set of output
A cluster the data point into subset p
P=Getcluster(v,RE){P=V/RE={A1,A2….As} where Ai[ai]
//A1,A2 or subset
Getreduced vertices(P)//shortest map path
GetEdges(P) // get all domain value
Create the reduced map or graph Gr=(Vr, Er,Fr,Rr)
For Ai belong p,|Ai|1 do

ISBN: 978-1-61804-318-4

Get Result // resume data

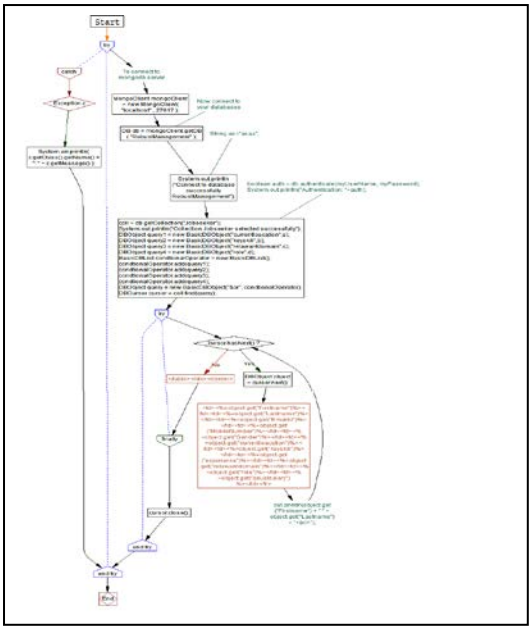


Figure 4. Flow Chart code Hierarchical k-means clustering

E. MongoDB(NoSQL Database)

MongoDB, are quickly grown to become a popular database for web applications and Mobile application, NoSQL is fast data access. [8] Is an entirely appropriate for a Node.JS applications let us write a JavaScript to any client, back-end, and the database layer. MongoDB is being used in many critical projects and products, besides, the inserted support for location queries is a bonus that's hard to dismiss, faster to enter and retrieve data's. Its schema-less kind is a more serious mate to our permanently evolving data structures in web applications and mobile application, relatively new to the database market. Data becomes grown so fast, and users upload generated from varied sources like Videos, texts, speeches, log files, images, etc. Our Data is Location Based; MongoDB has special built-in functions, so finding related data from particular locations is fast and accurate.

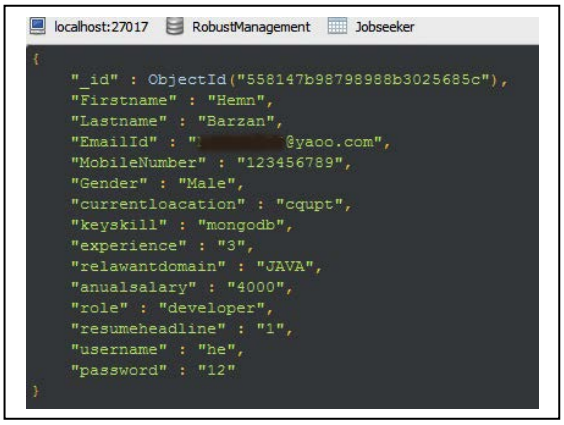


Figure 5. Data details in MongoDB server

In figure 4 MongoDB server, we used the db.collection. Find() method for retrieve or shown our data's features from a collection.

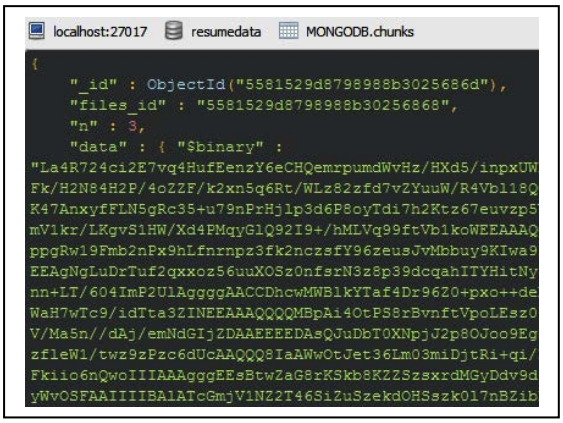


Figure 6. GridFS file details

In figure 5 MongoDB server are shown our saving CV files and storing and retrieving files that exceed the BSON-document with large files.

II. ARCHITECTURE

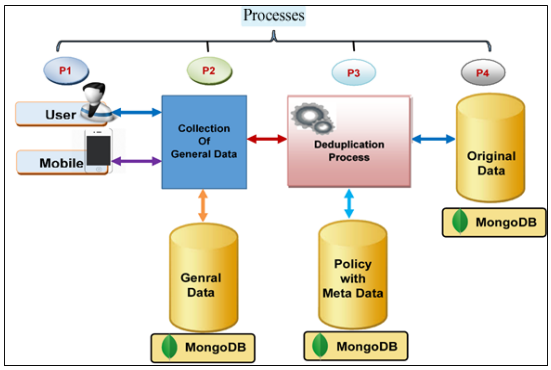


Figure 7. Architecture system diagram

In figure 6 fourth representation processes:

Process 1: User or Mobile User call as real world object, in this object, works to gather to access the database or Storage server.

Process 2: Now her performing data collected from the user or data object, in this General information or information will Store database (MongoDB).

Process 3: Object data store before that perform the deduplication process; deduplication Process shows the new policy and technique apply MongoDB.

Process 4: after it completing duplication process, it's ready to store original data (Now use the mapping as well as clustering).

ISBN: 978-1-61804-318-4

III. Out Put and Experimental Setup

We implemented the Robust and Efficient Data Management on Deduplication Process without SQL and Query as a Mobile application system using Java language with JSP. The Mobile application interface for our system forms used open-source JavaScript library, we used MongoDB for saved our Data's information details. All experiments were run using Android system for the mobile application and machine with Intel Core 3CPU @2.9GHz, 4096G main memory, and running on Windows 8.1 Pro.

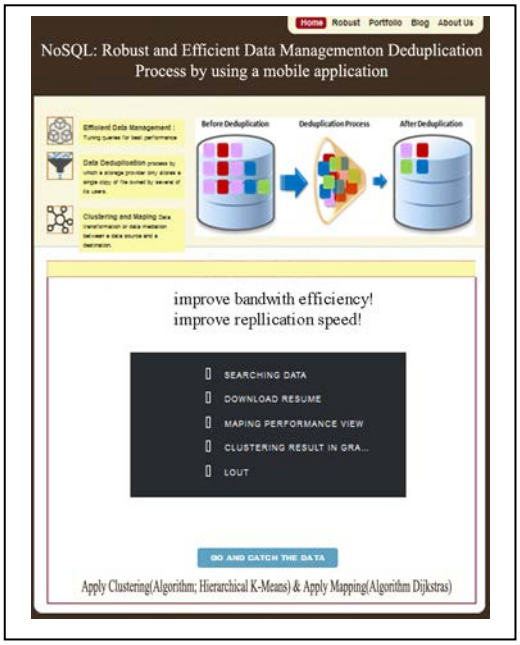


Figure 8. Our interface system

In figure 7 representation in our system application how users execution and run our method, first users try to use user ID with password after first step user can upload his\her CV details, second steps our data details automatically save in MongoDB server with run Deduplication algorithm for reduce the value of costs and remove duplicate value, third steps data from MongoDB that converts into single group or group object by clustering, fourth steps Mapping process for value mappings between two return data models. Any times when users want to retrieve or find the his\her document after input his phone number directly can get his\her own CV files or by an administrator user.

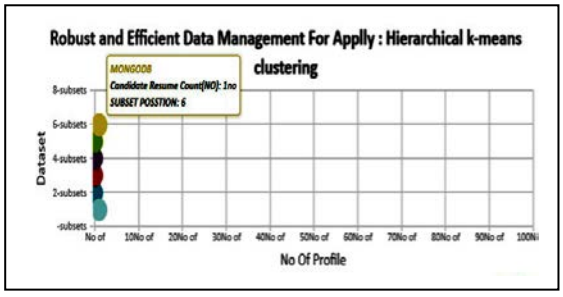


Figure 9. Clustering result in graph view

In figure 8 representation of clustering or explaining the clustering the resume data (resume dataset). We consider the resume database is in this database that called as a dataset, so dataset has a several subset (cluster level and cluster point cluster point is call as group dataset). Finally in this graph output show the how many profile is there based on the domain (example MongoDB have 20 no profile) her MongoDB is one of the subsets (cluster or group of data).

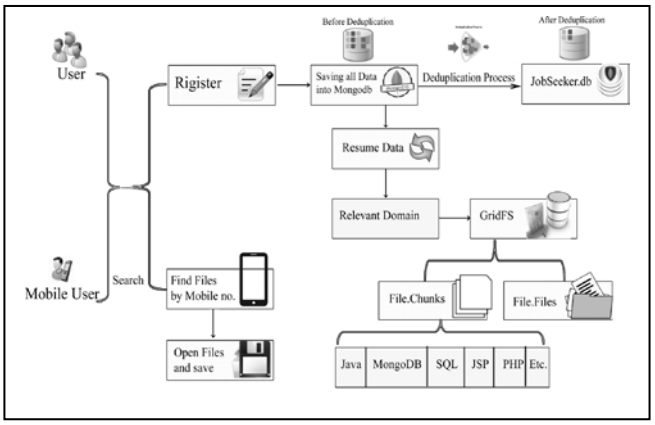


Figure 10. Implementation system diagram

IV. Scope

Our project scope provides a database to a very efficient data access or getting the output and also provides high security that becomes used to Deduplication technique. In this process reduce duplicated record entry as well as delicate storage also. After applying the clustering method to cluster our database to store like use clustering algorithm database divided into sub-database like subset value and also apply mapping methodology and map the resume data various keyword (call node value Ex: domain name ,experience and current location..etc). Moreover, we use MongoDB in this database is documented database MongoDB is high performance and also the large storage capacity that is I have to store the more than 10mp file also store.

V. Conclusion

The number of users can access the database inefficiently because they apply the Deduplication method, meanwhile, original content only share or store the database (it adjusts the dimension target table so that those sets of Deduplicates already in the target table are reduced to single records). Finding the data point and using clustering technique, so we used Hierarchical K-Means clustering over our paper to implement the resume data and apply the concept group files based domain (domain name is subset of resume data set). The group the data or data object into the database as of now use to resume database they get high-performance data it most familiar word for MongoDB (NoSQL). Finally, we apply the mapping technique map the cluster subset value and access the resume data based Dijkstra's algorithm and get fast data or dataset retrieval.

ISBN: 978-1-61804-318-4

ACKNOWLEDGMENT (HEADING 5)

This research work is supported by University Innovation Team Construction Plan of Chongqing, the National Science Foundation of China under Grant No. 61301124.

REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. (references).

[2] J. Mark W. Storer Kevin Greenan Darrell D. E. Long Ethan L. Miller, Secure Data Deduplication, StorageSS’08, October 31, 2008, Fairfax, Virginia, USA. Copyright 2008 ACM 978-1-60558-299-3/08/10

[3] R. Cattell, “Scalable sql and nosql data stores,” SIGMOD.

[4] V. Benzaken, G. Castagna, K. Nguyen, and J. Siméon, “Static and dynamic semantics of NoSQL languages,”SIGPLAN Not., vol. 48, no. 1, pp. 101–114, Jan. 2013.

[5] F. Cruz, F. Maia, M. Matos, R. Oliveira, J. a. Paulo, J. Pereira, and R. Vilaça, “MeT: Workload aware elasticity for NoSQL, book title = Proceedings of the 8th ACM European Conference on Computer Systems, series = EuroSys ’13, year = 2013, ISBN = 978-1-4503-1994-2, location = Prague, Czech Republic, pages = 183–196, num pages = 14, publisher = ACM, address = New York, NY, USA.”.

[6] J¨urgen Kaiser, Dirk Meister, Andre Brinkmann, Sascha Effert, Design of an Exact Data Deduplication Cluster, 978-1-4673-1747-4/12 c 2012 IEEE.

[7] Jia-Lien Hsu and Hong-Xiang Yang, A Modified K-means Algorithm for Sequence Clustering, 978-0-7695-3745-0/09 © 2009 IEEE.

[8] Dileepa Jayathilake, Charity Sooriaarachchi, Thilak Gunawardena, Buddhika Kulasuriya and Thusitha, A Study Into the Capabilities of NoSQL Databases in Handling a Highly Heterogeneous Tree, 978-1-4673-1975-1/12©2012 IEEE.

Secret Sharing in Visual Cryptography using NVSS and Data Hiding techniques.

Ms.Misha Ann Alexander
Student, Computer Engineering Department
Sinhgad Institute of Technology
Lonavala, India
annalexander33@gmail.com

Mr. Sanjay B. Waykar
Asst Prof, Computer Engineering Department
Sinhgad Institute of Technology
Lonavala, India
sbwaykar@gmail.com

***Abstract*— Visual Cryptography is a special unbreakable encryption technique that transforms the secret image into random noisy pixels. These shares are transmitted over the network and because of its noisy texture it attracts the hackers. To address this issue a Natural Visual Secret Sharing Scheme (NVSS) was introduced that uses natural shares either in digital or printed form to generate the noisy secret share. This scheme greatly reduces the transmission risk but causes distortion in the retrieved secret image through variation in settings and properties of digital devices used to capture the natural image during encryption / decryption phase. This paper proposes a new NVSS scheme that extracts the secret key from randomly selected unaltered multiple natural images. To further improve the security of the shares data hiding techniques such as Steganography and Alpha channel watermarking are proposed.**

***Index Terms*—Natural Visual Secret sharing, natural images, noisy share, pixel swapping, encryption, decryption.**

I. INTRODUCTION

We are in a digital world where digitization has touched industries, governments, education, research, trade etc. This has basically caused a large amount of high-risk data to be transacted over the Internet which is an insecure medium of data exchange. Cryptography is an encryption technique widely used in electronic communication to provide security in transmission. This technique converts plain text to cipher text so that only the intended people can read the content. Visual Cryptography is a new simple, easy to implement encryption technique developed by Moni Naor and Adi Shamir that encrypts the visual information such as text, pictures or written data and uses human visual system for decryption [2].The transmission of the visual shares in Visual Cryptography is called as visual secret sharing (VSS) Scheme [1]. In conventional cryptography, the secret is scattered into multiple shares and transmitted through multiple modes to a set of quantified users. When these shares are stacked together it reveals the secret content. The noisy textured share fulfills the security constraint but at the same time it causes the attackers attention. The increasing number of noisy shares makes the VSS scheme less user-friendly and difficult to manage. Natural Visual Secret Sharing (NVSS) provides an effective solution to these problems of VSS by reducing the number of transmitted shares and enhancing the user friendliness of the shares by using the QR code technique to hide the secret share. NVSS scheme extracts the features from natural images either printed or digital in nature captured with the help of digital devices having different settings, make and configurations. These extracted features are applied to the secret image which transforms it to an unidentifiable share. Theses natural shares are innocuous in nature and hence there is very less probability of the secret being intercepted during transmission [1]. To make the scheme more secure NVSS makes use of diverse carrier media to transmit the shares, but it suffers due to the reason that the secret image is distorted at the receiver end during decryption. The regeneration of similar natural images with same dimensions and settings during the encryption and decryption is truly a difficult task due to the use of dissimilar electronic devices. To reduce the distortion of the retrieved secret, a new NVSS scheme is proposed in which the secret key is extracted by processing the randomly selected natural meaningful images either selected from the database or some websites. The key generated along with chaotic equations are used to map the original pixels of a secret image onto new locations. To further enhance the security of the noisy shares during transmission it is hidden behind meaningful cover images.

This paper proposes algorithms for Key Extraction and Encryption /Decryption of Secret images. The remainder of the paper is as follows Section II contains the related work Section III presents the Proposed Scheme Section IV and V contains the evaluation and results of proposed work finally Section VI concludes the work.

II. RELATED WORK

The current research in visual cryptography basically focus on the VSS where the shares transmitted are noisy in nature [1][2].The noisy shares are not user friendly and hence the researchers tried to improve the user friendliness of the shares and hence the quality of the shares by adding a cover image to the noisy meaningless shares[4].Even though the quality of the shares improved but the recovered image had a problem of pixel expansion. To further improve the quality, user friendliness and security of the secret image researchers adopted the technique of steganography along with the VSS [9]. The Steganography is a technique where the secret is

embedded in the cover image the embedded image is the Stego share. The stego share are also identified by the steganalysis technique [6].Thereafter the researchers tried to use natural images to share the secret content but, however, the system failed due to the textures of the natural images were visible on the secret share [15].In [1] authors attempted to extract the features from some natural images and encrypted the secret image based on the features of the natural image selected. The natural images consist of both the digital image and the printed image. The printed image can be some handwritten data or some pictures and equipment like the cameras were used to capture the images. Even though natural images had a very high level of security due to the use of different equipment with variety of settings and features the image captured during the encryption and the decryption would vary in its size, resolution and other important parameters and hence this can cause the retrieved image to be distorted.

This paper extends the previous work of the authors by proposing a new NVSS technique which will eliminate the natural image preprocessing phase as well the security, user friendliness and manageability of the shares are increased by using one noisy share with a cover image and data hiding techniques like digital watermarking and Steganography to further improve the security of the system.

III. PROPOSED SCHEME

A. Background

Cryptography is a network security tool that provides confidentiality, integrity, and security. It makes use of encryption to enable confidentiality. One Time Password (OTP) was developed by Gilbert Vernam in 1917 and it is a very secure unbreakable technique which makes use of dynamic or random passwords each time [2]. Visual Secret Sharing (VSS) scheme is a technique which delivers the secret shares to the quantified users. These shares when stacked together it reveals the secret. These shares cause the attention of hackers due to its noisy nature. These shares are meaningless, unmanageable and not user-friendly in nature due to its noisy texture. The Visual Cryptography supports various secret sharing schemes like 2 out of 2 scheme, k out of n and n out of n scheme. In the 2 out of 2 scheme, two shares are generated one the cipher text and the other is the key. Every pixel in the original secret image is divided into sub pixels depending on whether it is a black or a white pixel. Decryption takes place by overlapping the two shares and revealing the secret image.

Natural Visual Secret Sharing (NVSS) Scheme is an improvement over the VSS scheme where the key is extracted from the randomly selected printed and digital images. These images can be photographs which can be captured by the digital equipment and hence the same settings are required during both the encryption and decryption phase.It is not easy to acquire the same image and settings during both encryption and decryption which leads to a distorted retrieved secret image.

The proposed new NVSS scheme extracts the key through multiple natural images either from the public internet or any other source, this key is given to the chaotic equation which scatters the original pixel positions. To further improve the security of the shares data hiding techniques are used. Compared to the previous NVSS scheme the new scheme improves the quality, manageability and user friendliness of the retrieved secret image as well as eliminates the natural image preprocessing phase.

B. Proposed New NVSS Scheme

The proposed new NVSS scheme has two major phases the key generation phase and the encryption phase.

- The Key is extracted from ‘n’ natural meaningful images. These natural images can be 24bit /pixel color images which are randomly selected from any websites on public Internet or photographs stored in the system. To extract the key from these images, it has to first undergo some preprocessing.
- The natural image has to be binarized first so that we can process the individual pixel values which can be either a black or white pixel. The 24-bit images are transformed into an 8-bit binary image or grayscale image.

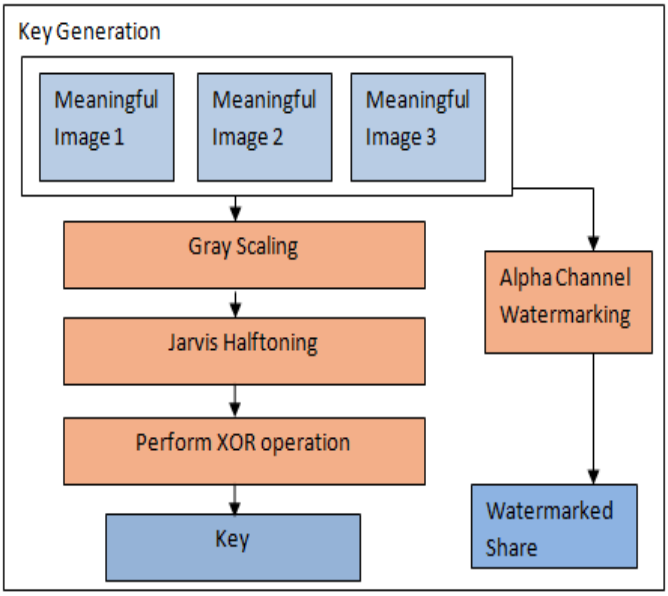


Fig.1. Key Extraction

- In a color image, every pixel is composed of three different color intensities i.e. Red, Green, and Blue. Gray scaling is a method that converts these three intensities into a single value. The conversion can be done by the average method.

$$G_{img} = R(x, y) + G(x, y) + B(x, y) \text{ ---------- (1)}$$

- The result is an 8-bit image. The 'n' natural images are converted into grayscale image and then further half toned.

- The grayscale natural images are then half-toned in which the continuous tone images are converted into black and white halftone image. Error diffusion can basically produce halftone images which are of better quality and is pleasing to human eyes rather than that of the other halftone methods.
- Every pixel (x, y) is compared with the threshold value 127 if the pixel value is greater than 127 then a white pixel will be generated or else a black pixel is generated. The resulting images can contain only two colors black and white. This process helps to differentiate between the background and foreground color in the natural images.

$$H(i,j) = \begin{cases} 1 \text{ if } IMG(i,j) >= \text{Threshold value} \\ 0 \text{ Otherwise} \end{cases} \quad \text{---------- (2)}$$

- Either a full black or a full white pixel is generated. An error is diffused by calculating the difference between the halftone and original natural image and then the difference is added to the next pixel.
- This will result into a matrix for each natural image. This matrix contains only 0's and 1's. The total number of 0's and 1's in each matrix are counted and an XOR operation is performed the result is the key.

Algorithm: - 1 KEY_GENERATION ( )

Input: - N1,N2,N3

Output: - KEY

1. Do for each image N1,N2,N3
2. For each pixel repeat 3-4
3. Calculate Gimg by equation (1).
4. Determine H (i, j) by equation (2).
5. End of Loop
6. KEY<- Calculate number of 0 and 1 perform XOR operation.
7. Store the watermark image in each natural images alpha channel.
8. Output KEY
9. End

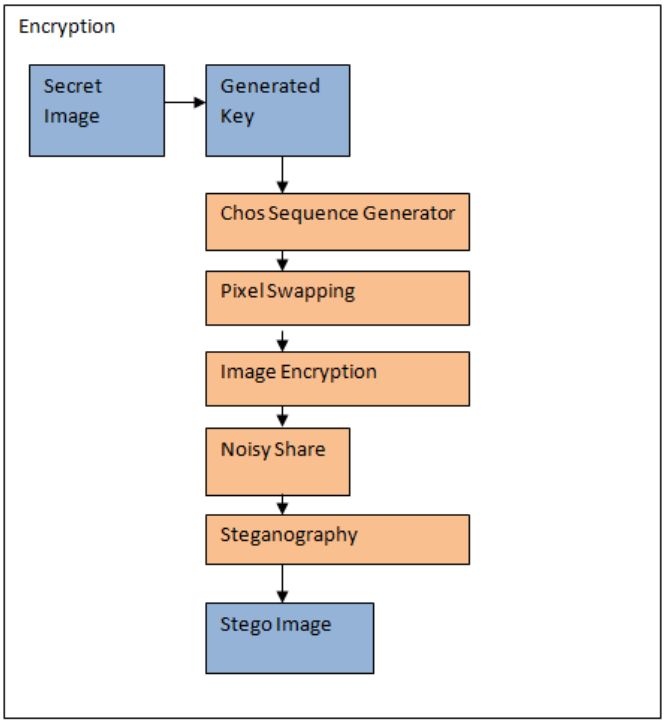


Fig. 2. Encryption Phase

Algorithm: - 2 ENCRYPTION ( )

Input: - KEY,SECRET_SHARE

Output: - SSHARE

1. NPIX <- Generate new pixel co-ordinates(x,y) using chaotic equations (3)
2. Add Ascii value of Key to the new co-ordinate value.
3. SHARE <- Shuffle the co-ordinates(x,y) to (x1,y1)
4. Calculate LSB of the cover image
5. SSHARE < - Replace the LSB bits of cover image with the bits noisy secret image.
6. Output SSHARE
7. End

- The chaotic equations carry a very dynamic behavior every pixel in the secret image is shuffled based on the values generated by the chaotic equations.
- To make it more secure the coordinates (x2, y2) are added with an offset which is the sum of ASCII value of the generated key.

ISBN: 978-1-61804-318-4

- The pixel (x, y) are placed at a different position (x_2, y_2) in such a way that the secret image is not visible to the human eyes.

$$X_{n+1} = 1-ax_n^2 + y_n$$

$$Y_{n+1} = bx_n \quad \text{---------- (3)}$$

- The share that is generated is noisy in nature and hence attracts the hackers. To safely transmit this noisy share it can be hidden within another cover image.

C. Data Hiding Technique

To improve the security of the share further we can make use of data hiding techniques like steganography. Cryptography and Steganography work very closely with each other to improve the security of the noisy share. A proper container or a cover image has to be selected in which the noisy secret can be embedded completely. There are various sizes of images like 8 bit and 24 bit. The larger the cover images the more bits can be stored in it. The various types of steganography in this system LSB (Least Significant Bit) technique is used and it's a very popular technique. In this technique, the LSB bits of the cover image are replaced with the bits of the noisy share.

If we want to encode A (ACSII value 65 or a binary value 01000001) in the below given carrier file.

01011101 11010000 00011100 10101100

11100111 10000111 01101011 11100011

After Embedding

01011100 11010001 00011100 10101100

11100110 10000110 01101010 11100011

D. Decryption

The stego image is transmitted at the receiver end .The noisy share is then retrieved from the cover image. The meaningful image is either transmitted to the receiver or their address in public Internet is sent to the receiver. The key is regenerated from the same natural meaningful images. This key as offset and the chaotic equation is used to map the pixels back to its original positions.

Algorithm: - 3 DECRYPTION ( )
Input: - SSHARE
Output: - SECRET

1. Stego Image is read
2. LSB bit is calculated
3. Noisy share is extracted
4. The Meaningful image is accessed from the appropriate websites
5. Generate the chaotic equation from the Meaningful images
6. Swap the pixels of the noisy share
7. End

evaluate the performance of the New NVSS scheme. The Secret image is the well-known image of Leena of dimensions 512 x 512 pixels. We also select three natural images from random websites which are 24-bit color images with various dimensions. These three images are passed to the Key Generation algorithm as input. This algorithm returns a unique key 715542 as output. Then the values of every pixel (x, y) is passed to the chaotic equations which exhibit a very dynamic behavior and the new coordinate with then added with the ASCII value of the key which is the new coordinate where the pixel (x, y) is mapped to. The noisy share is then embedded in a cover image and then transmitted. During decryption phase, the key is regenerated from the three same natural images extracted from the same websites. This key is then applied with the chaotic equations to the noisy share to recover the secret image. The time complexity of the algorithm is O(ncd hw) the loop in the algorithm executes for every pixel depending on the height and the width(hXw) of the selected images and the color depth (cd) whether 24bit or 8 bit image.

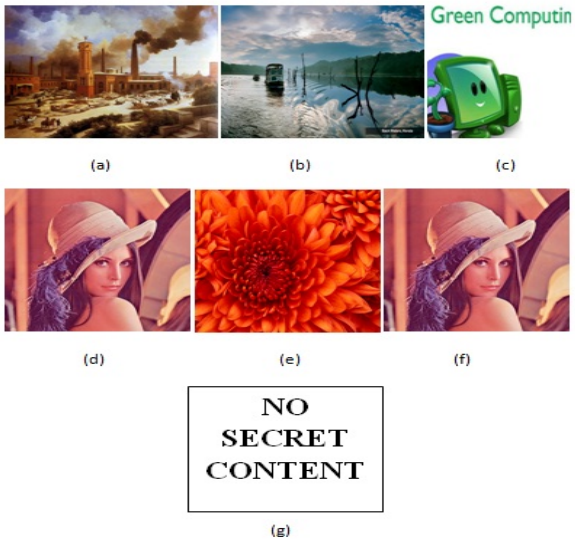


Fig 3. Experimental Results of New NVSS Scheme. a) Natural Image 1. b) Natural Image 2. c) Natural Image 3.d) Secret Image. e) Stego Image f) Recovered Secret Image g) Experimental result when the natural image is noisy or changed

.

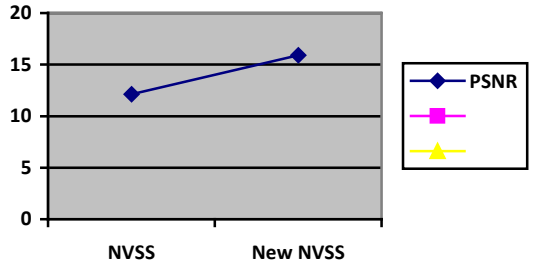


Fig. 4. Graph for PSNR

V. Comparative Study

This section shows the result of the proposed system with the existing NVSS scheme based on the PSNR (Peak Signal to Noise Ratio). The results of new NVSS scheme are more significant than the existing scheme.

Secret Share Dimensions	**NVSS scheme**	**New NVSS scheme**
Leena 256 x 256	---------	17.54db
Leena 512x512	12.12 db	15.92db

Parameter	**NVSS Scheme**	**Secured NVSS Scheme**
Type of Share	Noisy share hidden below QR code or natural image	Noisy share hidden below cover image.
Transmission Risk	Low risk	Very Low risk
No of Shares	One Share	One Share
Data Hiding techniques	QR code and Steganography is used.	Alpha channel watermarking and Steganography will be used.
Quality of Retrieved Share	Distorted Share	No Distortion

VI. Conclusion

This paper proposes a new NVSS scheme that generates a secret key from the natural meaningful images. This scheme uses the unaltered natural shares from random websites or from databases which reduce the distortion in the retrieved secret image. To further improve the security of the noisy shares steganography data hiding technique is used. The natural images can also be transmitted securely in the network using alpha channel watermarking.This study has contributed to the previous work of the authors by effectively reducing the transmission risk and this attempt has reduced the distortion of the retrieved secret share.

ACKNOWLEDGEMENT

I would extend my sincere thanks to my project guide Prof.S.B Waykar ,Assistant Professor Sinhgad Institute Of Technology,Lonavala for his guidance and unending support .I would also like to thank our HOD , ME Coordinator and all other staff members of my college, my parents, friends and all the other individuals for willingly helping me out continuously in the development of my project.

REFERENCES

[1] Kai-Hui Lee,Pei-Ling Chiu, “Digital Image Sharing by Diverse Image Media”IEEE transactions on Information Forensics and Security , vol 9 no 1 ,January 2014 , pp 88-98.

[2] Moni Naor,Adi Shamir “Visual Cryptography”Eurocrypt ,1994,pp1-11.

[3] K. H. Lee and P. L. Chiu, “An extended visual cryptography algorithm for general access structures,” *IEEE Trans. Inf. Forensics Security*, vol 7 no1, Feb. 2012,pp. 219–229,.

[4] K. H. Lee and P. L. Chiu, “Image size invariant visual cryptography for general access structures subject to display quality constraints,” *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 3830–3841, Oct. 2013.

[5] Inkoo Kang,Gonzalo R.Arce,Heung-Kyu Lee “Color Extended Visual Cryptography using error diffusion” , *IEEE Trans. Image Process.*, vol. 20, no. 1, , Jan. 2011,pp. 132–145.

[6] X. Wu, D. Ou, Q. Liang, and W. Sun, “A user-friendly secret image Sharing scheme with reversible steganography based on cellular automata,”*J. Syst. Softw.*, vol. 85, no. 8 , Aug. 2012 pp 1852–1863.

[7] Sadan Ekdemir,XunXunWo , *Digital Halftoning,Project in mputational Science Report,* January 2011,pp1 34.

[8] Natapon Pantuwong , Nopporn Chotikakamthorn,“Alpha Channel Digital Image Watermarking Method”, IEEE ICSP Proceedings,2008,pp 880-883.

[9] Andrew D. Ker “Steganalysis of LSB Matching in Grayscale Images”, IEEE Signal Proceedings,vol 12,no 6,June 2005,pp 441-444.

[10] M.Natarajan , Gayas Makhdumi,”Safegaurding the digital contents:Digital Watermarking” ,DESIDOC Journal of Library and Information Technology ,vol 29 no 3,May 2009,pp. 29-35.

[11] Pradosh Bandyopadhyay,Soumik Das,Atal Chaudhuri,Monalisa Banerjee,”A new Invisible Color Image Watermarking Framework through Alpha Channel”,March 30 2012,pp. 302-308.

[12] Zhongmin Wang,Gonzalo R.Arce,Giovanni Di Crescenzo, “Halftone Visual Cryptography via error diffusion”, *IEEE Trans. Inf. Forensics Security,vol 4no 3,September 2009,pp.383-396.*

[13] Weiqi Luo , Fangjun Huang , Jiwu Huang , *Edge Adaptive Image Steganography Based on LSB Matching Revisited,IEEE Trans Inf. Forensics Security,* vol 5no 2,June 2010,pp 201 214.

[14] A. Nissar and A. H. Mir, *Classification of steganalysis techniques: A study , Digital. Signal Process”’*, vol. 20, no. 6, Dec. 2010,pp. 1758 1770.

[15] P.L.Chiu K.H.Lee,K.W.Peng and S.Y. Cheng,”A new color image sharing scheme with natural shadows,”in Proc.10th WCICA ,Being,China,Jul .2012,pp 4-15.

A Prediction Mobility Scheme in Delay Tolerant Networks

Il-kyu Jeon, Young-jun Oh, Kang-whan Lee

***Abstract*—** In this paper, we propose an algorithm that predicts moving route in selecting relay nodes for efficient routing. In existing Delay Tolerant Networks(DTNs) algorithms based prediction, the node moves with orbit or predict probabilistic movement of node according to the schedules of the node. However, when the networks need prediction of moving schedules of nodes or the orbit of node, these algorithms have a low reliability. To solve this problem, this proposed algorithm predicts moving route of node through context information. This metrics contain the information of node properties for predicted analysis in the change rate of node moving information. To Compared the existing routing algorithm with DTN, the simulation results show that enhancement performance rather than epidemic as like overhead, average latency and average delay time.

***Keywords*—**Delay Tolerant Network, Prediction, Context-awareness, Mobility

I. INTRODUCTION

Delay Tolerant Networks(DTNs) is network architecture designed to enable communication even in an unstable network connectivity between end[1]. Nodes can't be pre-selected for a route because they can't know the information of the entire network. Therefore, the relay node should be selected effective to improve network performance. So the movement prediction scheme[2][3][4] of the node is needed for this. In this paper, we propose a movement prediction algorithm applying change rate of property information about the node. Compared to existing DTN routing algorithm [5] Simulation results show that enhances performance like overhead, average latency.

This work was supported in part by the U.S. Department of Commerce under Grant BS123456 (sponsor and financial support acknowledgment goes here). Paper titles should be written in uppercase and lowercase letters, not all uppercase. Avoid writing long formulas with subscripts in the title; short formulas that identify the elements are fine (e.g., "Nd–Fe–B"). Do not write "(Invited)" in the title. Full names of authors are preferred in the author field, but are not required. Put a space between authors' initials.

F. A. Author is with the National Institute of Standards and Technology, Boulder, CO 80305 USA (corresponding author to provide phone: 303-555-5555; fax: 303-555-5555; e-mail: author@ boulder.nist.gov).

S. B. Author, Jr., was with Rice University, Houston, TX 77005 USA. He is now with the Department of Physics, Colorado State University, Fort Collins, CO 80523 USA (e-mail: author@lamar. colostate.edu).

T. C. Author is with the Electrical Engineering Department, University of Colorado, Boulder, CO 80309 USA, on leave from the National Research Institute for Metals, Tsukuba, Japan (e-mail: author@nrim.go.jp).

II. SYSTEM MODEL AND METHODS

We proposed algorithms use change rate of property information of a node for prediction movement of a node. The first step approximate property information of a node using index. . The property information of a node calculates difference from each index and equation is expressed as follows.

$$A_{ij}(t) = |\, A_i(t) \quad A_j(t)\,| \qquad IA_i(t) = index \min_{\forall j} A_{ij}(t) \tag{1}$$

We proposed Where, A_{ij} (t) is difference of property information between node i and the index j at time t. IA_i (t) is approximation index has minimum A_{ij} (t) for all indexes at time t. The property information of the node is expressed as follows.

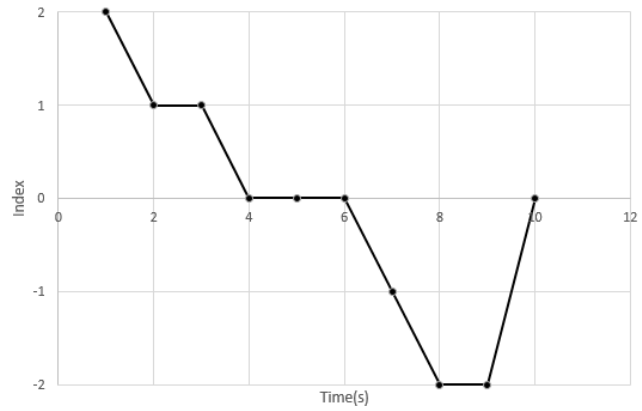
$$A_i = \{n \,|\, n \in Dir_i, V_i\} \tag{2}$$

Where, Dir_i is direction of a node I and V_i is a velocity of a node.

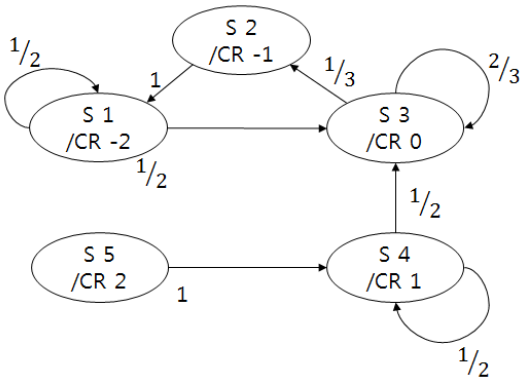
The second step calculates change rate of indexes and form probability matrix P. If the change of rate indexes is calculated, change rate of indexes transfer state. For example, if the number of indexes N sets 2 and change rate of indexes are -2, -1, 0, 1, 2, the change rate of indexes are converted state 1,2,3,4,5. The probability matrix P is defined as follow when the number of index set 2.

$$P = \begin{array}{c} \text{state} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} P_{1,1} & P_{1,2} & P_{1,3} & P_{1,4} & P_{1,5} \\ P_{2,1} & P_{2,2} & P_{2,3} & P_{2,4} & P_{2,5} \\ P_{3,1} & P_{3,2} & P_{3,3} & P_{3,4} & P_{3,5} \\ P_{4,1} & P_{4,2} & P_{4,3} & P_{4,4} & P_{4,5} \\ P_{5,1} & P_{5,2} & P_{5,3} & P_{5,4} & P_{5,5} \end{pmatrix} \end{array}$$

As shown in Fig 1, it is an example of change rate and diagrammed matrix, the number of index is set to 4 and the number of interval times is set to 10. In Fig 1 (a), graph shows change rate of indexes over time. In Fig 1(b), it is a diagram showing a matrix formed by using the change rate of indexes. Where, S is the state and CR is the change rate of the index. If P is 0, it did not mark on the diagram.



(a) An example of Change rate of index



(b) An example of diagramed matrix

Fig. 1 An example of Change rate of index and diagrammed matrix

III. Result

A network overhead and average latency were simulated in order to prove the efficiency of the proposed algorithm. Existing algorithms were used as epidemic routing algorithm based on non-prediction for comparison. Simulation environments are shown in the following Table 1.

Table 1. Simulation Configuration

Parameter	Value
Area(m×m)	3000 x 1500
Velocity (m/s)	2 ~ 20
Transmission range(m)	10, 25, 50, 100, 250
Number of nodes	50
Simulation time (Sec)	100000

From Fig. 2, it shows the average latency that compared Epidemic routing algorithm to proposed algorithm. The average delay time of Epidemic algorithm is better than the proposed algorithm. Although the transmission range could make the environment condition more than complexity(for example, a large transmission coverage etc.), the average latency of the proposed algorithm converged compare to the epidemic.

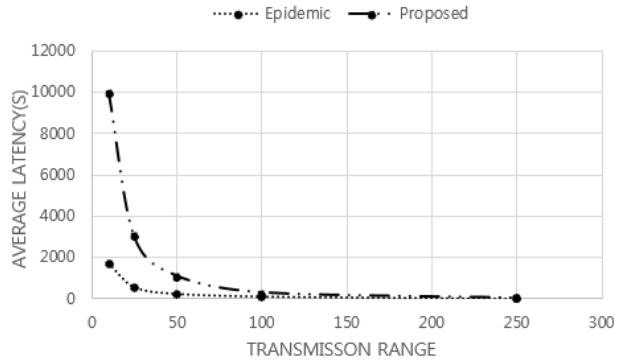


Fig. 2 Average latency and Overhead with 50 nodes

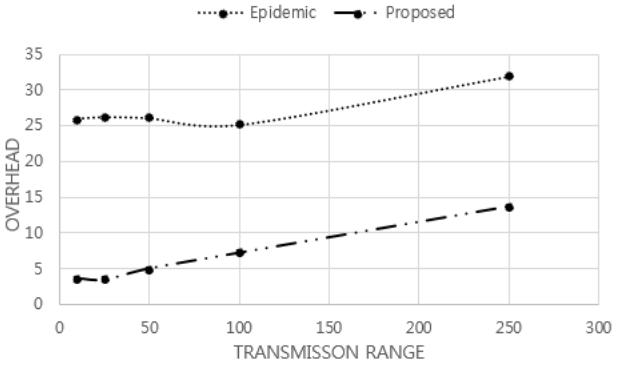


Fig. 3 Average Overhead with 50 nodes

From Fig. 3, it shows the average overhead that compared Epidemic routing algorithm to proposed algorithm. The average overhead increases along transmission range.

IV. Discussion and Conclusions

In this paper, we proposed the movement prediction algorithm at DTNs. The proposed algorithm approximates properties information of node and creates a probability matrix in order to predict movement of a node. As a results of simulation, our scheme decrease network overhead, but average latency is increased. In a later study, it is necessary to prove efficiency by comparing other algorithm based on movement prediction.

From Fig. 2, it shows the average latency and network overhead that compared Epidemic routing algorithm to proposed algorithm. The average delay time of epidemic algorithm is better than the proposed algorithms. Although the transmission range could make the environment condition more than complexity(for example, a large transmission coverage etc.), the average latency of the proposed algorithm converged compare to the epidemic.

Acknowledgment

" This research was supported by the MSIP(Ministry of Science, ICT and Future Planning(2014H1C1A1066391), Korea, under the Specialized Co-operation between industry

and partially supported by the Education and Research Promotion Program of KUT".

REFERENCES

[1] Delay Tolerant Networking Research Group [Internet]. Available: http://www.dtnrg.org
[2] A.oria, and O.Scheln, "Probabilistic routing in intermittently connected networks", the Fourth ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc 2003), 2003.
[3] En Wang, Yongjian Yang, Bing Jia, and Tingting Guo, "The DTN Routing Algorithm Based on Markov Meeting Time Span Prediction Model", Hindawi Publishing Corporation International Journal of Distributed Sensor Networks vol. 2013, 2013
[4] Q. Yuan, I. Cardei, and J. Wu, "Predict and Relay: An Efficient Routing in Disruption-Tolerant Networks," Proc. 10th ACM. Mobile Ad Hoc Networking and Computing (MobiHoc '09), 2009.
[5] A. Vahdat and D. Becker, "Epidemic Routing for Partially connected Ad hoc Networks", Technical Report CS-2000-06, Duke University, 2000.

Il-kyu Jeon He received his B.S degree from Korea University of Technology and Education, Cheonan, Korea, in 2013. He is a M.S degree candidate of Computer Science Engineering at Korea University of Technology and Education, Korea. His main research is Ubiquitous computing, Wireless Sensor Network, Ad-hoc network, International Mobile Telecommunications -Advanced, Wireless SoC.

National Culture and E- Government Services adoption Tunisian case

Allaya Aida
Faculty of Economic and management of Tunis
Tunisia
aida.allaya@yahoo.fr

Mellouli Majdi
Faculty of Economic and management of Sfax
Tunisia
majdi.m@hotmail.fr

Abstract—The main purpose of this paper is to examine empirically national culture impact on E-Government development in Tunisia. We used partial least squares path modeling (PLS) applied to variable in E-government Development Index from the United Nations E-government Survey 2012 and Hofstede's national culture dimension. We found that national culture influences attitudes of citizens to adopt E-Government services.

This paper can help Tunisian government policy and decision makers design and implement policies and strategies to improve E-Government services and their overall development. The study not only provides empirical support and validates the findings of previous research but also updates the results of similar studies in the study field.

Keywords- E-Government service, national culture, ICT (Information and Communications Technology), E-Government development index.

I. INTRODUCTION

E- Government services have increased around the world over the past decade. A recent study presented by United Nations E-Government survey (2010) found that Tunisia was ranked first in Maghreb and Africa and 66th in the world out of 192 countries in terms of "E-Government", thus moving 58 places compared to 2009 when it was ranked 124th.

In its latest report "2012 UN Global E-Government Readiness Survey," index of E-Government development, under United Nations (UN), highlights the degree of application of ICT by governments in order to improve their services. This index takes into account the use of Internet, telecommunications infrastructure and human resources.

Even better, the 9th report on the global information technology and communication 2009-2010 (GITR), published by the World Economic Forum in Davos and the European Institute of Business Administration (INSEAD), has classified Tunisia first in Africa and the Maghreb and 39th globally on a total of 133 countries. GITR is one of the most credible internationally on the impact of ICT on the development process and competitiveness of nation's evaluation reports.

All these distinctions make Tunisia an international industrial and technological destination and that still stands into account international standard.

The development of the E- government services in Tunisia indicates the capacity and the willingness of the public Tunisian sector to deploy ICT (information and communications technology) for improving knowledge and information in the service of the citizen. This development is a function of not only a country's state of readiness but also its technological and telecommunication infrastructure and the level of human resource development. It's widely acknowledged that national culture have a significant effect on consumer behavior and technology diffusion [1].

However how does national culture can be influenced development of E-Government? And what's the influence of national culture on citizen adoption E-Government services?

To answer the research question, we will refer on national cultural dimension presented by [2] and [3]. We used method PLS (Path latest squares) to examine the correlation between national culture and E-Government services.

ISBN: 978-1-61804-318-4

II. LITTERATURE REVIWE

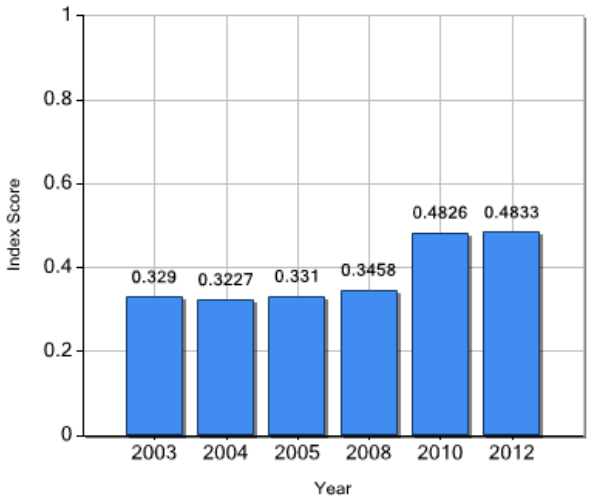
A . E-Government implementation

In the literature there are many definitions of E-government. The World Bank has defined E-Government as the use by government agencies of TIC (area networks, the internet, and mobile computing). It transforms relations with citizens (G2C), businesses (G2B), and other arms of government (G2G).

In this paper, we limit our study by the relation between E-government services to citizen only (G2C). This relation (G2C) explain interactions with business and industry, citizen empowerment through access to information, and transparency government management [4].

Corresponding to the rapid development of E-Government in the world, Tunisian government have presented the importance of providing government services and information via the Internet and world-wide-web to improve the efficiency, cost and quality of the government information and services provided to the public. Table I showed E-government developing index between 2003-2012.

Table I_: Tunisia e-Government Development Index

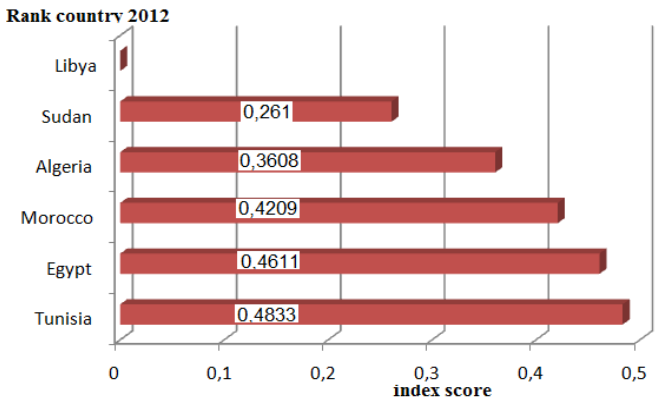


Source : UN Public Administration Program(2012)

In Table_I, we can observe the increasing of E- Government index development in Tunisia. The study of [5] concluded that the degree of E-Government service adoption could be explained by the perceived administrative benefit from adopting E-government services, the political nature of online applications, the government's organizational capacity in adopting new information technology, and the diffusion effect of E-Government service technology.

Additionally, Tunisia is ranked among the first African countries to adopt E_ government Service.

Table II_: Level E-Government Data in Northern Africa



Source UN Public Administration Program (2012)

Table II showed that Tunisian country is classified first in Northern Africa for adoption E-Government services. In spite of rapid globalization there's a difference between countries to adopt E-Government services. Several studies have concluded that national culture is the causal factor.

.

B . National culture and E-Government implementation

[6] define national culture that is a system of shared norms, values and priorities, that taken together, and constitute a design for living a people. Importantly, national culture have learned as previously stade and provides meaning to "how things ought to be done" for persons in a country [6,7]. These shared beliefs are acquired early in life through a person's primary socializing in families, in school [7].

However, National culture is a source of shared norms and behaviors. It influences expectations, preferences, attitudes of public towards e-government. For example, e-government is a new concept can lead to conflicts attitude against the dominant group norm. Hofstede's model of cultural indexes is the most widely used. In empirical study of IBM employees in 40 countries, Hofstede has identified four national culture dimensions:

Power distance: explain the extent to which a society accepts the fact that power in institutions and organizations is unequally disturbed. People in countries have high power distance accepted hierarchical order would have a negative attitude toward implementing and use services of E-government[8].

H1. There is a direct effect between power distance and E-Government services adoption.

Uncertainty avoidance: explain the extent to which members of society feel threatened by unknown situations. However member of society have strong uncertainty avoidance preferred structures situation and they accepted new of ITC [8].

.

ISBN: 978-1-61804-318-4

H2. Ther'a a direct effect between uncertainty avoidance and E-Government services adoption.

Individualism/collectivism: Explain the extent to which individuals are integrated into groups. High collectivism people consider a group as the source of identity. On the other hand individualistic countries used E- Government services more than collectivism people country. Technology in individualistic culture help people to perform time management however it could be concluded that country with high individualistic culture have a positive e attitude to adopt E- Government service[10,11].

H3. There's a direct effect between individualism/collectivism and E-Government adoption .

Masculinity/Femininity: explain the distribution of emotional roles between the genders. Culture with high on feminity, prefer relationships, caring for the weak and the quality of live.in other words, the feminity dimension could have positive attitude toward implantation government services [12].

H4. Ther'es a direct effect between masculinity/ feminity and E-Government adoption.

III. MATERIALS AND METHODS

This research tests the effect of national culture on e-government adoption. We have relied two data sets: the United Nations e-government development index (EGDI) and Hofstede's national culture dimensions.

A- The United Nations e-government development index (EGDI)

The United Nations Department of Economic and Social Affairs has published five surveys on E-Government development in its Member States since 2003. It introduced significant changes to the previously used survey instrument, "focusing more on how governments are using web sites and web portals to deliver public services and expand opportunities for citizens to participate in decision-making".

E-Government Development is measured by the Telecommunication Infrastructure index and the Human Capital index

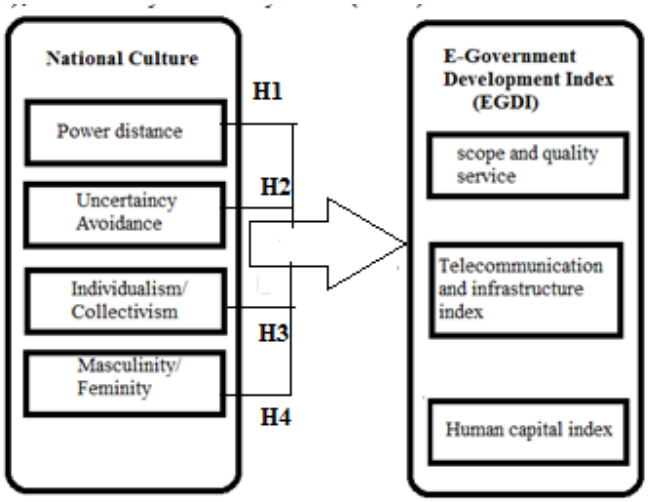
1) **Scope quality of online services: indicate:** Indicates national website and the websites of the ministries of (education, labour, social services, health and finance……) was visited to assign values to survey responses

2) **Telecommunication infrastructure index:** indicates a country's economic and ICT development.

3) **Human capital index:** aggregate level of education

B- National culture

In our statistical analysis, we have used Hofstede's dimensions index and data base UN Public Administration Program (2012). A questionnaire survey was conducted to identify characteristic of Tunisian national culture. The four dimension indexes are power distance index (PDI), uncertainty avoidance index (UAI), individualism-collectivism index (IDV), masculinity/femininity index (MAS).



IV- RESULTS

A PLS (path latest square) model has been proposed to evaluate the effect of national culture in adaptation E-Government services by citizens. The internal reliability of the four constructs (PDI, UAI, IDV, MAS) model was improved by eliminating variables with law factor loadings that were not statistically significant at $\alpha = 0.05$. Table 3 shows results of partial least squares path modelling

TABLE- 3: hypotheses validation of regression

	Beta	T-Statistics	Validation
H1	0,544	6,2943	Not supported
H2	0,369	3,375	Supported
H3	0,247	2,4209	Supported
H4	0,041	1,7386	Not supported

ISBN: 978-1-61804-318-4

V- DISCUSSION

Based on results of Table 3 cultural variables only, H2 and H3 are supported .

Tunisian national culture is characterized by a great power distance [13]. Tunisian people accept hierarchy and this attitude is explained that paternalism is presented current form of resolution problem. Thus in Tunisian family the culture of obedience is important, it's father who have absolute power [15]. Children grow up in atmosphere marked by respect. Despite high power distance in Tunisian culture, citizens adopt E-government services (H1 not supported). This attitude is explained that government have great consideration and responsiveness of the service. So citizen adopt available government service and obey instructions. Government have power and for Tunisian citizen power is right and good [15].

According to [16], Tunisian national culture is characterized by high uncertainty avoidance. Tunisian economic system is bureaucratic so citizen adopts E- Government services because they do not preferred the traditional method face to face due to the long routine. Using E-Government services, citizen become an active participant [17]. Through this interaction, user can control everything. So E-Government services can be appreciated by citizen with high uncertainty avoidance. Government is not the only ones to have the power to decide and instead of the individual who may better suit him. The user can also decide on a multitude of parameters: the information received, the duration of consultation, frequency of visit, site content, etc [18,19].

The individualism defines the relation between member of society. In Tunisia people consider the group as the main source of identity (high collectivism). They judge E-government service like the behavior group. Religion is also considered as one of the main determinant of internet usage in these countries [20]. People in the Arab world find the internet as an approach to break up the limitations of the traditional and social life [21]. Tunisia is an Arab-Muslim country. Family and religion are the origins of social relation. Values and norms of Tunisian society accorded importance to the family, school etc… [22]. On other hand, an individualistic culture would have a positive attitude toward implantation E-government service. It pay more attention to the performance of the individual and E- Government service would be highly regarded and quickly accepted because technology could help individual to perform more efficiently.

Tunisian society is characterized by male values, such as the desire to achieve personal performance [23]. The preference in Tunisian society is for achievement and material success. However it was argued that Tunisian country with high masculinity has a positive attitude toward e government services because these technologies increase the chance of success and support competition, which are the key of masculine culture.

VI- REFERENCES

[1] I. Kazakhstan., NY.M.Middelon " approches to evaluate of websites for public sector services in P. Kommers (ED) in proceeding of the IADIS Internatinal Conference, e-Society, Lisbon, Portugal, pp .279-248, Lisbon IADIS, 2007.

[2] G. Hofstede, " Culture's Consequences: International Differences in Work-Related Values". Beverly Hills CA: Sage Publications, 1980

[3] http://unpan3.un.org/egovkb/ProfileCountry.aspx?ID=175

[4] eGovernment ITU e-Government Implementation Toolkit http://www.itu.int/ITUD/cyb/app/docs/eGovernment%20toolkitFINAL.pdf

[5] J .Gant, "Interview. Director of Human Resources Ministry of Finance - Government of Kazakhstan. Ithaca, NY " *Journal of International Business Studies*, vol. 29, pp, 729-747 ,2007.

[6] G; Hofstede, "Cultures, Organizations: Software of the Mind," London: McGraw-Hill, 1991.

[7] P.L .Berger,., T .Luckmann,. "The social construction of reality: a treatise in the sociology of knowledge" ,1967.

[8] V . Tepestra.David .K 1991" the culture environment of business "3 rd ed cinicinati south-western.

[9] O.Furrer, , B.Shaw-Ching Liu, , D.Sudharshan,. "The relationships between culture and service quality perceptions: Basis for cross-cultural market segmentation and resource allocation ". Journal of Service Research vol.2, pp.355-371, 2000.

[10] J.L Chandon, M.S Chtourou, "Webmarketing, révolution ou évolution ?, Les Cahiers du Numérique ", n°.6, pp15-26 ,2000.

[11] D .Fink. and R.Laupase " Perceptions of Web site design characteristics: a Malaysian/Australian comparison, Internet Research:" Electronic Networking Applications and Policy, 10, 1, 44-55, 2000.

[12] S.J Simon. "A cross cultural analysis of web site design: an empirical study of global web users, Seventh Cross-Cultural Consumer and Business Studies Research Conference, Mexico, 1999.

[13] B .Sackmary, L.M Scalia "Cultural patterns of World Wide Web business sites: A comparison of Mexican and US companies", Seventh Cross-Cultural Consumer and Business Studies Research Conference, Mexico, 1999.

[14] L.Ben Slimane, A.El Akremi, M . Touzani., "Les domaines motivationnels de l'inventaire des valeurs de Schwartz: une analyse confirmatoire", 2ième Journées de la Recherche en Sciences de Gestion, Février 2002 ;

[15] A. Eddakir "Etude de la relation culture nationale-pratiques de management: cas du Maroc" , Actes de colloque AIREPME, Agadir, octobre 2003.

[16] G.Hofstede, "Cultures and Organizations: Software of the Mind".1st edition, McGraw-Hill USA, 1997

[17] D.L Hoffman. , T.P Novak., "Marketing in hypermedia computer mediated environment: conceptual foundations ", *Journal of Marketing*, 60, July, p. 50-68, 1996.

[18] Langar R., "Les préalables culturelles à la réussite de la gestion de la qualité totale ", Mémoire de DEA en Gestion, FSEG de Tunis 1998.

[19] J. Steuer "Defining virtual reality: dimensions determining telepresence", Journal of Communication, 42, 4, 73-93 1992).

[20] J. Deighton, The future of interactive marketing", Harvard Business Review,vol. 74, 6, p.151-161, 1996.

[21] E.Gama, "The Internet in the Arab world: A new space of repression? "The Arab Network for Human Rights Information (http://www.hrinfo.net/en/reports/ net2004).

[22] R Zghal. "Culture sociétale et culture d'entreprise", Les Cahiers de l'ERG, Publications de la FSEG de Sfax, octobre 1992 ;

[23] R. Zghal "Culture et comportement organisationnel, quelques problèmes méthodologiques et résultats de recherche ", Les cahiers du mirs, n°1, février 1991.

[24] Ben Fadhel A., " Dynamique séquentielle : culture-Gestion. Fondements théoriques et analyse empirique du cas tunisien ", Thèse d'Etat en Sciences de Gestion, Université de Nice, 1992.

A Novel Algorithm for Evolution to Cellular Green communications

Jahangir Dadkhah Chimeh
Communication Technology Dept.
Iran Telecommunication research center
Tehran, Iran
Dadkhah@itrc.ac.ir

Abstract—Green communications is a new concept in the communication networks which studies the power consumption and their environment effects. Therefore we should reduce the power of wireless networks somehow that the network preserves its QoS and coverage. Besides, most of the consumed energy in the mobile networks (60%-80%) is pertinent to the radio access network (RAN). Thus this section has an important role in the cellular networks. Besides, there is a relation among power consumption, carried traffic and the coverage area. We study these relations and in addition, algorithms to approach green communications in this paper. This study is based on the number of base stations (BSs) and their configurations or topologies.

Keywords—Green communications, Power consumption, QoS.

I. INTRODUCTION

Today, increasing the power consumption and the environment pollutions are two important problems in the human society. ICT based networks consume 2-10 percent of electric energy and 2% of annual generated CO2 in the world. Green communications is a key method for improving the cellular network and reducing the power consumption in the forward and backbone networks somehow that the network preserves the QoS parameters e.g. data rate, coverage, delay and coverage [1]. To minimize the power consumption we emphasize to the number and location of BSs and their configuration. 2-10 percent of electric energy and 2% of annual generated CO2 in the world are pertinent to ICT networks, e.g., 1% consumed energy in Italy (2 Twh) is pertinent to communication company which has second rank in that country and Vodafone England consumes 50MW energy per day [2]. Renewable energy sources have also an important role in the Green communications which we recommend to use it as well [3]. The present paper has been divided to 5 parts. Parts 2 to 6 has been devoted to power consumption in cellular networks, BS power consumption, temporal-space relation of the carried traffic, power reduction methods in mobile networks and Green communications algorithms. In the end we draw the conclusion.

II. POWER CONSUMPTION IN CELLULAR NETWORKS

There are 5 million antennas sites in the world now which consume 17.5 Gw energy which may be generated by nine 2Gw ordinary power plants or 15 nuclear power plants. These plants produce a pollution of 15 million automobiles emissions or 150000 Paris to New York annual round-trip flight [4]. RAN and core network (CN) are two main parts of the cellular networks. Since the volume of the traffic is related to the connected links to the users, thus the number of ports and users has relation to the consumed power. The core network includes the backbone network and transport, aggregator network and transport and servers. 120000 BTSs are installed annual which give service to 400 million users in the world. A medium size cellular network (about 12-15 thousand cells) consume energy of 170000 homes [5].
Power consumption in the various networks is depicted in Fig. 1a. as it is shown about 60% of power in the cellular networks is consumed in the BSs, about 20% power consumption is in the switch, about 7% power consumption is in the servers and the rest 3% is in the miscellaneous elements. Therefore, power consumption in the BSs is an important factor in the cellular networks. The upper part of the Fig.1b depicts the annual generated CO2 by a user which is about 11.6Kg and the down part of the Fig.1b depicts that the manufacturing or embodied energy is a much larger component in the mobile handset than in the base station [7].

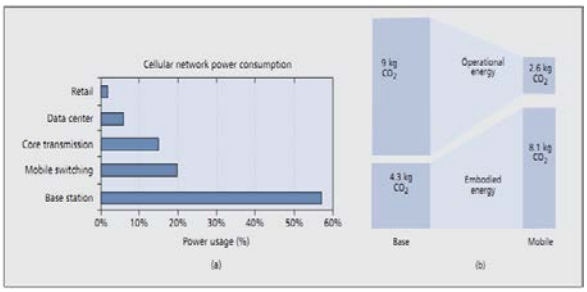


Fig. 1 power consumption and the generated CO2 in a mobile network [1][7]

III. BS POWER CONSUMPTION

Table I shows BS energy consumption. As it is shown GSM consumed energy is more than UMTS/HSPA and totally the consumed energy is diminished versus time and generation. A GSM BS consumes about 800w energy which cause an energy consumption of 8MWh annual which is equivalent to a home consumed energy. A 3G BS consumes about 500w energy which cause an energy consumption of 4.5MWh annual which is lower than a GSM BS.

ISBN: 978-1-61804-318-4

TABLE I. BS CONSUMED ENERGY [1]

	GSM	WCDMA
2007/08	800W	500W
Target (2010)	650W	300W

Fig. 2 depicts architecture of a BS. Main section of a BS is consisted of baseband, feeder and radio unit. 80% of the BS energy is consumed in the radio unit. Besides, 50% of the radio unit energy is consumed in its power amplifier. BS may consist of more than one RRU, BBU and feeder. RRU is radio unit hardware which is assigned to one sector. BBU is responsible for controlling the base band, switch and communicating to RNC. Other peripheral sections include power supply and environment indicators. P_{tx} is the power consumed in BBU, RRU and feeder sections and P_{misc} is power consumed of peripheral elements such as power supply, fans [8].

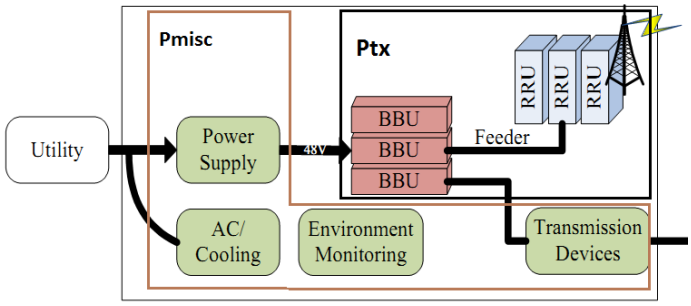


Fig. 2 BS architecture in 3G [8]

Therefore, BS power consumption P is composed of two parts P_{tx} and P_{misc} as $P = P_{tx} + P_{misc}$.
Experimental measurements reveal that we may use a linear models for the power consumed in BS.
As is shown in [8] the more the carried traffic grows, the more the consumed power grows. The shift in lines depends on the vendors and the number of BBUs and RRUs. If we show two main components of P_{tx}, i.e., RRU and BBU as P_a and P_b we have $P_{tx} = L.P_\alpha + P_\beta$.
where L is the traffic load factor. When the traffic is heavier, RRU should consume more power for more active links. On the other hand, BS performs baseband processes for all the BS frequencies carriers and the number of active links doesn't affect that processes and the consumed power depends on the number of frequency carriers. Besides, as it is shown in the sleep mode BS consumes a fixed energy and in active mode it increases due to the increase of the carried traffic [9].

IV. TEMPORAL-SPECAIL DEPENDENCY OF THE TRAFFIC

Traffic characteristics own fast statistical variations which depend on the user behavior in the different times. Traffic contains regular periods during a week. The traffic volume is low at nights and conversely it is high during a day. Besides, the traffic has low amplitude at the weekends [10].
Fig. 3 illustrates the traffic distribution in a city. Fig3a reveals that %30 of sectors convey %80 of traffic while %70 of sectors only convey %20 of the traffic. Fig. 3b reveals that %10 of sectors carry more than 1750 Mbytes during a day. Vodafone has declared that only %5 of its sites has been utilized %95.

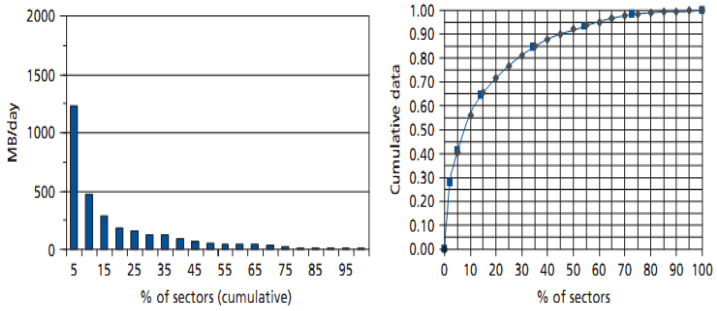


Fig.3 traffic distribution versus sectors in a metropolitan city [10]

V. POWER DIMINISHING IN THE CELLULAR NETWORKS

To reduce the power consumption in a cellular network we may divide it to 5 following categories [11]:

A. *Data centers in backhaul*

Today the number of subscribers is high. Beside that some of the services are very complex and thus the number of data centers in backhaul is also increasing. To realize the green communication in this situation we should consider the following two items:

- Since the consumed energy is a function of the traffic load pattern, thus data centers resources should be assigned or released according to the dynamic of requests, i.e., energy is not consumed if there isn't any request.
- Virtualization is a technology which is based on the simultaneously use of two or more OSs on a unique server somehow that make more exploitation of hardware and software and other resources. Indeed, we use more than one virtual machine, i.e. OS, to make machines, i.e. servers, work more efficient. Here, a physical machine gives services to some OSs by the name of virtual machine. Most of OSs uses %10 of hardware capability, thus virtualization allows using some different OSs on a unique system somehow that the hardware utilization improves.

B. *Macrocells*

Since about %60 of energy of a cellular network is consumed in the macrocells, this part of the network has high importance. To achieve to a good power reduction we consider three following components:

- BSs include schedulers which operate according to the traffic load dynamic. BSs may be turned on and off according to the dynamic of traffic load which makes a suitable power saving in the network.
- Cell zooming strategy is to define a cell size. Cell zooming is a function of traffic load, channel conditions, data rates and

the traffic demands. It is a leverage for the tradeoff between power saving and call blocking probability.

- About %50 of energy in the radio units consumes in the power amplifier (PA). Today, power in 3G networks changes linearly as a function of the bandwith and signal quality which requires a high quality amplifier.

C. Femtocells

Femtocells are used in the HSPA, WiMax and LTE networks today. Femtocells are small and low power sites which are connected to the users from the forward and to a wideband link from the backward. Since femtocells are small cells, they consume a very small energy from one hand and improves the network capacity from the other hand. The femtocell radius size is about 10m. Femtocells are cells with low power and small BSs. To consider green communications in femtocells, we should consider two following tasks:

- Power control in femtocells which is very important and mekes the intra-cell interference enough low.
- Interference between femtocells and macrocells which is an important problem in current mobile networks and should be considered more in this area.

D. End hosts

End hosts embrace ordinary mobile stations, smart phones, PCs and tablets which are very effective in the power consumption of mobile networks. To apply the green communications in this area we consider three following components:

- End hosts energy profile which is composed of all requests of the hosts for the energy consumption which is a function of the traffic pattern and end user behavior. This energy profile may be reduced by %35 through adaptive regulation of the end hosts sleep timers.
- Today, end users are equipped with some different radios such as bluetooth, WiFi, 3G which each one consumes a specific energy. For example data transfer by bluetooth radio consumes lower energy than by WiFi and data transfer by WiFi radio consumes lower energy than by 3G. Thus a suitable use of each radio makes the energy consumption to be reduced.
- Transmission mode is the most consumed mode of a host since it transmits/receives data in this mode. Thus, this mode should be designed very well. End hosts have DTX/DRX modes which causes a good energy saving in the host.

E. Services and applications

Recent services require more signaling exchange rather than the previous services. Besides, new applications also generate more traffic than the previous applications. Thus we should provide the better QoS for those. The important aim of the mobile networks is to present services like video and voice calls, online gaming, VoIP, web browsing, etc. Since the generated traffic has now grown much more than the previous traffic (up to 26 times rather than the traffic in 2010) thus the energy efficiency is very important now. Here, the three following objects are important to consider:

- We should consider the software architecture of services from the energy saving point of view when it is designing. Besides, we should consider good data compression, adaptive codecs, etc. to improve the QoS.
- We should use the predictive and learner algorithms. These algorithms cause more energy saving specially in mobile gaming.
- We should use cache servers which improves system performance, specially with a heavy generated traffic. A caching proxy server causes the response time of the requested services to be reduced by dispatching the stored contents of the previous alike requests of the same users or the other users. This server saves a copy of the repeated requested data and transmits those upon requesting those data which causes a cost, bandwidth and energy saving. Besides, it causes that the whole of the contents is transferred to the mobile in a short time.

VI. GREEN METHODS/ALGORITHMS

We categorize these algorithms as follows:

A. Site surveying

This method is based on the following three steps:
First step: we select a BS, then if any of the sectors in this BS is covered completely by any adjacent cell we candidate that selected BS to be turned off.
Second step: if the coverage isn't complete we may change the tilt of the antenna of adjacent cell to cover the selected cell completely. Besides, we care of the unwanted lobes of that antenna.
Third step: if the selected BS is not the last, we return to the first step and select next BS [12].

B. Analysis the minimum energy [2] [12]

Assuming N users in a region, we select M BSs somehow that the energy consumption is minimum. We constitute the following equation:

$$\begin{aligned} &minimize \quad (P_c + P_t) \\ &subject\ to\ QoS\ requirement \\ &power\ constraint \\ &\textstyle\sum_{j=1}^{M} P_{ij}/L_{ij} \geq P \\ &P_{ij} \leq P_{up} \end{aligned} \tag{1}$$

in which Pc is consumed power in BSs, Pup is maximum power of BS antenna in downlink, P_{ij} is radiated power of BS_i to user BS_j and L_{ij} is the path loss of BS_i to user BS_j.

C. Grid cell method

We grid the cell network into some grid cells and partition all equivalent BSs in each grid cell according to the following equations:

$$r_i + d(i,j) \leq R_j, \qquad r_j + d(i,j) \leq R_i \tag{2}$$

d(i,j) is the distance between BS_i to user BS_j. r_i and r_j are the normal coverage radiuses in the network. R_i and R_j are

maximum coverage radiuses which may be 200m-1km and 1km-5km in urban and suburban respectively. Two adjacent cells are equivalent if their ranges satisfy in the above equations. Equivalent cells are grouped in a distinct grid cell and all cells that are provided in one grid cell may be transformed to one cell (Fig. 4).

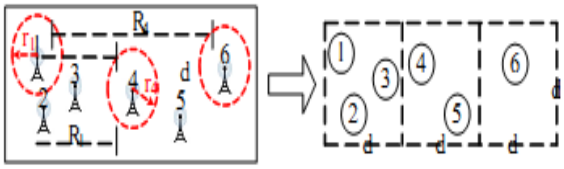


Fig. 4 Grid cell method for site surveying [8]

We start from one edge, e.g., west north and go to another edge, e.g., east west. Afterwards, we divide the whole of the region to some grid cells somehow that all the equivalent BSs will be resolved in each grid cell. For example in Fig. 4 three grid cells have been constituted and BSs 1, 2, 3 has been joint the left grid cell, BSs 4, 5 have been joint to the middle grid cell and BS 6 has been joint to the right grid cell. We need to BSs coordinates and coverage range in this method. BS coordinates may be found by GPS and the BS coverage area which depends on the type, location and power of antenna and terrain of the area may be found by the site surveying.

D. Introduced method

We use this method in the urban network which carries heavy traffic, however, we may use it in sub-urban and rural areas. The traffic volume is low at nights and high during the day. The introduced method has been accomplished during the following two distinct phases:

- Phase1or design phase, i.e., when we want to design the network and define the location and size of BSs ranges. Totally, design phase is accomplished based on the coverage region, the volume of the predicted traffic and the terrain.
- Phase 2 or site surveying phase, i.e., when the network is active.

We may apply our introduced methods in both of the above phases. We use hierarchical/umbrella network design in both phases. Cellular networks are composed of macrocells, microcells, picocells and femtocells. Macrocells are used for overall coverage and the other cells are used for providing enough capacity (Fig. 5).

Macrocells, microcells, picocells and femtocells have up to 20km, 2km, 100m and 10m radius respectively. Since a macrocell may cover up to 20/2=10 microcells we may inactivate up to 10 microcell in this method, a microcell may cover up to 2/0.1=20 picocells we may inactivate up to 20 picocell, and picocell may cover up to 100/10=10 femtocells we may inactivate up to 10 femtocells in this method.

To provide a green network, in this method, if we are in the design phase, i.e., if we want to establish a new network, designer should implement the macrocell first, then to carry the additional traffic he should utilizes the microcell and other sub-cells in the next steps. The aforementioned method has priority over the previous methods since when the network has low traffic, operator is capable to turn off the sub-cells easily without affecting on the other cells. If we are in the second phase, i.e., when the network is operating we may implementing new umbrella coverage in the network, i.e., we first establishes a new macrocell in the networks somehow that covers some microcells and other sub-cells. Then, we may turn off the sub-cells when the traffic is low.

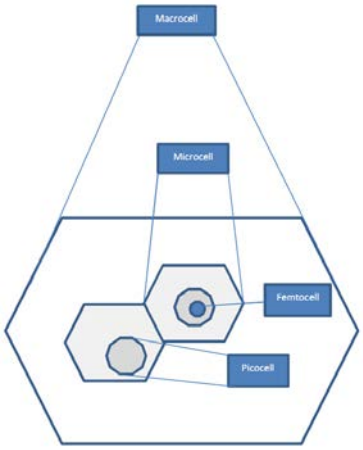


Fig. 5 Hierarchical/Umbrella network

VII. CONCLUSION

Green communications is a great necessity for the present networks. We may approach it by various methods. To utilize it the green equipment is one the approaches. The next method is changing the cell antenna characteristics somehow that makes us capable to tune the adjacent cell coverage. This method may create new blind spots and inter-cell interferences. The other approach is categorizing the network to some grid cells and transferring the equivalent BSs to one of them. Designing and utilizing new energy efficient services and applications is the other approach. Using the new hierarchical cell planning is also an effective approach. Therefore, to provide a green communications network we may select a distinct region firstly, e.g. a city, then survey it and provide the traffic characteristics as a function of time and location secondly, find the equivalent BSs and maintain only one of them.

REFERENCES

[1] Pei-Jung Chun, Green Radio-The case for more efficient cellular base stations, Mobile VCE, 2014.
[2] A. Baumgartnet, T. Bauschert, Greening cellular radio networks: A numerical method for selection of detachable base stations in low traffic scenarios, TIWDC, 2013.
[3] P. Nema et all, Pre-feasibility study of PV-Solar/Wind hybrid energy system for GSM type mobile telephony base station in central India, ICCAE, 2010.
[4] Alberto Conte, GreenTouch and Green Wireless Networks, Alcatel-Lucent Bell Labs, 2012.
[5] C. Lange et al., Energy Consumption of telecommunication networks, ECOC, 2009, Austria.
[6] C. Lange et al., Energy Consumption of telecommunication networks and related improvement options, IEEE JOURNAL OF SELECTED TOPICS IN QUANTUM ELECTRONICS, VOL. 17, NO. 2, MARCH/APRIL 2011.
[7] C. Han et al., Green Radio: Radio Techniques to Enable Energy-Efficient Wireless Networks, IEEE Communications Magazine, May 2011

[8] Chuny Peng et al., Traffic-Driven Power Saving in Operational 3G Cellular Networks, MobiCom, 2011.
[9] Jozif Lorincx et al., Measurements and Modeling of Base Station Power Consumption under Real Traffic Loads, Sonsors, 2012.
[10] O. Blume et al., Energy Savings in Mobile Networks Based on Adaptation to Traffic Statistics, Bell Labs Technical Journal 15(2), 77–94 (2010) © 2010 Alcatel-Lucent. Published by Wiley Periodicals, Inc.
[11] X. Wang, A Survey of Green Mobile Networks Opportunities and Challenges, Mobile Netw Appl, Springer, 2011.
[12] P. Gonzales et al., Base Station Location Optimization for Minimal Energy Consumption in Wireless Networks, IEEE VTC Spring, 2011.

New Framework for Constructing a Virtual Routing Table in the IGP Networks

Radwan Abujassar,
School of Computer Faculty, Bursa Orhangazi University,
Turkey, Bursa
Radwan.abujassar@bou.edu.tr

Abstract

The interaction between layer2 and layer3 can be generated a new protocol with high technique for detecting failure through layer2, and then layer3 make each node on the primary path find an adjacent node to re-route traffics through it in case failure. Our protocol consist some constrains on the adjacent node, which is capable for avoiding such problems could be happened on the network. Hence, all adjacent nodes will re-route the traffics through a disjoint path with primary path to the destination. The algorithms in our protocol based on Dijkstra algorithms and increment edge procedure. The aims of our protocol are decreasing end-to-end delay, reduce recovery time, and avoid congestion and loss of packets.

Keywords: OSPF, OA&M, Recovery, ECMP.

I. Introduction

Open Shortest Path First (OSPF) is used a dynamic link state protocol for TCP/IP traffic. The OSPF distribute a link state advertisement for construct the routing table, and uses the Dijkstra algorithms to compute the best shortest path. The traffic will pass from source to destination through the shortest best path after computed by the Dijkstra algorithm. There are many techniques have been published for improving recovery mechanism with guarantee loop free on the network. Our protocol does not need to encapsulate packets twice and return them to the source. The backup path or alternative path has improved the network reliable and display a significant role when the primary path is down. In section Π we discuss some related works about the re-route mechanism and then in section III will describe our proposed algorithm. In section IV we will show a flowchart for describing and in section V we explain more about the proposed algorithm and finally our future work and conclusion will be in section VI.

II. Related work.

Efficient routing protocol algorithms have been built for achieving the robustness and fast convergence within a short time, in case failure. In [3], the author indicates about the cost of links in the network and traffic engineering. As we mention above, the links cost consider one of the important parameters to determine the best path through the routing protocol algorithm. The OSPF protocol based on the dijkstra algorithm to compute the shortest path; the minimum path cost will determine by compare it with other candidate path. The packet will re-route from the backup path through the routing protocol, the problem here when the backup path pass other traffic

then the load will become high and the congestion lead to drop the packet. Therefore, the traffic engineer is coming to solve this problem by allocate the traffic through the equal cost path with less utilization.

The OSPF optimize multiple path protocol OSPF (OMP) emphasis to solve the load on the path [4][13], and achieve optimal distribution load balance in the network in case failure. However, there are two drawbacks of OMP mechanism. First, it needs more memory size to store vectors. Second: the link load message information will generate without any deterministic and unpredictable. However, if we assume there is more than one path has the similar cost with primary shortest one, then we can shift and divide the traffic from all ECMP to decrees the utilization on the primary path, and the load balancing will be achieved. In addition, the ECMP will avoid the loop in the network.

In our mechanism, we measure the load balancing is measured by [10][11]:

$$\text{Load Link Metric} = \frac{\text{traffic size}}{\text{link capacity} * \text{Time period}};\ \text{Link Cost (utilization)} = \text{link cost} * w * \text{utilization}$$

There are two kinds of the dijkstra algorithm; first: dijkstra algorithm to compute the best path by removing the links with bandwidth less than the threshold. Second, on demand dijkstra algorithm, which is generate the shortest path tree to the pre-computation mode, and depend on the bandwidth request the node will be added in the tree[5].

The IP recovery emphasise on two cases: first, the time to detect the failure. Second the time to compute the shortest path. Hence, we will discuss about some techniques that contribute to improve the recovery time by achieving a desirable result. Failure Insensitive technique is one of the efficient techniques in the IP recovery. When the failure occurs, FIR mechanism will inform the source node about failure through encapsulate the packet (encapsulate in encapsulate) with a special header and return back the packet to the source, the source node will notice about the failure through the new header, then will send packet to another path, which is disjoint with the primary path [12]. FIR mechanism will avoid loop in the network, but the recovery time will be not desirable. The drawback in this technique, when the packet will encapsulate two times that will lead to consume the bandwidth, delay and congestion in the network [16][18]. On the other hand, IPFRR is an applicable technique, it includes the LFA, U-turn and not-via address [8][9]. The draw back in IPFRR technique is the loop free not guaranteed. In addition, not-via address need to encapsulate/de-capsulate, which effects

on network performance as we indicate above [14].

III. Our Proposition

The optimization between layer 2 and layer 3 in our protocol will make the network more reliable through divided the roles between them. In our protocol, we will focus on layer 2 for detecting failure regarding to the minimum time it will be taken, which is a few millisecond. In addition, layer 2 will detect failure in two ways as below:

- *Hardware: loss of lights, loss of signal*
- *Software: CCMs, LBMs*

The detection in hardware is faster than software. In OA&M Ethernet, there are many features have been supplementary such as CCMs, LBMs to reduce detection time, delimitation the location of failures and lock all nodes temporary to send the traffic through that path who is affected from failure until it goes up. On the other hand, layer 3 through our protocol will construct additional routing table to store the alternative path and keep it prepared when a failure occurs.

In our protocol we will assume that the all nodes on the primary connect with an adjacent node has achieved all our protocol constrains as follow:

1. *The adjacent router should not be affected by the failure (it is not on the primary path between source and destination).*
2. *The adjacent node must have a path to the destination, which is disjoint with primary path.*
3. *When link failure, each node is connected directly with that link will know about the failure through layer 2.*
4. *The new path should have an enough capacity to tolerate additional packets from the other node in the case of failure.*
5. *The delay for each link in the topology must be less or equal to the delay on the primary path.*

Each node on the primary path asks all adjacent nodes if there is one validity path to the destination with regarding to the algorithm constrains (discussed above), the validity path will accumulate in the routing table as a backup.

IV. Algorithm Flowchart

We summarized the mechanism of our protocol in a flowchart (*see figure-1).* However, the mechanism of our protocol based path protection through an adjacent node. The preliminary step is to route the traffic demand based on efficient routing protocol with some modifications on the dijkstra algorithm, and also identify the primary path. Hereafter, each node on the primary path select an adjacent node to identify a set of candidate path to destination with consider all constrains. Next step is to calculate the best and shortest path, and mark it as a primary path between source and destination. After the primary path has been selected, the increment edge procedure will adjoin the total weight $\sum W0$ to the primary path. In that case, the adjacent node will eliminate that path from all calculations to determine a backup path to destination.

ISBN: 978-1-61804-318-4

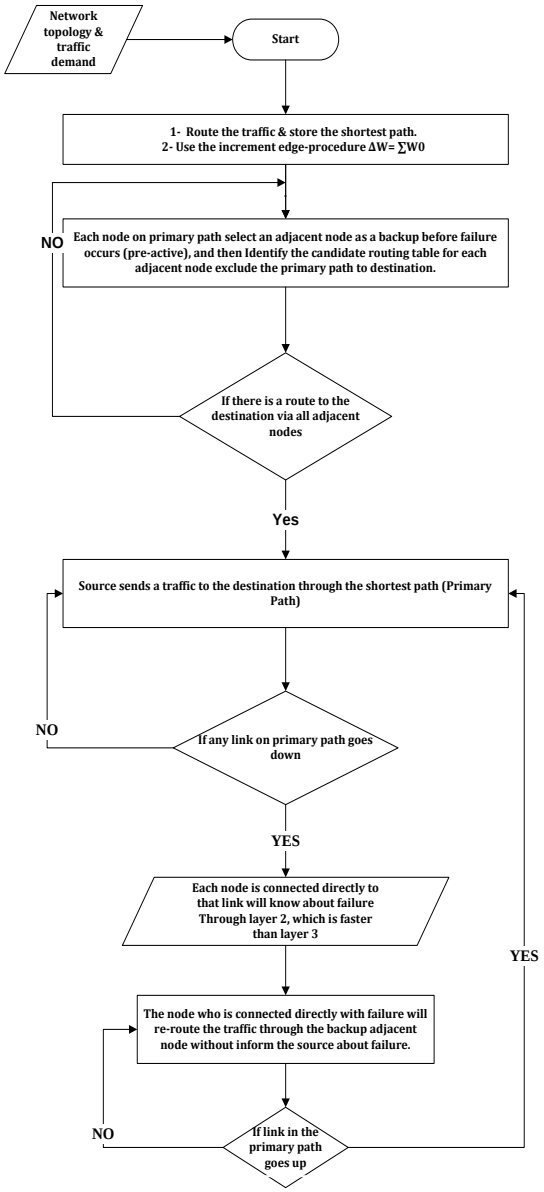


Figure-1 Flow chart of new protocol based path protection.

We can apply our protocol in two cases:

1- Equal Cost Multiple Path
2- Not Equal Cost Multiple Path

In ECMP, our protocol does not require to use the increment edge procedure, so each node will ask only for validity path to destination without change the cost for the primary path. In ECMP, loop freeness can be defined as below:

Defention-1*, when source sends packets via any next-hop nodes towards the destination the packet for no reason returns back to the source.*

There are three conditions have been proposed for loop freeness in the network:

(1) C1(v; d) < C1(s; d)
(2) Cj(s; d) + c(s;NHj(s; d)) < C1(s; d)
(3) Cj(s; d) = C1(s; d)

In condition-1, the source asks its adjacent node for best path better than its own path. Condition-2 the source will find an alternate path with loop free without questioning his neighbour. On the other hand, condition-3 it's for ECMP topology.

V. Proposed Algorithm

Figure 2.1 explain how new our technique works in case failure. Each node on the primary path connects with more than one adjacent node, it will select one of them as a backup, and how each adjacent node construct an additional routing table to accumulate the backup path on it to destination. Each router on the primary path is able to re-route traffic through a provision routing table in the adjacent node. In the link state protocol, dijkstra algorithm computes the shortest path between source and destination through edges weight. When the weights of edges become heavier that will reduce the property to be consisting them in the backup path. We perform a simple topology, which acts a good example to discuss our protocol. In this topology, each node has at least two neighbours that can re-route packets through either one of them when failure occurs. There are many paths between S→D as follow:

{{S→1→3... →D}, {S→1→4→.... →D}, {S→1→5→7→... →D}}

The total weights for all the topology ∑W0=X;

X: Total weights for all arcs on the topology.

Then: ΔW=∑W0=X;
If we assume the shortest path between S→D after the dijkstra algorithm has computed it as follow:

$$\{P_{S\to D}\ \{S\to 1\to 5\to 7\to \ldots..\ \to D\}\}$$

Hence, our new technique will add the total cost to the primary path that will lead to exclude all edges on the primary path to be chosen in the backup path as follow:

Step 1: in case fig.2.1, the primary shortest path {Ps→D {S→1→5→7→.... →D}} is calculated, and it is selected as the first acceptable path.

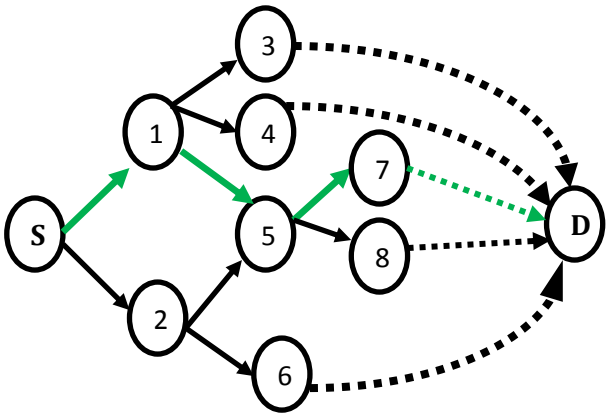


Figure 2.1 each node has at least two adjacent node

Step 2: In case fig.2.2, every edge on P has its weight increased by ΔW= W0 = X+ W0. Then, we re-invoke the backup path in the provision routing table for re-routing the traffic from node which is connected directly with link failure. *Node 5 will detect failure through layer2 and then re-route traffic through Node 8 without informing S about it.*

{ Ps→D {S→1→5→8→.... →D}}

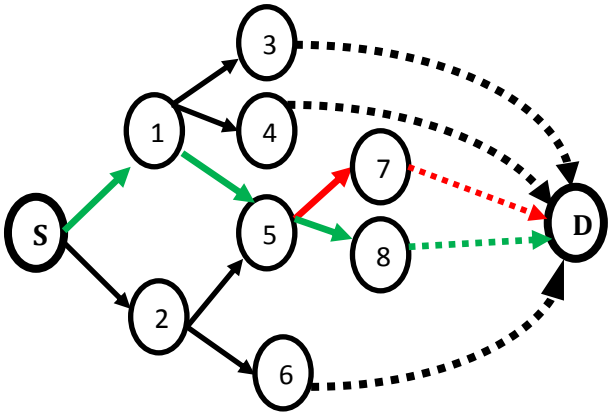


Figure 2.2 Node 5 re-route the traffic through disjoint path with primary path.

Figure 2.3 illustrates when failure occurs at different locations then all nodes on the primary path can re-route the traffic without any additional calculations

$$\{P_{S\to D}\{S\to 1\to 3\to \ldots..\ \to D\}\}$$

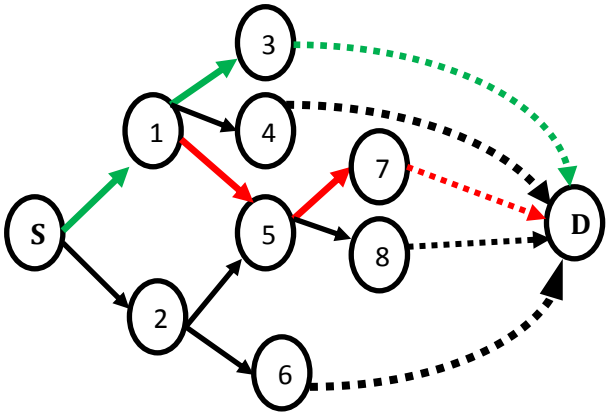


Figure 2.3 Node 1 re-route the traffic through an adjacent Node 3 without informing the source node.

In figure-3 show the provision routing table for each adjacent node to the node on primary path.Example-2 illustrates how the adjacent node constructs an additional routing table for the destination.

D	S	X

Node S: Construct candidate routing table For S to the D
Route via {s→>1 or 2}

D	1	N

Node 1: Construct candidate routing table for neighbour-1 to the destination.
Route via {s→1→3 or 4 or 5→D} depend on the cost and which is non on the primary path

D	2	Y

ISBN: 978-1-61804-318-4

Node 2: Construct candidate routing table for neighbour-2 to the destination.
Route via{s→5 or 6→D}

D	3	Z

Node 3: Construct candidate routing table for neighbour-2 to the destination.
Route via {s→1→3→N or Y→D}

I	D	NH	C

1- *D: Destination*
2- *NH: Next Hope*
3- *C: Cost.*

VI. Conclusion and Future work

In our future work, we will divide it into two parts: first, we will continue implement our mechanism to get a desirable result and compare my work with other works. The first implementation concerns on backup and re-route packet through an adjacent node as we discuss above. The implementation will show results about the recovery time and delay between end-to-end points. Hereafter, we will prove the mechanism through doing some modelling to prove it. Second part; we will concerns about layer-2 with OAM messages to make the detection failure faster in wired network. However, the implementation will be considering implementing a trigger to inform layer-3 about failure. The trigger will be as a cross layer, which is between layer 2 and layer 3.

References

[1] Aman Shaikh, Albert Greenberg, and OSPF monitoring: architecture, design and deployment experience, p.5-5, March 29-31, 2004, San Francisco, California

[2] Shaikh and A. Greenberg, "Experience in black-box OSPF measurement," in Proc. ACM SIGCOMM Internet Measurement Workshop (IMW), Nov. 2001, pp. 113–125.

[3] Shaikh, A. & Greenberg, A. (May 2002), "Optimizing OSPF/IS–IS Weights in a Changing World", *IEEE*, vol. 20, no. pp. 4.

[4] Banerjee, D. Sidhu," Path Computation for Traffic Engineering in MPLS Networks", Proceedings of IEEE ICN 2001.

[5] W. Paper, "Ethernet Operations, Administration, and Maintenance," Service Management, 2007, pp. 1-15.

[6] Hiroshi Ohta. "Standardization status on carrier class Ethernet OAM. IEICE Transactions on Communications", E89-B (3):644–650, March 2006.

[7] Aoyama and C. Technology, "A New Generation Network: Beyond the Internet and NGN," IEEE Communications Magazine, 2009, pp. 82-87.

[8] D'Amboise, J. (2009), "40 Gigabit Ethernet And 100 Gigabit Ethernet: The Development Of A Flexible Architecture ", IEEE Communications Magazine, Mar, pp.10-13.

[9] Ohta, H. (2008), "Standardization Status of Carrier-Class Ethernet", vol. 6, no. 2, pp.

[10] Grover, Wayne, Mesh Based Survivability Networks, Prentice Hall, and ISBN-10: 0-13-494576-X.

[11] W.D. Grover, S.M. Iee, D. Stamatelakis, and M. Iee, "Cycle-Oriented Distributed Reconfiguration:" Cycle, 1998, pp. 537 - 543.

[12] Network Simulator NS2, http://www.isi.edu/nsnam/ns.

[13] Sally Floyd and Van Jacobson." Random Early Detection gateways for Congestion Avoidance". IEEE/ACM Transactions on Networking, Vol.1 (4):pp. 397–413, August 1993.

[14] O. Wittner, "Link and Node Protection using Hamiltonian P-Cycles found by Ant-like Agents," Science and Technology, 2004, pp. 1-12.

[15] M. Shand, S. Brayant; 2008; IP Fast Reroute Framework; draft-ietf-rtgwg-ipfrr-framework-08.txt; [online] Available from: http//www.ietf.org.

[16] Shaikh, A. Goyal, M. & Rajan, R. (2002), "An OSPF Topology Server Design and Evaluation", vol. 733.

[17] L. S. Buriol et al., "A Memetic Algorithm for OSPF routing," Proc. INFORMS Telecom, 2002, pp. 187–88.

[18] M. Ericsson, M. Resende, and P. Pardalos, "A Genetic Algorithm for the Weight Setting Problem in OSPF Routing," J. Combin. Optim, vol. 6, no. 3, 2002, pp. 299–333.

An Implementation of Adaptive Multipath Routing Algorithm for congestion control

N.Krishna Chaitanya, Research Scholar, ECE Department, JNTUCE, KAKINADA

S.Varadarajan, Professor, ECE Department, S V University College of Engineering, TIRUPATI

Abstract:- ***This paper proposes a better adaptive multi path routing technique for routing the data packets effectively from source to destination under congestion at a router. In traditional adaptive multi path routing techniques, if congestion occurred at a router then the route is changed from source to destination. In a single path routing algorithm, all the data packets transmitted through a single path, where the time taken to transmit the packets is more. This drawback is eliminated by using multi path routing technique, where the packets are transmitted through different paths. The proposed method provides a better solution for minimizing the congestion by rerouting the data packets over other paths, which are not utilized by the same in multi-path routing. This method avoids the unnecessary dropping of packets at a congested router and improves the network performance.***

Keywords: Congestion, multi-path routing, Packets, router

I. INTRODUCTION

Most of the routing techniques in a network are based on a single path. As the number of data packets transferring increases, the data traffic increases in the network, as a result congestion will occur. To avoid this, multi-path routing [5] is preferred. In multi-path routing, the total available data is split and transferred among several paths.

Many multi-path routing protocol techniques have been proposed in networks. Some of the multi path routing techniques are Simultaneous Multi Path Communication (SMPC) [1], and Distribution and Congestion Minimized Multipath (DCMM) routing [2]. The existing methods are used to reduce the congestion in multipath routing.

In multi-path routing, still there is every possibility of occurrence of congestion. This paper proposes a method to avoid the congestion occurring in multipath routing. It reduces the unnecessary retransmissions and delay for data packets, which will effect on the performance of the network. In order to avoid congestion, multi path routing along with load balancing is used [3, 4].

The rest of the paper organized as follows. Section 2 provides the overview on the existing multipath routing techniques. In Section 3, we introduce the proposed Adaptive Multipath Routing for Congestion Control (AMR-CC). The Section 4 discusses the flow chart used in this method. Section 5 presents the simulation results, and section 6 concludes the paper.

II. EXISTING METHODS

One of the multipath routing techniques is Simultaneous Multi-Path Communication [3]. There are two types of SMPC's available; they are (i) SMPC-I & (ii) SMPC-P. Here both the techniques are based on bandwidth control.

In SPMC-I [1], all paths for communication are treated equally. The bandwidths of each path are controlled independently. In this technique, it is possible to control the bandwidth for each path with no information of any other path.

In SMPC-P [1], the priority will be given to the paths that are used for data transfer. If the total communication bandwidth used for data transfer is greater than the available bandwidth, it uses priority control scheme. In this, the communication bandwidth is controlled by decreasing the bandwidth of one of the paths in ascending order of priority level among the paths having a lower priority.

In these methods, still there exists a problem because of reducing the transmission bandwidth in the network, which will increase the data transmission delay and reduces the network performance.

Another existing method is Distribution and Congestion Minimized Multipath (DCMM) routing. Here, in this method, the routing decisions minimize network congestion, routing decisions address link congestion avoidance topology and maximum flow optimization. Here, number of paths in multi path routing is reduced, and as a result, it is unable to minimize the congestion [5].

Because of the limitations in the available techniques, a new technique has been proposed to minimize the network congestion and to improve network performance.

III. PROPOSED METHOD

A method was proposed with an algorithm, flow chart and presented by Chaitanya, N. Krishna, S. Varadarajan, and P. Sreenivasulu[6]. In this method the drawbacks in existing methods of multipath routing are eliminated. In this, Adaptive Multipath Routing for Congestion Control (AMR-CC), multiple paths are chosen and the load is distributed among the paths. Here, all the paths may not have the same capacity and capability. However, the load is equally distributed, because of the insufficient resources at a router in a path, but there is a possibility of congestion at that router. This can be minimized by this method. If congestion occurred at a router then it verifies the status of its neighbours. If any one of the neighbours is available as free then the congested router forwards the data packets to that router. This will be able to avoid the unnecessary retransmissions from the sender and reduces the overall transmission delay of data packets. As a result, the performance of the network increases. The delay to transfer the packet from source to destination depends on message size, transmission speed, link propagation delay, per hop processing delay and number of hops. General formula for total delay calculation based on the above parameters is indicated as

$$Total\ delay = Propagation\ time + transmission\ time + Processing\ time + store\ and\ forward\ time$$

Let us consider a network with N nodes that are connected by L links. This situation is indicated as

$$N = \{n1, n2, n3\ldots\}$$

$$L = \{l1, l2, l3\ldots\}$$

Based on these, a number of paths are chooses from source to destination in order to transfer the data packets. By using distance vector routing algorithm, a number of paths are calculated. These are paths are arranged on delay. The paths are

$$P = \{p1, p2, p3\ldots\}$$

Where, $p1 < p2 < p3\ldots$

The major problem with the single path (p_k) routing is, finding the single path from sender to destination by reducing the total network congestion. This is also going to affect on the links of selected route. To overcome this problem multipath routing is preferred. In this method, all the paths are well utilized; as a result, its performance also increases. But the drawback is, if any node or link fails, then the data is lost and is retransmitted by sender. To avoid this data lost and retransmission, here proposing an algorithm for re-routing of data packets through its neighbours under link or router failure cases in order to avoid the congestion in the path. Alternate path is chosen through its neighbour by sending a request. If a neighbour is accepts the request then the packets are routed it. Thereby, reducing unnecessary discarding and retransmissions.

Algorithm for the proposed method is

Algorithm
Step1: finding the best paths from source to destination P = {p1, p2, p3…pi}, where 1< i < n
Step2: Sort all the paths based on their performance metric delay.
Step3: Now choose a subset of paths from P P' ⊆ P
Step4: Distribute the load based on traffic in the network.
Step5: if congestion occurs request (neighbours j)
Step6: if request 'accepted' re-route the data through the path.
Step7: calculate the delay
Step8: evaluate its performance

IV Simulation Results and Discussions

Proposed algorithm is simulated by using NS-2 with version ns-2.35. Simulation results are compared for single path, multi path without congestion control, multipath with congestion control and adaptive multipath with congestion control. The proposed method works quite well and its performance is compared with all other techniques. For simulation we considered the following topology.

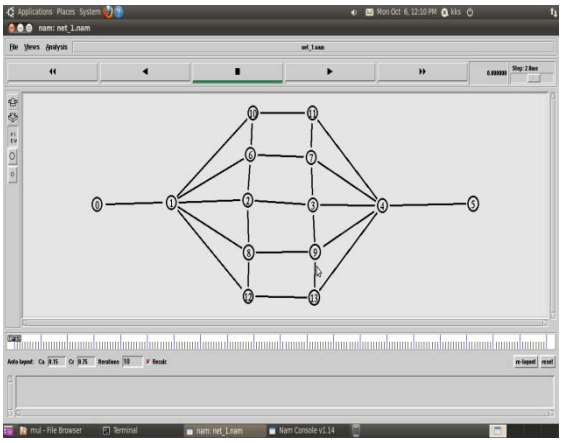


Figure 1: Example topology taken for simulation

From figure1, number of routers taken source to destination are 12. Assuming that all the channels have the same capacity and all the routers has the same capacity.

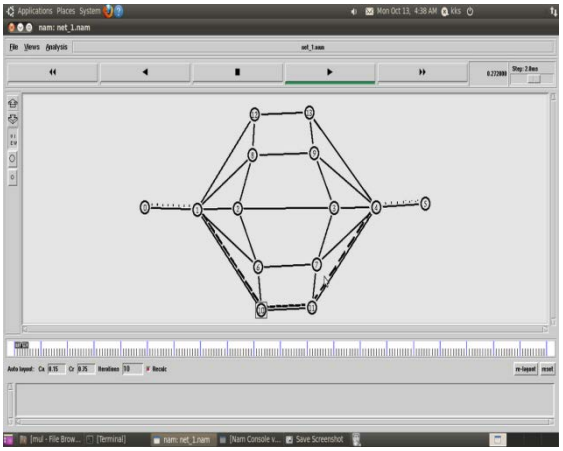


Figure 2: Single path packets transfer

First we analyzed the packet transmission through single path, where the time required to transfer the data is more. It is a basic method for routing of packets, but it is least preferred method now a days.

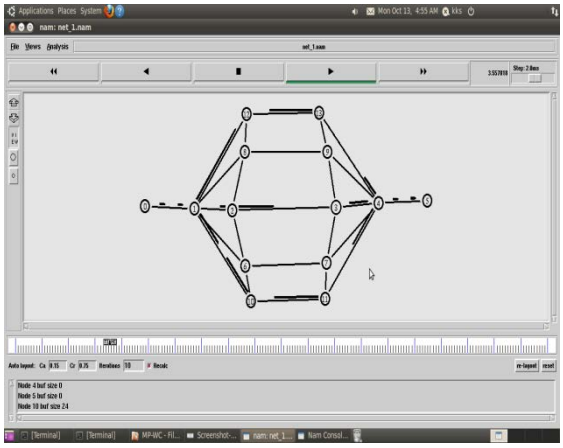


Figure 3: Multipath routing without congestion control

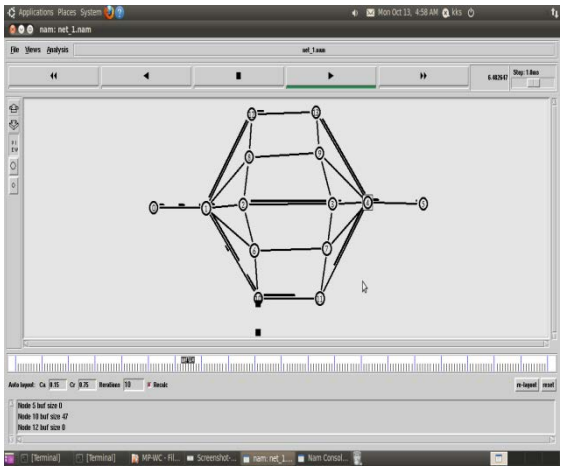


Figure 4: Packets dropping at a congested router in multipath routing with congestion control

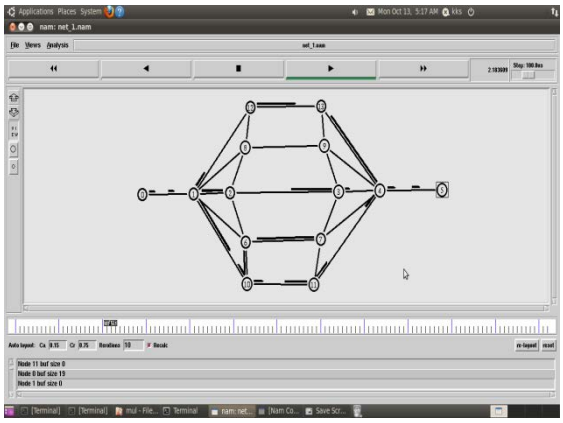


Figure5: Proposed method

ISBN: 978-1-61804-318-4

Table 1: Comparison of packet transmission techniques

Technique	packets sent	packets received	packets dron	packet delivery fraction (%)	average end-end delay (s)
Single path	**1000**	**891**	**109**	**89.1**	**1.09**
Multi path without congestion	**1000**	**1000**	**0**	**100**	**0.75**
Multi path with congestion	**1000**	**931**	**69**	**93.1**	**1.01**
Proposed method	**1000**	**1000**	**0**	**100**	**0.159**

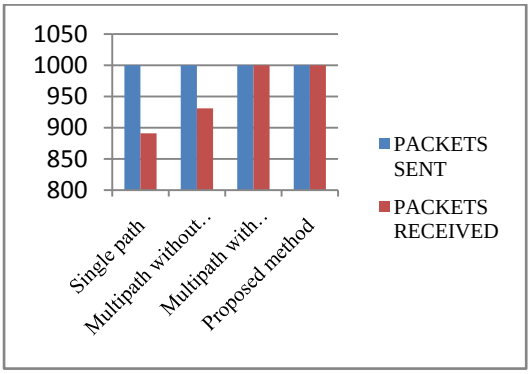


Figure 6: Packets sent versus packets received

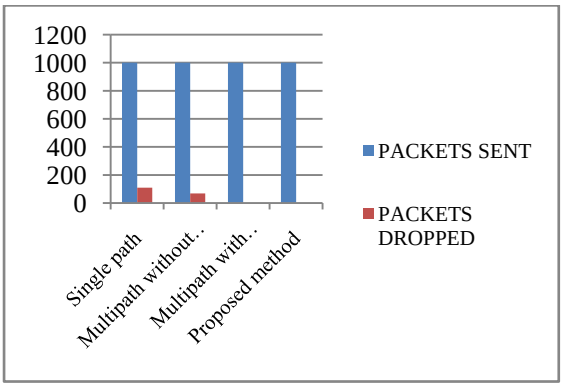


Figure 7: Packets sent versus packets dropped

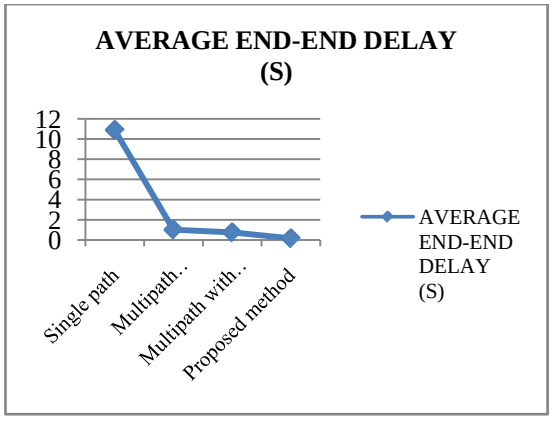


Figure 8: comparison of average end to end delay for various techniques

It is very clear from the diagram that, the proposed method out performs well compared to the existing methods in basic multi path routing techniques. As the table1 shows the number of packets sent and received for various methods are compared. In our proposed method, number of packets dropped is zero. In addition, the delay for transferring the packets in case of congestion is reduced because of adaptive multipath routing.

V CONCLUSIONS

In this paper, we simulated and verified packet delays, packet delivery rates by using NS-2. Simulation results are observed for different methods such as single path data transfer, multipath data transmission with and without congestion control. From the results we analyzed that the delay is deceased and the packet delivery rate is increased compared with traditional routing techniques. This method is preferred to achieve to more packet delivery rate and reduced delay.

REFERENCES

[1] Akiji Tanaka “Effects of length and number of paths on simultaneous multi-path communication” IEEE Society, IEEE International symposium on applications and internet, pp 214-217, 2011.

[2] Guo Xin, Zhang Jun, Zhang Tao, “A Distributed Multipath Routing Algorithm to minimize Congestion”, IEEE Journals and magazines, pp 7.B.2-1 – 2-8, 2009.

[3] Peter Key, Laurent Massoulie and Don Towsley “Multipath Routing, Congestion Control And Dynamic Load Balancing” IEEE Proceedings, pp 21-24, 2009.

[4] Soundararajan.S, R.S. Bhuvaneswaran “Adaptive Multipath Routing for Load Balancing in Mobile Ad Hoc Networks”, Journal of Computer Science, pp 648-655, 2012.

[5] Banner R, Orda Ariel, Multipath Routing Algorithms for Congestion minimization, IEEE/ACM Transactions on Networking, Vol.15, No.2, pp 413-424, 2007.

[6] Chaitanya, N. Krishna, S. Varadarajan, and P. Sreenivasulu. "Adaptive multi-path routing for congestion control." In *Advance Computing Conference (IACC), 2014 IEEE International*, pp. 189-192. IEEE, 2014.

A privacy protection and anti-spam model for network users

Yuqiang Zhang
School of Software Engineering
Beijing University of Technology
Beijing 100124, China
yuqzhang@emails.bjut.edu.cn

Jingsha He
School of Software Engineering
Beijing University of Technology
Beijing 100124, China
jhe_bjut@163.com

Jing Xu
School of Software Engineering
Beijing University of Technology
Beijing 100124, China
hxj@emails.bjut.edu.cn

Bin Zhao
School of Software Engineering
Beijing University of Technology
Beijing 100124, China
hejs2004@163.com

Abstract—

In recently network interaction, some sensitive information of users needs to inform the interactive party in order to ensure smooth interaction. Especially the e-mail address which network users commonly used must be provided to the interactive party nearly in all interactive processes. In this situation, many unsafe factors in network lead to leakage of user e-mail address easily, and cause a lot of e-mail problem which puzzle user frequently, affect mailbox's normal use, junk mail problem which need to be solved urgently is getting more and more serious. To solve the above problem, this paper present a new privacy protection model, this model keeps user's email address as a secret information, through changing the interactive pattern, not only fundamentally prevents the email leakage problem and protects user sensitive information, but also solves the junk mail problem from the source.

Keywords- Network interaction; privacy protection;e-mail address; spam

Ⅰ. INTRODUCTION

Along with the network development, the network interactive pattern also becomes more and more complex; people are often forced to supply their privacy or sensitive information to the network interactive party to begin an interaction. There are already plenty of techniques and tools to search and obtain user's online information, user private information submitted online are easily leaked to third party, this will bring hidden safe problem to user. Once the information is obtained by the malicious side, will bring the severely injure or even the loss which will be unable to recall to the user. With the fast development in internet application, users' privacy protection question arouses people's universal interest; simultaneously the user also set a higher request to the individual privacy information's initiative domination.

The privacy has three kind of basic shapes: Individual private affair, individual information, and individual domain. Individual information which uses in the network environment is one kind of privacy shape; it includes a very extremely wide scope, all the individual data of user contains in individual information.

The mailbox address as the individual information is one kind of the user’s privacy information, in current network interaction, users’ e-mail address must be provided to the interactive party to begin an interaction.

Many unsafe factors in network environment, such as the leakage in network transmission or interactive party negligence in the management, users' online information is accessed, stored, data mined, shared, manipulated, bought and sold, analyzed, stolen or misused by countless

* This work is supported by Beijing Natural Science Foundation (Grant No. 4142008) ， National Natural Science Foundation of China (Grant No. 61272500) ， National High-tech R&D Program (863 Program) (Grant No. 2015AA017204). Shandong Natural Science Foundation (Grant No. ZR2013FQ024).

corporate without users knowledge or consent. In these information, once commonly used mailbox address is divulged, has created junk mail being in flood in the network, brings the massive Trojan Horse, creates the subscriber's premises network information security problem, brings the puzzle which for the network user gets rid with difficulty.

In this paper, we describe a new method for the protection of user privacy; the method is based on changing the pattern of internet interactive to protect users` e-mail address, uses a sub-mailbox address code to replace the commonly used mailbox address, the sub-mailbox address has the date of expiry and user can manage it flexibly. In this case, this model also has anti-spam function.

The rest of paper is organized as follow. In the next section, we review some background information about the pattern of interaction between network users use their e-mail address in current network interactive, and describe the problem such as spam because of the network user's email address leakage, and also summarize the existing solution technology as well as its existence deficiency. In section Ⅲ we describe our new privacy protection model in details, contains architecture and so an. In section Ⅳ we use an example to illustrate how our privacy protection model works. We conclude this paper in section Ⅴ which we also present our future works.

Ⅱ BACKGROUD INFORMATION

When users visit the website, the first time interacting with service site, they often need to register personal information to request website to provide the services or visit the Web site for more content. In these log-on messages, User's mailbox address information is very important one item[1]. The email address be used to transmit the activated information about users new account number to user by website, or If the user and serves website to achieve a transaction, the email address will be used to receive service information which provide by website. This facilitates the interaction between users and Web sites the mailbox address has also become alternately essential tool in network interaction. But present's interactive pattern is users just can provide their commonly used mailbox address directly for serves website, only then can the website send the service information Customized by user to the user` mailbox.[2-3]

Not only the user's email address for a Web site used, but also all the communication party, such as their classmates and friends or other interactive website use the same email address to communicate with user. Due to many technical and administrative reasons, there are some unsafe factors in the network, User the real mailbox address which fills in each kind of service website is very easily gained by the malicious side. Once the user commonly used mailbox address reveals, the junk mail question is also following which influence user using commonly used mailbox normally[4]. Users can not change this status that commonly used email address has been obtained by other independent party, and have no way to let their email address information be secret again, also cannot afford to reject the spam by their selves, can only place hopes in the anti-spam technology on which the service provider adopts, or no longer uses this mailbox address to apply for a new one directly. These two methods have the flaw and the insufficiency, cannot fundamentally solve the problem from user angle.

The first method about the technology that adopted by Service providers is the mainstream method of anti-spam. Filtration technology is its major technique[5-6], the filtering technology distinguished from the role including MAT filtration technology, MDA filtration technology and MIJA filtration technology; the filtering technology distinguished from the method are the key character-based filtering technology, based on the white list filtering, blacklist-based filtering technology, reverse DNS query technology,

rule-based filtering, content-based filtering technologies and other mail filtering technology. To some extent, these technologies effectively inhibited the spam[7], but spammers are constantly updated its anti-filtering technology to deceive filters, and these techniques is to control the spam received in the destination, cannot stop spam from the source.

If user adopt another method that do not use the leakage email address again, and apply a new one. This method seems simple and feasible, but When the user re-visit the website or the network interactions, the same problem will appear again, the new email address have the same probability to be stolen by malicious party[8]. Second, the process of apply a new mailbox address is easily, but after that, Because users no longer use the old email address, so all the user`s classmates, friends, commence partner or other interactive website who communicate with user use the old mailbox before can no longer contact with user use the old email address again. Only to inform their new application's mailbox address to all good friends and correspondence partners one by one, and re-establish the new address book, users can use the new mailbox for normal daily communication. This work not only consuming time, very tedious, and often lead to some important communication party loses contact[9].

Therefore, in order to fundamentally solve the problem of the proliferation of spam, the most effective way is to put the user's email address as privacy information to protect, not leaking to any network interaction side. Of course, to make sure that the network interaction can run smoothly, the users mailbox be protected in secret are meaningful. But the method about anti-spam which research from this source is very few now.

This paper use a new method of privacy protection, regarding user's mailbox address as privacy information to effectively protect, using the e-mail address code substitute user's mailbox address and using technical measures to ensure that only the user and a particular interaction party can communicate by this e-mail address code. The users can freely flexible control the e-mail address code and independently distinct junk mail. Using this new model, users can easily find out the spam sender, effectively prevent from receiving spam and report the spammers promptly.

III. The New Model of Privacy Protection

In this section, we briefly introduce the working pattern of our new model different from the former network interactive, and present the key method be used in the new model. And then mainly describe the entity architecture and workflow about the new privacy model.

A. Working Pattern and Method

Be different with the traditional way, when users interacting with websites in our new model, users does not need to tell the interactive website their commonly used e-mail address directly, use a special and flexible e-mail address code replace the commonly used e-mail address. through this e-mail address code, the website who interact with the user can send the service information to user`s commonly used e-mail box. Using this method not only protect the user's email address and complete the whole process of interactive services.

Email address code is user-centered design, with temporary, management flexibility and other characteristics, sub-mailbox is the mailbox used by the user generated, user can replace the email address used to send and receive mail. Email address code is user-centered design, with temporary, management flexibility and other characteristics. The email address code is generated by the user`s commonly used mailbox, it can replace the commonly used mailbox to send and receive emails. But it has a special-purpose characteristic, can be used as a temporary email address, and it`s period of validity can be stetted freely by user, After email

ISBN: 978-1-61804-318-4

address code expiration, the user commonly used mailbox address can be used normally.

The working pattern show in figure 1

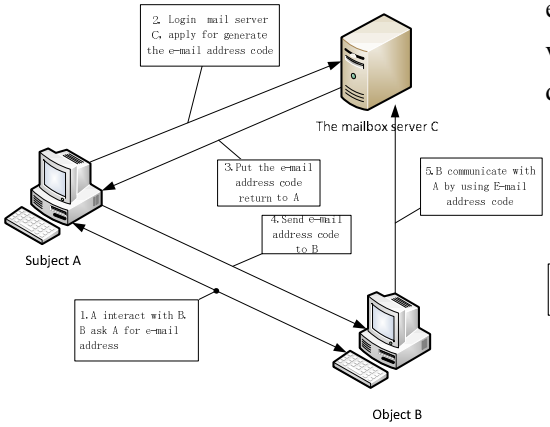


Figure 1 the working pattern of the model

B. Interactive Entities

This model has three main interaction entities: subject A - user, objects B - user interaction party and the mailbox server C.

Subject A: The owner of privacy information, which sponsor the interactive process.

Object B: The user`s interaction partner who requires users to provide the e-mail address to complete the interactive services.

The mailbox server C: A server provides users with mail services and as the mailbox addresses privacy information providers for user.

User's mailbox address as privacy information share with server, but object B is the mailbox address information irrelevant third party. As long as guaranteeing the interactive process complete smoothly, the user is unnecessary provide the real mailbox address to object B,

C. Main Archtecture

This main structure of the model contains four parts: generation, storage, management and verification module (as in figure 2). Generation modules based on users apply information to generate the e-mail address code. Storage module store e-mail address code and relevant information. Users through the management module to manage the e-mail address code, for example, modify the validity period, open or close the e-mail address code, or cancel the e-mail address code etc. verification module can verify the interactive party ID information, to decide receiving or reject the message.

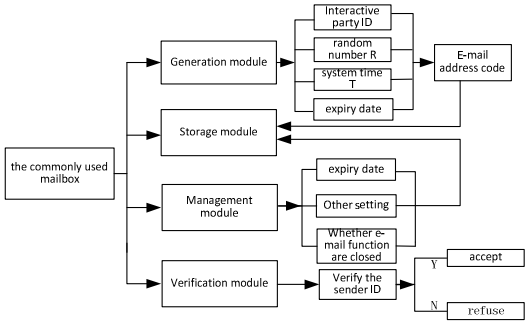


Figure 2 the main archtecture of the model

Generation module (producer): This module including the user hands in the application in the commonly used mailbox. It generates the e-mail address code based on the information such as ID number of the interactive party, random number R ， system time T and expiry date which is set depending on the situation about the interaction with the service website.

Storage module (store): This module is used to store all the information of the e-mail address code after the e-mail address code and the related information have produced, and makes the new e-mail address code binding with the user`s commonly used mailbox, Enables the information which transmits through thee-mail address code to be possible received by the user`s commonly used mailbox smoothly.

Management module (manager): this module is established for the user to manage the e-mail address code information. Through this module, the user may revise the e-mail address code information parameter according to different situation interacting with each website. Especially, when the user receives the junk mail, may adjust the sub-mailbox function parameter and the interactive pattern with the website through the administration module according to the special details.

ISBN: 978-1-61804-318-4

Verification module (verifier): This module is used to verify the ID information of the email sender who sends the email through the e-mail address code to the user. If the sender`s ID information is the same with the ID information of interactive party which be stored during the e-mail address code generate, the email will be accepted. Otherwise, the email will be rejected.

D. Workflow of the Model

Implementation of the model took the following technical solutions, base on the anti-spam privacy protection methods. this method entire frame including the user`s commonly used mailbox, the mailbox server and other people or website interactive with user. the user`s commonly used mailbox is subject A, other people or website interactive with user is object B, the mailbox server is C. The method includes the following steps to achieve process:

1) When browses a service site object B and needs object B to provide the services, Subject A musts to fill out the registration information, and needs to supply the commonly used e-mail address in order to complete the interaction smoothly.

In our new privacy protection model, subject A will do something follow and finish the registration.

2) Subject A login the mail server C and apply for a e-mail address code. according to object B 's email address and random number and the current time of system, using algorithms SHA-1 to generate the message digest, and using algorithms RSA to digital signature, mail server C generate the e-mail address code.

Simultaneously this mailbox address code will be bind with the commonly used mailbox of subject A by mail server C. In order to facilitate subject A to management the e-mail address code based on the situation in the process of the interaction with object B, all the information of the e-mail address code will be stored in the subject A`s commonly used mailbox.

3) Subject A obtains the mailbox address code which is generated by mail server C based on the information supplied by subject A, and can set the information of the e-mail address code initially: Establishment e-mail address code date of expiry, the short name for object B, the key words of Interactive service and other security parameters.

4) Subject A uses the new application E-mail address code to replace the commonly used mailbox address in registration information providing to object B, this e-mail address code is used just only for the interaction between A and B.

5) Object B use the e-mail address code interact with subject A. the letter transmits through the e-mail address is received by user`s commonly used mailbox which generate this e-mail address code, and subject A send the letter to object B also through the e-mail address code. In the object B receiver terminal, the received message shows that the sender address is e-mail address code not the subject A`s commonly used e-mail real address. Of course, to subject A the e-mail address code is transparent in the process of communications.

6) Because in the process of e-mail address code production contains ID information of object B, the mail service C can verify the ID information of the message sender who sends the message through the e-mail address code, if the ID information is consistent with the ID information of interactive party which be stored during the e-mail address code generate, the message will be received. Otherwise, the message will be rejected.

Ⅳ CONCLUSION

The merits of our privacy protection model and method are as follows:

◆ The real address information of user`s commonly used mailbox will not divulge for any other interactive sides. Users do not need to replace the mailbox frequently because of the mailbox address divulging question. The

absolute safety of user`s commonly used mailbox address can guarantee normal communication between user and its good friends.

◆ Users can determine which messages are spam freely depending on their own preferences, and can set parameters of e-mail address code flexible depending on the Specific interactive situation.

◆ Users can clearly determine the source of the mailbox problem such as the spam mail generated, and take appropriate measures flexibly.

◆ In this model , the mailbox address code management is flexible, users may momentarily reduce, lengthen its date of expiry, or open, close, cancel its receiving and dispatching mail function and so on according to their own need.

This mailbox address privacy protection method fundamentally solved the problem that user`s mailbox address is often obtained by the irrelevant side in the network causing a lot of mailbox problems to trouble the user. Because mailbox address's divulging creates the mailbox question has existed and urgently waits to be solved, but did not have a very good solution now, this article in view of these questions, proposed this privacy protection model. This privacy protection model is good at detecting and preventing spam and protecting the security and confidentiality of the user`s commonly used email address as user`s sensitive and privacy information from source， has a very good use value and practical significance. In the later research, the privacy protection model we proposed in this paper needs to be further refined, the performance and functionality of the application of this model need to be in-depth analyzed.

REFERENCES

[1] Lai.G.H, Chen.C.M, Laih.C.S and Chen.T"A collaborativeanti-spam system", Expert Systems with Applications, v 36, n 3PART 2, p 6645-6653, April 2009

[2] Liu.Y.Q, Cen.R.W, Zhang.M, Shao.P.M and Ru.LY"IdentifyingWeb Spam with User Behavior Analysis", 4th InternationalWorkshop on Adversarial Information Retrieval on the Web(AIRWeb 2008), April 22April 22, 2008

[3] T. Burghardt, E. Buchmann, J. M¨uller, and K. B¨ohm, "Understanding user preferences and awareness: Privacy mechanisms in location-based services," In OnTheMove Conferences (OTM), 2009.

[4] F. Xu, K.P. Chow, J.S. He, X. Wu, "Privacy Reference Monitor –A Computer Model for Law Compliant Privacy Protection," Proc. The 15th International Conference on Parallel and Distributed Systems (ICPADS'09), ShenZhen,China, Dec.8-11, 2009.

[5] D. Warren and L. Brandeis, "The Right to Privacy," Harvard LawRev., vol. 45, 1890.

[6] Marsono.M.N, El.K,, M.W and Gebali.F"A spam rejection schemeduring SMTP sessions based on layer-3 e-mail classification",Journal of Network and Computer Applications, v 32, n 1, pp.236-257, January 2009

[7] Junejo.K.N; Karim.A: PSSF"A Novel Statistical Approach forPersonalized Service-side Spam Filtering, Proc. theIEEE/WIC/ACM International Conference on Web Intelligence,2-5 November, 2007

[8] Shawkat.A, A.B.M., Yang.X" Spam Classification Using AdaptiveBoosting Algorithm", Proc. 6th IEEE/ACIS InternationalConference on Computer and Information Science, ICIS 2007;

[9] Li.K, Zhong.Z.Y, Ramaswamy.L"Privacy-Aware CollaborativeSpam Filtering", IEEE Transactions on Parallel and DistributedSystems, v 20, n 5, p 725-739, 2009

Authors Index

ISBN: 978-1-61804-318-4

ISBN: 978-1-61804-318-4