



ICWMC 2013

The Ninth International Conference on Wireless and Mobile Communications

ISBN: 978-1-61208-284-4

July 21 - 26, 2013

Nice, France

ICWMC 2013 Editors

Dragana Krstic, University of Nis, Serbia

Robert Bestak, CVUT in Prague, Czech Republic

ICWMC 2013

Foreword

The Ninth International Conference on Wireless and Mobile Communications (ICWMC 2013), held between July 21 and July 26, 2013 in Nice, France, continued the previous events on advanced wireless technologies, wireless networking, and wireless applications.

ICWMC 2013 addressed wireless related topics concerning integration of latest technological advances to realize mobile and ubiquitous service environments for advanced applications and services in wireless networks. Mobility and wireless, special services and lessons learnt from particular deployment complemented the traditional wireless topics.

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

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

We hope that ICWMC 2013 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the field of wireless and mobile communications.

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

ICWMC 2013 Chairs:

ICWMC Advisory Committee

Dragana Krstic, University of Nis, Serbia
Magnus Jonsson, Halmstad University, Sweden

ICWMC Industry/Research Chairs

Yan Zhang, IMEC- NL, The Netherlands
Christopher Nguyen, Intel Corp., USA

ICWMC Special Area Chairs

Security

Georg Neugebauer, RWTH Aachen University, Germany

ICWMC 2013

Committee

ICWMC Advisory Committee

Dragana Krstic, University of Nis, Serbia
Magnus Jonsson, Halmstad University, Sweden

ICWMC Industry/Research Chairs

Yan Zhang, IMEC- NL, The Netherlands
Christopher Nguyen, Intel Corp., USA

ICWMC Special Area Chairs

Security

Georg Neugebauer, RWTH Aachen University, Germany

ICWMC 2013 Technical Program Committee

Jemal Abawajy, Deakin University - Victoria, Australia
Mohammed Abdel-Hafez, UAE University-Al-Ain, United Arab Emirates
Seyed Reza Abdollahi, Brunel University - London, UK
Javier M. Aguiar Pérez, Universidad de Valladolid, Spain
Mari Carmen Aguayo-Torres, Universidad de Malaga, Spain
Chang-Jun Ahn, Chiba University, Japan
Ahmed Akl, LAAS/CNRS - Toulouse, France
Hamed Al-Raweshidy, Brunel University - Uxbridge, UK
Erick Amador, Intel Mobile Communications - Sophia Antipolis, France
Karine Amis, Institut Mines-Telecom/Telecom Bretagne, France
Radu Arsinte, Technical University of Cluj-Napoca, Romania
Mohammad M. Banat, Jordan University of Science and Technology, Jordan
Alessandro Bazzi, IEIT-CNR, University of Bologna, Italy
Norman C. Beaulieu, University of Alberta, Canada
Mounef Benmimoune, Université du Québec à Trois-Rivières, Canada
Ezio Biglieri, Universitat Pompeu Fabra - Barcelona, Spain
Alireza Borhani, University of Agder - Grimstad, Norway
David Boyle, Imperial College London, UK
Maurizio Bozzi, University of Pavia, Italy
Maria Calderon, University Carlos III of Madrid, Spain
Juan-Carlos Cano, Universidad Politécnica de Valencia, Spain
Pedro Castillejo Parrilla, Universidad Politécnica de Madrid, Spain
Sammy Chan, City University of Hong Kong, Hong Kong
Ajit Chaturvedi, Indian Institute of Technology Kanpur, India

Abdellah Chehri, University of Ottawa, Canada
Hsing-Lung Chen, National Taiwan University of Science and Technology - Taipei, Taiwan
Yunfei Chen, University of Warwick - Coventry, UK
Char-Dir Chung, National Taiwan University, Taiwan, R.O.C.
Silviu Ciochina, Universitatea Politehnica din Bucuresti, Romania
Hugo Coll Ferri, Polytechnic University of Valencia, Spain
Ana Collado, Centre Tecnologic de Telecomunicacions de Catalunya (CTTC) - Barcelona, Spain
Nicolae Crisan, Technical University of Cluj-Napoca, Romania
Danco Davcev, University "Ss Cyril and Methodius" - Skopje, Macedonia
Claudio de Castro Monteiro, Federal Institute of Education, Science and Technology of Tocantins, Brazil
Javier Del Ser Lorente, TECNALIA-Telecom - Zamudio (Bizkaia), Spain
Karim Djouani, Pretoria, South Africa / University Paris Est-Creteil (UPEC), France
Trung Q. Duong, Blekinge Institute of Technology, Sweden
Péter Ekler, Budapest University of Technology and Economics, Hungary
Karim El Defrawy, Hughes Research Labs (HRL) - Malibu, USA
Wael M. El-Medany, University of Bahrain, Bahrain
Ghaïs El Zein, IETR - INSA Rennes, France
Armando Ferro Vázquez, Universidad del País Vasco / Euskal Herriko Unibertsitatea - Bilbao, Spain
Ana-Belén García-Hernando, Universidad Politécnica de Madrid, Spain
Sorin Georgescu, Ericsson Research, Canada
Apostolos Georgiadis, Centre Tecnologic de Telecomunicacions de Catalunya (CTTC) – Barcelona, Spain
Nawel Gharbi, University of Sciences and Technology, USTHB, Algeria
Mikael Gidlund, ABB, Sweden
Lim Wee Gin, University of Nottingham Malaysia Campus, Malaysia
K. Giridhar, Indian Institute of Technology Madras, India
Michele Girolami, ISTI-CNR, Italy
Javier Manuel Gozalvez Sempere, University Miguel Hernandez of Elche, Spain
Amaç Güvensan, Yildiz Technical University, Turkey
Xiang Gui, Massey University, New Zealand
Frederic Guidec, IRISA-UBS, Université de Bretagne-Sud, France
Christian Hägerling, TU Dortmund University, Germany
Gerhard Hancke, Royal Holloway / University of London, UK
Mohamad Sayed Hassan, Institut Telecom / Telecom Bretagne, France
Laurent Herault, CEA-Leti - Grenoble, France
Unai Hernández-Jayo, Deusto Institute of Technology - DeustoTech., Spain
Fumio Ishizaki, Nanzan University, Japan
Muhammad Ismail, University of Waterloo, Canada
Yasunori Iwanami, Shikumi College Nagoya/Institute of Technology -Nagoya-shi, Japan
Tauseef Jamal, University Lusofona - Lisbon, Portugal
Ali Jemmali, École Polytechnique de Montréal, Canada
Michel Jezequel, Telecom Bretagne - Brest, France
Jehn-Ruey Jiang, National Central University - Jhongli City, Taiwan
Hu Jin, University of British Columbia, Canada
Magnus Jonsson, Halmstad University, Sweden
Yunho Jung, Korea Aerospace University, Korea
Adrian Kacso, University of Siegen, Germany
Mohamed Abd Rabou Kalil, Ilmenau University of Technology, Germany
György Kálmán, ABB AS - Akershus, Norway

Riad Kanan, University of Applied Sciences Western Switzerland (HES-SO), Switzerland
Subrat Kar, Indian Institute of Technology Delhi - New Delhi, India
Ghassan Ali Kbar, King Saud University - Riyadh, Saudi Arabia
Mounir Kellil, CEA, France
Zeashan Khan, GIPSA Lab - Grenoble, France
Timo O. Korhonen, Aalto University, Finland
Dragana Krstic, University of Nis, Serbia
Alexey Lagunov, M.V. Lomonosov - The Northern (Arctic) Federal University, Russia
Zihua Lai, Ranplan Wireless Network Design Ltd., UK
Louise Lamont, Communication Research Centre, Canada
Kin K. Leung, Imperial College, UK
Jingli Li, TopWorx - Emerson, USA
Xi Li, Beijing University of Posts and Telecommunications, China
Xun Li, Alcatel-Lucent Shanghai Bell Labs, China
Qilian Liang, Wuhan University, China
Fidel Liberal Malaina, University of the Basque Country, Spain
Justin Lipman, Intel R&D China, China
Donggang Liu, Wuhan University, China
Andreas Löffler, Friedrich-Alexander University of Erlangen-Nürnberg, Germany
Valeria Loscrí, University of Calabria, Italy
Jonathan Loo, Middlesex University - London, UK
Lin Luo, University of South Australia - Mawson Lakes, Australia
Christian Maciocco, Intel Corporation -Santa Clara, USA
D. Manivannan (Mani), University of Kentucky - Lexington, USA
Muneer Masadeh Bani Yassein, Jordan University of Science and Technology - Irbid, Jordan
Barbara M. Masini, CNR - IEIIT, University of Bologna, Italy
Daniel Massicotte, Université du Québec à Trois-Rivières, Canada
Catherine Meadows, Naval Research Laboratory - Washington DC, USA
Hamid Menouar, QUWIC - Qatar University Wireless Innovation Center, Qatar
Fabien Mieyeville, Institut des Nanotechnologies de Lyon, France
Makoto Miyake, M-TEC Company Limited / Mitsubishi Electric Corporation, Kamakura-City, Japan
Klaus Moessner, University of Surrey, UK
Mohamed M. A. Moustafa, Egyptian Russian University, Egypt
Lorenzo Mucchi, University of Florence, Italy
David Navarro, INL - Lyon Institute of Nanotechnologies, France
Wolfgang Narzt, Johannes Kepler University - Linz, Austria
Renato Negra, RWTH Aachen University, Germany
Marek Neruda, Czech Technical University in Prague, Czech Republic
Georg Neugebauer, RWTH Aachen University, Germany
Christopher Nguyen, Intel Corp., USA
Homayoun Nikookar, Delft University of Technology, The Netherlands
Ronit Nossenson, Jerusalem College of Technology, Israel
Loutfi Nuaymi, Telecom Bretagne - Rennes, France
Shigeaki (Aki) Ogose, Kagawa University, Japan
George S. Oreku, TIRDO/ North west University, South Africa
Abdelkader Outtagarts, Alcatel-Lucent Bell Labs, France
Tudor Palade, Technical University of Cluj-Napoca, Romania
Carlos Enrique Palau Salvador, Polytechnic University of Valencia, Spain

Asier Perallos, University of Deusto, Spain
Matteo Petracca, National Inter-University Consortium for Telecommunications, Italy
Salvatore Flavio Pileggi, Universidad Politécnica de Valencia, Spain
Thomas Plos, TU-Graz, Austria
Anastasios Politis, Technological Educational Institute of Serres, Greece
Carlos Pomalaza-Raez, Purdue University, USA / University of Oulu, Finland
Iwona Pozniak-Koszalka, Wroclaw University of Technology, Poland
Anand R. Prasad, NEC Corporation, Japan
Ion Emilian Radoi, The University of Edinburgh, U.K.
Hani Ragab Hassen, University of Greenwich, UK
Yusnita Rahayu, Universiti Malaysia Pahang (UMP), Malaysia
Christian Renner, University of Luebeck, Germany
Teng Rui , NICT, Japan
José Antonio Sánchez Fernández, Universidad Politécnica de Madrid, Spain
David Sánchez Rodríguez, University of Las Palmas de Gran Canaria (ULPGC), Spain
José Santa Lozano, Universidad de Murcia, Spain
Reijo Savola, VTT, Finland
Hans-Otto Scheck, Nokia Siemens Networks, Germany
Riccardo Scopigno, Istituto Superiore Mario Boella - Torino, Italy
Zary Segall, Royal Institute of Technology (KTH), Sweden
Jean-Pierre Seifert, Technische Universität Berlin & Deutsche Telekom Laboratories - Berlin, Germany
Sandra Sendra Compte, Polytechnic University of Valencia, Spain
Jun Bae Seo, University of British Columbia, Canada
Ali Shahrabi, Glasgow Caledonian University, UK
Adão Silva, University of Aveiro / Institute of Telecommunications, Portugal
Sivakumar Sivaramakrishnan, AUT University of Auckland, New Zealand
Wojciech Siwicki, Gdansk University of Technology, Poland
Mariusz Skrocki, Orange Labs - Warszawa, Poland
Vahid Solouk, Urmia University of Technology, Iran
Himanshu B Soni, G.H. Patel College of Engineering & Technology, India
Kuo-Feng Ssu, National Cheng Kung University, Taiwan
Razvan Stanica, INSA Lyon, France
Yasihisa Takizawa, Kansai University, Japan
Fatma Tansu Hocanin, Eastern Mediterranean University, Turkey
Thomas Ußmüller, University of Erlangen-Nuremberg, Germany
Václav Valenta, Ulm University, Germany
K. Vasudevan, Indian Institute of Technology - Kanpur, India
Vincent Verdot, Alcatel-Lucent Bell Labs, France
Natalija Vlajic, York University - Toronto, Canada
Baptiste Vrigneau, IUT - R&T Châtelleraut, France
Robert Weigel, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany
Yean-Fu Wen, National Chiayi University, Taiwan
Martin Werner, Ludwig-Maximilians-University Munich, Germany
Joachim Wilke, Karlsruhe Institute of Technology (KIT), Germany
Ouri Wolfson, University of Illinois at Chicago, USA
Qishi Wu, University of Memphis, USA
Pei Xiao, University of Surrey, UK
Mee Loong Yang, AUT University, New Zealand

Erkan Yüksel, Istanbul University, Turkey
Sven Zacharias, University of Limerick, Ireland
Sherali Zeadally, University of the District of Columbia, USA
Yuanyuan Zeng, Wuhan University, China
Hans-Jürgen Zepernick, Blekinge Institute of Technology, Sweden
Yan Zhang, IMEC- NL, The Netherlands

Copyright Information

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

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

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

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

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

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

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

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

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

Table of Contents

MAC Protocols and Mobility Management for Real-time Applications Using Wireless Sensor Networks <i>Muhsin Atto and Chris Guy</i>	1
The Application and Improvement of Temporally Ordered Routing Algorithm in Swarm Network with Unmanned Aerial Vehicle Nodes <i>Zhongqiang Zhai, Jun Du, and Yong Ren</i>	7
Towards the Definition of a Mobility-based Clustering Environment for MANET <i>Aida Ben Chehida, Ryma Abassi, and Sihem Guemara</i>	13
Creating Frame Structures for Position-based Hybrid Routing Algorithm <i>Resul Kara, Ibrahim Ozcelik, Huseyin Ekiz, and Murat Iskefiyeli</i>	19
An Improved Extended Kalman Filter for Localization of a Mobile Node with NLOS Anchors <i>siamak yousefi, Xiao-Wen Chang, and Benoit Champagne</i>	25
An Overview of Underwater Sensor Networks <i>Christina Peach and Abdulrahman Yarali</i>	31
Smart Relay Protocols for Throughput Optimization using AMC in LTE-Advanced Systems <i>Saransh Malik, Sangmi Moon, Bora Kim, Daejin Kim, Youngil Kim, Kunmin Yeo, and Intae Hwang</i>	37
Second Best Codeword for MIMO Broadcast Channels with Limited Feedback <i>Mouncef Benmimoune, Daniel Massicotte, and Sebastien Roy</i>	44
Further Results on Channel Estimation Based on Weighted Moving Average in T-DMB Receivers <i>Sung Ik Park, Heung Mook Kim, Jin-Ho Chung, and Jeongchang Kim</i>	48
Adaptive Reliable Routing Protocol for Wireless Sensor Networks <i>Nourhene Maalel, Pierre Roux, Mounir Kellil, and Abdelmadjid Bouabdallah</i>	52
Design for a UWB Down-Conversion Mixer for Multi-Band Applications <i>Wu-Shiung Feng, Chin-I Yeh, Prasenjit Chatterjee, and Chen-Yu Hsu</i>	56
A Microwave Triplexer Based on Coupled Resonators Approach <i>Lidiane S. Araujo and Antonio J. B. de Oliveira</i>	61
Spectral Resource Sharing for Uncoordinated Networks <i>Ayman Assra and Tricia Willink</i>	65

Performance Analysis of Coordinated Multi-Point with Scheduling and Precoding schemes in the Heterogeneous Network <i>Bora Kim, Saransh Malik, Sangmi Moon, Daejin Kim, Youngil Kim, Kunmin Yeo, and Intae Hwang</i>	70
A New Approach for Multistage Analog Complex Filter Design <i>Daniele Grasso and Giuseppe Avellone</i>	76
QoS-aware Traffic Offloading in 4G/WiFi Multi-RAT Heterogeneous Networks: Opnet-based Simulations and Real Prototyping Implementation <i>Ahmed Ibrahim, Ahmed Darwish, Shady Elbassiouny, and Hani Elgebaly</i>	82
A Top-Down Heuristic for TCFA Problem in WAN <i>Roza Goscienc, Iwona Pozniak-Koszalka, Leszek Koszalka, and Andrzej Kasprzak</i>	89
Frequency and Time Resource Allocation for Enhanced Interference Management in a Heterogeneous Network based on the LTE-Advanced <i>Sangmi Moon, Bora Kim, Saransh Malik, Cheolsung Kim, Intae Hwang, Kunmin Yeo, and Youngil Kim</i>	95
A New No Reference Metric for Estimation Video Quality Based on Blur Effect <i>Jean Araujo and Claudio Monteiro</i>	100
Traffic Steering Framework for Mobile-Assisted Resource Management in Heterogeneous Networks <i>Anton Dogadaev, Aleksandra Checko, Andrijana Popovska Avramova, Anna Zakrzewska, Ying Yan, Sarah Ruepp, Michael Berger, Lars Dittmann, and Henrik Christiansen</i>	107
A Proposal of Prediction of Peak-Signal-to-Noise-Ratio based on QoS Networks <i>Carlos Junior and Claudio Monteiro</i>	113
Effective ARP Based Call Admission Control (CAC) Scheme For LTE Systems <i>Abhishek Nayak, Sauarbh Verma, and Anurag Jain</i>	119
Performance Limitation of HSDPA MIMO by Pre-Coding Induced Phase Distortion <i>Wolfgang Aichmann</i>	123
Cooperative Communications Using Multilevel Bit-Interleaved Coded Modulations <i>Hadi Sawaya</i>	129
An Improved MU-MIMO OFDM Scheme for the 3GPP LTE-Advanced Standard <i>Yuechao Wang, Zhanji Wu, and Guanghao Wu</i>	134
Carrier Frequency Recovery for Oversampled Perfect Reconstruction Filter Bank Transceivers <i>Siavash Rahimi and Benoit Chanpagne</i>	140

A Semi-Analytical Performance Prediction of Turbo Coded SC-FDMA <i>Fatima Ezzahra Naamane, Mohamed Et-tolba, and Mostafa Belkasm</i>	146
The Effect of Human Bodies on Path Loss Model in an Indoor LOS Environment <i>Young-Hoon Kim and Seong-Cheol Kim</i>	152
Equal Gain Combining SC-FDMA Performance over Correlated Shadowed Rice Land Mobile Satellite Channel <i>Jyoti Gangane, Mari-Carmen Aguayo-Torres, Juan Sanchez-Sanchez, and Francisco Blaquez-Casado</i>	157
Linear Node Movement Patterns in MANETS <i>Mohammed Alzaylaee, John DeDoure, and Przemyslaw Poche</i>	162
A Proposed Real-Time Scheduling Algorithm for WiMAX Networks <i>Nada Elshennawy, Mostafa Youssef, Mohamed El-Derini, and Mahmoud Fahmy</i>	167
On the Performance of Selective Decode and Forward Relaying over Imperfectly Known Fading Relay Channels <i>Ali Zarei Ghanavati and Daniel Lee</i>	173
Fast Association Scheme over IEEE 802.15.4 based Mobile Sensor Network <i>Pranesh Sthapit, Yeon-Sang Choi, Goo-Rak Kwon, Jae-Young Pyun, and Suk-seung Hwang</i>	179
An Application of Second-Order Reed-Muller Codes for Multiple Target Localization in Wireless Sensor Networks <i>Thu Lam Ngoc Nguyen, Jaejin Lee, and Yoan Shin</i>	185
Bit Error Rate Analysis of MIMO Schemes in LTE Systems <i>Ali Jemmali, Jean Conan, and Mohammad Torabi</i>	190
The Performance Analysis of Complex SSC/MRC Combiner in Rice Fading Channel <i>Dragana Krstic, Petar Nikolic, Aleksandar Stevanovic, and Goran Stamenovic</i>	195
Intercell Interference Coordination for the ePDCCH in LTE-Advanced Macrocellular Deployments <i>David Gonzalez G, Mario Garcia-Lozano, and Silvia Ruiz Boque</i>	200
Modeling and Simulation of Radio Signals Attenuation Using Informed Virtual Geographic Environments <i>Mehdi Mekni</i>	209
A New Physical Cell Identifier Structure in Femtocell Networks <i>Robert Bestak</i>	217
A New Bio-Crypto System Approach to Enhance Security Of 4G Mobile Networks <i>Saeed Ashry, Abdulatif Elkouny, Hesham Elbadawy, Salwa Elramly, and Ahmed El-Sherbini</i>	222

The Advantage of Moving Nodes in Formations in MANETs and M2ANETs <i>Abdullah Alshehri, Przemyslaw Pochec, and John DeDourek</i>	228
BER Performance of BICM-coded Cooperative Networks with Selection Decode-and-Forward Relaying over Nakagami-m Fading Channels <i>Tsang-Wei Yu and Wern-Ho Sheen</i>	233
High-Level Energy Saving Strategies for Mobile Location-Based Services on Android Devices <i>Wolfgang Narzt</i>	238
Efficient Mobility Management in 6LoWPAN Wireless Sensor Networks <i>Abdelwahed Berguiga and Habib Youssef</i>	244
Mobile Ad-Hoc Search and Retrieval in the iTrust over Wi-Fi Direct Network <i>Isai Michel Lombera, Louise Moser, Peter Melliar-Smith, and Yung-Ting Chuang</i>	251
Proposal to Reduce the Computational Cost of Wireless Network Selection Process <i>Francirley Borges and Claudio Monteiro</i>	259
Challenges of Operationalizing PACS on Cloud Over Wireless Networks <i>Ricardo Ferraro de Souza, Carlos Becker Westphal, Daniel Ricardo dos Santos, and Carla Merke Westphall</i>	265

MAC Protocols and Mobility Management for Real-time Applications Using Wireless Sensor Networks

Muhsin Atto and Chris Guy
University of Reading, Reading, United Kingdom
{kp003919, c.g.guy}@reading.ac.uk

Abstract— Wireless Sensor Networks detect events using one or more sensors, then collect data from detected events using these sensors. This data is aggregated and forwarded to a base station (sink) through wireless communication to provide the required operations. Different kinds of MAC and routing protocols are needed for WSN in order to guarantee data delivery from the source nodes to the sink. GinMAC has been simulated for real-time applications, where energy saving, delay and reliability need to be considered. A Mobility module for GinMAC has been implemented. It has been showed that GinMAC can be applied for real-time applications in both mobile and static scenarios. However, further improvements of its performance are needed. Some future plans to investigate this are given at the end of this paper.

Keywords— WSN; Real-time Applications; GinMAC; Mobility; Castalia.

I. INTRODUCTION

Wireless Sensor Networks (WSNs) detect events using one or more sensors, then collect data from detected events using these sensors. This data is aggregated and forwarded to a base station (sink) through wireless communication to provide the required operations. There are a lot of challenges that need to be considered before designing protocols for WSNs. Different kinds of Medium Access Control (MAC) and routing protocols need to be designed for WSN in order to guarantee data delivery from the source nodes to the sink in different applications.

The initial applications supported by WSNs were mostly in environment monitoring, such as temperature monitoring for a specific area, house alarms, and so on. The main objectives in such applications only involved simple data processing. Energy consumption needed to be considered for specific applications, so little attention was taken on data delivery and reliability related issues such as in [1][2][3].

WSNs have been extended and their design have been advanced to support more hard design and complex applications, such as security, military, fire detection and health care related applications. In these applications, data delivery and reliability must be taken as important parameters in addition to energy efficiency, because data must be collected from the sources of events and be forwarded to the sink in real time with high reliability, otherwise the application will be useless.

In this paper, an implementation of GinMAC [4] including the proposed mobility management module is described for real-time applications, where energy saving, delay and reliability for end to end data delivery over multi hop WSNs need to be considered. An example of the proposed real-time application in this paper is healthcare, where data needs to be collected from the body of the patients and then sent to a sink, for example [5].

The rest of the paper is structured as follow. Motivations for the paper will be given in Section II, then the implementation of GinMAC for both static and mobility applications will be described in Section III. Simulation scenarios and required parameters with figures showing the results for GinMAC implementation for the proposed applications will be given in Section IV. The simulation results and some discussion is in Section V. A conclusion and future work are presented in Section VI.

II. MOTIVATIONS

Most of the recently proposed protocols for WSNs consider either energy saving or reliability for the target applications, none of them have considered both performance metrics at the same time [1]. However, some applications may need to guarantee both energy saving and reliability at the same time without any errors, otherwise the applications may be useless. Therefore, in order to provide this, new and very efficient MAC protocols need to be designed. Previous work has shown that GinMAC is the only protocol, which can be used for real-time applications to provide the required performances as shown in [6]. The novel motivations for this paper are the following:

- Design MAC protocols for real-time applications where the required energy saving, reliability and delay for data delivery need to be considered.
- Design mobility management modules for real-time applications.
- Adapt GinMAC to add new features to improve its applicability to real-time applications which require mobility, such as healthcare applications as described in [5].
- Simulate a GinMAC implementation including the proposed mobility management module for real-time applications.

III. MAC PROTOCOLS FOR REAL-TIME APPLICATIONS

It was concluded in [6] that GinMAC is a possible MAC protocol for use in real-time applications, where reliability, energy saving and delay can be guaranteed. Challenges and requirements that need to be considered before designing any MAC protocols for such applications are also described in the same paper. The implementation of GinMAC including a mobility management module is described in this Section.

A. Implementation of GinMAC for Real-time Applications

GinMAC [7] is a Time Division Multiple Access (TDMA) based MAC protocol, so energy saving and reliability with bounded delay can be achieved. However, an efficient synchronization and slot allocation algorithms need to be designed in order to allocate the required slot time for each node in the network and let the radio of the nodes be turned on only in the allocated time. In this case, each node needs enough slots of time to transmit data toward a sink, including control messages such as messages for slots permission, mobility and topology control related messages. GinMAC has been modified to add new features to improve its applicability to applications which require mobility, such as healthcare applications. The GinMAC implementation in [7] does not support mobility while this one does. Topology management and time synchronization for GinMAC in this implementation are described below.

1) *Slot allocations in GinMAC*: GinMAC is a TDMA based protocol and assumes that data is forwarded hop by hop toward a sink using a tree based topology, consisting of n nodes. Time in GinMAC is divided into a fixed length called *Epoch E*, each E is subdivided by $n*k$ time slots so that each node allocates k slots for transmitting data toward its parent until it reaches a sink. Each node is assigned k exclusive slots with four different types, which are basic slots (TX, RX) for data transmitting and receiving, additional slots (RTX, RRX) for re transmitting, broadcast slots ($BROD$) for topology control between nodes in the network and unused slots (U) for saving energy (if any). More details about how these slots are used can be found in Figure 1.

Additional slots are used only for re transmission to perform required reliability for the target applications, these slots are used even in the case when no data is available for transmission, as described in [4]. Unused slots are used for saving energy when data can not be delivered using basic and additional slots. This implementation for GinMAC does not contain unused slots, but they may be used in the future for increasing the lifetime of the network. Broadcasting slots are used for topology control. Slots for each node need to be allocated according to the defined topology so that the required performance can be achieved.

2) *GinMAC Topology Control Management*: GinMAC is a tree based WSN topology so that each node transmits its data toward a sink in its allocated slots and sleeps for the rest of the time. The current static topology that is proposed is a WSN with 13 nodes with static slot allocation, each node has enough slots of time to transmit all data from its children and its own, including control messages toward a sink. GinMAC

supports mobility for leaf nodes and this will required to design new topology control and management algorithms to provide connectivity between static and mobile nodes in the network. It is assumed that the Base Station (BS) has adequate power to reach all nodes in the network using down-link slots. However, the sensor nodes cannot always do this because of their limited power supply.

A node added to the network must determine in which slots it must become active before it can transmit or receive data. The steps used to achieve this are described below. After a node is switched on, it must first ensure time synchronization with the rest of the nodes in the network. Both control and data messages transmitted in the network can be used to obtain time synchronization. The node continuously listens to overhear a packet from the sink. After overhearing one message, the node knows when the GinMAC frame starts as each message carries information about the slot in which it was transmitted.

As a next step, the node must find its position in the topology which must stay within the defined topology envelope. For this purpose, the new node listens for packets in all slots. Transmitted data packets from a sink use a header field in which a node that is ready for transmission can find its information and then according to this information starts and stops data transmission toward its parent. A node may be configured with a list of valid nodes or clusters that it is allowed to attach to when mobility is supported. This might be necessary to ensure that a node will only attempt to join the network using known good links as determined by measurements before the deployment to provide the required performance.

3) *Synchronization Messages for GinMAC*: At the start of each frame, the sink needs to broadcast a synchronization packet which it is denoted as *SYNCH* into the network. This packet holds the start time, end time and slot numbers for each node in the network. When nodes receive a *SYNCH* packet from the network, they will extract their information from the *SYNCH* packet and then discard it. In this case, CSMA is used by the sink to synchronize nodes in the network and nodes use TDMA to transmit their data to their parents. After nodes receive their slot information from the sink, they need to ask permission for data transmission from their parents. Then, after slots related information has been received by a node, it has to handshake with its parent and then can start to transmit data. After a node uses its allocated slots, it can go to sleep and wake up at the same time in the next frame. Each node in this case will access the channel using their unique start time, so this will avoid any chance of collision with transmissions from other nodes in the network.

GinMAC lets nodes and their parents be active at the same time so that data can be transmitted between them. This synchronization algorithm for GinMAC synchronizes nodes with their parents so that parents can be active in the same time when one of their children is active. This time synchronization algorithm is good enough to deliver packets with the required performances for the applications described in Section IV.

The core idea behind this GinMAC implementation is to

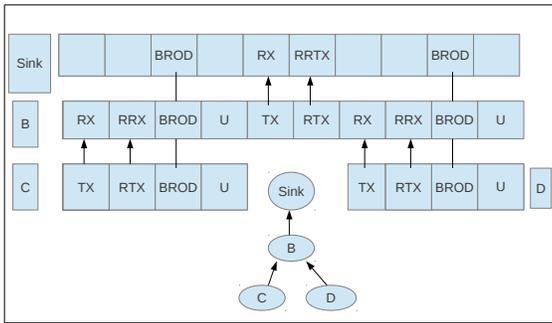


Fig. 1. Slot allocations and Synchronizations for nodes in the GinMAC

let nodes sleep as much as possible without effecting data delivery and required maximum delay, and this can only be performed using a TDMA based technique. The static topology is designed to let nodes have enough slots to transmit their data and in the rest of the frame go to sleep. The slot allocation and synchronization for GinMAC can be found in the Figure 1.

B. Mobility for Real-time Applications Using GinMAC

A new challenge is posed when mobility needs to be considered in a WSN for different applications such as healthcare, vehicle and monitoring people related applications. In this case topology control, resource management and performance control need to be designed to provide good connectivity between static and mobile nodes in the network and provide the required performance. Mobility and topology control for critical applications using WSNs are described in [8][9][10]. The proposed mobility management module in this paper follows the same messages and concepts as in the above papers.

1) *Mobility Management Module for GinMAC*: There may be cases when moving from one location to another in the network effects the connectivity of the network and then reconfiguration algorithms are needed. In order to support mobility for real-time applications, control messages which need to be transferred between static and mobile nodes to find a better attachment have been defined. Some of the possible control messages are Advertisement(ADV), join (JOIN), and join acknowledgement JOIN ACK messages. Static nodes are called *Clusters*.

When clusters switch on their radios, they need to send ADV to the network and then wait some time, then when mobile nodes receive these ADV messages they will ask to join the network. When clusters receive JOIN messages from the mobile nodes they will send back a JOIN ACK message to let the mobile node know that request to join has been accepted. So using these control messages connectivity between mobile nodes and cluster nodes will be established. In the proposed application only leaf nodes are allowed to be mobile nodes and all other nodes are clusters. The mobility module lets nodes move across a line between mobile nodes and a sink.

Mobile nodes may have more than one cluster they could join so they have to decide which cluster will be selected for transferring data toward their parents. In this GinMAC implementation, the cluster with maximum Receiver Signal Strength Indicator (RSSI) is considered the best one to be selected for the new attachment. Cluster nodes send ADV including available positions over time and when mobile nodes receive ADV, they compare the RSSI from their current parents to the received RSSI from the current ADV messages. In the case that a new cluster has better RSSI, mobile nodes need to leave their current parents and attach to this new cluster which is included in the currently received ADV message. When a new attachment is selected then a join request needs to be sent to that cluster. Upon receiving the JOIN request from a mobile node, JOIN ACK needs to be sent by the selected clusters.

Slots in the each frame need to be updated according to the new attachments, mobile nodes need to release the first tree position after it is attached to the second tree address, so in this case slots allocated for the new clusters need to be increased and slots allocated for the old clusters need to be decreased. A new algorithm for updating slots is needed for GinMAC to balance allocated slots for nodes according to the different attachments. A new algorithm has been designed to update channel according to new movements and changes in the topology of the network. Table II shows results from this mobility module describing different attachments between a mobile node and different clusters for transmitting packets toward a sink.

2) *Move Detection in GinMAC*: There are some cases when nodes can move without being detected. For instance, clusters may be unaware of leaving mobile nodes and then will keep space in the channel for that particular node. This will consume more energy and reduce the reliability of the network. There may be cases when clusters are not available for attachment any more without letting mobile nodes know. So an additional two control messages for this new mobility module for the proposed MAC protocol have been used, which are denoted by *KEEPLIVE* and *NODEALIVE*. *KEEPLIVE* control message is used by clusters to let its currently attached mobile nodes know that this cluster is still available and *NODEALIVE* message is used by mobile nodes to let their attached clusters know that they are still available for attachment. Mobile nodes wait for a specific interval to receive messages from the attached clusters, if they do not receive anything during that interval, a *NODEALIVE* message needs to be sent, to let a cluster know that they still want to use that cluster. If no reply is received then mobile nodes need to search for a new address to make a new attachment.

IV. SIMULATION SCENARIOS FOR GINMAC IMPLEMENTATION

The GinMAC protocol is compared with TMAC [11] in terms of energy saving, reliability and delay for end to end data delivery for WSNs having static and mobile nodes. More details about simulation parameters and scenarios are given below.

A. Simulation Scenarios and parameters

Castalia has been used in this work, because of its capabilities for simulating protocols for WSNs based on the real data, as shown in [12]. Both MAC protocols were simulated according to the application requirements given in the following sections, using different packets rates. We define packet rate by R_p = packets per second, so as we can see from the graphs in our simulation results that $R1$ means nodes send 1 packet per second, $R2$ means nodes send 2 packets per second and so on. More details about the topology of the deployed WSN, MAC protocols and other parameters can be found in Table I.

TABLE I
SIMULATION PARAMETERS

Parameter	Value
MAC Protocols	GinMAC and TMAC
Network Dimensions(in meters)	90 X 90
Distance Between pair of nodes	20 meters
Simulation Duration	10 minutes
Measurement Metrics	Life time, delay and reliability
Number of Nodes	13
Packet Rates (packet(s) per second)	1-5
mobility speed(meters in seconds)	5
mobility interval(in minutes)	1
Advertisement interval (in seconds)	15
Initial Energy(in Joules)	18720
Real Radio	CC2420

B. Proposed Applications for GinMAC Simulation

A simple application where all nodes send data towards a sink using different packet rates is used. The proposed application in this paper is healthcare related, where data needs to be collected from the body of patients and then sent back to a base station. Some threshold values have been selected for the proposed application, which need to be achieved before deciding if these MAC protocols can be used for the target application or not.

1) **Reliability:** The reliability of given protocols is defined as the ratio between total packets generated and sent by source nodes and total packets received by a sink or a final destination. So Reliability = (Received Packets/Sent Packets). The threshold value for reliability needs to be high enough in order to achieve the requirements for the proposed application. When mobility is supported this threshold may be reduced, therefore, reliability is highly dependent on the applications. Information from the patients needs to be accurate and delivered to a sink as soon as possible to make the required operations. Based on this, it has been assumed that at least a reliability of 0.99 for the static scenarios and a reliability of 0.98 for the mobility scenarios must be achieved.

2) **Energy Saving and Lifetime:** Energy saving is one of the most important challenges needed to be considered in any applications when designing WSNs. Each node has a small battery with limited power which cannot be recharged very often, so each protocol needs to consider consumed energy for each node in order to extend the life time of the entire

TABLE II
DIFFERENT ATTACHMENTS USING THE PROPOSED MOBILITY MODULE FOR GINMAC SEE SECTION III-B1

Time(sec)	RSSI(dBm)	Cluster	current RSSI(dBm)	Required Action
0.0356	-76.10	10	-100	new attachment
15.0051	-86.05	1	-76.10	no action
15.0081	-88.02	5	-76.10	no action
15.0224	-84.24	9	-76.10	no action
15.02667	-75.05	6	-76.10	new attachment

network. The lifetime of the networks is the maximum days that a WSN can survive, whilst spending energy at the given rate. Let consumed energy by each node be denoted by C joules, initial energy by E joules and current simulation time by T seconds, then the lifetime of given MAC protocols for each node in the network has been calculated as follow:

$$LifeTime(indays) = ((E/C) * T)/86400 \quad (1)$$

where 86400 is number of seconds in each day. We define life time as how many days each node will be alive consuming C joules as described in the above equation. It has been assumed that nodes in the proposed healthcare application can be recharged every week.

3) **Delay Calculation:** Delay is defined as the difference between the time when each packet is sent from its source node to the time when the same packet is received by its final destination. Delay in real-time applications needs to be measured so that all data must be delivered within bounded delay, i.e., each packet that is delivered after this delay is considered to be lost and will be ignored. All data needs to be collected from the source nodes and then delivered to the sink within a bounded delay. It was decided that a delay of more than 10 seconds for delivering data to a sink may risk the health of the patients.

C. Simulation for Static and Mobility Scenarios

GinMAC and TMAC have been simulated with different packets rates as shown in each graph, the WSN topology is also given below. The results graphs from running simulation for both static and mobility scenarios are shown below.

V. SIMULATION RESULTS AND CONCLUSION

1) **Packets Delivery and Reliability:** It is shown in the Figure 3 that GinMAC can offer the applications requirements in term of reliability (as defined in Section IV) at various packet rates. From packet rates 1 to 4, GinMAC delivers more than 0.99 of packets from source nodes to a sink. Thus, it can be said that GinMAC can be used for the proposed applications. However, TMAC cannot offer a reliability of more than 0.91 in both high and low packet rates using the same parameters as shown in the Figure 3.

GinMAC cannot offer reliability of more than 0.97 for packets rates more than 4 and if it is needed to increase this reliability for such packet rates, the number of slots for nodes in the static topology may be increased. But, GinMAC performs better than TMAC in both high and low packet rates

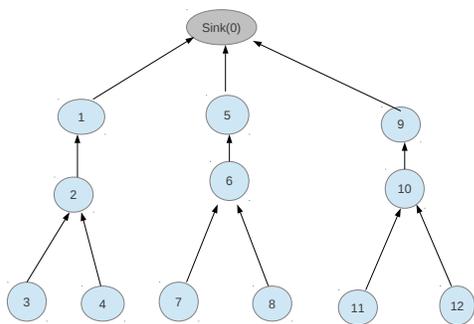


Fig. 2. Topology of the WSN in our Simulation

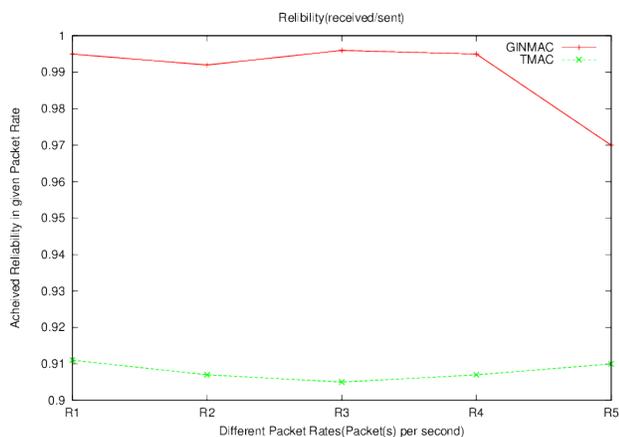


Fig. 3. Performances in term of Reliability using TMAC and GinMAC for Static scenario using different packet rates see section IV-A

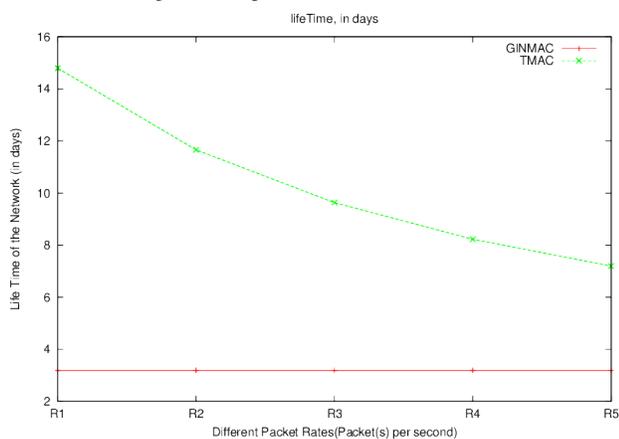


Fig. 4. Performances in term of life time of the nodes in networks using GinMAC and TMAC for Static scenario using different packet rates see section IV-A

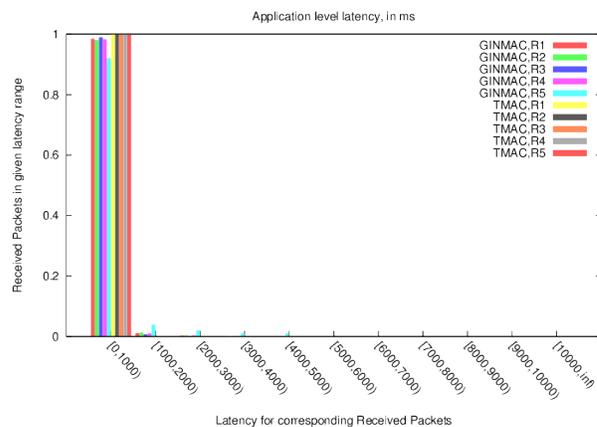


Fig. 5. Latency for delivered packets using TMAC and GinMAC for Static scenario using different packet rates see section IV-A

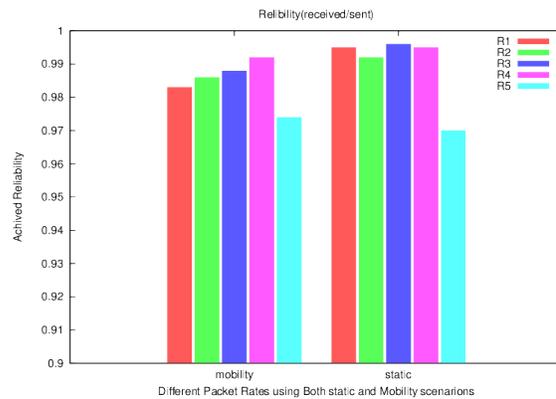


Fig. 6. Performances in term of Reliability using both mobility and static scenarios for GinMAC using different packet rates see section IV-A

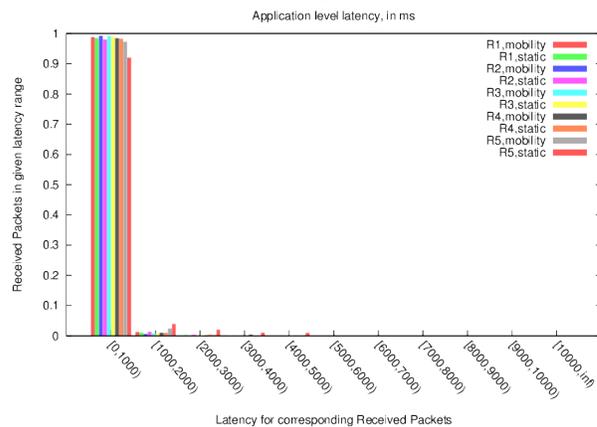


Fig. 7. Latency for delivered packets using both mobility and static scenarios for GinMAC using different packet rates see section IV-A

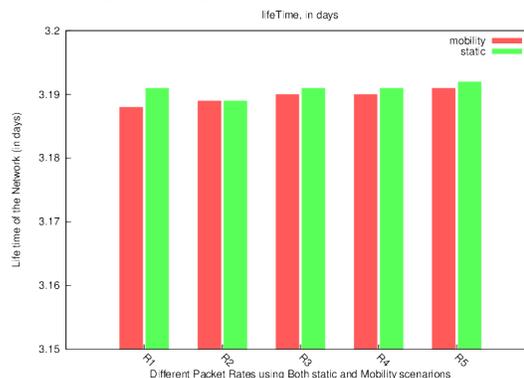


Fig. 8. Life time of the nodes in the network using both mobility and static scenarios for GinMAC using different packet rates see section IV-A

as shown in the Figure 3. Therefore, the conclusion is that TMAC can not be applied for applications where reliability of more than 0.91 needs to be guaranteed at any packet rates. This is expected to be so since GinMAC uses static slots allocation and no interference can occur. TMAC uses contention based techniques and hence collisions are more likely to happen.

2) *Energy Saving and Lifetime*: Figure 4 shows the average life time of the nodes in the network and shows that TMAC performs better than GinMAC in terms of energy saving in both low and high packet rates. This is expected to be so, because of the adaptive related techniques for TMAC, which depend on the on-going traffic in the network nodes which need to be active. In this case, TMAC lets nodes sleep a lot at low packet rates, however, GinMAC does not consider traffics for sleeping nodes. GinMAC consumes the same amount of energy for all given packet rates. This is expected to be so, because according to the GinMAC design specification given in [4] GinMAC does not consider traffic for sleeping nodes and nodes are active in their allocated slots even in the case where no packets are available for transmission to achieve the required reliability and bounded delay. Hence, nodes will consume nearly the same amount of energy at both high and low packet rates. A WSN using GinMAC can survive 3.19 days which is not enough to be used for these applications without improvement.

3) *Delay for Data Delivery*: In Figure 5, *MaxLatency* defines the bounded latency that all packets need to be delivered, which represents the threshold for latency in the proposed applications and *MaxColumns* defines the number of columns to be used for measuring the latency for given MAC protocols. Any delivered packets after the last column are considered to be lost and may be discarded. GinMAC does not perform better than TMAC in term of latency as shown in the Figure 5, however this performance is enough to be used for the proposed applications. According to the results from Figure 5, most of the packets (which is about more than 0.98 of received packets) are received before first 5 seconds from packet rates 1 to 5, and the rest of packets are received within the last 5 seconds and hence all packets are received within 10 seconds.

A. Results from the Mobility Scenario

Mobility module for GinMAC uses the RSSI to select a better attachment. It has been assumed that there is only one mobile node however, more than one node can be mobile for the leaf nodes in the proposed mobility management module. Figures 8 and 6 show that GinMAC offers nearly the same reliability and lifetime for both static and mobility scenarios. Figure 7 shows that latency is also the same as in static topology and the required delay performance is offered using the mobility module.

VI. CONCLUSION AND FUTURE WORK

An implementation of GinMAC including the proposed mobility module for real-time applications were described in this paper, where the required performance needs to be

guaranteed. It has been shown that GinMAC can be used for the target applications, except that the life time needs to be improved. A mobility module has been designed and simulated for GinMAC. The results from the mobility module have also showed that GinMAC implementation can give nearly the same performance in both mobility and static scenarios for the proposed applications. GinMAC can not provide the required routing for the proposed applications, hence efficient routing protocols need to be designed in the future to cooperate with GinMAC in order to provide the required routing and extend the life time of the network for the proposed applications.

ACKNOWLEDGEMENT

The authors would like to thank Mr James Brown from the University of Lancaster in UK, for his help for providing the basic ideas about the GinMAC and its specification. The authors would also like to thanks Dr Athanssions Boulis for information about Castalia simulator which helped a lot.

REFERENCES

- [1] P. Suriyachai, U. Roedig, and A. Scott, "A survey of mac protocols for mission-critical applications in wireless sensor networks," *Communications Surveys Tutorials*, IEEE, vol. 14, no. 2, 2012, pp. 240 – 264.
- [2] J. Lotf and S. Ghazani, "Overview on routing protocols in wireless sensor networks," in *Computer Engineering and Technology (ICCT)*, 2010 2nd International Conference on, vol. 3, April 2010, pp. 610 – 614.
- [3] M. Soyuturk and D. Altılar, "Reliable real-time data acquisition for rapidly deployable mission-critical wireless sensor networks," in *INFOCOM Workshops 2008*, IEEE, April 2008, pp. 1 – 6.
- [4] P. Suriyachai, J. Brown, and U. Roedig, "Time-critical data delivery in wireless sensor networks," in 6th IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS '10). IEEE, June 2010, pp. 216–229.
- [5] H. Huo, Y. Xu, H. Yan, S. Mubeen, and H. Zhang, "An elderly health care system using wireless sensor networks at home," in *Sensor Technologies and Applications, 2009. SENSORCOMM '09*. Third International Conference on, June 2009, pp. 158 –163.
- [6] M. Atto and C. G. Guy, "Wireless sensor networks: MAC protocols and real time applications," in *The 13th Annual Post Graduate Symposium on the Convergence of Telecommunications, Networking and Broadcasting (PGNet2012)*, Liverpool, UK, United Kingdom, June 2012, pp. 1 – 6.
- [7] J. Brown and U. Roedig, "Demo abstract: Ginlite - a mac protocol for real-time sensor networks," in *In Proceedings of 9th European Conference on Wireless Sensor Networks (EWSN'12)*, Feb 2012.
- [8] Z. Zinonos, R. Silva, V. Vassiliou, and J. Silva, "Mobility solutions for wireless sensor and actuator networks with performance guarantees," in *Telecommunications (ICT)*, 2011 18th International Conference on, May 2011, pp. 406–411.
- [9] Z. Zinonos and V. Vassiliou, "Inter-mobility support in controlled 6lowpan networks," in *IEEE Globecom 2010 Workshop on Ubiquitous Computing and Networks (UbiCoNet2010)*, Miami, Florida, USA, Dec 2010, pp. 1718–1723.
- [10] R. Silva, Z. Zinonos, J. Sa Silva, and V. Vassiliou, "Mobility in wsns for critical applications," in *Computers and Communications (ISCC)*, 2011 IEEE Symposium on, July 2011, pp. 451–456.
- [11] Y. Tselishchev, A. Boulis, and L. Libman, "Experiences and lessons from implementing a wireless sensor network mac protocol in the castalia simulator," in *Wireless Communications and Networking Conference (WCNC)*, 2010 IEEE, April 2010, pp. 1–6.
- [12] A. Boulis, "Castalia simulator: A simulator for wireless sensor networks and body area networks," in <http://castalia.npc.nicta.com.au>, NICTA, [retrieved: April, 2013], pp. 1 – 120.

The Application and Improvement of Temporally Ordered Routing Algorithm in Swarm Network with Unmanned Aerial Vehicle Nodes

Zhongqiang Zhai, Jun Du, Yong Ren

Department of Electronic Engineering
Tsinghua University
Beijing, China

elezzq@gmail.com, du-j11@mails.tsinghua.edu.cn, reny@tsinghua.edu.cn

Abstract—In recent years, the research and application of Unmanned Aerial Vehicle (UAV) network has become a significant topic. The mobile ad hoc network established by UAV nodes can be more efficient to complete various tasks in a harsh environment. Plenty of research focuses on the routing protocol, which is an important factor to play group advantage of UAVs. In the swarm network with UAV nodes, nodes failure or mobility may cause routing failure, which results in communication failure or longer delay. The existing routing-repair mechanisms are accompanied by a great deal of control overhead, which cannot solve the problem mentioned above effectively. This article analyses the advantages and disadvantages of Temporally Ordered Routing Algorithm (TORA), and proposes a routing protocol named Rapid-reestablish Temporally Ordered Routing Algorithm (RTORA). RTORA adopts reduced-overhead mechanism to overcome adverse effects caused by link reversal failure in TORA. The simulation results using OPNET demonstrate that RTORA has less control overhead and smaller delay than TORA.

Keywords-UAV; swarm; routing failure; RTORA; OPNET.

I. INTRODUCTION

With the development of Unmanned Aerial Vehicle (UAV) and its application in various fields, some deficiencies using single UAV to perform tasks are exposed, such as limited scope of target search and monitoring, single task load and low task timeliness. In the past few years, establishing a mobile ad hoc network by multiple UAVs became a method to solve the problems above. In such a network, UAVs can sense environmental information in a wide range, and complete a number of different tasks at the same time through information exchange and dynamic task allocation. Furthermore, the remaining UAVs can continue to perform the task when several UAVs are damaged, which can improve the task timeliness. As an important factor to play groups advantage of UAVs, the routing protocol has been one of the key points in the UAV network. Swarm network with UAV nodes is a network that high-density UAVs implement a saturated search, detection or attack as a swarm through mutual cooperation. Such high-density networks are typically deployed in harsh environments. The probability of transmission link failure will greatly increase, which is caused by damaged relay nodes or neighbor nodes moving beyond the transmission range. Rediscovering and

reestablishing new routes will bring longer delay. Therefore, a strategy to solve the problem of routing failure is needed.

One way to alleviate the issue above is to prevent link failure. Mobility Aware Ad hoc On-demand Distance Vector Routing (MA-AODV) is proposed to deal with link failure caused by mobility in [1]. MA-AODV periodically calculates mobility, and chooses the transmission route with smaller-mobility nodes. But this method cannot avoid link failure completely. Additionally, the period of calculation also needs to be considered. Too big value cannot guarantee the accuracy of mobility, and too small value will greatly increase the control overhead.

Another approach is to look for a new route after the current route is damaged. Some classic on-demand routing protocols use this method, e.g., Dynamic Source Routing (DSR), Ad hoc On-demand Distance Vector Routing (AODV), Temporally-Ordered Routing Algorithm (TORA), etc. The node detecting link failure sends a Route Error (RERR) message to the source node in DSR [2] and AODV [3]. Then, the source node discovers another route to destination by broadcasting a Route Request (RREQ) packet. TORA [4] repairs the transmission route in the local area where the link is interrupted. Some previous studies indicate that TORA is adapted to dense ad hoc networks [5][6], which will better meet the demand of the swarm network.

According to the above situation, we propose Rapid-reestablish TORA (RTORA) routing protocol based on TORA to solve routing failure in the swarm network with UAV nodes. The simulation results show that RTORA has smaller delay and lower control overhead than TORA.

The paper is organized as follows: Related research is presented in Section II. Section III analyzes the TORA protocol. In Section IV, we describe the mechanisms of RTORA. Section V presents the simulation results of TORA and RTORA. The conclusion and future work is given in Section VI.

II. RELATED WORK

With the focus on Quality of Service (QoS), many improved routing protocols about routing failure in the Mobile Ad hoc Network (MANET) are proposed. Z. Cheran et al. [7] design the Enhancement of Fault-Tolerant AODV (ENFAT-AODV) routing protocol, which uses the backup route to solve the problem of current routing failure in Wireless Sensor Networks (WSN). In [8], the AODV with

Reliable Delivery (AODV-RD) protocol reestablishes the route by the mechanisms of link failure prediction and alternate nodes selection. Multiple routes are created to solve the problem of link failure in [9]. In those three protocols, nodes periodically broadcast HELLO messages to acquire the link-status and topology changes, which is the same as AODV whose control overhead is huge due to periodic HELLO packets. Furthermore, the backup routing establishment and maintenance, as well as reasonable alternative nodes selection, will add to the broadcast of control packets. In [10], the researcher uses weight hop based packet scheduling for AODV routing protocol to reduce the queue length caused by link failure in the network.

The cost of these routing protocols is additional control packets, which is advantageous in a sparse network yet brings some shortcomings in dense swarm network. Because each node has many neighbor nodes in high-density network, there will be a large number of control packets in the entire network when a node broadcasts a routing update packet, which may cause congestion and increase delay. So, the control overhead of routing update must be as small as possible.

Researchers improved TORA considering the network life, bandwidth demand and load balance respectively in [11][12]. These protocols utilize the advantage of link reversal mechanism in TORA. The link reversal can be defined as follows. At the beginning, the link from node X to node Y is allowed to transmit data packets, which is known as a downlink. But the link from node Y to node X for transmission is prohibited. After the change of node status, the link from node Y to node X becomes the downlink. The allowable transmission direction of data packets is reversed. The link failure is the condition that may cause the change of node status. These protocols based on TORA can get the anomalies of routes and establish a new transmission path quickly. The overhead of updating is controlled in the local area, which reduces the overhead and delay. But all of the protocols assume that link reversal is successful. In fact, link reversal may fail due to the harsh environment, which will result in a lot of useless control overhead and increase the time to reestablish the transmission.

The proposed routing protocol in this paper not only keeps the advantage in TORA, but also reduces the control packets caused by routing reestablishment.

III. TORA PROTOCOL

As an on-demand routing protocol, TORA responds to link failure in the network quickly by link reversal. The mechanism limits the overhead at a local scale and avoids the control message flooding. Link reversal can increase bandwidth utilization and reduce the delay of routing reestablishment, which is suitable for high-dynamic mobile networks. Moreover, multiple paths from the source to the destination are created in TORA, which supports link reversal and multipath transmission.

A. Height Mechanism and Routing Establishment

A Directed Acyclic Graph (DAG) from a source to the destination is established by the “height” of nodes in TORA.

Data packets are only allowed to be transmitted from a higher node to a lower node, which is called downlink. The height is a 5-tuple (t, oid, r, v, id) , of which the first three values (t, oid, r) is called “reference level”. The meanings of the various parameters are as follows:

- 1) *Time t*; the time of creating the reference level;
- 2) *Identification oid*; the identity of the node creating the reference level;
- 3) *Reflection r*; the value of the reflection bit;
- 4) *Arab values v*; the hop-count information relative to the destination node;
- 5) *Identification id*; the identity of a node itself.

The reflection bit is a binary value: “r=0” means the reference level has not been reflected; “r=1” means that the reference level created is reflected back.

The “height” of a node is very important for routing establishment and link reversal. The value of the “height” is the basis to determine the downlink. The strategy to compare the “height” of $h_i(t_i, oid_i, r_i, v_i, i)$ and $h_j(t_j, oid_j, r_j, v_j, j)$ is shown in Fig. 1.

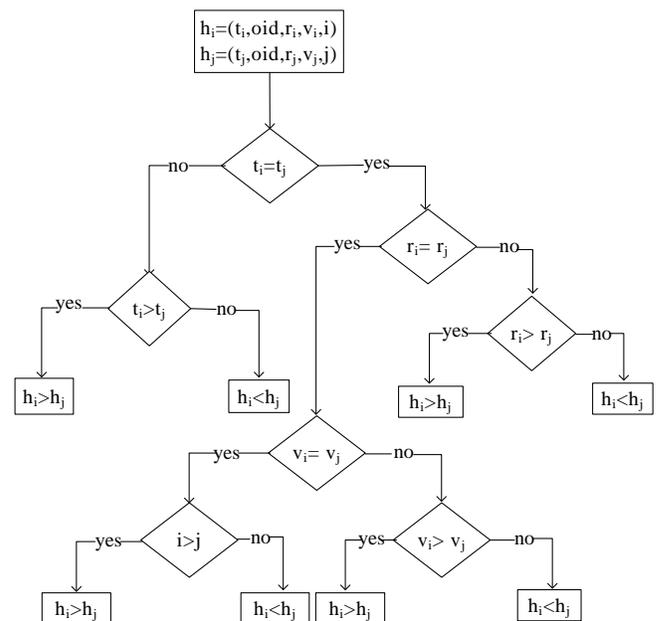


Figure 1. The comparison of height in TORA

Fig. 2 shows a simple network topology. If the source node S prepares to send data to the destination node D, it will broadcast the routing Query packet (QRY) to its neighbors. Each node that receives a QRY while not being the destination node will relay the control packet. When the destination node D receives the QRY from the source S at T_0 , node D creates the height $(T_0, D, 0, 0, D)$, then broadcasts an Update packet (UPD). Each node receiving UPD defines a height relative to the destination node D, such as node C and E in Fig. 2. After node C and E define their heights, they also send a UPD to indicate their heights. As a result, their neighbors can define the heights too. Finally, all nodes will define a height relative to the destination and get the height of its neighbors. Each node will also know the downlinks. In the storage of a node, its own height, the height of its

neighbor nodes and the link-status information will be saved. The DAG is formed as shown in Fig. 2. The source node S can find two different shorter routes to transmission: S-A-B-C-D and S-A-B-E-D.

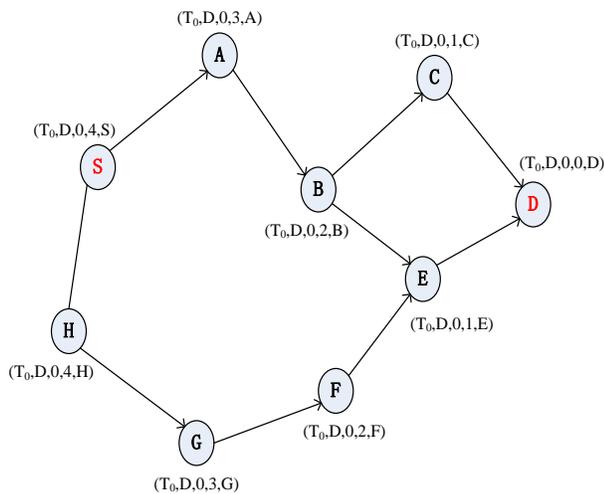


Figure 2. DAG of TORA

B. Strength and Weakness of Link Reversal in TORA

The nodes losing the last downlink will perform link reversal to update the route. We assume that the current transmission path is S-A-B-C-D in Fig. 2. If the link from node C to node D is interrupted at T_1 , node C will lose the downlink to node D, which results in the change of the topology. Firstly, node C replaces previous height with new height $(T_1, C, 0, 0, C)$, and then sends a UPD with the new height. The downlink from node B to node C reverses. Secondly, node B has another downlink from node B to node E, so it neither changes the height nor generates a UPD. The new DAG is shown in Fig. 3.

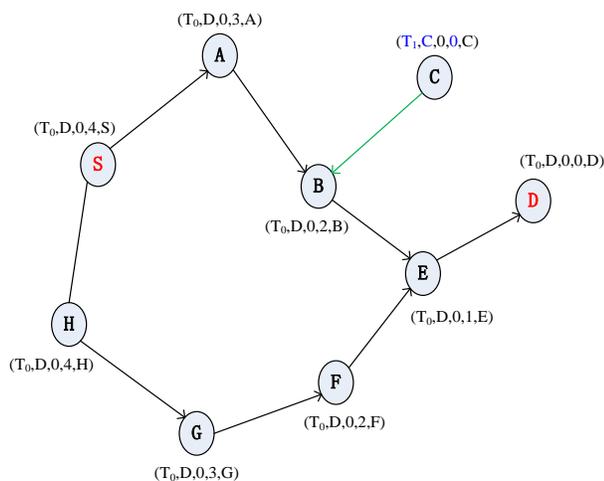


Figure 3. DAG of TORA with new height of node C

In the above case, the link reversal greatly reduces control overhead, and quickly finds another transmission

path (S-A-B-E-D). Although the multipath in TORA supports the link reversal, link reversal failure will still occur under some conditions. In Fig. 3, for example, the downlink from node E to node D is interrupted at T_2 , and the route will be updated as follows:

- Node E changes its height: $(T_2, E, 0, 0, E)$, and then sends a UPD.
- Node B and F lose the last downlink, and update the height: B $(T_2, E, 0, -1, B)$, F $(T_2, E, 0, -1, F)$, and broadcast a UPD severally.
- Node C, A, G, H perform the same operation: C $(T_2, E, 0, 0, C)$, A $(T_2, E, 0, -2, A)$, G $(T_2, E, 0, -2, G)$, H $(T_2, E, 0, -3, H)$.
- Node S changes the reflection bit of the reference level and updates the height $(T_2, E, 1, 0, S)$, and becomes the highest node in the local area. All related nodes change the height successively: H $(T_2, E, 1, -1, H)$, A $(T_2, E, 1, -1, A)$, G $(T_2, E, 1, -2, G)$, B $(T_2, E, 1, -2, B)$, F $(T_2, E, 1, -3, F)$, C $(T_2, E, 1, -1, C)$.
- When receiving a UPD from node B and node F, node E realizes that the reference level created by itself is reflected back. Then node E changes the height into NULL $(-, -, -, -, E)$, and broadcasts a cleared packet (CLR). Other nodes that receive a CLR delete the relational routes and set their heights NULL.

As we see, the link reversal is invalid. A large number of control packets are flooding the network, which results in more control overhead and greater delay of routing reestablishment. Hence, it has a significant impact on network transmission.

IV. RTORA PROTOCOL

In order to solve the problem resulting from the link reversal failure, we must prevent useless control packets from flooding. The RTORA protocol adopting reduced-overhead mechanism can ease this problem effectively.

Compared with TORA, RTORA also applies the height mechanism and the link reversal. Each node stores the information including the height list of neighbors and the link-status list. RTORA has some differences from TORA: The height of a node is defined as a 4-tuple (t, oid, v, id) , which deletes the reflected bit. Once a node that is not the source node performs the link reversal, its height will be changed into NULL. The nodes with the height of NULL cannot participate in the routing updating until a new QRY arrives. Furthermore, the nodes receiving a CLR decide what to do according to various situations instead of direct deletion in TORA.

The main principle of routing maintenance is shown in Fig. 4. We call it reduced-overhead mechanism. All nodes receiving a UPD or CLR update the link-status list and the height list of their neighbors, and then check the link-status list. If another downlink exists, the nodes do nothing. Otherwise, the link reversal will be operated. If the node preparing to perform the link reversal is not the source node, it will change the height into NULL and send a CLR. On the other hand, the node having other neighbors without NULL will update its height to become the highest in the local area

and send a UPD; the node having no other neighbors will create and send a new QRY.

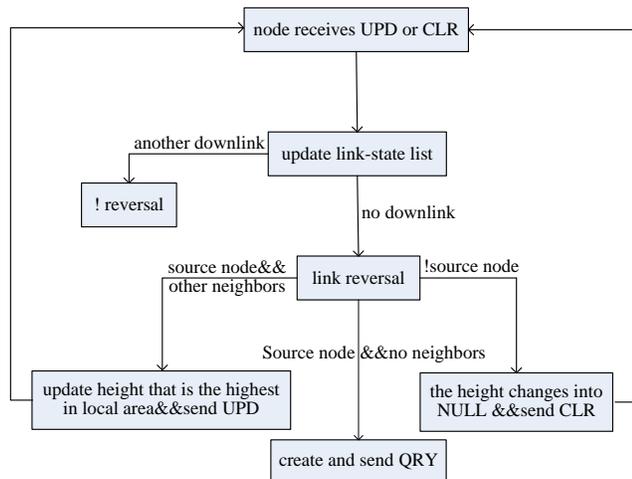


Figure 4. The reduced-overhead mechanism of RTORA

We still consider the example in Section III. The downlink from node C to node D is interrupted in Fig. 2. Node C changes the height into NULL and broadcasts a CLR. Because node B losing the downlink to node C has another downlink to node E, it deletes the information concerning node C in the link-status list without changing its own height or sending a CLR. The new topology is shown in Fig. 5.

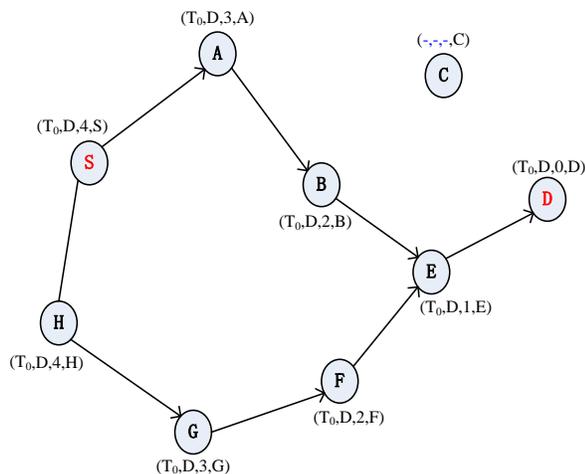


Figure 5. DAG of RTORA with new height of node C

If the downlink from node E to node D fails later, the sequence of events will happen as follows:

- Node E will replace the height by NULL, and send a CLR.
- Node B, F, C, A, G, H will change their heights into NULL.
- The source node S has no neighbors whose heights are not NULL, so it will delete routing information and broadcast a QRY to establish a new DAG.

The source node becomes a special node due to the change of the routing maintenance strategy. If we do not pay

attention to this situation, a bad influence will appear. As shown in Fig. 5, we assume the downlink from S to A fails. Node S will change the previous height into NULL and send QRY to establish a route. But, in fact, node S only changes the height into S ($T_0, D, 5, S$), which is the highest in the local area. Then, a new transmission path (S-H-G-F-E-D) will be found quickly.

V. SIMULATION AND ANALYSIS

The network model is built and simulated by OPNET [13][14]. We assume that 30 UAV nodes are randomly deployed in an area of 4 square kilometers. In order to build a dense network, the communication radius of UAV is configured with the value of 400 meters. This will ensure that there are enough neighbors for every UAV node. The main parameters of TORA and RTORA routing protocols are given in Table I. These values are most often adopted in TORA simulation research.

TABLE I. MAIN PARAMETERS OF TORA AND RTORA

parameters	value
Mode of Operation	On demand
Beacon Period	3 second
Max Beacon Timer	30 second
Max Retries	3

A. Traffic Configuration and Mobility Model

Low-resolution video conferencing service is configured in the simulation. The source node generates a frame (128×120 pixels) at a rate of 10 frames per second. The values of main parameters are shown in Table II. The start-time is allowed for routing initialization. The random waypoint (RWP) mobility model is used where nodes move randomly in an area of 100 meters radius around its current position at the speed of 20 meters per second. The value of pause time is set to 0 second.

TABLE II. MAIN PARAMETERS OF TRAFFIC

parameters	value
Frame Interarrival Time Information	10 frames per second
Frame Size Information	128×120 pixels
Start Time	100 second
Duration	End of Simulation

B. Performance Metrics

In the swarm network with UAV nodes, node mobility will cause link failure. The routing reestablishment necessarily increases control overhead and delay. Furthermore, the congestion caused by the broadcasting of dense nodes will increase delay too. So we choose control overhead and average end-to-end delay as the global statistics to consider.

1) *Control overhead*; the number of bytes generated by routing discovery, routing establishment and routing maintenance in one second.

2) *Average end-to-end delay*; the delay caused by data transmitted from source nodes to destination nodes, including the waiting time of routing establishment, transmission and propagation delay.

C. Simulation Results

In order to guarantee the accuracy of the experimental results, five experiments with different simulation seed (100, 256, 512, 800, 1000) are run. The simulation time is set to 600 seconds. The results are shown in Fig. 6 and in Fig. 7, whose simulation seed is set to 512 (similar to other results with different simulation seed).

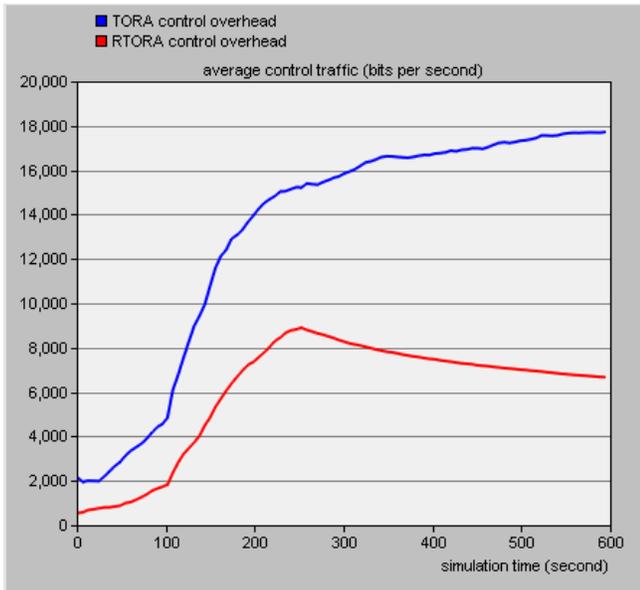


Figure 6. The control overhead of TORA and RTORA

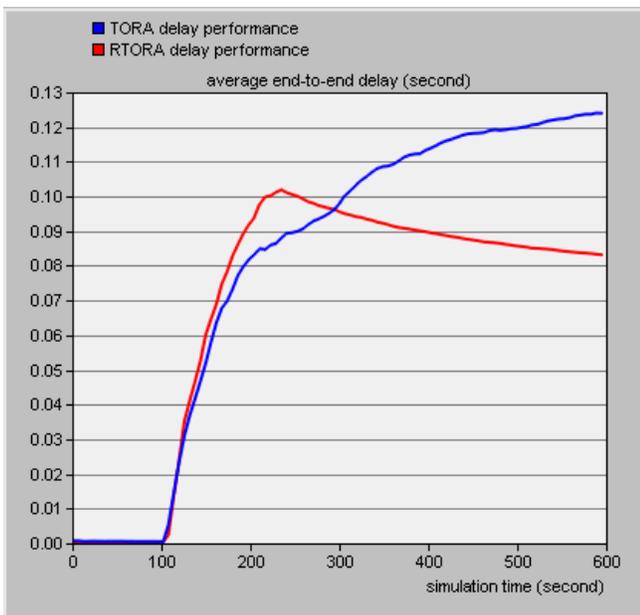


Figure 7. The average delay of TORA and RTORA

Fig. 6 shows that the control overhead of RTORA is reduced by about 56% compared with TORA, which can effectively improve network bandwidth utilization and reduce the probability of congestion caused by the

communication overload. In Fig. 7, the delays of two protocols are close in the beginning, because the link reversal is successful when there are multiple paths. But with more link failures, the probability of link-reversal failure increases. In this situation, RTORA can quickly inform to the source to reestablish route and effectively reduce the control overhead. So the average delay is reduced by about 25% compared with TORA.

According to the above results, the routing maintenance strategy in RTORA plays the advantages of link reversal better. Regardless of success or failure of the link reversal, the protocol can quickly reestablish a new transmission path. Especially in the highly-dynamic networks, the probability of link failure and link-reversal failure greatly increases. If link reversal fails, the reduced-overhead mechanism in RTORA shortens the time of routing reestablishment and minimizes the control overhead. The routing update strategy in RTORA can avoid the flooding of control packets, so the probability of transmission congestion will be lower. Thus, RTORA further improves the delay performance of the network.

VI. CONCLUSION AND FUTURE WORK

In this paper, we analyze the characteristic of swarm network with UAV nodes, and discuss the shortcomings of current routing protocols. Based on the link reversal in TORA, we propose the RTORA protocol that adopts the so-called reduced-overhead mechanism and thus solves the problem resulting from the link-reversal failure in TORA. The simulation results show that RTORA has lower control overhead and better end-to-end delay performance in the harsh environment assumed. We will consider RTORA protocol using location-based information in the future work.

REFERENCES

- [1] Y. Khamayseh, O. M. Darwish, and S. A. Wedian, "Ma-aodv: Mobility aware routing protocols for mobile ad hoc networks," in *Systems and Networks Communications, 2009. ICSNC'09. Fourth International Conference on*, 2009, pp. 25–29, doi: 10.1109/ICSNC.2009.80.
- [2] D. B. Johnson and D. A. Maltz, "Scalable support for transparent mobile host internetworking," *Wireless Networks*, 1995, vol. 1, pp. 311–321, doi: 10.1007/BF01200849.
- [3] C. E. Perkins and E. M. Royer, "Ad-hoc on-demand distance vector routing," in *Mobile Computing Systems and Applications. Proceedings. WMCSA'99. Second IEEE Workshop on*, 1999, pp. 90–100, doi: 10.1109/MCSA.1999.749281.
- [4] V. D. Park and M. S. Corson, "A highly adaptive distributed routing algorithm for mobile wireless networks," in *INFOCOM'97. Sixteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, 1997, vol.3, pp. 1405–1413, doi: 10.1109/INFCOM.1997.631180.
- [5] M. Kassim, R. Ab Rahman, and R. Mustapha, "Mobile ad hoc network (MANET) routing protocols comparison for wireless sensor network," in *System Engineering and Technology (ICSET), 2011 IEEE International Conference on*, 2011, pp. 148–152, doi: 10.1109/ICSEngT.2011.5993439.
- [6] P. Kuppasamy, K. Thirunavukkarasu, and B. Kalaavathi, "A study and comparison of OLSR, AODV and TORA routing

- protocols in ad hoc networks,” in *Electronics Computer Technology, 2011 3rd International Conference on*, 2011, vol. 5, pp. 143–147, doi: 10.1109/ICECTECH.2011.5941974.
- [7] Z. Che-Aron, W. F. M. Al-Khateeb, and F. Anwar, “ENFAT-AODV: The fault-tolerant routing protocol for high failure rate Wireless Sensor Networks,” in *Future Computer and Communication (ICFCC), 2010 2nd International Conference on*, 2010, vol. 1, pp. V1–467, doi:10.1109/ICFCC.2010.5497747.
- [8] J. Liu and F. Li, “An improvement of AODV protocol based on reliable delivery in mobile ad hoc networks,” in *Information Assurance and Security, 2009. IAS’09. Fifth International Conference on*, 2009, vol. 1, pp. 507–510, doi: 10.1109/IAS.2009.78.
- [9] V. Ramesh and P. Subbaiah, “Preemptive AOMDV routing for mobile Ad-hoc networks,” in *Sustainable Energy and Intelligent Systems (SEISCON 2011), International Conference on*, 2011, pp.622–625, doi: 10.1049/cp.2011.0436.
- [10] D. K. Sharma, C. Kumar, S. Jain, and N. Tyagi, “An enhancement of AODV routing protocol for wireless ad hoc networks,” in *Recent Advances in Information Technology (RAIT), 2012 1st International Conference on*, 2012, pp. 290–294, doi: 10.1109/RAIT.2012.6194522.
- [11] D. Dharmaraju, A. Roy-Chowdhury, P. Hovareshti, and J. S. Baras, “INORA-a unified signaling and routing mechanism for QoS support in mobile ad hoc networks,” in *Parallel Processing Workshops. Proceedings. International Conference on*, 2002, pp. 86–93, doi: 10.1109/ICPPW.2002.1039716.
- [12] F. Yu, Y. Li, F. Fang, and Q. Chen, “A new TORA-based energy aware routing protocol in mobile ad hoc networks,” in *Internet, 2007. ICI 2007. 3rd IEEE/IFIP International Conference in Central Asia on*, 2007, pp.1–4, doi: 10.1109/CANET.2007.4401666.
- [13] M. Rajput, P. Khatri, A. Shastri, and K. Solanki, “Comparison of Ad-hoc reactive routing protocols using OPNET modeler,” in *Computer Information Systems and Industrial Management Applications (CISIM), 2010 International Conference on*, 2010, pp. 530–534, doi: 10.1109/CISIM.2010.5643454.
- [14] X. Chang, “Network simulations with OPNET,” in *Simulation Conference Proceedings, 1999 Winter, 1999*, vol. 1, pp. 307–314, doi: 10.1109/WSC.1999.823089.

Towards the Definition of a Mobility-based Clustering Environment for MANET

Aida Ben Chehida, Ryma Abassi, Sihem Guemara El Fatmi
Higher School of Communication, Sup'Com
University of Carthage
Tunis, Tunisia

Emails : {bechehida.aida, ryma.abassi , sihem.guemara } @supcom.rnu.tn

Abstract— A MANET is a collection of mobile nodes cooperating using low bandwidth wireless links. Its use is increasingly growing because it can be rapidly and easily deployed, and allows users to access and manipulate data anytime and anywhere. Because of these characteristics, classical routing protocols cannot be applied in such environment. To be efficient, routing protocols in MANET should in fact manage mobility, handle nodes energy dissipation and ensure security. Paradoxically, these constraints are not really taken into account. This paper presents a first step of our global objective towards the definition of a complete and generic security environment for MANET routing protocols. Its main contribution is then to propose a mobility-based clustering algorithm for MANET constituting the basis for our security architecture. From a given configuration of mobile nodes, we focus our attention on the definition of two major phases of the clustering process life cycle: (1) the *setting up phase* where clusters are generated and cluster-heads are elected and (2) the *clusters maintenance phase* where the organization of the clusters is maintained in the presence of mobility. This phase reacts to all network topology changes that may occur in the network such as the displacement of a node, the failure of a node, or the arrival of a new node. Simulation experiments are conducted to evaluate the performance of our algorithm in terms of the number of cluster-heads, the node lifetime and the packets delivery ratio.

Keywords-MANET; clustering algorithm; energy; mobility.

I. INTRODUCTION

A Mobile Ad hoc NETWORK (MANET) is a set of battery-powered mobile nodes connected by low bandwidth wireless link and cooperating with each other to route packets [1]. Because routing process is essential for a successful MANET deployment, efficient routing protocols are necessary. Unfortunately, because nodes are mobile and behave simultaneously as hosts and as routers, classical routing protocols cannot be applied to this type of networks. This is why some specific routing protocols have been proposed for MANET. The main disadvantage of the majority of these protocols is that they lose their efficiency when the network size is large or the mobility of nodes is high [1]. One way to overcome this problem is to organize a MANET into clusters. Clustering in MANET is used to organize nodes into groups (clusters) characterized by cluster-head (CH) and member nodes [2]. This organization minimizes the amount of storage for communication information, makes the routing process easier, optimizes the use of the network bandwidth, etc. Clustering is generally deployed using two phases: setting up

and maintenance. In the first phase, some nodes are chosen to act as coordinators (CHs) and each CH is associated to some members' nodes, the whole making one cluster. CHs are responsible for coordination among the nodes within their clusters (intra-cluster coordination) as well as communicating with other CHs (inter-cluster communication). Because the network topology changes over time (displacement, failure, arrival or departure of a node), a clustering maintenance is required to update the cluster organization.

In this paper, we propose a novel clustering approach for MANET taking into account nodes mobility. This proposition is the first step towards the definition of a complete and generic security environment for MANET routing protocols. This security environment that will be based on a Watchdog mechanism imposes to cluster members' nodes to be 'one-hop neighbors'. In our context, this constraint improves the intra-communication quality due to the interference and reduces the energy consumption. In the initial phase of the algorithm, a cluster is composed of one-hop members and its CH is the node having the smaller weight according to the mobility and energy parameters. The maintenance phase is performed according to the several kinds of changes that may happen i.e. link failure and/or new link appearance. The remaining part of this paper is structured as follows. Section 2 reviews some most related clustering protocols. Section 3 introduces the proposed clustering algorithm. Section 4 focuses on the clustering maintenance phase. Simulation results are presented in Section 5. Finally, Section 6 concludes this paper.

II. RELATED WORKS

In the literature, several clustering algorithms were proposed for MANET. These algorithms can be classified into proactive (most commonly used) and reactive ones [4].

Proactive algorithms differ from each other by the CH selection criteria in the setting up phase. In the lowest-ID algorithm [5], the node with the lowest ID within its closed neighborhood is selected as a CH. The cluster is then formed by the selected CH and all its neighbors. This criterion might be an inconvenient since every time node IDs are reshuffled, the neighboring list of all the nodes needs to be changed. In the CONnectivity-based clustering algorithm CON [6], nodes broadcast their identifier (ID) and according to the number of received IDs, every node computes its degree. The node with the highest degree is selected as CH. The major drawback of this algorithm is that the node degree may change very

frequently which, may inhibit CHs to play their role for very long. In the previous algorithms, the election of CHs does not take into consideration the node's energy, which may lead to battery drainage. This is all the more unacceptable, given that CHs battery drainage causes a frequent re-invocation of the clustering algorithm. The lowest MOBility Clustering algorithm MOBIC [7] is another proactive approach that ameliorates the previously cited works by considering mobility during CH election. However, it does not take into account the energy as an election criterion making possible a rapid failure among the CHs. Globally, proactive algorithms lead not only to useless exchanged clustering messages (high computational overhead) but also to frequent updates in the clusters structure due to the periodicity of clustering process invocation. An original optimization for the proactive clustering algorithms was proposed by M. Elhdhili et al. [4]. In this work, a local reactive update is specified following the mobility of the node and the mobility of the CH.

M. Chatterjee et al. [8] described a reactive protocol called Weighted Clustering Algorithm (WCA). In WCA, CHs are selected based on a collection of attributes: the ideal number of nodes it can support, mobility, transmission power and battery power. The node with the minimum weight is selected as CH. To maintain the cluster organization, the CH chooses new CHs for its member nodes going far from it. If a mobile node cannot reach any existing CH, it re-invokes the clustering algorithm to form new clusters. This might be an inconvenience, especially for high mobility and it generates an important computational overhead.

To our best knowledge, there is no existing work dealing with clusters with one-hop members. This characteristic is however useful for our work because the clustering process will be used as a basis for the security environment that we plan to define in a future work. For the moment, we will limit ourselves to the introduction of a mobility-based clustering algorithm for MANET. Our algorithm is designed to react to the unpredictable topology changes that may occur, i.e. link failure, new link, etc.

III. THE CLUSTERING SETTING UP PHASE

In this section, we describe the properties of the proposed mobility-based clustering algorithm and present the first phase: the setting up.

A. Basic properties

The mobility-based clustering algorithm has the following properties and/or conditions:

- It is completely distributed. Each node decides its own role (CH or member).
- It is adaptive to the changes in network, due to the node displacement, addition or failure.
- Each cluster is fully-connected in that sense that the members and the CH are one hop neighbors.
- Each node must belong to one and unique cluster.
- The clustering algorithm is based on a combined weight used in the decision of selecting a CH. The node having the lowest weight is elected as CH. The parameters of the weight are: the mobility and the residual energy of the node. The elected CH

should have the lowest relative mobility and the highest remaining battery power to allow the better stability to the clusters.

- ✓ The node mobility (M): To calculate the metric M for each node, the simple heuristic mechanism MOBIC [8] is used. Once calculated, M is included in the HELLO message sent by each one of the cluster nodes.
- ✓ The node residual energy (E): E can be easily retrieved from the node at a given time.

The setting up phase deployment requires that each node in the network discovers its one-hop and two-hop neighbors, computes its weight and broadcasts it. This is done during the pre-processing phase.

B. Pre-processing phase

Pre-processing involves the following two steps: *neighbors discovery* and *weight computing and exchanging*.

1) Neighbor discovery

Each node i finds its direct neighbors DN_i by periodically broadcasting a HELLO message including its identifier (ID) as well as its mobility index M_i and by waiting for an ACK_HELLO message. Formally, DN_i (1) is the set of i 's neighbor, let us say j , such that the distance between i and j is lesser than i 's transmission range $TXrange(i)$. V is the set of nodes in the network.

$$DN_i = \{j \in V / dist(i, j) < TXrange(i)\} \quad (1)$$

Note that ACK_HELLO message contains j 's neighbors. These latter are stored in the neighbor table as two-hop neighbors. These messages are summarized in Table I.

2) Weight computing and exchanging

Each node i computes its combined weight value W_i using the average mobility metric (M_i) and the node residual energy (E_i) by the following formula.

$$W_i = w_1(1 - E_i) + w_2 M_i \quad (2)$$

where w_1 and w_2 are the weights and $w_1 + w_2 = 1$. Once computed, each node i diffuses its weight to all its one-hop neighbors using a WEIGHT message as defined in Table I.

C. Setting up phase

The setting up phase consists of two components: *cluster identification* and *cluster-head election*.

1) Cluster identification component

This component is used to generate the restricted (one-hop) neighborhood. Because this latter may not be unique, a choice can be made according to the cluster mobility in order to ensure stable clusters.

a) The identification of the Restricted neighborhood

Each node i computes its restricted neighborhood RN_i . Each two nodes that belong to the same RN_i must be one-hop neighbors. Formally, RN_i is the set of i 's neighbor, let us say j , such that for each node k in RN_i , k and j are neighbors.

$$RN_i = \{j \in DN_i / k \in RN_i \rightarrow j \in DN_k\} \quad (3)$$

This identification may generate several sets of RN .

b) The choice of the restricted neighborhood

The chosen RN is the one having the least mobility. Knowing that each node receives the relative mobility metric M of its neighbors (in the HELLO message), it can calculate the mobility average of each generated RN . The chosen RN represents then the node's cluster.

Cluster identification is completed by broadcasting the RN message defined in Table I. Each node receiving this message is aware about the cluster appartenance of its neighbors, i.e. whether a neighbor belongs to the same cluster as it.

2) Cluster-head election component

Upon receiving the weights from all RN members, the nodes with the lowest weight among their RN neighbors declare themselves as CHs by multicasting (to all RN members) a CH message. All RN neighbors of the elected CHs join them as members by broadcasting JOIN messages as defined in Table I.

TABLE I. MESSAGES EXCHANGED AND NOTATIONS: CLUSTERING SETTING-UP

Message	Meaning
HELLO (my_ID, my_M)	Notify neighbors about my ID and my relative mobility M .
ACK_HELLO (my_ID, list_my_neighbors)	Notify neighbors about my ID and my one-hop neighbors.
WEIGHT (my_ID, my_W)	Notify neighbors about my ID and weight.
RN (my_ID, my_RN)	Notify neighbors about my ID and RN.
CH (CH_ID, CH_Member)	Notify RN neighbors about my role: I am a CH, my ID is CH_ID and my members are CH_Member.
JOIN (my_ID, CH_ID)	Notify neighbors that I am going to join the cluster whose CH's ID is CH_ID.

If two nodes have the same weight, then the node with the smaller ID becomes the CH.

IV. THE CLUSTERING MAINTENANCE PHASE

As explained previously, the main contribution of this work concerns mobility handling in clustering environment. In this section, we introduce the proposed algorithms. Note that, when describing the procedures of our algorithm, we assume that each node has already performed the pre-processing phase, i.e., discovered its one-hop and two-hop neighbors, computed its weight and broadcasted it.

A. Initialization

Initialization is used either during the setting up phase to form clusters and to elect CHs or after the clustering setting up when a novel node joins the network or when a node is detached from its cluster. Such a node executes the procedure INIT in order to determine their own role (setting up phase) or to know in which cluster they will belong (a new arriving node or a moving node). Note that this is the first procedure that is executed by each node in the network.

Before the clustering setting up, if a node i joins the network, it participates to the cluster identification and CH election as presented in Section III.

After the clustering setting up, if a new arriving node i joins the network or a moving node i detaches from its cluster, it can be detected by a CH or by a member node.

If a cluster-head j detects a node i , it checks if i is neighbor with all its cluster members (j knows the i 's neighbors). If that is the case, j sends to i the NEW_CH message containing its ID , its cluster members and a flag set to 0. This flag indicates to i that it receives the message NEW_CH from a CH and that it is neighbor with all the j 's members. When receiving this message, i may join the j 's cluster by broadcasting the JOIN message. If node j receives the JOIN message from i , it adds this node into its cluster.

However, if i is not neighbor with some j 's cluster members, node j sends to i the NEW_CH message with the flag set to 1. This flag will indicate to the node i that it receives the message NEW_CH from a CH but it is not neighbor with all the j 's cluster members. If node i sends the JOIN message to j , node j creates a new cluster with i (and eventually some members nodes which are neighbors with i in order to equilibrate the two clusters), leaves its function of CH to one of its member node and informs its members by multicasting the ELECTION message. In both cases (i is neighbor with all j 's cluster members or not), j does not send the NEW_CH message if the number of its members is less than two.

If a member j detects the node i , it first ensures that its CH k is not neighbor with i . In such case, it sends MEMBER message to its CH k and waits for the CH decision. If k authorizes node j to create a cluster with i , node j sends to i the NEW_CH message with the flag set to 2. This flag will indicate to node i the reception of a NEW_CH message from a member node. If node i broadcasts the JOIN message, node k removes the node j from its cluster and node j creates a cluster with node i . In both cases (i is detected by a CH or by a member node), if node i detects more than one NEW_CH message, it joins the node that send the NEW_CH with flag set to 0. However, if node i does not detect any NEW_CH message with flag set to 0, it joins the node's cluster that send NEW_CH message with flag set to 1. Else, node i joins the node that send NEW_CH message with flag set to 2. If node i detects more than one NEW_CH message with flag set to 0, 1 or 2, the node i joins the cluster with the lowest CH weight. The messages exchanged are defined in Table II.

TABLE II. MESSAGES EXCHANGED AND NOTATIONS: CLUSTERING MAINTENANCE

Message	Meaning
NEW_CH(CH_ID, CH_Member, Flag)	This message is sent by a CH. It includes its ID, its cluster members and a flag.
MEMBER (my_ID, new_node_ID)	This message is sent by a member node to its CH to indicate the detection of a new node.
ELECTION (my_ID, New_CH_ID)	This message is sent by a CH. It includes its ID and the new elected cluster-head ID.

The following algorithm depicted in Fig. 1 illustrates the procedure INIT.

```

Procedure INIT;
Begin
    If (Receivej NEW_CH (j, Memberj, "0"))
        Then Send JOIN (i, j);
        Else If (Receivej NEW_CH (j, Memberj, "1"))
            Then
                Send JOIN (i, j);
            Else If (Receivej NEW_CH (j, Memberj, "2"))
                Then
                    Send JOIN (i, j);
        Else (i Form RNi)
            If (  $\forall x \in RN_i, w_i < w_x$  ) Then
                Begin
                    SendRNi CH (i, Memberi);
                    clusterheadi := i;
                End
            Else If (Receivej CH (j))
                Then
                    If (  $\forall x \in RN_i, w_j < w_x$  ) Then
                        Begin
                            clusterheadi := j;
                            Send JOIN (i, j);
                        End
                    End
End;
    
```

Figure 1. Initialization algorithm

The second step concerns link failure handling. It is detailed in the following subsection.

B. Link failure handling

When a node i detects the failure of node j , there are two possibilities: (1) j was a cluster member and i its CH (2) j was a CH. In the first case, j is dropped from i 's cluster where in the second case, i compares its weight with its RN_i neighbors weights. If i has the lowest weight, it considers itself as CH and informs the rest of the RN members by sending a CH message. Else, node i simply waits for this message from another node. This is illustrated by the following algorithm depicted in Fig. 2.

```

Procedure Failure_Node;
Begin
    If (clusterheadi == i) And (j ∈ cluster)
        Then clusteri := clusteri / {j};
    Else if (clusterheadi == j)
        Then
            If (  $\forall x \in RN_i, w_i < w_x$  ) Then
                Begin
                    Sendj CH (i, Memberi);
                    clusterheadi := i;
                End
            Else If (Receivek CH (k, Memberk))
                Then
                    If (  $\forall x \in RN_i, w_k < w_x$  )
                        Then
                            Begin
                                clusterheadi := k;
                                SendRNi JOIN (i, k);
                            End
                    End
End;
    
```

Figure 2. Failure node handling algorithm

C. New link handling

As detailed by the following procedure depicted in Fig. 3, when a node j moves inside the network (j is detached from its cluster), it may be detected by a CH or a member i . For the first case, i checks if j is neighbor with all its cluster members then unicasts a NEW_CH message such as presented previously. In the second case, i notifies to its CH k the existence of a new node j using a MEMBER message. Note that the decision of integrating this node in an existing cluster is left to the CH.

```

Procedure New_link;
Begin
    If (clusterheadi == i) Then
        If (j is_neighbor RNi) Then
            FSendj NEW_CH (i, Memberi, "0");
        Else
            Begin
                SendRNi ELECTION (i, k) / k ∈ RNi;
                Sendj NEW_CH (i, Memberi, "1");
            End
        Else
            Begin
                Sendk Member (i, j) / clusterheadi == k;
                Sendj NEW_CH (i, Memberi, "2");
            End
    End;
End;
    
```

Figure 3. New link detection algorithm

In the rest of this paper, some simulations and results are presented in order to show the feasibility of our proposition.

V. SIMULATION AND RESULTS

The aim of the following section is to study the performance of our proposition by a simulation work. Several simulations were achieved on a 100m*1000m grid while considering different network sizes varying between 10 and 40 mobile nodes, a transmission range varying between 1 and 250 m and a node speed between 500 and 3000 m/s. The simulation parameters used are listed in Table III.

TABLE III. SIMULATION PARAMETERS

Parameter	Meaning	Value
N	Number of nodes	10-40
X*Y	Network size	1000m*1000m
TR	Transmission range	1m-250m
S	Node speed	500-3000 m/s
T	Data traffic	CBR (Constant Bit Rate); data payload=512 bytes Rate= 4 packets/s
B	Node bandwidth	2 Mbps
Run Time	Time of simulation	100 sec

To measure the performance of the proposed clustering algorithm, we consider the following performance parameters:

- The node lifetime: the duration from the beginning until the node runs out of its battery power.
- The number of clusters, also known as number of CHs, defining the number of logical partitions formed in the network with the mobile nodes.
- The Packet Data Ratio (PDR): the ratio between the number of data packets received by the destination node and the number of data packets transmitted by the source node. It represents the reliability of packet transmissions in a network.

Fig. 4 depicts the numbers of CHs with varying number of mobile nodes in the network. Obviously, formed clusters number depends of the number of existing nodes. However, such as depicted by the figure, this dependence is linear since we constructed clusters optimally while trying to have some equality between numbers of nodes per cluster. For instance, with thirty nodes five clusters were constructed while with forty five nodes six clusters were created.

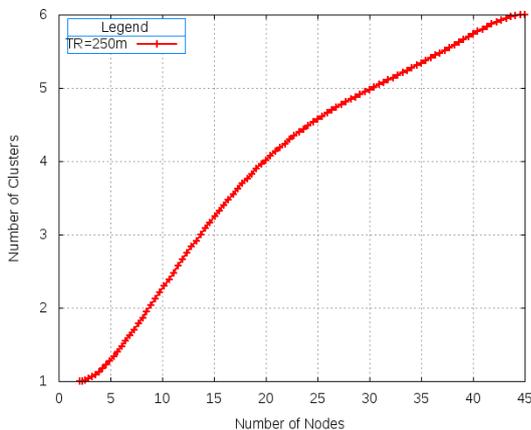


Figure 4. Number of formed clusters with varying number of nodes in the network

The second simulation scenario depicted in Fig. 5 shows the evolution of node lifetime according to the number of nodes in the network and for two different transmission ranges: 10m and 125m.

We observe that node lifetime decreases slightly when the number of nodes increases. In fact, it ranges between 98s and 80s when there are twenty more nodes. This can be explained by the fact that the increase of the nodes number leads to a larger amount of traffic which contribute to a faster depletion of nodes battery. Regarding transmission range modification, we observe a low variation of values. Hence, transmission range has a small impact on node lifetime.

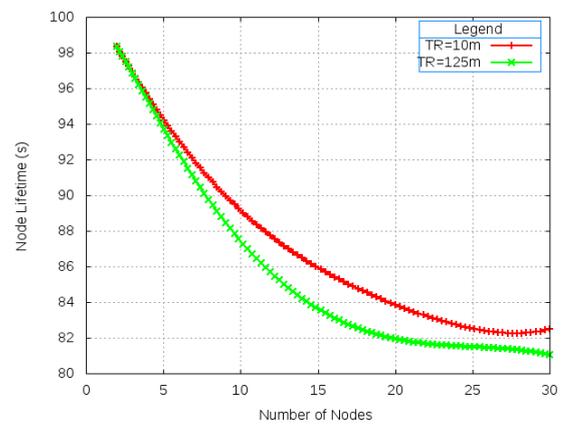


Figure 5. Node lifetime with varying number of nodes in the network

The third simulation scenario evaluates the number of clusters according to the transmission range variation and for different network densities (N=20, N=30 and N=40 nodes). As depicted by Fig. 6, we notice a large number of clusters for small transmission range because this latter leads to limited node visibility and consequently to a small clusters covering area. Moreover, we obtain a reduced number of clusters for large transmission range because in such case clusters cover larger areas. Note that this result is substantially the same for all three considered network densities. In fact, the number of clusters varies in a small interval almost [4, 8] for 20 nodes, [5, 13] for 30 nodes and [6, 13] for 40 nodes.

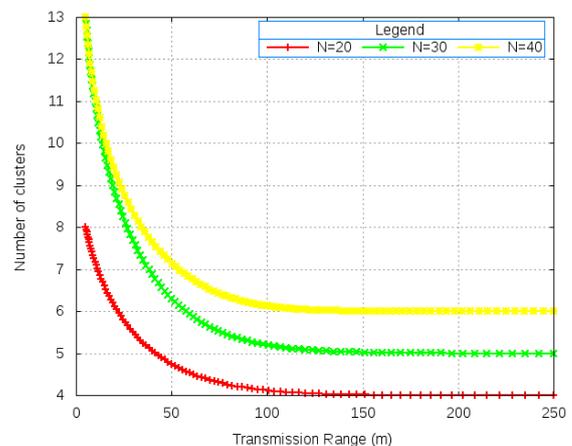


Figure 6. Number of clusters with varying number of nodes in the network

Fig. 7 depicts a simulation scenario considering the variation of a node lifetime with respect to a varying transmission range. We notice that node lifetime decreases as transmission range increases. This may be explained by the fact that when the transmission range increases, node visibility increases too leading to more exchanged traffic that contributes to lower the node lifetime.

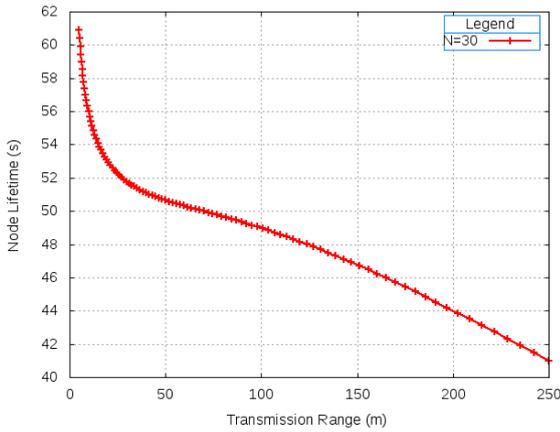


Figure 7. Node lifetime with varying transmission ranges

Fig. 8 depicts the variation of the number of clusters with respect node speed. We notice that this number is stable. In fact, with increase in speed, cluster members change clusters frequently; however, the number of clusters shows a low variation in value.

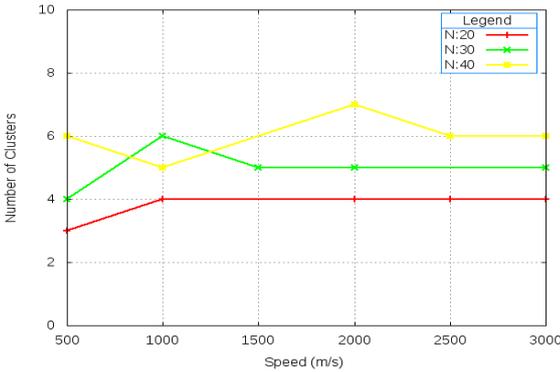


Figure 8. Number of clusters with varying nodes speed

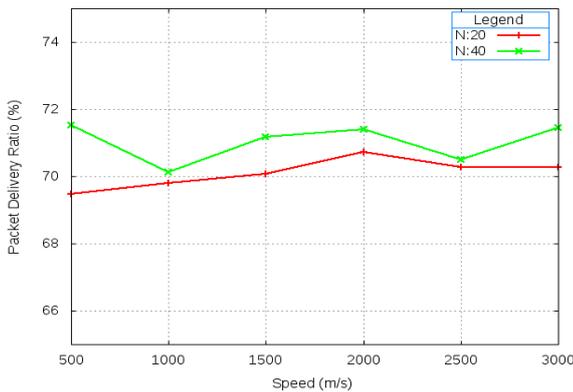


Figure 9. PDR with varying node speed

The last simulation depicted in Fig. 9 is concerned with the variation of the PDR when varying speed of nodes in the network. Since we have a normal environment without attackers target, nodes receive almost all generated data packets from the source node.

The results obtained demonstrate that our mobility based clustering algorithm selects optimal cluster heads and consequently achieves stability, reliability, and low maintenance. This is all the more interesting that our algorithm dealt with some specific initial constraints needed, as explained previously, for our trust management process enforcement.

VI. CONCLUSION AND FUTURE WORK

The dynamic topology of MANET, as well as the limited node capability and bandwidth pose significant problems for wide networks. Generally, and in order to resolve this problem, several propositions based on clustering techniques have been addressed. The major drawback of these techniques is that they do not take into account the mobility issue in an optimal way. This drawback has led us to propose a mobility-based clustering algorithm to form and maintain more stable clusters. The proposed clustering algorithm is based on two phases: the setting up and the maintenance. The first phase forms the clusters then elects CHs with the minimum weight computed through two parameters: mobility and residual energy. The second phase maintains the organization of clusters in the presence of mobility due to the displacement of a node, the failure of a node, or the arrival of a new node.

Several simulations were conducted in order to evaluate our algorithm performances in terms of number of CHs, nodes lifetime and PDR. These latter were evaluated with varying speeds, nodes number and transmission range.

In further work, we aim to establish a trust process based on Watchdog, delegation and the proposed clustering algorithm.

REFERENCES

- [1] Z. Xing, L. Gruenwald, and K. Phang, "A Robust Clustering Algorithm for Mobile Ad-hoc Networks", Handbook of Research on Next Generation Networks and Ubiquitous Computing, IGI Global, Editor Samuel Pierre, Chapter 18, December 2008.
- [2] A. Nassuora and A. Hussein, "CBPMD: A New Weighted Distributed Clustering Algorithm for Mobile Ad hoc Networks (MANETs)", American Journal of Scientific Research ISSN, 1450-223X, Issue 22, 2011, pp. 43-56.
- [3] I. Shayeb, A. Hussein, and A. Nassuora, "A Survey of clustering schemes for mobile Ad-hoc Network (MANET)", American Journal of Scientific Research ISSN, 1450-223X, Issue 20, 2011, pp. 135-151.
- [4] M. Elhdhili, L. Azzouz, and F. Kamoun, "Lowest weight: reactive clustering algorithm for ad hoc networks", In Proceedings of the 11th IFIP International Conference on Personal Wireless Communications, pp. 135-146, September 2006, Spain.
- [5] A. Ephremides, J. Wieselthier, and D. Baker, "A design concept for reliable mobile radio networks with frequency hopping signaling", In Proceedings of the IEEE 75, January 1987, pp. 56-73.
- [6] A. K. Parekh, "Selecting routers in ad hoc wireless networks", In Proceedings of the SBT/IEEE International Telecommunications Symposium, August 1994, pp. 420-424.
- [7] P. Basu, N. Khan, and C. Little, "A mobility based metric for clustering in mobile ad hoc networks", In Proceedings of IEEE ICDCS workshop on wireless networks and mobile computing, April 2001, pp. 413-418, Phoenix, AZ.
- [8] M. Chatterjee, S. Das and D. Turgut, "WCA: A weighted clustering algorithm for mobile ad hoc networks", Journal of Cluster Computing, vol. 5, no. 2-4, April 2002, pp. 193-204.

Creating Frame Structures for Position-based Hybrid Routing Algorithm

Resul Kara

Computer Engineering Department
Faculty of Engineering, Duzce University
Duzce, Turkey
resulkara@duzce.edu.tr

Ibrahim Ozelik

Computer Engineering Department
Comp. & Information Sci. Fac., Sakarya University
Adapazari, Turkey
ozcelik@sakarya.edu.tr

Huseyin Ekiz

Suleyman Sah University
Istanbul, Turkey
hekiz@ssu.edu.tr

Murat Iskefiyeli

Computer Engineering Department
Comp. & Information Sci. Fac., Sakarya University
Adapazari, Turkey
miskef@sakarya.edu.tr

Abstract— Ad hoc wireless networks are composed of mobile nodes that communicate with each other without any infrastructure like a base station or access device. Routing algorithms of ad hoc networks can be classified as table-driven, on-demand and hybrid algorithms. Position-based Hybrid Routing Algorithm (PBHRA) is a hybrid routing algorithm that based on nodes position information. In this study, protocol implementation of PBHRA has been done. Frame structures of protocol have been formed. The protocol was simulated using these created frames. The simulation results show that the proposed frame structures suitable for ad hoc networks.

Keywords-routing protocol; wireless network; routing; ad hoc network

I. INTRODUCTION

Infrastructureless wireless networks, also known as Mobile Ad Hoc Network (MANET), are composed of randomly moving mobile nodes without a central control like a base station or a predefined infrastructure. At the present time, these mobile nodes -that may take part in airports, ships, trucks, automobiles, very small devices and on people- are prevalently used in many industrial and commercial applications. This mentioned place necessitates the mobility of nodes. Therefore, mobile nodes in ad hoc networks must use routing algorithm in order to communicate each other.

Routing algorithms for MANET can be classified as table-driven, on-demand and hybrid algorithms. PBHRA that has been developed by us a hybrid routing algorithm depends on the node position information. PBHRA and its performance analysis have been presented in [1].

It is required to establish packet structures for the hardware and protocol realization of the routing algorithms used for ad hoc and sensor networks. This requirement involves determination overhead of packet structure to effect of performance. It is important for performance examination of algorithms. Accordingly, many researchers have studied on determination of frame structure and protocol realization for MANET routing algorithm [2-9].

In [2], software is developed and an application environment is set to apply Ad Hoc On Demand Distance Vector (AODV) algorithm as protocol. A protocol realization is done on Ns-2 simulator for Dynamic Source Routing (DSR) and AODV algorithms and the comparisons related to the functioning of these two algorithms as protocol are done [3].

Protocol design and realization of Fisheye Routing Algorithm that aims to decrease the consumption of band width studied in [4]. Holter et al. [5] made an implementation over Quagga Routing software protocol for the realization of wireless Open Shortest Path First (OSPF) algorithm that aims to minimize protocol overloads [5]. A protocol realization is done on ns-2 simulator for DSR and AODV algorithms and the comparisons related to the functioning of these two algorithms as protocol are done [6].

Working steps of Optimized Link State Routing (OLSR) protocol, the selection of detailed multi point relay sets, frames used in protocol realization, and their structures are given in [7]. The general packet and heading structures have been described that is needed to be used for the packet exchanges of MANET routing algorithms in routers [8]. Since MANET routers themselves are the wireless nodes, these structures are also used for protocol realization. In another study, Neighborhood Discovery Protocol (NHDP) is introduced and frame structures are given [9]. NHDP uses a local exchange of HELLO messages so that each router can determine the presence of, and connectivity to, its 1-hop and symmetric 2-hop neighbors.

In this study, the proposed frame structures for PBHRA will be presented as an original study because they are different from the algorithms of PBHRA that are implemented above.

Depending on all these, in this study, PBHRA algorithm is Summarized in section II. In Section III, the frame structures for implementing PBHRA are given. In Section IV, according to evaluation consideration given in RFC 2501 [10], the parameters/values of PBHRP, which is developed in Matlab 7.01, are given; according to performance issues,

which are specified in same RFC, their analyses are also presented.

II. FRAME STRUCTURE OF POSITION BASED HYBRID ROUTING PROTOCOL (PBHRP)

PBHRA algorithm that is suggested in [1] entails to provide GPS in every node to determine the positions of the other nodes in the network like other position-based algorithms. According to the algorithm, the node in the center of the network behaves as a master node. Master node realizes route determination process using position, node density, and battery life information of the other nodes in the network.

PBHRA has the following steps [1]:

- a. The first activated node in the network is assigned as master node.
- b. Master node announces master node announcement packet periodically to make the network know that it is a master node.
- c. The other nodes in the network send update packets in which their position information exists to the master node.
- d. Master node constitutes position matrix using update information.
- e. Master node constitutes distance matrix after it calculates the distances in between every node by using position information.
- f. The node that corresponds to the smallest valued row after the row elements of position matrix are summed is determined as central and candidate master node is fixed.
- g. Current master node proposes to be the new master node by sending proposal packet to the master node.
- h. New master node sends master node announcement packet to the network.
- i. Other nodes send their update packets to new master node if necessary.
- j. Master node calculates cost criterions of every node by fuzzy logic method using distance and position matrices, and then fuzzy matrix is constituted [11].
- k. Optimization is done using fuzzy matrix to determine the paths between nodes with the lowest cost.
- l. Nodes ask for the shortest path when they send data to another node by sending path demand packet to the master node. Master node tells the shortest path to the node which demands according to the optimization result.
- m. Master node circuits its role to another appropriate node if it moves away from the center or its battery life decreases to the below of the threshold value.
- n. Other nodes in the network keep only the identity number and the position of the master node in their memories.

The flow chart of PBHRA is given in Figure 1 [1].

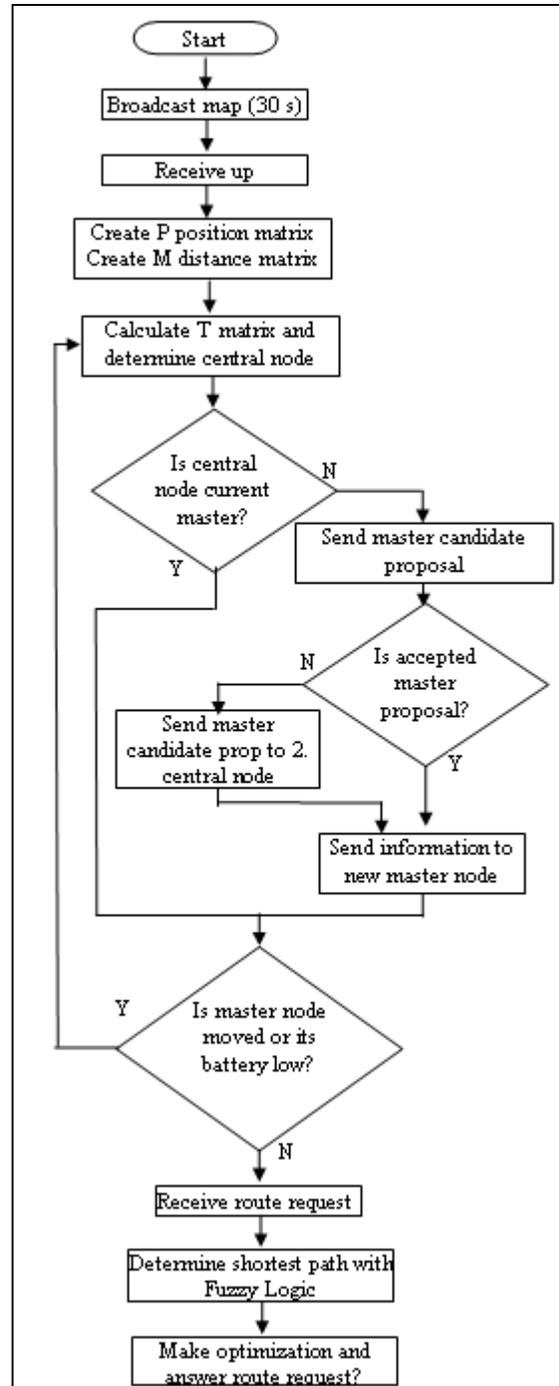


Figure 1. The flow chart of PBHRA [1]

III. FRAME STRUCTURE OF POSITION BASED HYBRID ROUTING PROTOCOL (PBHRP)

The frame structure of PBHRA whose working steps are given above is determined for the protocol realization. Designed frames for PBHRP have been given below.

Master node announcement frame (*maf*) that is to be sent to the other nodes in the network by master node according to *step b* of the algorithm is given in Figure 2.

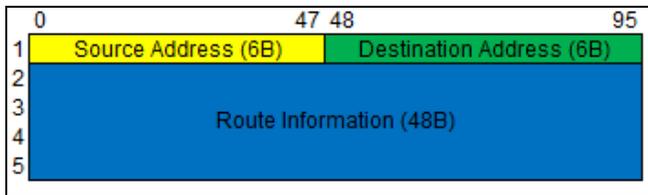


Figure 2. Master node announcement frame

According to the structure in Figure 2, the source address space of size 6 Bytes is the physical address of the master node and the destination address space of size 6 Bytes is the broadcast address to enable every node in the network for being a receiver. The route information space of size 48 Bytes is used to be filled by every node which transfers master node announcement frame to its neighbors along itself. With regard to this structure, the addresses of at most 8 addresses are stored except source and destination addresses. maf is transferred with broadcasting method in this algorithm.

Every node except master node sends update frame (*uf*) that includes position, battery, and density conditions according to *step c* of PBHRA. The structure of the update frame is given in Figure 3.

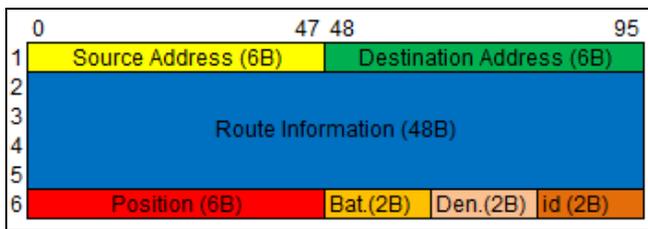


Figure 3. Structure of update frame

The source address space of the update frame given in Figure 3 is the physical address of the node which sends data, and its destination address space is the physical address of the master node. Route information space is the inverse of the route information of *maf*. So, *uf* is to be send along the way where *maf* is sent. There exist latitude, longitude, and height information gathered from the GPS inside the node in the position space. Battery life (as percentage) is given in the battery space, buffer fullness rate (as percentage) is in density space, and an integer value that is incremented by 1 in each sending exists in identity space. The value in identity space is used by master node to determine whether the update frame is old or the new one. *uf*'s are transferred to the network after their destinations and routes are determined. They are unicast and source-routing frames.

According to *step d* of PBHRA, master node stores update information, taken from other nodes, into position matrix. According to *step b* of the algorithm, master node calculates the distances between every node by using the position information matrix.

Master node offers a candidate after it determines the node at the center of the network by adding the elements of the row or the column of the distance matrix. Structure of

master node candidate suggestion frame (*mcsf*) is given in Figure 4.

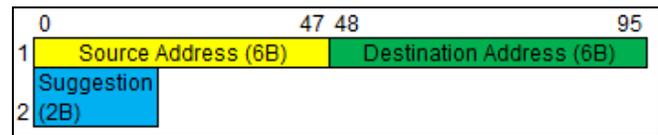


Figure 4. Master node candidate suggestion frame

The source address space is the physical address of master node, and the destination address involves the address of the candidate master node with regard to *mcsf* given in Figure 4. Suggestion Data space includes "01" value in base 10. The node which receives the suggestion declares if it accepts the suggestion response frame (*srf*) or not. Data in Suggestion Data space of *srf* given in Figure 5 is made "00" to deny the candidate suggestion, and it is made "11" to accept the suggestion.

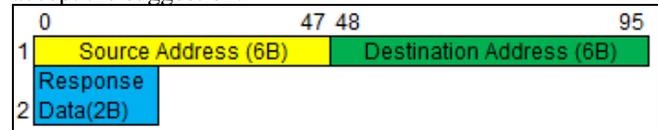


Figure 5. Structure of master suggestion response frame

According to *step j* of the algorithm, master node of the network stores the cost values, calculated by fuzzy logic, of each node to the others in fuzzy matrix.

According to *step l* of PBHRA, a node with a demand to send data asks the route through the destination node to master node via route demand frame (*rdf*). Based on the structure of *rdf* given in Figure 6, source address space is the physical address of the node which demands, and the destination address space is the physical address of the master node. Also, demand destination address space is the physical address of the node to which the data is intended to send.

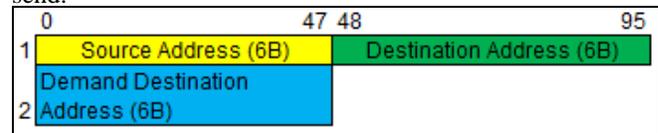


Figure 6. Structure of route request frame

Master node responds with a route response frame (*rrf*) after it receives route request frame. In Route Data space of *rrf* given in Figure 7, there exist physical addresses of nodes that include the path from source to destination. There are also physical addresses of eight intermediate nodes except source and destination. Master node uses the cost data of fuzzy matrix for this operation. If need more than eight inter node in order to routing data packet, PBHRA proposes node clustering [1].

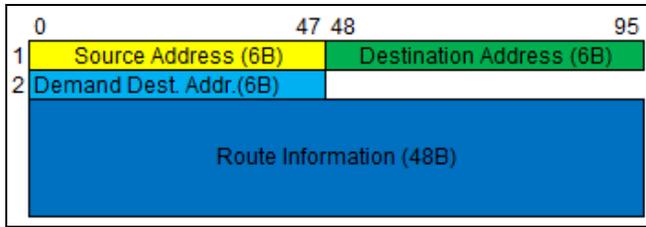


Figure 7. Structure of route demand frame

According to PBHRA, a node which will send data sends its data using data frame (*df*) given in Figure 8 after it learns about the path of its destination from master node.

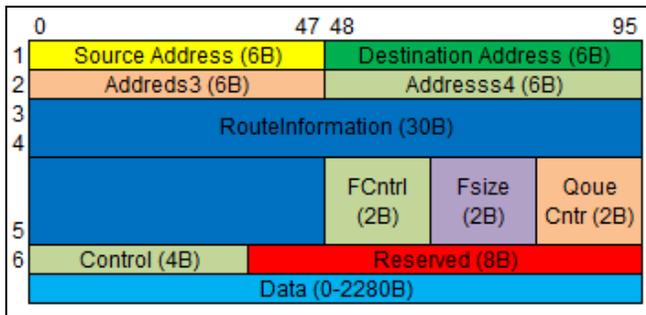


Figure 8. Structure of data frame

Control and data frames needed to send data in the network are formed. In the following sections, some simulations on sending data in network are performed by using these frames.

IV. PERFORMANCE EVALUATION OF PBHRP

A software is developed via MATLAB 7.01 evaluate the performance of the developed protocol. In the software, some functions are written to achieve the purposes and the processes, and a modular software environment is set. Evaluation environment and the parameters provided in RFC 2501 [10] is used for the performance evaluation.

The parameters below are used for the simulation.

Number of Nodes: 10, 20, 50, and 100-noded networks are used for the simulation.

Environment Size: A random distribution is used in the environment having 5000 m² per node in both horizontal and vertical axes, and 10 m height.

Packet Size: Fixed-sized data packets of size 512 and 1024 Bytes are used.

Number of Sources: 5, 10, 20, and 30-sourced networks are considered. Source number denotes the nodes that are able to send data packets simultaneously.

Action Speed: Random-waypoint model is used for the nodes acting with random speeds between 0-20 m/s.

Stagnation Time: 0, 10, 20, 50, and 100 seconds are considered as stagnation times. Stagnation time is the time in which nodes stay stagnant. 0 second tells that the node is always in action, and 100 seconds express that the node waits 100 seconds stagnantly after an action.

Propagation distance: Coverage area of the nodes is a circle with radius 100 m.

Simulation Time: Simulation models are run for 100 seconds.

Number of Iterations: Each different network handled with simulation software is run 10 times and average results are gathered.

Packet Delivery Rate -one of the criterions used for performance evaluation of ad hoc routing protocols- is a performance evaluation criterion which determines the access rate to packets when they are sent by data sender nodes in the network. It is expressed as the rate (as percentage) of the packets that reach to their destination. If the packets sent by the sender nodes in the network don't reach to their destination, then the packet delivery rate is low. If the number of packets that access to their destination increase, then the packet delivery rate increases, too.

Expression in (1) is used to determine packet delivery rate.

$$PDO = \frac{n_h}{n_t} \cdot 100 \tag{1}$$

n_h denotes the number of packets that reach to their destination, n_t is the total number of data packets sent in this expression.

Simulation software is run on 10, 20, 50, and 100-noded networks with different stagnation times and different source numbers to determine packet delivery rate and the results given in Table I are obtained.

TABLE I. PACKET DELIVERY RATE VALUES FOR 10, 20, 50, AND 100-NODED NETWORKS WITH DIFFERENT STAGNATION TIMES AND DIFFERENT SOURCE NUMBERS

Stagnation Time (s)	0	10	20	50	100
10 Nodes-5 Sources	99,52	100	100	100	100
10 Nodes-10 Sources	100	100	100	100	100
20 Nodes-10 Sources	97,30029	96,35286	96,72257	97,81829	100
20 Nodes-20 Sources	99,5	99,75	98,20567	99,43567	99,062
50 Nodes-10 Sources	100	100	100	100	100
50 Nodes-20 Sources	100	100	100	99,759	99,345
50 Nodes-30 Sources	100	100	100	100	100
100 Nodes-10 Sources	100	100	100	93,473	99,623
100 Nodes-20 Sources	86,872	100	96,397	100	100
100 Nodes-30 Sources	98,663	100	100	100	100

According to the values in Table I, the networks with high mobility have less packet delivery rate than the networks with less mobility. The change in the number of nodes in network has not much effect on the performance.

Proposed PBHRP is compared with DSDV, DSR, and AODV protocols, which are the leading in performance evaluation in ad hoc networks in terms of packet delivery rate. 50-noded and 20-sourced networks with different stagnation times are compared and the results are given in Table II.

According to the values in Table II, the proposed PBHRP has better packet delivery rate values than AODV, DSDV, and DSR protocols.

Another performance evaluation criterion for routing protocols of wireless ad hoc networks is normalized routing load. Normalized routing load is determined dividing by routing packets to total packets.

TABLE II. PACKET DELIVERY RATE VALUES OF 50-NODED AND 20-SOURCED NETWORKS IN PBHRP, AODV, DSDV, AND DSR PROTOCOLS

Stagnation Time (s)	0	10	20	50	100
PBHRP (PDR %)	98,62	98,86	98,88	99,34	99,54
AODV (PDR %)	96,97	98,83	98,83	99,07	99,53
DSDV (PDR %)	71,39	73,72	67,44	79,06	99,53
DSR (PDR %)	83,72	93,25	92,09	92,09	99,53

Simulation software is run on 10, 20, 50, and 100-noded networks with different stagnation times and different source numbers to determine normalized routing load and the results given in Table III are obtained.

TABLE III. NORMALIZED ROUTING LOAD VALUES FOR 10, 20, 50, AND 100-NODED NETWORKS WITH DIFFERENT STAGNATION TIMES AND DIFFERENT SOURCE NUMBERS

Stagnation Time (s)	0	10	20	50	100
10 Nodes-5 Sources	0,32062	0,29104	0,27102	0,25795	0,2357
10 Nodes-10 Sources	0,35207	0,33053	0,30061	0,2933	0,28889
20 Nodes-10 Sources	0,33659	0,29459	0,26316	0,24513	0,25496
20 Nodes-20 Sources	0,35704	0,33475	0,30337	0,30967	0,29706
50 Nodes-10 Sources	0,34729	0,24106	0,18958	0,15848	0,14292
50 Nodes-20 Sources	0,24075	0,18157	0,16203	0,1385	0,12903
50 Nodes-30 Sources	0,21231	0,17666	0,15236	0,14248	0,13818
100 Nodes-10 Sources	0,5685	0,46341	0,35674	0,30115	0,25547
100 Nodes-20 Sources	0,44066	0,33952	0,29364	0,26131	0,24
100 Nodes-30 Sources	0,3923	0,30951	0,27837	0,25099	0,23836

According to the values in Table III, in the networks with same number of nodes, the networks with large number sources have higher normalized routing load.

TABLE IV. NORMALIZED ROUTING LOAD VALUES OF 50-NODED AND 20-SOURCED NETWORKS IN PBHRP, AODV, DSDV, AND DSR PROTOCOLS

Stagnation Time (s)	0	10	20	50	100
PBHRP	0,2408	0,1816	0,162	0,1385	0,129
AODV	0,42	0,35	0,38	0,41	0,15
DSDV	0,37	0,48	0,54	0,37	0,12
DSR	1,3333	1,3333	1,2427	1,2941	1,0161

According to the values in Table IV, the proposed PBHRP has better normalized routing load values than AODV, DSDV, and DSR protocols.

V. CONCLUSIONS

Protocol realization of one of the position-based routing algorithms, PBHRA, which is developed to be used in mobile ad hoc networks, was presented in this study.

First of all, the frame structures to be used by the protocol are formed. The control frames that the protocol will use to control the network and the data frames intended to be used to send data are created.

To test the usability and the performance of the proposed protocol, a simulation is performed with MATLAB 7.01. With this simulation software, PBHRA is run for different node number, source numbers, and different mobility scenarios. According to the gathered results, the increase in the number of nodes in network does not make any change in packet delivery rate. In case of high mobility, the packet loss increases 1.6 % in the average.

It is shown that PBHRP performs better with 16% in case of high mobility when it is compared with other protocols in terms of packet delivery rate.

The obtained performance results show that the proposed protocol and frame structures are a new alternative for mobile ad hoc networks.

REFERENCES

- [1] R. Kara, I. Özçelik, and H. Ekiz, "A new Routing Algorithm in MANETs: Position Based Hybrid Routing", Scientific Research And Essays, vol. 5, pp. 328-338, 2010.
- [2] I.D. Chakeres and E.M. Belding-Royer, "AODV Routing Protocol Implementation Design", Proceedings of the International Workshop on Wireless Ad Hoc Networking (WWAN), Tokyo, Japan, 2004.
- [3] A.K. Saha, K.A. To, S. PalChaudhuri, S. Du, D.B. Johnson, "Physical Implementation and Evaluation of Ad Hoc Network Routing Protocols using Unmodified Simulation Models", Proc. ACM SIGCOMM Asia Workshop, April 2005.
- [4] A.C. Sun and A. Ahmad, "Design and Implementation of Fisheye Routing Protocol, Msc. Thesis in Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 2000.
- [5] K. Holter, A. Hafslund, F.Y. Li, and K. Ovsthus, "Design and Implementation of Wireless OSPF for Mobile Ad Hoc Networks", Scandinavian Workshop on Wireless Ad-hoc Networks (ADHOC 06), Stockholm, Sweden, May 2005.

- [6] T. Clausen and P. Jacquet, "Optimized Link State Routing Protocol (OLSR)", <http://www.ietf.org/rfc/rfc3626.txt>, May 2013.
- [7] T. Clausen, C. Dearlove, J. Dean, and C. Adjih, "Generalized Mobile Ad Hoc Network (MANET) Packet/Message Format", <http://www.ietf.org/rfc/rfc5444.txt>, May 2013.
- [8] T. Clausen, C. Dearlove, and J. Dean, "Mobile Ad Hoc Network (MANET) Neighborhood Discovery Protocol (NHDP)", <http://www.ietf.org/rfc/rfc6130.txt>, May 2013.
- [9] S.S. Kumar, M.N. Kumar, and V.S. Sheeba, "Implementation of Hybrid Ad-Hoc Routing Protocol Advances in Recent Technologies in Communication and Computing (ARTCom)", 2010 International Conference on, 2010, pp. 151 – 154.
- [10] S. Corson and J. Macker, "Mobile Ad hoc Networking (MANET): Routing Protocol Performance Issues and Evaluation Considerations", IETF, <http://www.ietf.org/rfc/rfc2501.txt>, May 2013.
- [11] R. Kara, I. Ozcelik, and H. Ekiz, "Analysis of Fuzzy Logic Usage in PBHRA Routing Algorithm in Wireless Ad Hoc Networks", 6th International Conference on Wireless and Mobile Communications (ICWMC-2010), 20-25 Sept 2010, Valencia, Spain, pp. 337 – 342.

An Improved Extended Kalman Filter for Localization of a Mobile Node with NLOS Anchors

Siamak Yousefi

Department of Electrical and
Computer Engineering, McGill University
Montreal, H3A0E9, QC, Canada
Email: siamak.yousefi@mail.mcgill.ca

Xiao-Wen Chang

School of Computer Science,
McGill University,
Montreal, H3A0E9, QC, Canada
Email: chang@cs.mcgill.ca

Benoit Champagne

Department of Electrical and
Computer Engineering, McGill University
Montreal, H3A0E9, QC, Canada
Email: benoit.champagne@mcgill.ca

Abstract—Tracking a mobile node using a wireless sensor network under non-line of sight (NLOS) conditions, has been considered in this work, which is of interest to indoor positioning applications. A hybrid of time difference of arrival (TDOA) and angle of arrival (AOA) measurements, suitable for tracking asynchronous targets, is exploited. The NLOS biases of the TDOA measurements and the position and velocity of the target are included in the state vector. To track the latter, we use a modified form of the extended Kalman filter (EKF) with bound constraints on the NLOS biases, as derived from geometrical considerations. Through simulations, we show that our technique can outperform the EKF and the memoryless constrained optimization techniques.

Keywords—Extended Kalman filter; localization; non-line of sight; ultra wideband.

I. INTRODUCTION

Ground-based wireless positioning has received great attention in indoor and dense urban areas due to the limitations of the global positioning system (GPS) [1]. The fine timing resolution of ultra-wideband (UWB) pulses makes them robust against multipath fading and thus, the methods based on time of arrival (TOA) can achieve good accuracy [2]. Localization techniques that directly exploit the TOA measurements require accurate target-anchor synchronization. However, if only anchor-anchor synchronization is maintained, the time difference of arrival (TDOA) data can be exploited for localization. Furthermore, the angle of arrival (AOA) measurement of UWB signal requires no tight synchronization among the anchors, and its accuracy improves with increasing the system bandwidth [3]. Although the AOA measurement of UWB signal is challenging in practice, some methods have been recently proposed for the joint measurement of TOA and AOA in UWB systems [4]. Indeed, exploiting a hybrid of angle and range information can reduce the number of required line of sight (LOS) nodes for localization.

One of the main challenges in network-based positioning remains the non-line of sight (NLOS) problem, which occurs due to the blockage of target-anchor view. In a NLOS scenario, the TOA of the first detectable signal at the receiver is larger than it would be for a direct line of sight (LOS) path, resulting in a positively biased range measurement. Also, the AOA measurements will be corrupted by a random perturbation distributed within $[0, 2\pi)$ [1]. The TOA-based localization in NLOS for a single time instant is summarized in [5]. More recently, the constrained optimization technique using

sequential quadratic programming (SQP) in [6] has shown a good performance for position and NLOS bias estimation. In [6], the lower bound on the NLOS bias is set to zero and the upper bound proposed in [7] is employed. When the target is moving, the history of the past measurements along with the state equation can help in estimating a better location over time. In [8] and [9], the TDOA-AOA data are exploited for cellular and UWB indoor tracking, respectively. In both papers, Kalman filter (KF) preprocessing is applied on the received TOA data for NLOS mitigation, and the variances of the range measurements are estimated. In [9], further NLOS mitigation is done by scaling the covariance matrix of the measurement noise in an extended Kalman filter (EKF).

While these methods enable the tracking of a moving node under NLOS conditions, they have some disadvantages. First, the KF preprocessing cannot mitigate the effect of NLOS effectively, especially if the nodes are initially in a severe NLOS situation. Second, the online variance calculation of the range measurement of a moving target might not be done with good accuracy for NLOS links. In [10], the state vector of the EKF is augmented with the NLOS biases in addition to position and velocity, which helps in improving the positioning accuracy under NLOS conditions. Similar state equations have been used in particle filters (PF) developed for UWB indoor tracking of a target equipped with an inertial measurement unit (IMU) in [11], [12]. However, the computational cost of the PF grow exponentially with the size of the state vector [12], hence it is less suitable for practical applications.

In this paper, we propose a constrained form of the EKF for localization of an asynchronous mobile node under mixed LOS/NLOS conditions. The state vector of the proposed filter, similar to [10], is augmented with the NLOS biases. The steps of the constrained EKF follow those of the EKF, except for the estimation of the state vector at each time instant, where the 2-norm of a linearized vector function is minimized subject to bound constraints on the NLOS biases. These constraints are linear and depend on the geometry and the type of the measurements exploited. Since here, we use TDOA-AOA measurements for localization of an asynchronous node, the upper bound given in [7] can not be exploited. Therefore, we propose novel lower and upper bounds on the NLOS bias by exploiting the underlying geometry of the angle and range difference equations. Through simulations, we show that the proposed constrained EKF outperforms the ordinary EKF in [10], and the memoryless constrained optimization technique

in [6]. Other advantages of the proposed approach are its simplicity and low computational cost, which makes it suitable for practical applications.

The remainder of this paper is organized as follows: In Section II, the system model is described, and the formulation of the problem is given. The derivation of the bounds on the NLOS biases and the constrained EKF technique are described in Section III. In Section IV, illustration of simulation results are presented. Finally, Section V concludes the paper.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. Range and Bearing Measurements in NLOS

We consider a 2-dimensional (2D) indoor plane on which a wireless sensor network (WSN) with M fixed anchors at known positions $\mathbf{p}_i = [X_i, Y_i]^T$ is deployed. The goal is to track a mobile target with unknown position $\mathbf{p}[k] = [x[k], y[k]]^T$ and velocity $\mathbf{v}[k] = [v_x[k], v_y[k]]^T$ at the k -th time instant. The synchronized anchors are placed near the boundary of the room in order to cover the area, as the anchor placement has a direct effect on the localization accuracy [5]. For the sake of simplicity, let the indices $i \in \{1, \dots, N\}$ denote the anchors that are in NLOS, and $i \in \{N+1, \dots, M\}$ denote the anchors with LOS measurements. We assume that the NLOS links can be identified accurately at every time instant. To this end, NLOS identification techniques based on hypothesis testing have been proposed in [8], [9], [13], [5]. Furthermore, NLOS identification based on the features of the received UWB timing pulse can yield a good result [14].

The range measurement for the i -th anchor, $z_i[k]$, is obtained by multiplying the corresponding TOA measurement with the speed of wave propagation. The set of range equations for the M anchors are

$$z_i[k] = d_i[k] + \epsilon[k] + b_i[k] + n_i[k], \quad i = 1, \dots, M \quad (1)$$

where $d_i[k] = \sqrt{(x[k] - X_i)^2 + (y[k] - Y_i)^2}$ is the true range, $\epsilon[k]$ is proportional to the relative clock offset between the target and the anchors (the clock offset is common among the anchors as they are synchronized), $b_i[k]$ is the NLOS bias which is a positive random variable for $i \in \{1, \dots, N\}$ and zero for $i \in \{N+1, \dots, M\}$, and $n_i[k]$ is the measurement noise. The distribution of the NLOS bias is location dependent and time varying; therefore, assuming a priori known distribution is not practical. However, as the target moves, the instantaneous value of the NLOS bias can be approximately modelled as a random walk

$$b_i[k+1] = b_i[k] + w_b[k], \quad i = 1, \dots, N \quad (2)$$

where the increment $w_b[k]$ is a uniform random variable with zero-mean and variance $\sigma_{w_b}^2$ [11], [12]. The measurement noise in (1) is modelled as a normal random variable with zero-mean and variance $\sigma_{n_i}^2[k]$. In UWB systems, the variance of this noise can be represented as

$$\sigma_{n_i}^2[k] = \sigma_0^2 d_i^\beta[k], \quad (3)$$

where σ_0 is a proportionality constant and β is the path-loss exponent, which might have different values based on the environment, e.g., $\beta = 2$ for LOS and $\beta = 3$ for NLOS have been considered in [15].

The AOA can be measured jointly with the TOA by means of UWB antenna arrays as done in [4]. The AOA of the NLOS anchors are discarded as done in [9], since they are corrupted by random perturbations uniformly distributed in the interval $[0, 2\pi)$. The AOA of the LOS anchors with respect to a reference coordinate system are

$$\theta_i[k] = \tan^{-1} \left(\frac{y[k] - Y_i}{x[k] - X_i} \right) + n_{\theta_i}[k], \quad i = N+1, \dots, M \quad (4)$$

where the measurement error, $n_{\theta_i}[k]$, is normally distributed with zero-mean and variance σ_θ^2 .

The clock offset parameter $\epsilon[k]$ is cancelled out by subtracting the measured range of each anchor in (1) from a reference anchor, which amounts to employing TDOA measurements. If all the anchors are in NLOS, i.e., $N = M$, then the range difference achieved through subtraction has either a positive or negative bias. For this scenario, the techniques in [8] and [9] can be employed, where KF preprocessing is used for smoothing and variance calculation. However, as mentioned earlier, KF smoothing can not mitigate the effect of the NLOS bias efficiently. In this paper, to overcome this limitation, we assume instead that there is at least one LOS range measurement available, i.e., $N < M$. In this case, the range difference measurements can be obtained with respect to a selected reference LOS anchor, which for simplicity is indexed by M . If multiple LOS anchors are available, then amongst them, we propose to select as reference anchor the one which is closest to the target, i.e., with respect to which all the LOS range differences are positive. Since a smaller measured range is usually less likely to be in NLOS compared to a larger range, this selection of the reference anchor is relatively robust against false alarm in NLOS detection. Furthermore, the choice of the reference anchor can change over time due to the transition from LOS to NLOS. If the i -th anchor changes from NLOS to LOS, then $b_i[k]$ becomes zero, while if the i -th anchor changes from LOS to NLOS then the bias changes from zero to a positive random value. Therefore, if the reference anchor goes through such a transition, then another anchor which is in LOS has to be selected as the reference. If there are two or more LOS anchors, then by rejecting the NLOS range measurements, the location can be estimated without ambiguity, however, the performance highly depends on the AOA measurement error, which might have a large error in some applications.

The range difference equations hence take the form

$$\Delta z_i[k] = \begin{cases} \Delta d_i[k] + b_i[k] + \Delta n_i[k] & i = 1, \dots, N \\ \Delta d_i[k] + \Delta n_i[k] & i = N+1, \dots, M-1 \end{cases}$$

where $\Delta z_i[k] = z_i[k] - z_M[k]$, $\Delta d_i[k] = d_i[k] - d_M[k]$ and $\Delta n_i[k] = n_i[k] - n_M[k]$.

B. Problem Formulation

Let the unknown state vector, which includes the biases, be defined as $\mathbf{s}[k] = [x[k], y[k], v_x[k], v_y[k], b_1[k], \dots, b_N[k]]^T$. The measurement equation can be expressed as

$$\mathbf{y}[k] = \mathbf{h}(\mathbf{s}[k]) + \mathbf{n}[k], \quad (5)$$

where

$$\mathbf{y}[k] = \begin{bmatrix} \Delta z_1[k] \\ \vdots \\ \Delta z_N[k] \\ \Delta z_{N+1}[k] \\ \vdots \\ \Delta z_{M-1}[k] \\ \theta_{N+1}[k] \\ \vdots \\ \theta_M[k] \end{bmatrix}, \quad \mathbf{h}(s[k]) = \begin{bmatrix} \Delta d_1[k] + b_1[k] \\ \vdots \\ \Delta d_N[k] + b_N[k] \\ \Delta d_{N+1}[k] \\ \vdots \\ \Delta d_{M-1}[k] \\ \tan^{-1} \left(\frac{y[k] - Y_{N+1}}{x[k] - X_{N+1}} \right) \\ \vdots \\ \tan^{-1} \left(\frac{y[k] - Y_M}{x[k] - X_M} \right) \end{bmatrix},$$

and $\mathbf{n}[k]$ is a normally distributed random vector with zero-mean and covariance matrix $\mathbf{R}[k]$. We assume uncorrelated range and angle measurements, i.e.,

$$\mathbf{R}[k] = \begin{bmatrix} \mathbf{R}_r[k] & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_\theta[k] \end{bmatrix}, \quad (6)$$

in which we let

$$\mathbf{R}_r[k] = \sigma_{n_M}^2[k] \mathbf{1}_{M-1} \mathbf{1}_{M-1}^T + \text{diag}(\sigma_{n_1}^2[k], \dots, \sigma_{n_{M-1}}^2[k]),$$

$$\mathbf{R}_\theta[k] = \sigma_\theta^2 \mathbf{I}_{M-N},$$

where $\mathbf{1}_{M-1} = [1, \dots, 1]^T$ of size $m \times 1$, and \mathbf{I}_m is the identity matrix of size $m \times m$.

The state equation is given by

$$\mathbf{s}[k+1] = \mathbf{A}_a \mathbf{s}[k] + \mathbf{B}_a \mathbf{w}_a[k], \quad (7)$$

where

$$\mathbf{A}_a = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_N \end{bmatrix}, \quad \mathbf{B}_a = \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_N \end{bmatrix}, \quad \mathbf{w}_a[k] = \begin{bmatrix} \mathbf{w}[k] \\ \mathbf{w}_b[k] \end{bmatrix},$$

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & \delta t & 0 \\ 0 & 1 & 0 & \delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0.5\delta t^2 & 0 \\ 0 & 0.5\delta t^2 \\ \delta t & 0 \\ 0 & \delta t \end{bmatrix},$$

$\mathbf{w}[k]$ is normally distributed with zero mean and covariance matrix $\text{diag}(\sigma_x^2, \sigma_y^2)$, $\mathbf{w}_b[k]$, which is uncorrelated with $\mathbf{w}[k]$, has a uniform distribution with zero mean and covariance matrix $\sigma_{w_b}^2 \mathbf{I}_N$, and δt is the time step duration.

The aim is to find the position and velocity of the target at the k -th time instant, i.e., $\mathbf{x}[k]$ and $\mathbf{v}[k]$, based on all the past and current measurements $\mathbf{y}[j]$ for time instants $j \in \{1, \dots, k\}$. A common approach is to employ the EKF as done in [10]. However, in this work, we improve the estimation accuracy by taking advantage of extra information, in the form of practical constraints on the NLOS bias $b_i[k]$, which is derived in the sequel.

III. PROPOSED TECHNIQUE

We first determine more accurate constraints on the NLOS biases through the geometry of the network and then propose the improved EKF algorithm.

A. Constraint Region

We are interested in tight lower and upper bounds on the NLOS bias $b_i[k]$ when TDOA-AOA measurements are exploited. Note that in UWB systems, the range measurement noise is small compared to the NLOS bias. Also the AOA measurement error of a LOS link is relatively low. Thus, in deriving these bounds we assume that the noise terms are negligible, i.e., $\Delta z_i[k] = \Delta d_i[k] + b_i[k]$ for $i \in \{1, \dots, N\}$ and $\tan(\theta_M[k]) = (y[k] - Y_M)/(x[k] - X_M)$. Herein, we consider the case where only one LOS anchor is available, although the bounds are still applicable when more LOS anchors exist.

For illustrative purposes, consider the physical setting of the anchors and target shown in Fig. 1. The measured AOA at the LOS reference anchor, $\theta_M[k]$, defines a line segment passing through the target position and denoted as L_M . Furthermore, the true (unbiased) ranges, $\Delta d_i[k]$, define a set of hyperbolas, denoted as H_{iM}^u and given by

$$\|\mathbf{x} - \mathbf{p}_i\| - \|\mathbf{x} - \mathbf{p}_M\| = \Delta d_i[k], \quad i = 1, \dots, N \quad (8)$$

Therefore, if we know $\Delta d_i[k]$, then the target position can be found by intersecting one of the hyperbolas H_{iM}^u with the AOA line segment L_M . Due to the NLOS condition, we measure $\Delta z_i[k]$, which is biased compared to $\Delta d_i[k]$. If $|\Delta z_i[k]| \leq \|\mathbf{p}_i - \mathbf{p}_M\|$, then all the points \mathbf{x} , for which the difference of the distances from the i -th and the M -th anchor is equal to $\Delta z_i[k]$ are located on a different set of hyperbolas, denoted as H_{iM}^b and given by

$$\|\mathbf{x} - \mathbf{p}_i\| - \|\mathbf{x} - \mathbf{p}_M\| = \Delta z_i[k], \quad i = 1, \dots, N \quad (9)$$

We note, however, that, if the bias is large, we might face a situation where $\Delta z_i[k] > \|\mathbf{p}_i - \mathbf{p}_M\|$, and then there is no point on the plane satisfying (9). In this case, it is convenient to redefine $\Delta z_i[k]$ by thresholding it to its maximum permissible value, i.e., $\Delta z_i[k] \leftarrow \min(\Delta z_i[k], \|\mathbf{p}_i - \mathbf{p}_M\|)$. If $\Delta z_i[k]$ is replaced by $\|\mathbf{p}_i - \mathbf{p}_M\|$, then (9) becomes a ray passing through \mathbf{p}_M . In the sequel, $\Delta z_i[k]$ refers to the range difference measurement after being readjusted in this way.

Since the bias is always positive, a trivial lower bound on $b_i[k]$ is 0. Also, since by definition $b_i[k] = \Delta z_i[k] - \Delta d_i[k]$, an upper bound on $b_i[k]$ can be found by finding a lower bound on $\Delta d_i[k]$. Invoking the triangle inequality, it follows that $\Delta d_i[k] \geq -\|\mathbf{p}_i - \mathbf{p}_M\|$. Thus, loose lower and upper bounds on $b_i[k]$ might be set as

$$0 \leq b_i[k] \leq \Delta z_i[k] + \|\mathbf{p}_i - \mathbf{p}_M\|. \quad (10)$$

However, we can achieve tighter lower and upper bounds on $b_i[k]$ by taking advantage of the geometry.

Let the hyperbola H_{iM}^b intersects with the line segment L_M at $\mathbf{q}_i[k]$, and define $r_i[k] = \|\mathbf{q}_i[k] - \mathbf{p}_M\|$. By expressing the equation of hyperbola (9) in polar coordinates, it follows that

$$r_i[k] = \frac{\|\mathbf{p}_i - \mathbf{p}_M\|^2 - \Delta z_i^2[k]}{2(\Delta z_i[k] + \|\mathbf{p}_i - \mathbf{p}_M\| \cos(\theta_M[k] - \varphi_i))}, \quad (11)$$

where φ_i is the angle between the horizontal axis and the line segment connecting the i -th and the reference anchors. If the bias does not exist in the range difference equation, then $r_i[k]$ is equal to $d_M[k]$. However, since $\Delta z_i[k] \geq \Delta d_i[k]$, $r_i[k]$ is biased compared to $d_M[k]$. It can be verified analytically that

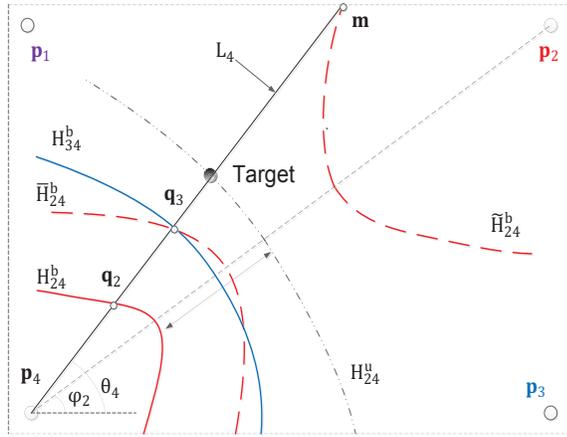


Fig. 1. Geometry of the nodes in 2-D. The anchor with index $M = 4$ is the reference LOS anchor and the other anchors are in NLOS. The time index k is omitted for convenience.

$r_i[k]$ in (11) is a monotonically decreasing function of $\Delta z_i[k]$ with upper bound

$$r_i[k] \leq d_M[k], \quad i = 1, \dots, N \quad (12)$$

Let us define $r_j[k] = \max\{r_i[k], i = 1, \dots, N\}$; then $r_j[k] \leq d_M[k]$, i.e., the distance of the target from the reference anchor has to be at least $r_j[k]$. Note that in case $\Delta z_i[k]$ is replaced by $\|\mathbf{p}_i - \mathbf{p}_M\|$, $r_i[k] = 0$ and (12) is satisfied. For the vector $\mathbf{q}_j[k]$, define in turn

$$\Delta \bar{z}_i[k] = \|\mathbf{q}_j[k] - \mathbf{p}_i\| - \|\mathbf{q}_j[k] - \mathbf{p}_M\|, \quad (13)$$

which means that $\mathbf{q}_j[k]$ is located on another set of hyperbolas, denoted as \bar{H}_{iM}^b , representing the points for which the distance difference from the i -th and M -th anchors is $\Delta \bar{z}_i[k]$. For example, in Fig. 1, \bar{H}_{24}^b passes through \mathbf{q}_3 where $j = 3$. Similar to (11), by intersecting \bar{H}_{iM}^b and the line segment L_M , we have that

$$r_j[k] = \frac{\|\mathbf{p}_i - \mathbf{p}_M\|^2 - \Delta \bar{z}_i^2[k]}{2(\Delta \bar{z}_i[k] + \|\mathbf{p}_i - \mathbf{p}_M\| \cos(\theta_M[k] - \varphi_i))}. \quad (14)$$

Since $r_j[k] \leq d_M[k]$ and $r_j[k]$ decreases monotonically with respect to $\Delta \bar{z}_i$, it follows that $\Delta \bar{z}_i[k] \geq \Delta d_i[k]$. Thus, we can define $\Delta \bar{z}_i[k] = \Delta d_i[k] + \bar{b}_i[k]$, where $\bar{b}_i[k] \geq 0$. Also, since $r_i[k] \leq r_j[k]$, then by comparing (11) and (14) it follows that $\Delta z_i[k] \geq \Delta \bar{z}_i[k]$, and hence $b_i[k] \geq \bar{b}_i[k]$. Although $\bar{b}_i[k]$ is unknown, the positive term $b_i[k] - \bar{b}_i[k]$, which can be used as a lower bound on $b_i[k]$, is known through $b_i[k] - \bar{b}_i[k] = \Delta z_i[k] - \Delta \bar{z}_i[k]$. Therefore, the lower bound on $b_i[k]$ is obtained as

$$b_i[k] \geq l_i[k] = \Delta z_i[k] - \Delta \bar{z}_i[k]. \quad (15)$$

For the upper bound, we assume that the target position is constrained to be in a closed area with arbitrary shape, where the coordinates of the border are known. In general it is not easy to handle complicated physical constraint on the position in terms of computation. Therefore, we use the geometrical constraints on the position to impose simple bound constraints on the NLOS bias. Let us assume $\mathbf{m}[k]$ is the intersection

of the AOA line segment and the border of the closed area. Similar to the analysis leading to the lower bound, define

$$\Delta \tilde{z}_i[k] = \|\mathbf{m}[k] - \mathbf{p}_i\| - \|\mathbf{m}[k] - \mathbf{p}_M\|. \quad (16)$$

According to (16), $\mathbf{m}[k]$ is located on a set of hyperbolas, denoted as \tilde{H}_{iM}^b and representing the points for which the distance difference from the i -th and M -th anchors is $\Delta \tilde{z}_i[k]$. The distance of $\mathbf{m}[k]$ from the reference anchor, i.e., $\tilde{r}[k] = \|\mathbf{m}[k] - \mathbf{p}_M\|$, satisfies

$$\tilde{r}[k] = \frac{\|\mathbf{p}_i - \mathbf{p}_M\|^2 - \Delta \tilde{z}_i^2[k]}{2(\Delta \tilde{z}_i[k] + \|\mathbf{p}_i - \mathbf{p}_M\| \cos(\theta_M[k] - \varphi_i))}. \quad (17)$$

Since $\tilde{r}[k] \geq d_M[k]$ and $\tilde{r}[k]$ decreases monotonically with respect to $\Delta \tilde{z}_i[k]$, then through a similar argument as used for (12), it follows that $\Delta \tilde{z}_i[k] \leq \Delta d_i[k]$. Also from (16) and the triangle inequality it follows that

$$\Delta d_i[k] \geq \Delta \tilde{z}_i[k] \geq -\|\mathbf{p}_i - \mathbf{p}_M\|. \quad (18)$$

Therefore, $\Delta \tilde{z}_i[k]$ is a tighter lower bound on $\Delta d_i[k]$, thus a better upper bound on $b_i[k]$ is achieved as

$$b_i[k] \leq u_i[k] = \Delta z_i[k] - \Delta \tilde{z}_i[k]. \quad (19)$$

B. Improved Extended Kalman Filter

The underlying concept of the constrained EKF is similar to the conventional EKF except that a constrained optimization problem is solved at every iteration. At time instant $k-1$, let us assume that $\mathbf{s}[k-1|k-1]$ and $\Sigma[k-1|k-1]$ are the estimated state and covariance matrix, respectively. Then at time instant k , we can do the prediction as

$$\begin{aligned} \mathbf{s}[k|k-1] &= \mathbf{A}_a \mathbf{s}[k-1|k-1], \\ \Sigma[k|k-1] &= \mathbf{A}_a \Sigma[k-1|k-1] \mathbf{A}_a^T + \mathbf{B}_a \mathbf{Q}_a \mathbf{B}_a^T, \end{aligned} \quad (20)$$

where \mathbf{Q}_a is the covariance matrix of $\mathbf{w}_a[k]$. Using the predicted position, i.e., $[x[k|k-1], y[k|k-1]]^T$, an estimate of the distance $d_i[k]$ between the target and each anchor is computed. These estimates, say $\hat{d}_i[k]$ are used to compute the measurement noise variance $\sigma_{n_i}^2[k]$ according to the model in (3) and hence $\mathbf{R}[k]$ is computed according to (6).

Next, we would like to make use of the new range difference and angle measurement vector at the k -th time instant, i.e. $\mathbf{y}[k]$ to obtain the filtered state $\mathbf{s}[k|k]$ and associated covariance matrix $\Sigma[k|k]$. To this end, we employ a linearization technique based on the Taylor series expansion of the nonlinear function $\mathbf{h}(\mathbf{s}[k])$ in (5) around the predicted state $\mathbf{s}[k|k-1]$:

$$\mathbf{h}(\mathbf{s}[k]) \approx \mathbf{h}(\mathbf{s}[k|k-1]) + \mathbf{H}[k](\mathbf{s}[k] - \mathbf{s}[k|k-1]), \quad (21)$$

where $\mathbf{H}[k] = \frac{\partial \mathbf{h}(\mathbf{s}[k])}{\partial \mathbf{s}[k]} \big|_{\mathbf{s}[k]=\mathbf{s}[k|k-1]}$ is the Jacobian matrix derived explicitly in [10]. Therefore, based on this first order expansion we define a linearized measurement vector

$$\begin{aligned} \tilde{\mathbf{y}}[k] &= \mathbf{y}[k] - \mathbf{h}(\mathbf{s}[k|k-1]) + \mathbf{H}[k]\mathbf{s}[k|k-1] \\ &\approx \mathbf{H}[k]\mathbf{s}[k] + \mathbf{n}[k] \end{aligned}$$

If we denote the prediction error in $\mathbf{s}[k|k-1]$ by $\tilde{\mathbf{w}}[k|k-1] = \mathbf{s}[k] - \mathbf{s}[k|k-1]$, which can be considered to be zero mean with covariance $\Sigma[k|k-1]$, then we can combine the linearized range and angular measurements and the predicted state as

$$\begin{bmatrix} \tilde{\mathbf{y}}[k] \\ \mathbf{s}[k|k-1] \end{bmatrix} \approx \begin{bmatrix} \mathbf{H}[k] \\ \mathbf{I}_{N+4} \end{bmatrix} \mathbf{s}[k] + \begin{bmatrix} \mathbf{n}[k] \\ \tilde{\mathbf{w}}[k|k-1] \end{bmatrix}, \quad (22)$$

where the covariance matrix of the noise vector is

$$\mathbf{C}[k] = \begin{bmatrix} \mathbf{R}[k] & 0 \\ 0 & \Sigma[k|k-1] \end{bmatrix}.$$

Now we find $\mathbf{s}[k|k]$, the filtered estimate of $\mathbf{s}[k]$ by solving the bounded least square problem

$$\begin{aligned} \min_{\mathbf{s}[k]} & \left\{ \left\| (\mathbf{C}[k])^{-\frac{1}{2}} \left(\begin{bmatrix} \tilde{\mathbf{y}}[k] \\ \mathbf{s}[k|k-1] \end{bmatrix} - \begin{bmatrix} \mathbf{H}[k] \\ \mathbf{I}_{N+4} \end{bmatrix} \mathbf{s}[k] \right) \right\|^2 \right\} \\ \text{s.t.} & \quad l_i[k] \leq b_i[k] \leq u_i[k], \quad i = 1, \dots, N \end{aligned} \quad (23)$$

Note that in ordinary EKF, (23) is minimized without constraints. In contrast to the unconstrained EKF, (23) is a linear least squares problem with additional bound constraints on the entries of the state vector, and can be solved with moderate computational cost. In our work, we use the Matlab routine `lsqlin`. We can use the nonlinear function $h(\mathbf{s}[k])$ in (23) instead of the first order Taylor series approximation, however, the computational complexity would be much higher. Moreover, we did not observe significant improvement in the localization accuracy in our simulation experiments.

To update the covariance, we use as an approximation the recursive formula of the ordinary EKF algorithm, that is

$$\Sigma[k|k] = (\mathbf{I} - \mathbf{G}[k]\mathbf{H}[k])\Sigma[k|k-1], \quad (24)$$

where $\mathbf{G}[k]$ is the gain of the EKF

$$\mathbf{G}[k] = \Sigma[k|k-1]\mathbf{H}^T[k](\mathbf{R}[k] + \mathbf{H}[k]\Sigma[k|k-1]\mathbf{H}^T[k])^{-1}.$$

At this stage, both the state and covariance matrix are corrected and the algorithm continues recursively.

C. Initialization and Algorithm Summary

An initial guess of the position and biases can be made using the single snapshot constrained optimization technique in [6], while the initial values of the velocity components are set to zero, thus $\mathbf{s}[0|0]$ is obtained. Since relatively decent estimates of the location and biases are achievable using [6], a relatively moderate diagonal error covariance matrix $\Sigma[0|0]$ is considered. The algorithm is summarized below.

Algorithm 1 Proposed EKF with Bound Constraints

- 1: **Initialization:**
 - 2: Initialize position and biases using [6] and set $\mathbf{v}[0|0] = \mathbf{0}$
 - 3: Set $\Sigma[0|0] = \alpha\mathbf{I}$ where $\alpha > 0$
 - 4: **for** $k = 1, 2, \dots, K$ **do**
 - 5: Prediction using (20)
 - 6: Estimate $d_i[k]$, compute $\sigma_{n_i}^2$ in (3) and update $\mathbf{R}[k]$
 - 7: Find $\mathbf{s}[k|k]$ via the constrained minimization in (23)
 - 8: Update the covariance matrix $\Sigma[k|k]$ using (24)
 - 9: **end for**
-

IV. SIMULATION RESULTS

We consider $M = 4$ anchors with fixed positions $\mathbf{p}_1 = [0, 10]^T$, $\mathbf{p}_2 = [10, 10]^T$, $\mathbf{p}_3 = [10, 0]^T$, and $\mathbf{p}_4 = [0, 0]^T$, where the units are in meters. The initial position of the target is selected randomly within the square region covered by the anchors. The direction of its velocity vector is initialized randomly while its norm (i.e., speed) is set to 0.4m/s. The

location of the target changes over time according to the dynamic model in (7) where the accelerations are normally distributed with $\sigma_x = \sigma_y = 10^{-3}\text{m/s}^2$. The trajectory is evaluated for $K = 200$ time samples with a step size $\delta t = 0.1\text{s}$. The reference anchor, located at \mathbf{p}_4 , is in LOS, while the other anchors are in NLOS. The exact range values are corrupted by an exponentially distributed NLOS bias with a mean value equal to 0.2 of the true target-anchor range [6]. The additive noise $n_i[k]$ is modelled according to (3) with $\sigma_0 = 10^{-4}$, $\beta = 2$ for the LOS anchor, and $\beta = 3$ for the other NLOS anchors. The true angle of the LOS anchor is also disturbed by a normally distributed error with zero-mean and $\sigma_\theta = 1\text{deg}$.

We simulate the proposed constrained EKF described in Algorithm I with bounding constraints on the NLOS biases as derived in Section III-A. We considered both sets of constraints, i.e., the loose bound (10) and the new tighter bounds in (15)-(19), and denote the corresponding algorithms as Constrained EKF I and Constrained EKF II, respectively. In our implementation of these algorithms, we use the true values of σ_x , σ_y , σ_0 and set $\sigma_w = 3\text{m}$, although the final performance is robust to some amount of mismatch in these parameters. We compare our result with the single snapshot constrained optimization technique in [6] and the EKF in [10]. The PF is not considered herein since the nonlinearity is not severe and the distribution of the noise is Gaussian. In fact, due to lacking a bound on the biases, the PF approach might perform only slightly better than EKF [10]. The method in [6], is implemented in Matlab by means of `fmincon` function with the choice of SQP technique. Again, we consider both sets of constraints (10) and (15)-(19), and denote the corresponding algorithms as Memoryless I and Memoryless II, respectively. The initialization of Constrained EKF I and Constrained EKF II are done using the Memoryless I and Memoryless II, respectively, while EKF is initialized with Memoryless II. For the initialization of the state covariance matrix, we use $\Sigma[0|0] = 10\mathbf{I}_7$, which appears to be reasonable for the considered area and the possible error in position components and biases.

To evaluate the comparative performance of the various tracking algorithms under study, we run Monte Carlo (MC) simulations in Matlab for 500 different starting points and trajectories, generated randomly as specified above. The performance is evaluated in terms of the root mean squared error (RMSE) between the true and estimated vectors. The RMSE in the velocity estimate is also investigated.

The RMSE in positioning is plotted in Fig. 2 for the different techniques under study. We generally find that the performance of the EKF in [10] is strongly dependent on the initialization and its convergence is not always guaranteed due to the absence of bounds on the NOLS biases. The constrained optimization technique in [6] can not yield smooth position estimates due to its memoryless nature, although on the average its performance is satisfactory. We emphasize that the performance of this algorithm is notably improved when the tighter bound given in (15) and (19) is used instead of the simpler one in (10). Initially, Constrained EKF I has a higher RMSE compared to the unconstrained EKF, since it is initialized with a less accurate position. However, within a number of iterations, its performance surpasses that of EKF and Memoryless I. Overall, constrained EKF II exhibits the

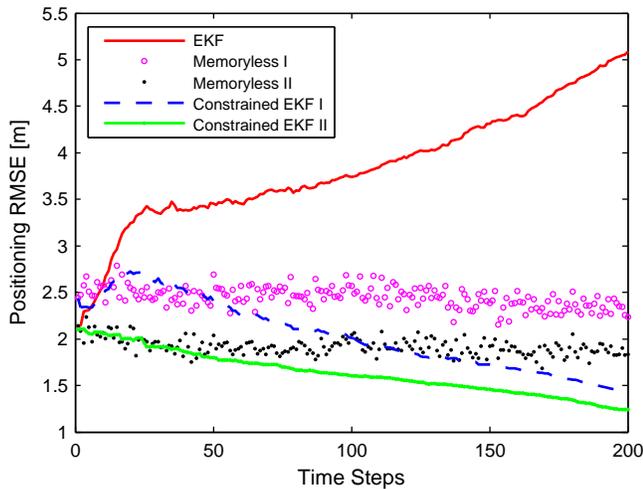
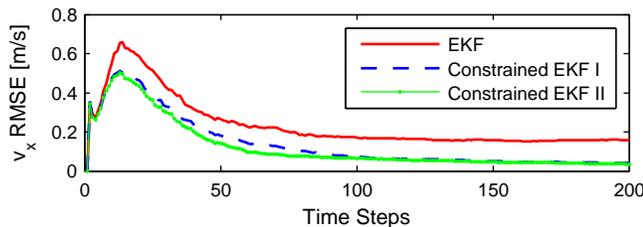


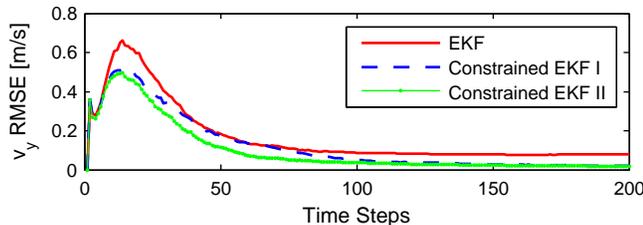
Fig. 2. RMSE of target position vs. time steps.

best performance by a significant margin, which proves the usefulness of our approach.

The RMSE in estimating the velocity components $v_x[k]$ and $v_y[k]$ in Cartesian coordinates is illustrated in Fig. 3. Again, the performance of the newly proposed EKF technique, with bounding constraints on the NLOS biases, exceeds that of EKF by a significant margin.



(a) RMSE in estimating v_x



(b) RMSE in estimating v_y

Fig. 3. RMSE of velocity components in logarithm scale vs. time steps.

V. CONCLUSION

A modified EKF algorithm for localization of a moving target was proposed in this paper, where novel bound constraints on the NLOS bias based on geometrical considerations were incorporated in the filtering (optimization) step of the EKF equations. The main assumption is that at least one LOS anchor remains available for referencing purposes. Through simulations, it was observed that our technique significantly outperformed EKF and memoryless constrained optimization techniques. Our technique shows good robustness against

NLOS biases, even when most of the nodes are initially (and remain) under severe NLOS conditions.

ACKNOWLEDGMENT

The authors would like to thank Dr. Kegen Yu of the University of New South Wales in Australia for his helpful comments. This work was supported by a grant from the Natural Science and Engineering Research Council (NSERC) of Canada and by the Fonds québécois de la recherche sur la nature et les technologies (FQRNT).

REFERENCES

- [1] K. Yu, I. Sharp, and Y. J. Gao, *Ground-Based Wireless Positioning*. John Wiley and Sons Ltd., 2009.
- [2] S. Gezici, Z. Tian, G. B. Giannakis, H. Kobayashi, A. F. Molisch, H. V. Poor, and Z. Sahinoglu, "Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks," *IEEE Signal Process. Mag.*, vol. 22, pp. 70–84, Jul. 2005.
- [3] D. Dardari, R. D'Errico, C. Roblin, A. Sibille, and M. Z. Win, "Ultrawide bandwidth RFID: The next generation?" *Proc. of the IEEE*, vol. 98, no. 9, pp. 1570–1582, Sep. 2010.
- [4] M. Navarro and M. Najar, "Frequency domain joint TOA and DOA estimation in IR-UWB," *IEEE Trans. on Wireless Commun.*, vol. 10, no. 10, pp. 1–11, Oct. 2011.
- [5] I. Guvenc and C. C. Chong, "A survey on TOA based wireless localization and NLOS mitigation techniques," *IEEE Communications Surveys Tutorial*, vol. 11, no. 3, pp. 107–124, quarter 2009.
- [6] K. Yu and Y. J. Guo, "Improved positioning algorithms for nonlinear-of-sight environments," *IEEE Trans. on Veh. Tech.*, vol. 57, no. 4, pp. 2342–2353, Jul. 2008.
- [7] S. Venkatraman, J. Caffery, and H. R. You, "A novel TOA location algorithm using LOS range estimation for NLOS environments," *IEEE Trans. on Veh. Tech.*, vol. 53, no. 5, pp. 1515–1524, Sep. 2004.
- [8] N. J. Thomas, D. G. M. Cruickshank, and D. I. Laurenson, "Performance of a TDOA-AOA hybrid mobile location system," in *Second Int. Conf. on 3G Mobile Communication Technologies*, 2001, pp. 216–220.
- [9] C.-D. Wann, Y.-J. Yeh, and C.-S. Hsueh, "Hybrid TDOA/AOA indoor positioning and tracking using extended Kalman filters," in *IEEE Vehicular Technology Conf. Spring*, vol. 3, May 2006, pp. 1058–1062.
- [10] M. Najar, J. M. Huerta, J. Vidal, and J. A. Castro, "Mobile location with bias tracking in non-line-of-sight," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 3, May 2004, pp. 956–959.
- [11] D. B. Jourdan, J. J. Deyst Jr., M. Z. Win, and N. Roy, "Monte carlo localization in dense multipath environments using UWB ranging," in *IEEE Int. Conf. on Ultra-Wideband*, Sep. 2005, pp. 314–319.
- [12] J. González, J. L. Blanco, C. Galindo, A. O. d. Galisteo, J. A. Fernández-Madriral, F. A. Moreno, and J. L. Martínez, "Mobile robot localization based on ultra-wide-band ranging: A particle filter approach," *Robot. Auton. Syst.*, vol. 57, no. 5, pp. 496–507, May 2009.
- [13] L. Cong and W. Zhuang, "Nonline-of-sight error mitigation in mobile location?" in *Twenty-third Annual Joint Conf. of the IEEE Computer and Communications Societies*, vol. 1, Mar. 2004, pp. 650–659.
- [14] S. Marañón, W. M. Gifford, H. Wymeersch, and M. Z. Win, "NLOS identification and mitigation for localization based on UWB experimental data," *IEEE J. Sel. A. Commun.*, vol. 28, no. 7, pp. 1026–1035, Sep. 2010.
- [15] S. Venkatesh and R. M. Buehrer, "Non-line-of-sight identification in ultra-wideband systems based on received signal statistics," *Microwaves, Antennas Propagation, IET*, vol. 1, no. 6, pp. 1120–1130, Dec. 2007.

An Overview of Underwater Sensor Networks

Christina Peach and Abdulrahman Yarali
 Telecommunication Systems Management
 Murray State University
 Murray, KY, USA
 {cpeach, ayarali}@murraystate.edu

Abstract—It is becoming clear that underwater sensor networks are needed to predict natural disasters, provide scientist research for marine life exploration, and to advance military capabilities. These underwater sensor networks pose great challenges to our existing technologies used in terrestrial sensor networks, mainly due to the acoustic waves used in water. These waves behave differently than the radio waves on land. The purposes of this research paper is to provide a better understanding of an underwater sensor network architecture, basic characteristic, applications, and examine the propagation limitations of acoustic communication. We also provide an overview of modulations and medium access controls.

Keywords—Underwater sensor; architecture; applications; propagation; modulation

I. INTRODUCTION

Underwater Sensor Networks' (UWSN) are made up of many sensor nodes and Autonomous Underwater Vehicles (AUVs). Together, these sensor nodes and AUVs allow researchers and the military to actively monitor the ocean. Underwater sensor networks make it possible to conduct search missions, record climate changes, provide pollution control, study marine life, conduct survey missions, tactical surveillance, and predict natural disturbances in the ocean. There has been an increased interest in building UWSN due to the fact that, these networks have the ability to monitor, investigate, and track underwater occurrences that are important to our national security and off shore exploration. Many characteristics of these networks make their construction challenging, and difficult to implement. These characteristics include: limited bandwidth capacity, high error probability, large propagation delays, fouling, corrosion, high bit rate errors, limited battery power, and temporary losses of connectivity. Another key characteristic is that, the sensor nodes and autonomous underwater vehicles must have the capability to self-configure themselves. The devices should be able to coordinate their configurations, location information, and movement information with each other. Data must also be sent back to an on-shore station. The technology that enables the application is wireless underwater acoustic networking. Underwater communications is not new. In 1945, there was a need to communicate with submarines in the United States. The development of an underwater telephone was the solution, as outlined by Akyildiz et al. [1]. However, underwater networking really has not been extensively

explored. Underwater sensor networks use acoustic communications based on acoustic wireless communication. Radio waves can travel long distances underwater, but must travel along very low frequencies. The frequency range typically used is 30-300 Hz. This frequency range requires extraordinary transmission power and very large antennas.

In this paper, we discuss the aspects of underwater sensor networks that create challenges in design, types of applications, underwater sensor architecture, modulation, and the available medium access control protocols.

II. APPLICATIONS

Underwater sensor networks are needed for several applications. The military can use this technology for mine reconnaissance. Mine reconnaissance is when autonomous underwater vehicles equipped with optical sensors perform assessments on objects that appear to be mines and detect whether they are in fact, mine like materials. They also use the networks for distributed tactical surveillance. In this application, sensors are used along with AUVs that work together to monitor an area for targeting, intrusion detection, surveillance, submarine warfare, and reconnaissance. A surveillance system is able to detect divers, submarines, and other vehicles based on data, sensors collect. These systems have higher accuracy rates than typical sonar systems.

There are also many geological applications that make use of an underwater sensor network. Sensor networks can monitor seismic activity along the ocean floor to measure earthquakes and predict an approaching tsunami. This data allows coastal areas to be notified and warned in time to evacuate the immediate danger area. Environmental monitoring applications allows for the detection of pollution, and how it is affecting marine life. The data can monitor chemical, biological, and even nuclear pollution in any body of water giving a more sophisticated analysis of water quality as outlined by Heidemann et al. [5].

Sensors can also monitor winds and ocean currents to predict the weather in a more reliable fashion. Climate change can also be monitored and studied along with the human effects on ecosystems, and biological tracking of fish. Applications to assist navigation can be useful to boat operators in shallow waters to determine potentially dangerous objects under water. Assisted navigation applications can also be used to find sunken boats and wreckage. Autonomous Underwater Vehicles (AUVs) are useful in many applications because they do not require

cables, tethers, or remote control. These devices are inexpensive and can reach any depth in the ocean and can be used in environmental monitoring and oceanography applications. One of the most common AUVs is small submarines armed with many sensors.

III. ARCHITECTURE

There are two types of architectures including: two-dimensional (see Fig. 1) architecture and three-dimensional architecture (see Fig. 2) as outlined in [1]. Static two-dimensional are used for environmental monitoring where sensor nodes are anchored to the ocean floor. These nodes are interconnected to underwater sinks. The sinks are in charge of relaying information from underwater to the surface station. They have a vertical acoustic transceiver and a horizontal acoustic transceiver. Horizontal transceivers communicate directly to the sensor nodes to send command, configurations, and collect monitored data. The vertical transceivers relay the data back to the surface. The on-shore station also has a transceiver that will handle multiple transmissions from the sinks. The sinks can be connected with underwater sensors by either a direct link or by a multi hop path. The easiest way to connect to the sensors is through a direct link, but is most likely not the most energy efficient way to connect. By using direct links, the throughput is also reduced due to acoustic interference. The acoustic interference comes from the high transmission power being used. The key difference in multi hop systems is that the information is relayed through intermediate sensors until it reaches the sink. This increases the network capacity and saves energy, which are two important aspects in underwater sensor networks.

The three-dimensional architecture is normally used when you want to identify or observe happenings that cannot be looked at with the sensor nodes attached to the ocean floor. In these networks, the sensor nodes float at various depths to observe any occurring phenomenon. These sensors could be placed on buoys with varying wire lengths to achieve different depths. Although this would be the easiest approach, floating nodes on buoys allows for problems involving tampering. For example, if the military is carrying out a surveillance of an enemy the nodes could be detected and disabled easily. The nodes are more susceptible to weather conditions, and multiple buoys could confuse boaters navigating the open waters. These reasons make this approach not very desirable.

Another approach would be to fasten nodes to the ocean bed. Once anchored, floating devices are attached that can be filled with air to achieve a variation in depth placement. The depth is determined by the wire length as in the first approach, but the nodes here have their own electric motors that can change the wire length on demand. There is still a disadvantage with this approach as well. Ocean currents could interfere with the system's ability to maintain the desired depths [1].

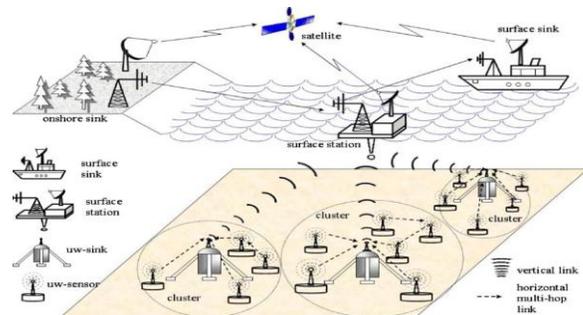


Figure 1. Architecture for 2D underwater sensor networks. [1]

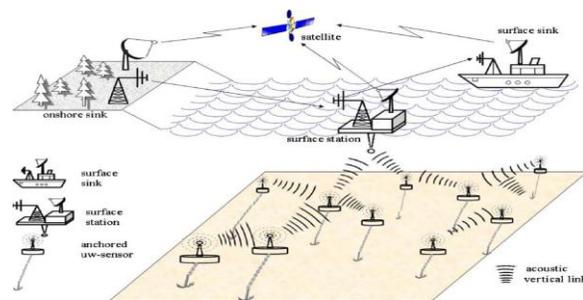


Figure 2. Architecture for 3D underwater sensor networks.[1]

IV. AUTONOMOUS UNDERWATER VEHICLES

AUVs are important to the network design. One objective for them is to become less dependent of on-shore communication and move toward relying solely on local intelligence. Recovering AUVs to recharge their batteries can be a cumbersome task; therefore, solar systems seem to be a viable option. Solar energy will increase the AUVs overall lifetime and they will be able to gather data continuously for several months before any human intervention is needed. Two kinds of AUVs used in explorations are called drifters and gliders. These are oceanographic instruments that do not need the same level of sophistication as small submarines. Drifters work with the current. They have the capability to move vertically, but drift along with the current. Their main purpose is to take various measurements at different preset depths. Gliders are powered by a battery and use hydraulic pumps create changes in their buoyancy to move through the water. They are equipped with a GPS device that allows them to be located once they climb to the surface. Typically, these devices can only operate for a few weeks to a few months, and can reach depths of 200 meters to 1500 meters. Gliders move very slowly through the water which allows them to conserve battery power and increase their lifetime [1].

V. UNMANNED UNDERWATER VEHICLES

The use of Unmanned Underwater Vehicles (UUV) is important to improve system efficiency. The Department of Defense, Homeland Security and the military are extremely interested in the capabilities and advantages that unmanned

underwater vehicles possess. The military has conducted several test deployments of this type of vehicle as outlined by Garreiro et al. [4]. Some of the test deployments included UUVs are: Remote Environmental Measuring Units (REMUS) [4], Mid-sized Autonomous Research Vehicle (MARV) [14], 21" Unmanned Underwater Vehicles (21UUV) [4], and Manta Test Vehicle (MTV) [4]. REMUS was 7.5 inches in diameter and was deployed for testing hundreds of times. It used acoustic communications with both oceanographic and chemical sensors. MARV is a mid-sized autonomous research vehicle that was 12.75 inches in diameter. It demonstrated various technologies under thruster based hover payload with low speed control. 21UUV is 21 inches in diameter and has been tested in the water over one hundred times. This vehicle demonstrated vision based navigation, photo mosaic, and side scan sonar imagery. This device was used for autonomous controller experiments where it was a test bed and weapon launch using MTV. MTV is a manta test vehicle that was tested in the water over ninety times for unmanned weapon launches. This vehicle had advanced network communications with ISR suites including RADINT, SIGINT, Optics, and IR. The military branch testing these UUVs is the NAVY. The military would like to have a family of not only unmanned underwater vehicles but also just unmanned sensor vehicles. These vehicles would include advanced sensors such as smart skins and Nano-sensors that work in conjunction with distributed underwater networks. The potential advanced payloads would allow for attacks to the enemy without being seen. Missile launching would be executed from underwater. Unmanned underwater vehicles might just be the weapons of the future, providing the military with high energy lasers and advanced weaponry [4].

The one thing that slows down the development of these weapons is the guidance laws that must be followed. These laws require vehicles to have onboard navigation guidance controls that enable vehicles to work with each other to achieve proper positions and orientation. Missile guidance requires line-of-sight guidance laws, command-to-line-of-sight guidance laws, proportional navigation guidance laws, and optimal guidance laws. The majority of the missile laws are already implemented in traditional missile weapon technologies; however, underwater unmanned vehicles must be programmed and tested thoroughly before building the weapons of the future as outlined by Sutton et al. [9].

VI. MODULATION

The electromagnetic spectrum controls all communications on land, because telecommunication companies have to constantly provide ways to communicate over long distances. This is made possible through radio and optical systems. These systems also allow for high bandwidth capacity on land, whether the power is high or low. In the physical layer of underwater sensor networks, the communication approach used is acoustic. Acoustic waves are the best choice for underwater systems, mainly because electromagnetic frequencies are diffused by the water. The acoustic communication system allows for better communication beyond ten meters as outlined in [5].

Acoustic communications have to deal with the network performance limits and poor bandwidth efficiency. To address these issues modulation schemes are used. The best modulation approach to achieve more bandwidth efficiency is Frequency Shift Keying (FSK). This approach is resistant to channel variations because carrier frequencies are chosen using information bits. At the next stage, measured power is compared at different frequency levels by the receiver to determine what was sent. Therefore, channel estimation is not needed because the receiver is using just energy detection. A disadvantage of frequency key shifting is the use of guard bands. Guard bands are inserted between the frequencies to prevent frequency spreading. Time spreading also causes interference, and guard intervals must be placed between consecutive symbol communications to promote channel clearing. These two introductions of guard bands and intervals slow down the data rate to an extremely low level as outlined in [1].

In order to use FSK advanced approaches would need to be added to improve the low data rate. One approach that could be implemented if the bandwidth is not under any constraints, would be frequency hopped FSK. Frequency hopped FSK removes the requirement of guard intervals to promote channel clearing. Another method of modulation that has had very little success is called multi-carrier modulation. This type of modulation divides the amount of bandwidth that is available into overlapping sub bands. This makes the duration of the waveform longer for each symbol. The sub-bands allow the system receiver to disregard inter symbol interference making channel equalization at the receiver simpler. This simplification has prompted attempts to integrate Orthogonal Frequency Division Multiplexing (OFDM) into the underwater channels. The problem with this type of multiplexing is that the channels themselves have large Doppler spreads allowing for a substantial introduction of interference in the subcarriers [5]. Phase coherent modulation includes Quadrature Amplitude Modulation (QAM) [12] and Phase Shift Keying (PSK) [12] is direct transmissions that can be used to promote high data rate communication in underwater networks, but also introduces inter symbol interference. The interference must be addressed at the receiving end of communication.

There are two modulation approaches that are promising for underwater sensor networks. These methods include: Code Division Multiple Access (CDMA) and Multi-Input Multi-Output (MIMO) techniques. CDMA is an up and coming technology that is important in acoustic networks. This is because this method allows for random, overlapping access to communication channels that are being shared in the network. CDMA is promising as a viable modulation approach to underwater sensor networks. Multi-input multi-output approaches require a system of multiple transmitters and receivers to increase channel capacities linearly. The degree in which channel capacities become linear depends on how many transmitters and receivers are added to the system. This technique looks to be extremely promising for underwater acoustic communication as outlined by Manjula et al. [6].

VII. PROPAGATION

Propagation in underwater sensor networks poses a major challenge for technologies we currently use in terrestrial sensor networks. Acoustic waves used for underwater networks have a lower loss than the traditional radio waves used in other sensor networks. The lower losses of these waves make them more suitable for underwater applications than radio waves. Acoustic waves travel at five orders of magnitude slower than radio waves presenting greater challenges that must be overcome with algorithms, protocols, and node placement [6].

Propagation is affected by several factors underwater. Hereafter we describe some of these factors influencing propagation. When data travels from one place to another, it will always experience some level of power loss. This power loss is called attenuation. In underwater acoustic channels, propagation delay limits the available range and frequency at which the system can function. When greater distances are needed to be covered by the system or if higher frequencies must be used will have an increase in attenuation. Attenuation is also caused when the ocean is rough at the propagation medium constraints.

Surface-bottom reflection and refraction occurs when the depth changes the speed at which the data is being transferred. Surface-bottom reflection and refraction is constantly present in underwater networks because the ocean's velocity changes frequently due to storms, tides, and waves. These changes create different levels of refraction that must be dealt with, at all levels of the system, but particularly in deeper waters [5]. This is especially true in large scale systems. Doppler effects are also created due to different levels of directional motion that introduce long delay spreads and phase distortion.

Noise is another factor in underwater sensor networks that affects the propagation. The propagation medium in these networks is constrained by the ocean floor and the surface. These constraints cause multipath propagation where echoes are created from unwanted reflections. The echoes appear as signal burst or as signal transmission replications. Multipath propagation allows each of the components throughout the underwater sensor network to potentially exhibit different forms and levels of attenuation as outlined in [5]. Man-made noise caused by boats, power plants, and oil pumps along with noise created from shipping activity can create propagation delays especially where there is lots of boating activity. Ambient noise is created by the natural occurrences in the ocean. Tides, waves, rain, wind, and seismic activity is categorized as this type of noise.

Acoustic waves carry noise farther than radio waves making noise, a more severe problem when dealing with acoustic communication. As one can see the variability of propagation requires the underwater sensor network to be robust in many areas, where in radio communication, the variations are more localized and easier to address. There are several different propagation models with different degrees of complexity and accuracy to simulate and analyze an underwater wireless sensor network. Designing an

efficient and reliable communication channel in underwater environment is not an easy task because of speed of acoustic signal, propagation delay, multipath interferences, path loss due to higher frequency, and high bit error rate due to phase and magnitude fluctuations. Absorption is also increased with the range, which causes the increasing drop of SNR when the range increases. Figure 3 shows the Signal-to-Noise Ratio (SNR) increase with distance [11].

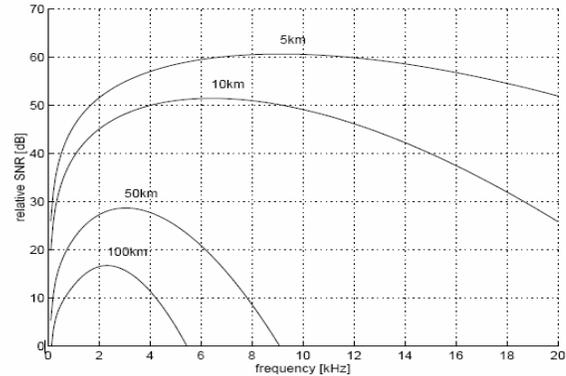


Figure 3. SNR vs. frequency and range for undersea. [11]

VIII. MEDIUM ACCESS CONTROL

Medium Access Control (MAC) takes place in the data link layer and manages access to the underwater networks acoustic communication. To keep the network performance at its best MAC protocols must be used. The main objective of MAC protocols is to prevent any collisions that could occur. These protocols also manage network throughput, scalability, adaptability, latency, and energy consumption (efficiency) as outlined by Pompili et al. [8]. Medium access control can either be contention free or contention based protocols. The contention free approach allocates different time slots, frequencies, or codes to all of the system nodes to avoid collisions in the physical layer. In contention based approaches the nodes must compete with one another to use the medium. This is all handled on a demand bases, and the protocol must work to keep overhead down while still preventing any form of collision as outlined by Otnes [7]. Some contention free MAC protocols include Frequency Division Multiple Access (FDMA) [13], Time Division Multiple Access (TDMA) [13], and Code Division Multiple Access (CDMA) [13]. FDMA divides the available bandwidth into several frequency bands and then places guard bands between them. The individual bands are assigned to each station where they can send data. Once these bands are assigned, they remain those stations until the end of communication. Station interference is handled with a band pass filter and guard bands that confine and separate the transmitter frequencies. FDMA is ideally used with stream data because it identifies preset frequency bands and uses them for the entire length of communication [3].

In an underwater sensor network, FDMA would allow each individual node to communicate the remaining nodes

simultaneously without interference with each other. However, because acoustic communication has a limited bandwidth capacity and multipath propagation, it creates spectral nulls. If the underwater network has several nodes it will require a very narrow frequency band for each node. Operating the system in this way produces extreme risks of complete fading to some of the nodes. For this reason FDMA is not considered a viable option for underwater networks according to [7]. FDMA could very well be used along with clusters using higher frequency bands to create a usable hybrid system. TDMA uses only one channel that is shared between all of the nodes in the system. This is accomplished by timesharing. Frames, or time-intervals, are split up into fixed length slots where each node in the system is assigned a slot [10]. The node can only transmit during its allocated time, and then wait until the frame is repeated cyclically back to their time slot before transmitting again. In this approach all of the systems nodes have to know when their time slot begins and ends which adds additional propagation delay into the system. These added delays can be compensated by the use of guard times, but in underwater systems where propagation delays are already high this is not ideal. TDMA does not allow the nodes to transmit data simultaneously, so in order to handle synchronization it normally adds preamble bits to the beginning of each slot [3]. The addition of guard times and preamble bits will also degrade the efficiency to some extent [7]. CDMA is typically used in cellular technologies. It is a spread spectrum technology that takes multiple telephone conversations and attaches a certain code that only the sender and receiver understand. It then cuts the information into bits and transfers them to the receiver to reconstruct [2]. CDMA in underwater sensor networks work in much the same way. They also use binary (bits) to modulate the data with the same spread spectrum concept. This MAC technology is a very promising technique for acoustic communication. CDMA's advantage over FDMA is its resistance to node fading, because the frequency band is used by the entire system of nodes. This is because each node uses a unique code with low cross-correlation that allows all of the nodes to send and receive data at the same time on the same band. There is a trade off when using low cross-correlations; they involve long codes. When long codes are being used; it lowers data rates that are already a challenge in acoustic communications. Doppler effects that are present in acoustic channels further reduce the correlation properties. Algorithms must be developed that are complex for demodulation and multi-user detection because CDMA gives network design issues back to the physical layer from the data link layer. Frequency hopping and direct sequence CDMA could be viable options for these modulations. Code division multiple access works best when the sound pressure levels are around the same magnitude as the systems receivers. If these magnitudes are not similar; the system could experience what is known as the near far problem. This is when nodes that are closer receive transmissions stronger than nodes that are farther away. The stronger nodes can project interference that can create issues with demodulation and detection of the weaker

nodes that are farther away. The near far problem would be a greater issue with underwater sensor networks residing in deeper waters. The advantage that CDMA has over TDMA is that it can transmit data simultaneously between nodes. However, a CDMA system that experiences the far near problem is closely comparable to TDMA systems [7]. Table 1 shows a comparative performance of access technologies for underwater networks.

TABLE I. UNDERWATER CHANNEL PERFORMANCE

Parameter	CDMA	FDMA	TDMA
Scalability	✓		
Complexity			✓
Security	✓		
Synchronization			✓
Throughput	✓	✓	

IX. CONCLUSION

The development of underwater sensor networks is driven by the potential applications. These applications will advance our military capabilities to new highs, help us to discover organisms we did not know existed, study climate changes, and allow for scientific data collection to predict natural disasters before they happen. These applications will protect our country and the people from disasters and war. Discovering new organisms from the data collection taken at the bottom of the ocean could possibly be the cures we need for diseases such as cancer and AIDS. The challenges we face in building these underwater sensor networks must be studied extensively to determine the correct technological approaches to use in creating them.

Most of the protocols that are in use have been used on land dealing with radio waves. Radio waves are very different than acoustic waves in water. The way to combat these differences is to find the right combinations of equipment and protocols to make hybrid systems to use. Hybrid systems can make use of the strengths in each of the forms of technology to combat all of the design challenges and limitations. Technology trends in underwater sensors and instrumentation have led to the miniaturization and increased energy efficiency of instruments. Use of advanced manufacturing technologies, such as Micro-Electro-Mechanical Systems (MEMS) and nanotechnology, has reduced the size of instruments. Increase of platforms functionality while reducing operational costs has been achieved by integration of multiple sensors on a single platform. The research on underwater networks continues to advance, and it is encouraging that one day these underwater sensor networks will become a reality that will advance our nation to new heights over the next several years.

REFERENCES

- [1] I.F. Akyildiz, D. Pompili, and T. Melodia, "Underwater Acoustic Sensor Networks: Research Challenges." *Ad Hoc Networks Journal*, 2005, vol. 3, no. 3, pp. 257-279, doi: 10.1016/j.adhoc.2005.01.004
- [2] J. Fitzgerald, and A. Dennis, "Business Data Communications and Networking". Wiley, 2012, 11th edition,

- [3] B.A. Forouzan, "Data Communications and Networking" McGraw-Hill, 2013, 5th edition
- [4] L.G. Garreiro, and A.A. Burke, "Unmanned Underwater Vehicles" 7th Annual SECA Workshop and Peer Review, Sep. 2006
- [5] J. Heidemann, M. Stojanovic, and M. Zorzi, "Underwater Sensor Networks: Applications, Advances, and Challenges." Philosophical Transaction of the Royal Society, Jan. 2012, vol. 370, no. 1958, pp. 158-175, doi: 10.1098/rsta.2011.0214
- [6] R.B. Manjula, and M.S. Sunilkumar, "Issues in Underwater Acoustic Sensor Networks" International Journal of Computer and Electrical Engineering, Feb. 2011, vol. 3, no.1, pp. 101-110
- [7] R. Otnes, "Underwater Acoustic Networking Techniques" Springer Briefs in Electrical and Computer Engineering, 2012, pp. 1-4, doi: 10.1007/978-3-642-25224-2_1
- [8] D. Pompili, and I.F. Akyildiz, "Overview of Networking Protocols for Underwater Wireless Communications" IEEE Communications Magazine, Jan. 2009, vol. 47, no. 1, pp. 97-102, doi: 10.1109/mcom.2009.4752684
- [9] R. Sutton, W. Naeem, M. Ahmad, and R.S. Burns, "A Review of Guidance Laws Applicable to Unmanned Underwater Vehicles" The Journal of Navigation, Jan. 2003, vol. 56, no. 1 pp. 15-29, doi: 10.1017/S0373463302002138
- [10] J. Lee, M. Turnipseed, and L. Brun "Market and Technology Trends in Underwater Sensors & Instrumentation" Marine Technology Reporter, Nov-Dec. 2012, vol. 55, no. 9, pp. 36-39
- [11] M. Stojanovic, "Underwater Acoustic Communications," entry in Encyclopedia of Electrical and Electronics Engineering, John G. Webster, Ed., John Wiley & Sons, 1999, vol. 22, pp. 688-698.
- [12] J.G. Proakis, "Digital Communications", McGraw-Hill, 2001, 4th edition
- [13] T.S. Rappaport, "Wireless Communications: Principles and Practice", Dorling Kindersley, 2009, 2nd edition
- [14] T. Fulton, "Simulation and In-water Testing of the Mid-Sized Autonomous Research Vehicle", Unmanned Maritime Vehicle T&E Conference, June 2005

In Figure 1, we describe the structure of the Adaptive-MCS system with RN operation. At the Evolved Node B (eNB), the data is coded, interleaved, modulated and then, transmitted through the channel. Once at the receiver the channel condition is estimated with an SNR criterion, and this information is sent back to the transmitter, which decides which MCS level to use. The previous channel condition parameters are stored in a buffer. When the signal arrives at the Relay Node (RN), we select the protocols for various scenarios by first analyzing the channel parameters given by the Channel Evaluator (CE) from the RN-UE link. As per the performance of CE, the suitable MCS level is chosen for the best average throughput performance. Choosing the MCS level means to select a specific code rate and modulation scheme according to the estimation of the channel conditions. Based on the idea of pre-evaluated channel quality, we select the favorable relay protocol. This data is then sent to the User Equipment (UE). The UE also analyzes the CE, based on the channel condition between RN-UE links. If the channel condition is favorable, a high order of modulation and code rate are used. Otherwise, a low order of modulation and code rate are selected. With the appropriate MCS level, AMC can obtain both excellent throughput performance and quality for a specific channel condition.

III. PROPOSED CRITERIA FOR ADAPTIVE MCS SELECTION AND PROTOCOL DESIGN

Adaptive Modulation and Coding is performed according to several SINR regions. Here, we first discuss the region boundary for the modulation regarding the modulation adaptation among various schemes of modulation as QPSK and 16-QAM, with a code rate of 1/3 and 3/4. Let Y_{sr} and Y_{rd} denote the received SINR of the SR and RD link. P_{SR} & P_{RD} can be the error probability for the Source-Relay link and Relay-Destination link, respectively. If the RN can obtain data correctly with the probability of $(1-P_{SR})$, the final errors are calculated from the detection of the combined SD and RD link, P_{SD} . When the relay cannot acquire the data correctly at the SR link the probability is given by, P_{SD} . Thus, the total BER for this state is given by (1)

$$P_e = (1 - P_{SR})P_{SD} + P_{SR} \quad (1)$$

We know that BER of M-QAM modulation can be obtained as (3)

$$P = \alpha Q\sqrt{\beta\gamma} \quad (2)$$

where, $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{u^2}{2}} du$, where α and β are decided by the modulation scheme. But, the above scheme is complex for inversion. So, to simplify the above design and performance analysis, we model the expression, where n is the MCS level, as

$$\text{BER}_n(\gamma) \cong \alpha_n \exp(-b_n \gamma) \quad (3)$$

Here, we analyze various characteristics of the scheme comprising the Adaptive-MCS with Relay.

A. Precoding Scheme

The pre-coding scheme is located at the eNB. This improves the system performance by using the estimated channel information calculated at the RN. There are several techniques used for pre-coding, such as Pre-Zero Forcing (ZF) and Pre-Minimum Mean Square Error (MMSE).

B. Relay Protocols with AMC

The relay protocols considered in our research paper are AF, DF and DMF. We will now evaluate various relay protocols with AMC. Figure 2 shows a flowchart of the Relay scheme with AMC. Firstly, the Channel State Information (CSI) is calculated based on the link condition and estimation. Then each protocol is selected as per the situation since all protocols have the same amount of maximum throughput. Suppose, we select the DMF protocol. Then, we check the given MCS level. In the MCS level, we then check the type of the code rate and modulation. When satisfied as defined by the condition, we again check the CSI for the next link of relay and the UE, as stored in the buffer at the relay node. Based on the estimation of the CSI, we make the error check of the present CSI and previous CSI. If the state is true we can then calculate the final throughput estimate as per the given MCS level; otherwise, we need to recheck the MCS level and append the new CSI value in the relay node. Then, we need to verify the MCS level for the RN-UE link. Once the throughput is estimated, the new data frame, which is needed to be transmitted to the UE with lower error probability, needs to be verified. The dotted part in the flowchart shows the main performance area in the algorithm.

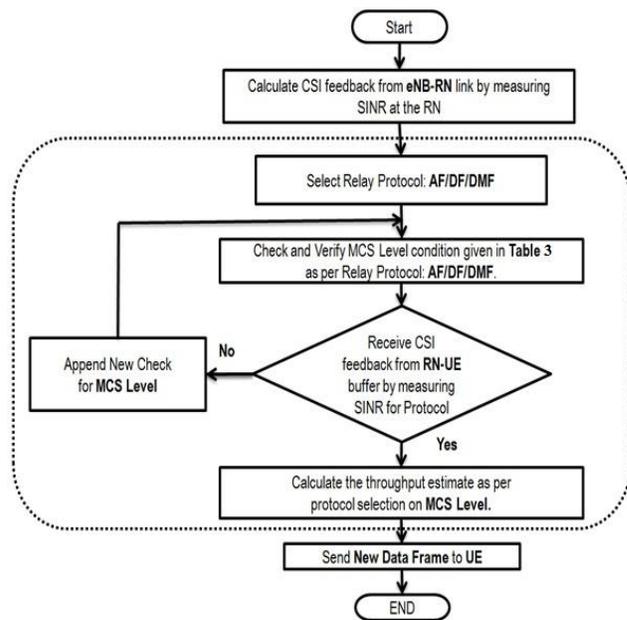


Figure 2. Flow Chart of AMC with Relay Scheme

C. AF with MCS

Simply called an AF protocol, the relay tends to scale the received version of the signal and transmits an amplified version of it to the destination or UE. It is the most basic type of fixed relay. It simply amplifies the received data and then forwards it to the UE. The relay is usually not capable of performing in a bad channel condition as it amplifies the noise factor of the received data vector. But in order to compare its performance with a new DMF protocol, we use AF for analysis as a reference case of the standard known protocol. When MCS is applied to the AF with various code rates and modulation schemes we can observe the progressive improvement in the case of adaptive MCS [10]. As given in Figure 1, when we select the AF protocol, then as per the channel condition, the SINR value is estimated from the channel evaluator, by using the precoding scheme at the RN and the UE.

D. DF with MCS

The relay node is for the relay to decode the received signal, re-encode it, and then retransmit it to the receiver. This kind of relaying is termed as a fixed decode-and-forward (DF) scheme, which is often simply called a DF scheme.

In the basic DF relaying scheme, we implement SINR estimation using precoding methodology, which actually evaluates the channel and helps to estimate the SINR value with channel quality. The decoding operation is repeated multiple times which helps in improving the system performance by noise and interference reduction.

E. DMF with MCS

DMF protocol signal processing is an alternative to DF signal processing to reduce receiver power consumption due to channel decoding at the relay as well as to minimize the overall delay at the destination. In the DF schemes previously described, the relay forwards the source’s message only if it is able to successfully decode. However, in many applications, channel decoding may not be desirable at the relays either due to limited transceiver capabilities or due to lack of knowledge of the channel codebook. In this case, the signals transmitted by the source can only be detected or demodulated on a symbol-by-symbol basis. At this position there is a need for a relay protocol capable of performing much better in decoding performance. So, we design a new protocol capable of performing in both cases whereby in the bad decoding case it is capable of maintaining the high error bits decoding with soft decoding and the higher modulation scheme.

IV. SIMULATION RESULTS

The simulation results are based on the link level Monte Carlo simulations. Noise components are the same at all channel links, but channel fading component changes increase and decrease, based on the links, as eNB-RN link and RN-UE link characteristics. Table I shows the

simulation parameters based on 3GPP LTE-Advanced 20 MHz bandwidth.

TABLE I. SIMULATION PARAMETERS FOR 20MHZ

Parameter	Value
Carrier Frequency	2 GHz
Bandwidth	20 MHz
Subcarrier spacing	15 KHz
Sub frame Duration	1 ms
FFT Size	2048
No. of subcarriers/PRB	12
Channel	EPA, EVA, ETU
Modulation scheme	QPSK,16 QAM
Noise	AWGN
Relay Node (RN)	1
Relaying Protocol	AF, DF, DMF

A. FER, SER and BER Analysis of AF,DF and DMF Protocols

We will now discuss the error performance and analysis of all three protocols in order to clarify their behavior in our adaptive MCS relay system.

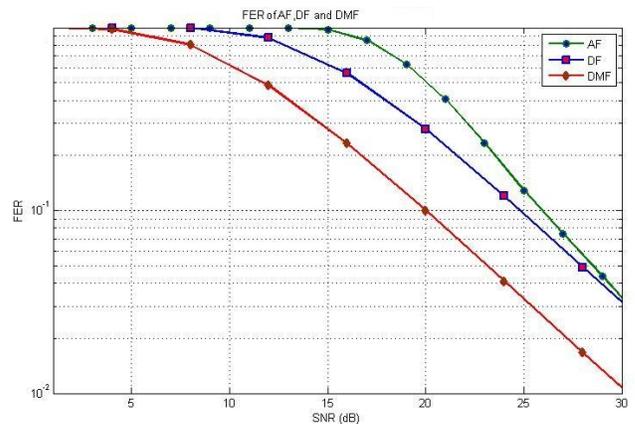


Figure 3. FER results with AF,DF and DMF protocols.

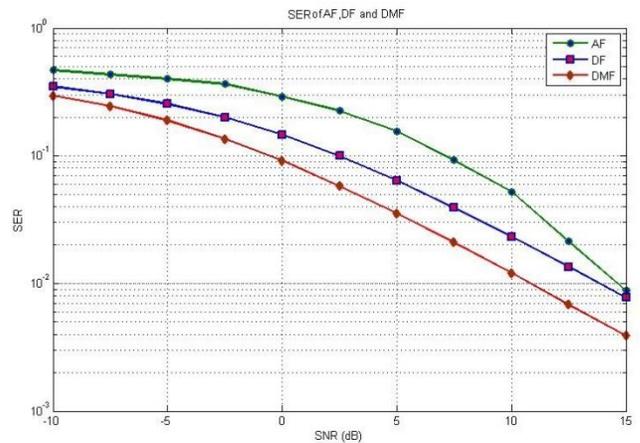


Figure 4. SER results with AF,DF and DMF protocols.

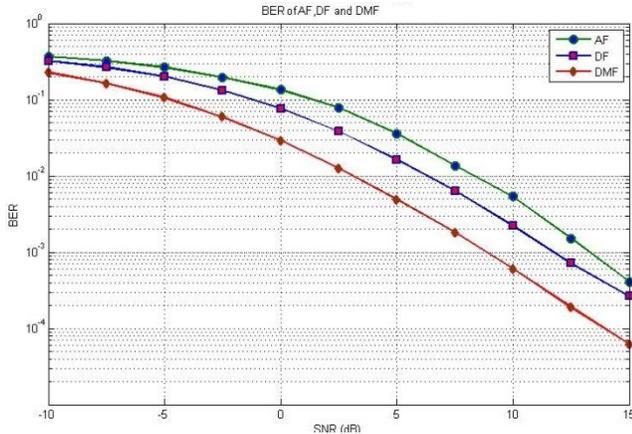


Figure 5. BER results with AF,DF and DMF protocols.

Figure 3 presents the Frame Error rate (FER) performance for the 3 kinds of protocols. We can clearly observe that the performance of the DMF protocol shows a very high trade off compared to AF and DF protocol. The tradeoff between AF and DF at higher SNR shows very little error performance advantage but on the other hand the DMF protocol shows explicit performance.

Figure 4 analyzes the Symbol Error Rate (SER) performance analysis of the AF, DF and DMF protocol. Compared to Figure 3 the performance of AF, DF and DMF show slight improvement in performance. The simulation examines the error probability based on each symbol transmission. Therefore, compared to the FER performance this figure demonstrates better results, as the error encountered in the symbol rate is much more reduced in comparison to the frame error of each case.

Figure 5 shows the Bit Error Rate (BER) analysis, which still shows better error performance for DMF compared to the AF and DF protocols. Here, we can observe that even at an improved error performance than compared than to the FER or SER, AF and DF performance are very close at lower SNR values, whereas, the performance of DMF still shows better results than the conventional protocols.

B. Non-Adaptive MCS with AF,DF and DMF Protocols

Table II shows the Non-Adaptive MCS level table, with various MCS level for AF, DF and DMF protocols based on the code rate, as turbo coding with 1/3 and 3/4. The modulation schemes followed in this case are QPSK and 16-QAM. We observed various values of maximum throughputs in different code rates and modulation schemes. The maximum throughput achieved by all the coding schemes and modulation schemes are different at various levels. The maximum throughput in all MCS is independent of the nature of protocol.

TABLE II. NON-ADAPTIVE MCS WITH RELAY

MCS Level	Protocol	Code Rate	Modulation	Max. Throughput (Mbps)
1	AF	1/3	QPSK	14.4
2	AF	3/4	QPSK	21.6
3	AF	1/3	16-QAM	28.8
4	AF	3/4	16-QAM	43.06
5	DF	1/3	QPSK	14.4
6	DF	3/4	QPSK	21.6
7	DF	1/3	16-QAM	28.8
8	DF	3/4	16-QAM	43.06
9	DMF	1/3	QPSK	14.4
10	DMF	3/4	QPSK	21.6
11	DMF	1/3	16-QAM	28.8
12	DMF	3/4	16-QAM	43.06

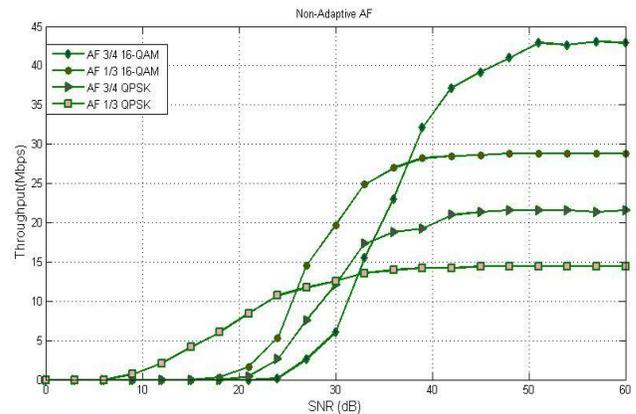


Figure 6. Throughput of Non Adaptive MCS AF Protocol.

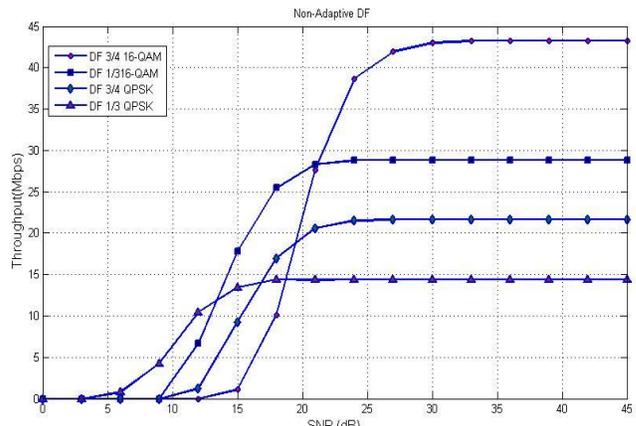


Figure 7. Throughput of Non Adaptive MCS DF Protocol.

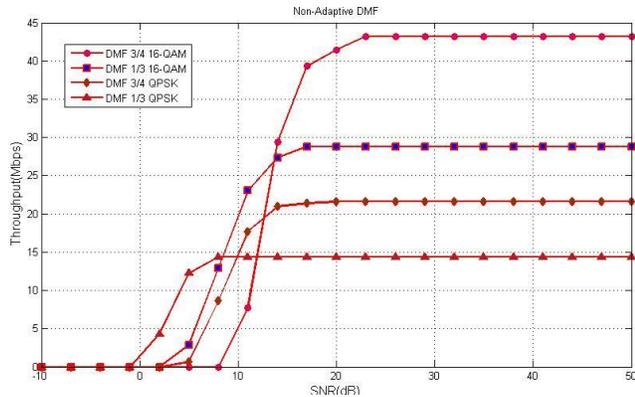


Figure 8. Throughput of Non Adaptive MCS DMF Protocol.

Figure 6 shows the maximum throughput values at different code rates for the AF protocol. The maximum throughput for code rate 1/3 with QPSK is observed, of approximately 14.4 Mbps. As the code rate is increased in the case of the same modulation, we can observe an increase in throughput at 21.6 Mbps. However, we have to compromise with SNR performance in this case. Similarly, with the increase in the modulation scheme, we can observe the increase in throughput rate. Finally, we observed that the higher the code rate and modulation, the higher the throughput but we have compromised the SNR performance which is gained at a very high SNR.

Figure 7 shows the performance analysis for the Non - Adaptive MCS DF protocol. The maximum throughput is the same in the case of AF and DF, as seen in Table 2. But, as we observe the values of SNR tradeoff we can see a considerable gain in the throughput case. We can observe that better decoding schemes improve the relay performance for relay code rate and modulation.

Figure 8 shows the throughput of the Non-Adaptive MCS DMF protocol, as the case of AF and DF. We can understand that the code rate and modulation is the same for all protocols and so is the maximum throughput as seen in Table 2. If we look at the SNR Performance we can see that the DMF protocol achieves higher throughput at lower SNR values.

C. Non-Adaptive MCS with AF,DF and DMF Protocols

Table III shows the values for the Adaptive MCS level for the AF, DF and DMF Protocol for various values for code rate and modulation scheme. We made observations and analysis on the basis of Signal-to-Noise-Ratio (SNR) and Average Throughput in Mega bit per second (Mbps).The rate are chosen based on the best code rate and best modulation scheme provided, with the highest throughput order [10].

TABLE III. ADAPTIVE MCS WITH RELAY

MCS Level	Protocol	Code Rate	Modulation	SNR (dB)	Avg. Throughput (Mbps)
1	AF	1/3	QPSK	21	12.384
2	AF	1/3	16-QAM	36	26.976
3	AF	3/4	16-QAM	51	42.984
4	DF	1/3	QPSK	12	10.468
5	DF	1/3	16-QAM	21	27.65
6	DF	3/4	16-QAM	30	42.984
7	DMF	1/3	QPSK	8	14.4
8	DMF	1/3	16-QAM	14	27.36
9	DMF	3/4	16-QAM	23	42.984

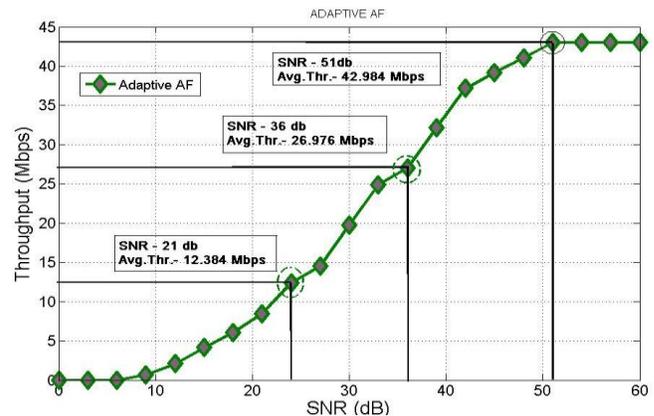


Figure 9. Adaptive Throughput of adaptive MCS AF Protocol

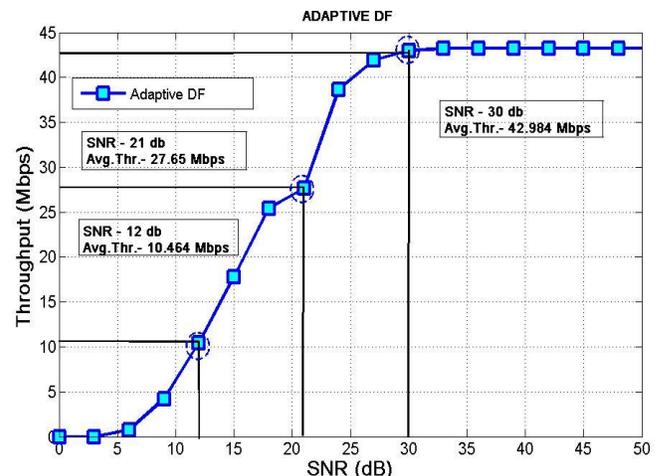


Figure 10. Adaptive Throughput of adaptive MCS DF Protocol

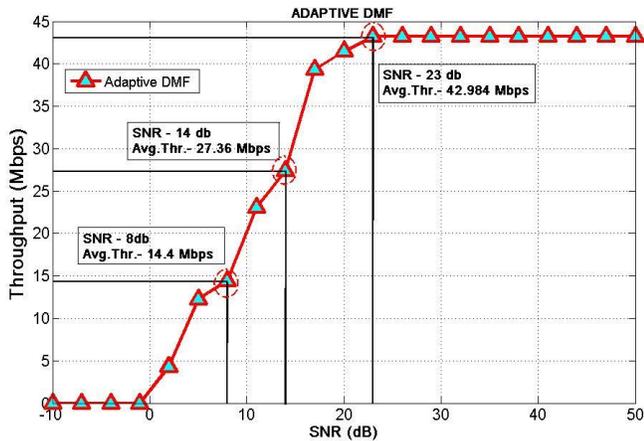


Figure 11. Adaptive Throughput of adaptive MCS DMF Protocol

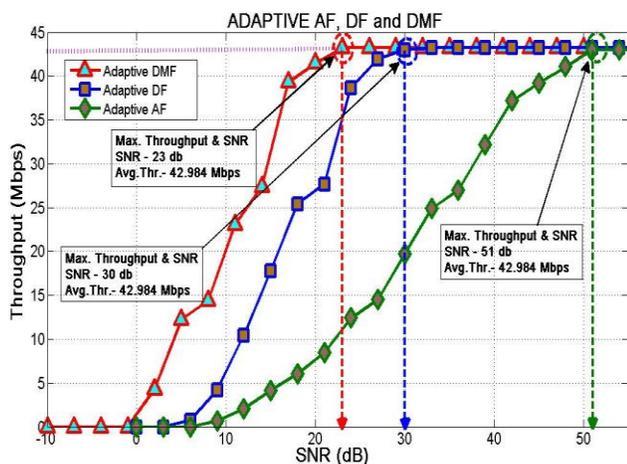


Figure 12. Average Throughput Analysis of Adaptive MCS Relay Protocols

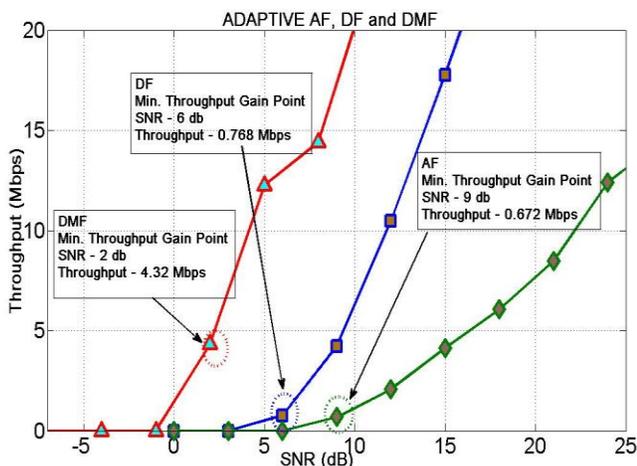


Figure 13. Maximum Throughput Analysis of Adaptive MCS Relay Protocols

Figure 9 shows the average throughput of the adaptive MCS AF protocol. We analyzed the new MCS level for the specified code rate and modulation schemes for the AF case to achieve the average throughput and maximum throughput

at the same time. The rapid increase in the throughput is calculated on the basis of maximum throughput achieved at lower code rates and lower modulation schemes. As analyzed points are shown in Figure 9 the MCS level 1 on SNR at an average throughput of 21 dB was 12.364 Mbps. Then the MCS level 2 is switched to SNR 36 dB point and an average throughput of 26.976 Mbps. Following this, MCS level 3 is switched until the maximum throughput is achieved at SNR 51dB with an average throughput of 42.984 Mbps.

Figure 10 shows the average throughput of the adaptive MCS DF protocol. The rapid gain in average throughput is observed similarly to the case of AF. As analyzed points are shown in Figure 10 the MCS level 4 is selected first on SNR at 12 dB at an average throughput of 10.468 Mbps, then the MCS level 5 is switched to SNR 21 dB point and average throughput of 27.65 Mbps, and then the MCS level 6 is switched till the maximum throughput is achieved at SNR 30dB and an average throughput 42.984 Mbps. As far as the maximum throughput is considered it is the same for all protocols.

Figure 11 shows the average throughput of the adaptive MCS DMF protocol. The rapid gain in average throughput is observed best in the case of DMF. As analyzed points are shown in figure 11 the MCS level 7 is selected first on SNR at 8 dB at an average throughput of 14.4 Mbps, then the MCS level 8 is switched to SNR 14 dB point and average throughput of 27.36 Mbps, and then the MCS level 9 is switched till the maximum throughput is achieved at SNR 23db and an average throughput of 42.984 Mbps. The maximum throughput is considered the same for all protocols but the SNR gain is best in the case regarding the DMF protocol

Figure 12 shows the Maximum throughput analysis comparing all the protocols in the case of the adaptive relay with MCS. Here, we observe that the DMF protocol shows a gain of 7 dB compared to the DF protocol. DMF attains the maximum throughput at the SNR of 23 dB and the DF gains maximum throughput at 30 dB. Similarly, as compared to the DF and DMF, AF gains maximum throughput at 51dB which is far worse than the case of DF and DMF clearly shows an SNR gain of 21 dB.

Figure 13 shows the performance of the AF, DF and DMF Adaptive MCS relay with minimum average throughput. This is in contrast with the Maximum throughput case where DMF and DF show very close performance but the performance of AF and DF show a dramatic change in gain and prove AF as the worst. In this case, DMF shows the Minimum throughput gain at 2 dB with 4.32 Mbps, whereas DF shows a throughput gain at SNR 6 dB with 0.768Mbps and AF shows SNR 9 dB and a throughput of 0.672 Mbps. This shows that in contrast of the maximum throughput, the minimum throughput demonstrates that the DMF protocol has consistent performance at high and low SNR values.

V. CONCLUSION

We propose an AMC scheme using relay protocols like AF, DF and DMF. The behaviors of these protocols are analyzed on parameters of FER, SER, BER, maximum throughput, average throughput and minimum throughput. We use the AMC scheme for improving throughput and reliability, because of the nature of different modulation and coding schemes. The simulation results of the proposed system with adaptive MCS prove that among the AF, DF and DMF protocols, the DMF protocol performs best specifically at a lower SNR value and also provides better average throughput. We observed that the proposed DMF protocol is capable of performing with the best performance in lower and high SNR values and with high consistency and provides the best throughput efficiency. The main consideration point in the proposed mechanism is the application of the DMF protocol, which when implemented with the AMC scheme shows outstanding results compared to the conventional AF and DF schemes.

ACKNOWLEDGMENT

This research was supported by the MSIP (Ministry of Science, ICT & Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (NIPA-2013-H0301-13-3005) supervised by the NIPA (National IT Industry Promotion Agency). This study was financially supported by Chonnam National University, 2012. This research is supported by Korea Research Council for Industrial Science and Technology under B551179-12-07-00.

REFERENCES

- [1] R. Pabst, B. H. Walke and D. C. Schultz, "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Commun. Mag.*, vol. 42, no. 9, Sep. 2004, pp. 80–89.
- [2] S. W. Peters, A. Y. Panah, Kien T. Truong, and R. W. Heath, Jr., "Relay architectures for 3GPP LTE-advanced," *EURASIP J. Wirel. Commun. and Network.*, vol. 2009, Article ID 618787, 2009, 14 pages.
- [3] S. Sesia, I. Toufik, and M. Baker, *LTE, The UMTS Long Term Evolution: From theory to practice*, Wiley & Sons, Feb. 2009.
- [4] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocol and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, pp. 3062-3080, Dec. 2004.
- [5] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity-part i: System description," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1927–1938, Nov 2003.
- [6] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity - part ii: Implementation aspects and performance analysis," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1939–1948, Nov 2003.
- [7] G. Holland, N. Vaidya and P. Bahl, "A rate-adaptive MAC protocol for multi-hop wireless networks," *Proc. of the 7th MobiCom*, pp.236-251, Rome, Italy, 2001.
- [8] S.G. Chua and A.J. Goldsmith, "Adaptive coded modulation for fading channels", *IEEE Transaction on communications*, May 1998, pp. 595-6022.
- [9] S.K. Lai, R.S. Cheng, K.B. Letaief, and R.D. Murch, "Adaptive trellis coded MQAM and power optimization for OFDM transmission", *IEEE Vehicular Technology Conference*, May 1999, pp. 290-295.
- [10] J.H. Lee, G.S. Yoon, I.S. Cho, C.W. Seo, S. Portugal and I.T. Hwang, "Design and Performance Analysis of a Communication System with AMC and MIMO Mode Selection Scheme", *Journal of The Institute of Electronics Engineers of Korea*, vol. 47-TC, No. 3, March 2010.
- [11] T. Keller, L. Hanzo, "Adaptive modulation techniques for duplex OFDM transmission", *IEEE Transaction on Vehicular Technology*, Sept 2000, pp. 1893-1906.
- [12] J. G. Proakis, *Digital Communications*, 4th ed. McGraw-Hill, New York, 2001.

Second Best Codeword for MIMO Broadcast Channels with Limited Feedback

Mouncef Benmimoune and Daniel Massicotte

Electrical and Computer Engineering Department
 Laboratory of Signal and System Integrations
 Université du Québec à Trois-Rivières
 3351, Boul. des Forges, Trois-Rivières, QC, CANADA
 Email: {mouncef.benmimoune, daniel.massicotte}@uqtr.ca

Sébastien Roy

Electrical and Computer Engineering Department
 Sherbrooke University
 2500, BLD de l'Université, Sherbrooke, QC, CANADA
 Email: s.roy@ieee.org

Abstract—In this paper, we propose a simple design for a multi-user MIMO precoder using a Grassmannian codebook. The precoder design is applied to both sides of the wireless link under the assumption of limited feedback and the same codebook. The proposed scheme aims to correct the precoding vectors choice in order to avoid more than one user having the same vector, a situation which degrades system performance. Simulation results show that our approach provides greater system performance enhancement in both error probability and sum rate.

Keywords—MIMO broadcast channels; precoding; codebook; codeword correction; limited feedback.

I. INTRODUCTION

Recently, Multi-Input Multi-Output (MIMO) technology has been introduced in several applications, such as wireless LANs and cellular telephony because of its high gain in both channel capacity and reliability. The MIMO concept consists in deploying multiple antennas at both the transmitter and the receiver in order to exploit the spatial dimension. This new dimension can be used in a single-user scenario (SU-MIMO), providing a system capacity that increases linearly with the multiplexing gain, regardless of the availability or otherwise of the channel state information at the transmitter (CSIT) [1]. Or it can be leveraged in a multiuser configuration (MU-MIMO), where several mobile stations (MS) communicate with a base station (BS) while sharing the same time-frequency resource.

In multi-user case, the situation vis-a-vis the sum capacity is substantially different, since the interference must be taken into account and balanced in a trade-off against the data rate [2]. As a result, the CSIT is required since it critically affects the multiplexing gain [3].

In information theory, it is well known that the optimal strategy for achieving sum capacity in a MIMO broadcast channel is dirty paper coding (DPC) [4]. However, deploying DPC in real-time systems is impractical due to the complexity of the encoder and the decoder. Moreover, it is sensitive to CSIT inaccuracy. For low complexity and suboptimal performance, several non-linear and linear precoders have been proposed [5], [6], [7], [8]. However, all these schemes need perfect CSIT, which is not practical.

In the literature, a considerable number of publications have assumed partial CSIT, where unitary precoding is the

main, denoted also as limited feedback (see, e.g., [9] and the references therein). Recently, this scheme was introduced in IEEE 802.16 e/m and LTE Advanced for its performance. In fact, it was shown in [10] that for SU-MIMO systems, even a few bits of feedback, of the order of the number of transmit antennas, is sufficient to achieve near-optimal performance. For MU-MIMO systems, however, the amount of feedback per user must grow linearly with the number of transmit antennas and the SNR in decibel [1].

The use of the same codebook in all MS causes a rank loss in the channel matrix since two or more users can choose the same codeword. Few solutions to avoid this situation are found in the literature. Indeed, most papers suggest using different codebooks (see, e.g., [1]). But this is not quite feasible. In Ding *et al.* [11], the authors propose to modify the codebook of each user by a rotation of the general codebook by a random unitary matrix that is known at the BS. However, this solution is impractical because the BS must know all unitary matrices. All the more, the latter requires a storage and more feedback overhead for matrix's identification. To address this problem, in this paper, we propose to make a correction of the precoding vectors choices at the BS while keeping the same codebook.

Throughout the paper, lower-case bold letters are used for vectors and upper-case bold letters for matrices; $\|\cdot\|_2$ denotes the Euclidean norm of the vector; $|\cdot|$ denotes the cardinality of a set or the absolute value of the scalar; $Tr(\cdot)$ denotes the matrix trace; and $(\cdot)^H$ and $(\cdot)^{-1}$ denote Hermitian and matrix inversion, respectively. The identity matrix is denoted by \mathbf{I} .

The rest of this paper is organized as follows. The MU-MIMO system model and the Grassmannian codebook construction are depicted in the next section. The problem formulation and the proposed precoding design under limited feedback is discussed in Section III. Simulation results are given in Section IV, and we conclude with Section V.

II. SYSTEM MODEL

A. Multi-User MIMO System with Limited Feedback

Let us consider a closed loop MU-MIMO broadcast channel with N_t antennas at the BS and N_r antennas at each of K MS. The vector data $\mathbf{s} \in \mathbb{C}^{K \times 1}$ is preprocessed by the

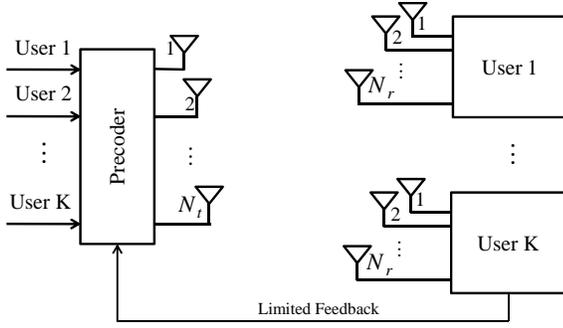


Fig. 1. Downlink Multi-User MIMO system

precoding matrix $\mathbf{W} \in \mathbb{C}^{N_t \times K}$ and then transmitted via flat-fading channel $\mathbf{H}_k \in \mathbb{C}^{N_r \times N_t}$ between the BS and the k^{th} MS. At the k^{th} MS, the equivalent baseband input-output relationship can be written as

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{W} \mathbf{s} + \mathbf{n}_k, \quad (1)$$

where $\mathbf{n}_k \in \mathbb{C}^{N_r \times 1}$ is the noise term which follows an independent complex Gaussian distribution with zero mean and N_0 variance, i.e. $\mathcal{CN}(0, N_0)$.

At the receiver, a linear MMSE detector is used to estimate the symbols transmitted, the decoding matrix of the k^{th} user being given by [12]

$$\tilde{\mathbf{g}}_k = \tilde{\mathbf{h}}_k^H \left(\tilde{\mathbf{H}}_k \tilde{\mathbf{H}}_k^H + \frac{KN_0}{P_t} \mathbf{I}_{N_r} \right)^{-1}, \quad (2)$$

where

$$\tilde{\mathbf{H}}_k = \mathbf{H}_k \mathbf{W}. \quad (3)$$

$$\tilde{\mathbf{h}}_k = \mathbf{H}_k \mathbf{w}_k. \quad (4)$$

In this case, the received signal-to-interference and noise ratio (SINR) for the k^{th} user is [12]

$$\text{sinr}_k(\mathbf{H}_k, \mathbf{w}_k) = \frac{\frac{P_t}{N_t} |\tilde{\mathbf{g}}_k^H \mathbf{H}_k \mathbf{w}_k|^2}{\sum_{i=1, i \neq k}^K \frac{P_t}{N_t} |\tilde{\mathbf{g}}_k^H \mathbf{H}_k \mathbf{w}_i|^2 + N_0 \|\tilde{\mathbf{g}}_k\|_2^2}, \quad (5)$$

where P_t represents the total transmit power which must satisfy the following constraint

$$E[\|\mathbf{W} \mathbf{s}\|_2^2] = \text{Tr}(\mathbf{W} \mathbf{W}^H) \leq P_t, \quad (6)$$

where \mathbf{W} is given by $\mathbf{W} = [\mathbf{w}_1 \mathbf{w}_2 \dots \mathbf{w}_K]$.

The downlink sum rate can be expressed as follows

$$r(\mathbf{H}_k, \mathbf{w}_k) = \sum_{k=1}^K \log_2(1 + \text{sinr}_k(\mathbf{H}_k, \mathbf{w}_k)). \quad (7)$$

B. Grassmannian Codebook

The problem of Grassmannian line packing is to find the packing of P lines in \mathbb{C} such that the distance between any pair of lines is maximal. This packing of P lines is captured by the matrix $\mathbf{V} = [\mathbf{v}_1 \mathbf{v}_2 \dots \mathbf{v}_P]$, where each column \mathbf{v}_i represents a unit vector with $\mathbf{v}_i^H \mathbf{v}_i = 1$ and $\mathbf{v}_i^H \mathbf{v}_j \neq 1 (i \neq j)$. The distance

between two lines produced from unit vectors \mathbf{v}_1 and \mathbf{v}_2 can be defined by the function $d(\mathbf{v}_1, \mathbf{v}_2)$, which represents the sine of the angle formed by these two lines [13], [12]

$$d(\mathbf{v}_1, \mathbf{v}_2) = \sin(\theta_{1,2}) = \sqrt{1 - |\mathbf{v}_1^H \mathbf{v}_2|^2}. \quad (8)$$

To maximize the precoding gain with codebook vectors is to maximize the minimum distance between any pair of lines spanned by the codebook vectors [12].

The minimum distance of a packing is the sine of the smallest angle between any pair of lines, i. e.

$$\delta(\mathbf{V}) = \min_{1 \leq k < l \leq P} \{\sqrt{1 - |\mathbf{v}_k^H \mathbf{v}_l|^2}\}. \quad (9)$$

The Grassmannian codebook is designed so that

$$\{\mathbf{V}\} = \arg \max_{\mathbf{X} \in \mathcal{U}_{N_t}^P} \{\delta(\mathbf{X})\}, \quad (10)$$

where $\mathcal{U}_{N_t}^P$ is defined as the set containing all unitary matrices.

III. MULTI-USER MIMO PRECODER DESIGN

In this paper, we address a problem that is common to all common codebook in limited feedback broadcast systems. As stated in the introduction, prior work has mainly focused on rotational schemes. Moreover, the standardization processes have so far ignored this problem. Thus, this paper comes up with a very simple and interesting solution.

In our proposal, we share the same codebook between the BS and all MS. To this end, we consider a Grassmannian codebook of size P . We consider a practical scenario where the partial CSI is provided at the BS through limited feedback. We assume that each user has a knowledge of its own CSI and no information about the channels of other users. In our study, we assume also that the feedback channel is error free and delay free and the Channel Quality Indicator (CQI) is sent directly (without quantization), since we are interested only in the effects of limited feedback on the precoding. However, the number of bits needed for the quantization of CQI is relatively low [14].

The detailed procedure of the proposed scheme is as follow :

1) First, each MS calculates the CQI, such as :

$$\mathbf{cqi}_k = \|\mathbf{H}_k \mathbf{v}_p\|_2^2; p = 1, \dots, P. \quad (11)$$

2) Each MS must find the index of the best precoding vector in the Grassmannian codebook that can maximize the CQI :

$$\mathbf{w}_k = \arg \max_{\mathbf{v}_p \in \mathbf{V}} \{\mathbf{cqi}_k \mid p = 1, \dots, P\}. \quad (12)$$

3) After obtaining the best precoding vector, each MS transmits its own index and the CQI to the BS through finite rate feedback.

4) The BS recovers the position of the precoding vector.

5) If two or more precoding vectors are identical, the BS asks the users with low CQI to transmit their higher CQI and respective index, excluding the first choice.

6) Else, the precoder combines the signal of each user with the selected \mathbf{w}_k .

IV. SIMULATION RESULTS

In this section, we compare the performance in terms of error probability and sum rate of the proposed scheme regarding the proposals in [11] and [12]. In the simulations, the approach of Ding *et al.* [11] implies strictly the codebook rotation procedure. Also, for the sake of simplicity we consider the scheme of Fang *et al.* [12] without user selection.

In all simulations, a MU-MIMO broadcast channel is considered where the BS is equipped with two antennas transmitting to two MS with two receiver antennas each. We also consider that the BS transmits one data stream to each user with the same power allocation, QPSK modulation is taken into account and linear MMSE detector is adopted.

Results in Fig. 2 shows the comparison of BER performance versus the SNR. According to Fig. 2, we can see clearly that our proposal has the best performance. A gain of 2 dB and 4 dB is measured at 10^{-1} and 10^{-2} of BER, respectively.

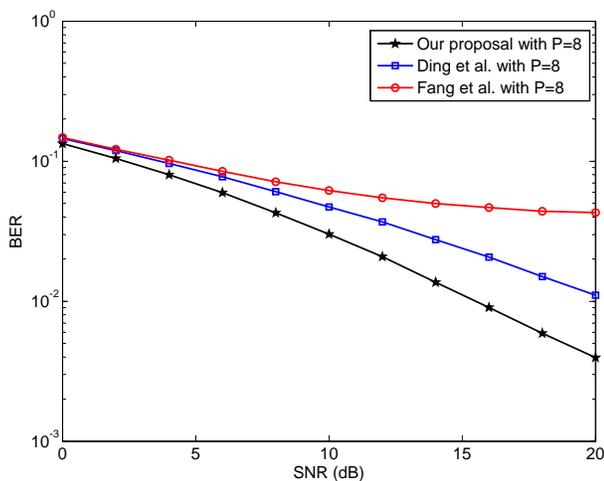


Fig. 2. BER comparison versus the SNR for 2x2 MU-MIMO system with $K = 2$ and $P = 8$.

Fig. 3 presents the sum rate CDF for the three schemes. The figure shows that the proposed scheme outperforms methods [11] and [12]. The figure shows also almost the same performance for methods [11] and [12]. It should be noted that the curves are plotted for SNR= 5dB.

The comparison of ergodic sum rate versus the SNR is depicted in Fig. 4. The figure shows that our proposal is superior regardless of the SNR level. As expected, a gain of 1 dB and 2 dB is observed for a low and high SNR, respectively.

V. CONCLUSION

In this paper, we proposed to use a simple method to correct the choice of the precoding vectors in MU-MIMO broadcast channels. The proposed precoder is designed for the assumption of imperfect channel knowledge at the transmitter through limited feedback. The simulation results show a significant performance gain with respect to existing methods. Compared to [11], our scheme uses the common codebook in all users.

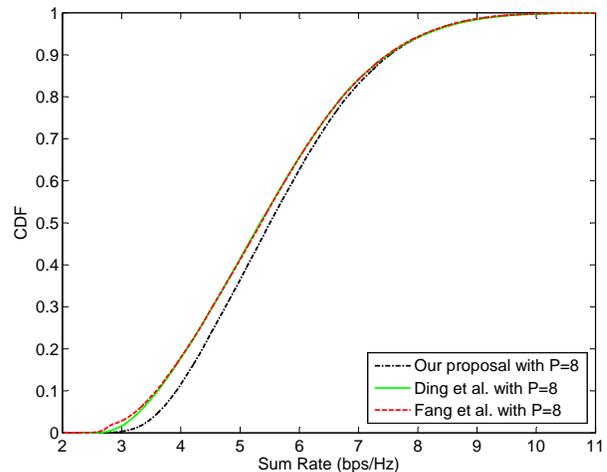


Fig. 3. Comparison of CDF of sum rate for 2x2 MU-MIMO system with $K = 2$ and $P = 8$.

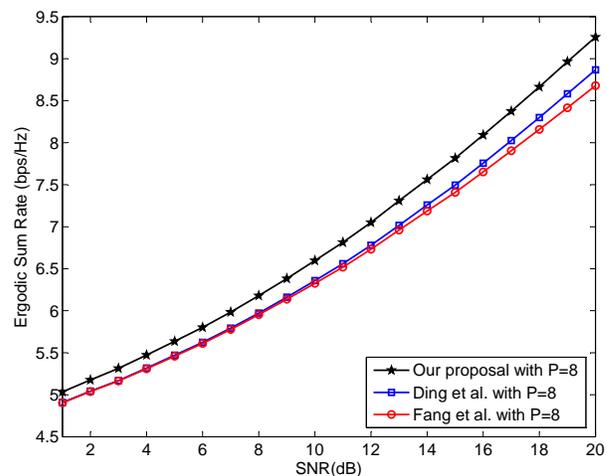


Fig. 4. Ergodic sum rate comparison for 2x2 MU-MIMO system with $K = 2$ and $P = 8$.

Moreover, the choice of precoding vectors is performed at both sides of the wireless link.

ACKNOWLEDGMENT

The authors wish to thank the Microsystems Strategic Alliance of Quebec and the Natural Sciences and Engineering Research Council of Canada for its financial support.

REFERENCES

- [1] N. Jindal, "MIMO broadcast channels with finite rate feedback," *IEEE Trans. Inform. Theory*, vol. 52, no. 11, pp. 5045-5059, Nov. 2006.
- [2] A. B. Gershman and N. D. Sidiropoulos, *Space-Time Processing for MIMO Communications*, Chichester, U.K.: Wiley, 2005.
- [3] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691-1706, Jul. 2003.
- [4] M. Costa, "Writing on dirty paper," *IEEE Trans. on Inf. Theory*, vol. 29, no.3, pp. 439-441, May 1983.
- [5] J. Jing, R. Buehrer, and W. Tranter, "Spatial T-H precoding for packet data systems with scheduling," in *IEEE Fall Veh. Tech. Conf.*, Vol. 1, pp. 537-541, 2003.

- [6] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Processing*, vol. 52, no. 2, pp.461-471, 2004.
- [7] C. Peel, B. Hochwald, A. Swindlehurst, "Vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization," *IEEE Trans. on Communications*, vol. 53, no.1, pp. 195-202, 2005.
- [8] L. U. Choi and R. D. Murch, "A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach," *IEEE Trans. on Wireless Commun.*, vol. 3, no. 1, pp. 20-24, 2004.
- [9] D. J. Love, R. W. Heath, V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341-1365, Oct. 2008.
- [10] D. J. Love, R. Heath, W. Santipach, and M. Honig, "What is the value of limited feedback for MIMO channels?," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 54-59, Oct. 2004.
- [11] P. Ding, D. J. Love, and M. D. Zoltowski, "Multiple antenna broadcast channels with shape feedback and limited feedback," *IEEE Trans. Signal Processing*, vol. 55, pp. 3417-3428, July 2007.
- [12] S. Fang, G. Wu, Y. Xiao, and S. Q. Li, "Multi-User MIMO Linear Precoding with Grassmannian Codebook," *International Conference on Communications and Mobile Computing*, vol. 1, pp.250-255, 2009.
- [13] D. J. Love, R. W. Heath, Jr., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2735-2747, 2003.
- [14] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna downlink channels with limited feedback and user selection," *IEEE J. Select. Areas Commun.*, vol. 25, no. 7, pp. 1478-1491, Sept. 2007.

Further Results on Channel Estimation Based on Weighted Moving Average in T-DMB Receivers

Sung Ik Park and Heung Mook Kim
Broadcasting System Research Department
Electronics and Telecommunications Research Institute
Daejeon, South Korea
{psi76, hmkim}@etri.re.kr

Jin-Ho Chung
School of Electrical and Computer Engineering
Ulsan National Institute of Science and Technology
Ulsan, South Korea
jinho@unist.ac.kr

Jeongchang Kim
Dept. of Electronics and Communications Engineering
Korea Maritime University
Busan, South Korea
jchkim@hhu.ac.kr

Abstract—In this paper, we describe novel channel estimation schemes for terrestrial digital multimedia broadcasting (T-DMB) receivers and present further results on their performance under fading channels. In order to improve the performance, novel channel estimation schemes utilize the weighted moving average. Numerical results show that the novel channel estimation schemes outperform the conventional scheme with approximately 2~3dB in the symbol error rate performance under fading channels.

Keywords—Channel estimation; moving average; orthogonal frequency division multiplexing (OFDM); T-DMB.

I. INTRODUCTION

In South Korea, the mobile broadcasting standard for the terrestrial digital multimedia broadcasting (T-DMB) is based on the Eureka-147 digital audio broadcasting (DAB) system [1]. In the T-DMB system, the orthogonal frequency division multiplexing (OFDM) is used for signal transmission [2][3]. The T-DMB system supports the broadcasting of moving pictures and their reception in harsh conditions such as places surrounded by high-rise buildings and highways where vehicles are moving at very high speed.

Recently, a coherent detection method has been developed to improve the reception performance of T-DMB receivers [4]. For the coherent detection, a one-tap equalizer in the frequency domain is used. The tap coefficients of the equalizer are obtained by estimating channel distortions and calculating the inverse of the estimated channel coefficients. However, the inverse of the channel coefficients can result in severe noise enhancement under deep fading channels.

In order to reduce the noise enhancement, several novel channel estimation schemes utilizing the weighted moving average were proposed in [3]. In this paper, we describe novel channel estimation schemes for T-DMB receivers and present further results on their performance under fading channels.

This paper is organized as follows. In Section II, the system model of the T-DMB system is presented. Section III describes the considered channel estimation schemes and numerical results are shown in Section IV. Finally, conclusions are drawn in Section V.

II. SYSTEM MODEL

In this section, we describe a system model for T-DMB systems. At the transmitter, binary bits are input and mapped into modulation symbols. After pilot insertion, the modulated signal $X[k]$ is input to an inverse fast Fourier transform (IFFT) block and transformed into time-domain signals. The cyclic prefix (CP) is inserted to prevent possible inter-symbol interference (ISI) between adjacent OFDM systems where a copy of the last part of the OFDM symbol is transmitted. The transmitted signal passes through the fading channel with additive white Gaussian noise (AWGN).

At the receiver, the CP is removed and the pilot based signal correction is performed after FFT [5]. There are two typical pilot patterns of the OFDM transmission systems [6]. For the block-type pilot pattern, the pilot signal is assigned to the specified OFDM symbols. Since the pilot signals are periodically sent in the time domain, the block-type pilot pattern is suitable to slow fading channels. For the comb-type pilot pattern, on the other hand, the pilot signals are uniformly distributed in the frequency domain for each OFDM symbol. Hence, the comb-type pilot pattern is suitable to fast fading channels. However, since the pilot signals are carried on the limited number of subcarriers, the channel state information for the data subcarriers should be estimated by interpolating with neighboring pilot subcarriers. As a result, the channel estimation performance with the comb-type pilot pattern is affected by the interpolation methods [8]. In the T-DMB system, the pilot signals are inserted at the fixed position in every 76 OFDM symbols. Hence, the pilot pattern in the T-DMB has the propensity of the block-type.

III. CHANNEL ESTIMATION SCHEMES

The channel estimation of the T-DMB system consists of two stages. First, the received pilot signals are extracted from the pilot subcarriers. Next, the channel coefficients are obtained by comparing the received pilot signals with the known pilot signals. The received pilot signals $Y_p[k]$ are extracted from the received signals $Y[k]$ after FFT, and the reference channel coefficients for the equalizer are estimated in the frequency domain by comparing the extracted pilot signals $Y_p[k]$ and the known pilot signals $S_p[k]$.

Assuming the perfect synchronization, the received pilot signals $Y_p[k]$ can be written as $Y_p[k] = H[k] \cdot S_p[k] + W[k]$ where $H[k]$ denotes the channel transfer function, and $W[k]$ is the AWGN for the k -th pilot subcarrier after FFT. Also, we assume that the transmitted pilot signal $S_p[k]$ is known to the receiver. Therefore, the frequency response of the channel on the k -th pilot subcarrier, $\hat{H}[k]$, can be simply estimated as $\hat{H}[k] = Y_p[k]/S_p[k] = H[k] + \tilde{W}[k]$ where $\tilde{W}[k] = W[k]/S_p[k]$. The channel estimation scheme is based on the least square method [7].

Since the tap coefficients of the equalizer are obtained by calculating the inverse of the estimated channel coefficients, the severe noise enhancement is induced under deep fading channels. In order to reduce the noise enhancement, we adopt novel channel estimation schemes based on the M -point weighted moving average [3].

Let $\hat{H}_w[k]$ be the estimated channel coefficients after the M -point weighted moving average at the k -th pilot subcarrier of the OFDM symbol. Then, $\hat{H}_w[k]$ can be obtained as

$$\begin{aligned} \hat{H}_w(k) &= \frac{1}{M} \sum_{j=-(M-1)/2}^{(M-1)/2} w_j \hat{H}[k+j] \\ &= \frac{1}{M} \sum_{j=-(M-1)/2}^{(M-1)/2} w_j \{H(k+j) + \tilde{W}(k+j)\} \end{aligned} \quad (1)$$

where M is odd number and w_j is the j -th weight for moving average [3].

In this paper, six weighted moving average schemes of [3] are considered. The considered weighting values of the respective schemes are given in Table I. The quadratic, cubic, and Lagrange weighting values are derived from the quadratic, 4-point cubic, 4-point Lagrange interpolation filters, respectively [8]. Note that the variance of the equivalent noise term in (1) can be computed as

$$\frac{\sigma^2}{M} \left(\sum_{j=-(M-1)/2}^{(M-1)/2} w_j^2 \right). \quad (2)$$

Hence, the variance of the noise decreases with the increasing value of M . However, channel coefficients $\hat{H}[k]$ used in the weighted moving average can be severely varied due to the channel variation. This results in the distortion of the estimated channel response and thus, the performance degradation. Therefore, it is required to determine the average number of points by considering the trade-off between the accuracy of estimated channel coefficients and the noise reduction. Note that the considered moving average schemes are performed after the channel estimation.

TABLE I. WEIGHTS FOR MOVING AVERAGE [3]

Schemes	Weight values
Moving average (MA)	$w_j = 1, j \leq (M-1)/2$
Selective MA	$w_j = \begin{cases} 0, & j = \arg \min_j \hat{H}(k+j) \\ \text{or } j = \arg \max_j \hat{H}(k+j) , & j \leq (M-1)/2 \\ 1, & \text{otherwise} \end{cases}$
Linear weighted MA	$w_j = 1 - \left \frac{2j}{M+1} \right , j \leq (M-1)/2$
Quadratic weighted MA	$w_j = \begin{cases} -18 \left \frac{j}{M+1} \right ^2 + 1, & j < \frac{M+1}{6} \\ 9 \left \frac{j}{M+1} \right ^2 - \frac{15}{2} \left \frac{j}{M+1} \right + \frac{3}{2}, & \frac{M+1}{6} \leq j < \frac{M+1}{2} \end{cases}$
Cubic weighted MA	$w_j = \begin{cases} 96 \left \frac{j}{M+1} \right ^3 - 40 \left \frac{j}{M+1} \right ^2 + 1, & j < (M+1)/4 \\ -32 \left \frac{j}{M+1} \right ^3 + 40 \left \frac{j}{M+1} \right ^2 - 16 \left \frac{j}{M+1} \right + 2, & (M+1)/4 \leq j < (M+1)/2 \end{cases}$
Lagrange weighted MA	$w_j = \begin{cases} 32 \left \frac{j}{M+1} \right ^3 - 16 \left \frac{j}{M+1} \right ^2 - 2 \left \frac{j}{M+1} \right + 1, & j < (M+1)/4 \\ -\frac{32}{3} \left \frac{j}{M+1} \right ^3 + 16 \left \frac{j}{M+1} \right ^2 - \frac{22}{3} \left \frac{j}{M+1} \right + 1, & (M+1)/4 \leq j < (M+1)/2 \end{cases}$

IV. NUMERICAL RESULTS

In this paper, we present further results on the performance of the novel channel estimation schemes for T-DMB systems under fading channels via computer simulation. The channel model used for the simulations is Brazil Channel A [9]. Note that the Brazil channel A founded by the laboratory test in Brazil, represents the outdoor channel and is widely used in simulations for terrestrial digital television systems. We assume the perfect synchronization and known pilot signals at the receiver. We compare performances of the weighted moving average and conventional schemes. The T-DMB system is modeled based on transmission mode I of the Eureka-147 DAB system without any error correction coding scheme. Since the

elementary symbol period T is $0.48828\mu\text{s}$ ($=1/2048$ ms) for the sampling frequency of 2.048 MHz [1], the maximum delay of Brazil Channel A corresponds to about 12 elementary symbols. Note that the number of subcarriers, subcarrier spacing, OFDM symbol duration, and CP duration are 1536, 1kHz, $2551 \cdot T$ (≈ 1.246 ms), and $504 \cdot T$ (≈ 246 μs), respectively.

The performance measure considered in this paper is the symbol error rate (SER) after one-tap equalization with the inverse of estimated channel coefficients. The transmitted signals are mapped with $\pi/4$ differential quadrature phase shift keying ($\pi/4$ -DQPSK) constellations. Fig. 1 shows the SER performance of the weighted moving average and conventional channel estimation schemes under Brazil Channel A with 10 Hz Doppler. As shown in Fig. 1, the weighted moving average schemes outperform the conventional scheme with approximately 2~3 dB under the fading channel. Note that the performances of MA, selective MA, linear weighted MA, and quadratic weighted MA schemes are similar. Compared to MA, selective MA, linear weighted MA, and quadratic weighted MA schemes, the performances of the cubic and Lagrange weight MA schemes are degraded due to the estimation error.

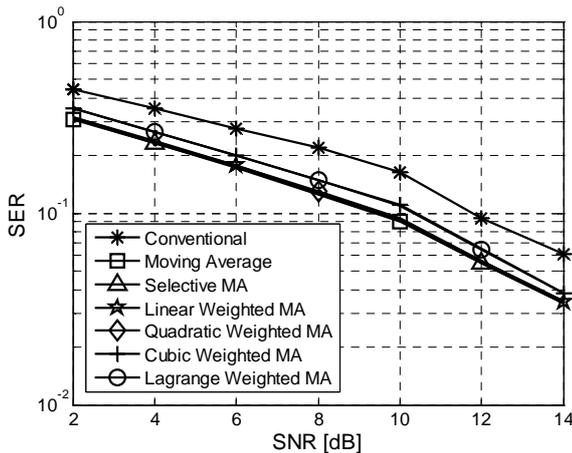


Figure 1. SER performance after equalization under Brazil Channel A with 10Hz Doppler

According to (1), the weighted moving average scheme requires M multiplications and $(M - 1)$ additions. As shown in Table I, since the weight values of five schemes except for the selective MA scheme can be precalculated and stored for a given M , the complexity for weight values can be negligible. Also, for MA and selective MA schemes, since the weight values are one and zero, the multiplication is not required. Therefore, MA and selective MA schemes outperform the cubic and Lagrange weight MA schemes with the reduced complexity. Note that the comparison test for the estimated channel coefficients is additionally required for the selective MA scheme.

Note that the noise enhancement can be decreased with increasing number of points used for averaging. However, the accuracy of the channel estimation can be degraded with

the large number of points for averaging due to the channel variation.

V. CONCLUSIONS

In this paper, we described novel channel estimation schemes based on the weighted moving average in order to improve the channel estimation performance in T-DMB receivers. The channel estimation schemes based on the weighted moving average outperform the conventional scheme with approximately 2-3 dB in SER performance after equalization under fading channels. Also, the channel estimation schemes with weighted moving average are expected to contribute to the performance improvement of the channel estimation in any OFDM systems.

Also, we are studying to further improve the channel estimation performance and reduce the complexity of the channels estimation by developing new interpolation schemes and optimizing the size of the moving window. Unfortunately, the Brazil A channel is the multipath profile for the static environment. Hence, we will use the well-known TU6 multipath channel [10] in order to consider a mobile environment in the urban area. Also, in this paper, the size of the moving window was not optimized for a mobile channel but was set to be values for static channels. Under a high speed mobile channel, since the channel is rapidly varied, the size of the moving window should be optimized.

ACKNOWLEDGMENT

This research was supported by the Ministry of Science, ICT and Future Planning, Rep. of Korea, under the support program supervised by the Korea Communications Agency (KCA) (Development of Terrestrial Broadcasting Technology for Next Generation DTV).

REFERENCES

- [1] ETSI, "Radio broadcasting systems; Digital Audio Broadcasting (DAB) to mobile, portable and fixed receivers," ETSI EN 300 401 v1.3.3, Sept. 2001.
- [2] G. Lee, S. Cho, K.-T. Yang, Y. K. Hahm, and S. I. Lee, "Development of terrestrial DMB transmission system based on Eureka-147 DAB systems," *IEEE Trans. Consumer Electron.*, vol. 51, no. 1, Feb. 2005, pp. 63-68.
- [3] S. I. Park, J. Y. Lee, H. Lim, H. M. Kim, H. Kim, and J. S. Lim, "A novel channel estimation in T-DMB receivers," in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, 2007, pp. 237-241.
- [4] L. Thibault, D. Taylor, L. Zhang, J. Y. Chouinard, and R. Boudreau, "Advanced demodulation technique for COFDM in fast fading channels," *International Broadcasting Convention*, 2003, pp. 416-422.
- [5] A. A. Hutter, R. Hasholzner, and J.S. Hammerschmidt, "Channel estimation for mobile OFDM systems," in *Proc. of IEEE Vehicular Technology Conference*, Amsterdam, Netherlands, Sept. 1999, pp. 305-309.
- [6] Y. Zhao and A. Huang, "A novel channel estimation method for OFDM mobile communication systems based on pilot signals and transform-domain processing," in *Proc. of IEEE 47th Vehicular Technology Conference*, vol. 3, May. 1997, pp. 2089-2093.
- [7] X. Wang, Y. Wu, J.-Y. Chouinard, S. Lu, and B. Caron, "A channel characterization technique using frequency domain

- pilot time domain correlation method for DVB-T systems,” IEEE Trans. Consumer Electron., vol. 49, Nov. 2003, pp. 949-957.
- [8] T. M. Lehmann, C. Conner, and K. Spitzer, “Survey: Interpolation methods in medical image processing,” IEEE Trans. Medical Imaging, vol. 18, no. 11, Nov. 1999, pp. 1049-1075.
- [9] Mackenzie, ABERT/SET Group, “General description of laboratory tests,” DTV Field Test Report in Brazil, July 2000.
- [10] COST207, “Digital land mobile radio communications (final report),” Commission of the European Communities, Directorate General Telecommunications, Information Industries and Innovation, 1989, pp. 135–147.

Adaptive Reliable Routing Protocol for Wireless Sensor Networks

Nourhene Maalel, Pierre Roux, Mounir Kellil
CEA, LIST, Communicating Systems Laboratory
Gif-sur-Yvette, France
{nourhene.maalel,pierre.roux,mounir.kellil}@cea.fr

Abdelmadjid Bouabdallah
UTC, Heudiasyc Laboratory, UMR CNRS 6599,
Compiègne, France
madjid.bouabdallah@hds.utc.fr

Abstract—Many Wireless Sensor Networks (WSN) applications success is contingent upon the reliable delivery of high-priority events from many scattered sensors to one or more sink nodes. In particular, WSN has to be self-adaptive and resilient to errors by providing efficient mechanisms for information distribution especially in the multi-hop scenario. To meet the stringent requirement of reliably transmitting data, we propose a lightweight and energy-efficient joint mechanism for packet loss recovery and route quality awareness in WSNs. In this protocol, we use the overhearing feature characterizing the wireless channels as an implicit acknowledgment (ACK) mechanism. In addition, the protocol allows for an adaptive selection of the routing path, based on a collective cooperation within neighborhood.

Keywords- WSNs; Reliable transport; Routing

I. INTRODUCTION

All WSN (Wireless Sensor Network) technologies have experienced an exponential increase in popularity mostly due to their potentially low cost of maintenance and deployment.

However, wireless sensor networks may face a number of challenges that can hamper their widespread exploitation [1]. A WSN has to be self-adaptive and resilient to errors by providing efficient mechanisms for information distribution especially in the multi-hop scenario. These requirements have to be achieved in a networking environment that is constrained by limited processing capability, scarce energy resources and unreliable communication channels [1]. In particular, in a typical harsh environment, the radio signal is often affected by interference: medium access conflicts, multipath fading, shadowing, etc. These problems may result in significant packet losses in WSNs. Moreover, the success of many applications (particularly mission-critical ones like life-care data and alarms) requires the delivery of high-priority events to sinks without any loss from the original sources to the final destination [2]. These constraints emphasize the need for an energy-efficient, scalable and reliable data transport system.

Data retransmission has been considered as one of the most common schemes [3] for improving transmission reliability in WSN. ACKnowledgment/ Negative ACKnowledgment (ACK/NACK) messages are the basic method used to assess the necessity of retransmission. Nevertheless, such a method generates an extra traffic

causing an additional overhead, which is not suitable in a highly constrained and error prone environments, like WSNs. Accordingly, an alternative solution should be found to deal with retransmissions without wasting bandwidth.

In this paper, we define a reliable and energy-efficient joint mechanism, for packet loss recovery and route quality evaluation in WSNs. In this protocol, we use the overhearing feature, characterizing the wireless channels [3], as an implicit ACK mechanism. In addition, the protocol allows for an adaptive selection of the routing path based on a link state metric.

The remainder of this paper is organized as follows: the next section highlights the need for reliable data delivery in WSNs, and reviews solutions aiming at providing it. Protocol description and analysis are given in Section 3, and finally, Section 4 concludes this paper.

II. BACKGROUND

The error control can be implemented as multipath routing by forwarding packets along several paths in order to improve the overall reliability [4]. Copies of the same packets can be forwarded randomly over multiple routes [4]. Another solution is to identify many paths and select one as the primary route while the other alternatives are used in case of problems in the primary path [5]. Maintaining multiple paths is usually costly in large scale WSNs.

Another traditional way to achieve reliable transmission is the Automatic Repeat reQuest (ARQ) mechanism based on the ACK/NACK messages [6]. However, this mechanism should be minimized because sensor nodes are severely resource constrained and data transmission is one of the most costly operations performed by sensors [6]. Moreover, the unreliable radio channel affects the acknowledgment delivery as well. If the sender does not receive any acknowledgment in the specified time interval, it retransmits the message even if the packet was properly delivered. In practice, the sender node makes a delimited number of trials to successfully deliver a message. Therefore, relying on explicit acknowledgement is not appropriate with regard to the constrained nature of WSNs.

More recently, a Multicast Protocol for Low power and Lossy Networks (MPL) called Trickle Multicast [7], was designed. Trickle multicast utilizes a sequence number in the data packet to cope with packet losses. Packets along with their respective sequence numbers are temporarily stored by

the nodes so as retransmission can be triggered when necessary. Trickle multicast, though is based on network flooding for data dissemination and storage. Given the resource constraint nature of WSNs, this flooding mechanism is not suitable to sensor networks.

We could identify two categories of transmissions Hop By Hop (HBH) and the End to End approach (ETE). According to She and al. [9], HBH is more energy efficient at the cost of large transmission delay compared to ETE. Nevertheless, HBH outperforms ETE on the delay metric for high bit error rate cases. Given that Zhao et al. [8] show that error rates of 10% or above in dense WSN may be experienced, HBH is the most suitable candidate for WSNs. Let's notice that the problem with ETE recovery is highly related to the harsh radio environments of deployment and to the multi-hop forwarding techniques, which favor exponential error accumulation [8].

Some researches proposed solutions to alleviate the retransmissions cost like PSFQ [10], which distribute data from a source node by sending data at a relatively slow speed but allowing nodes that come across data loss to recover any missing segments from their local immediate neighbors. This protocol is efficient for fast recovery but if packet loss occurs in an intermediate node towards the sink, buffer must standby until packet re-transmission is done. This causes buffer overflow and increases data transmission delays. Blagojevi and al. [11] presented a probabilistic acknowledgement mechanism switching between explicit and implicit acknowledgement depending on the current path reliability. For this solution, path reliability is determined by measuring the Received Signal Strength Indication (RSSI) which is proved to not always be a good indicator to estimate the link state [12]. Messina and al. [13] proposed a solution where the protocol achieves reliability through caching and retransmission. As mentioned, this solution requires each one hop neighbor to cache the data until the success of its transmission. Once a packet loss is detected, all the one hop neighbors will act on the behalf of the node which experiences the loss by retransmitting the packet and performing its routing task. Such a practice leads to extra energy consumption and may fasten nodes "battery depletion".

III. PROPOSED PROTOCOL

A. Overview of the mechanism

Our solution seeks into elaborating an efficient error control mechanism with implicit acknowledgments to face the link failure and packet loss problem in WSN. When a sensor node transmits a packet, nodes of its neighborhood overhear its packet transmission even if those are not the intended recipients [3]. This arises from the broadcast nature of the wireless channel.

Our solution uses this overhearing characteristic instead of the acknowledgment messages to guarantee reliability on networks. Moreover, when a packet loss is detected, retransmission is carried out by the most reliable link between the node which sent the (lost) packet and its one-

hop neighbors. The reliability of links is defined according to a metric which will be detailed in the next section. Our algorithm relies on a spanning tree for ordinary routing operations, and resorts to exploiting alternative paths only when a malfunctioning is detected.

B. Protocol operation

1) Considered architecture

We consider a dense and randomly distributed WSN. Before discussing the details of the protocol, we need to clarify our assumptions:

- All nodes have sufficient resources to carry their sensing, computing, and transmission/reception operations.
- Data packet is generated by sensors and transmitted to the sink node.
- Each sensor node is stationary for its lifetime and is able to record the link performance between itself and its neighboring node in terms of number of lost packet / number of sent packets.
- We adopt a routing scheme in which the routing decision takes the shortest path towards the sink. Each node is assigned a rank corresponding to the hop-distance to the sink and data is carried rank by rank towards the sink. In this sense, the node B in Figure 1 has rank N and its neighbors have rank N-1, N, or N+1. We assume that each node is aware of its own rank (in respect to its neighbors) as well as the ranks of its neighbors.
- Each node of rank N (Figure 1) classifies its k neighbors of rank N-1 from index 0 to k with 0 corresponding to the most reliable node according to our metric defined in the next section. The node with index 0 is the elected one to carry out the retransmission task when packet loss is detected.

2) Index assignment

Our protocol relies on its routing metric to assign indexes to nodes. As mentioned above, index 0 corresponds to the most reliable link (the higher metric). The index assignment is used to choose the best next hop for the packet retransmission hence its importance. The metric component of our protocol evaluates links according to the Link Quality Indicator (LQI) and the probabilistic history model. The LQI is a metric of the current received signal quality. This measurement is reported with each received packet in the MAC header by the used 802.15.4 standard [14]. The use of LQI ensures adaptability to the environmental conditions by expressing the real quality of the link. Besides, LQI experiences frequent fluctuations in highly interfered environment. Hence, we consider statistics (average number of lost packet per link) as a basis to assess the reliability of links.

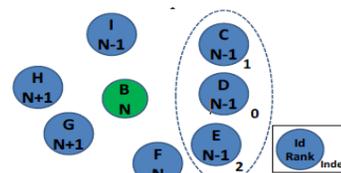


Figure 1. Rank assignment among the neighborhood

For this reason, we have decided to weight the metric by the link failure probability given by our probabilistic history model, $P_{hist BC}$. Since the channel state is binary (packet received: Up, packet lost: Down), a simple count of the number of state is sufficient to fully describe the history:

$$P_{hist BC} = \frac{nUp}{nUp + nDown} \quad (1)$$

Therefore, even if the last LQI value recorded does not match the real state of the link we can correct it. Let's precise that $P_{hist BC}$ is set to 1 in the beginning (during a fixed time T_{init}) before getting effective history. In fact, in the establishment of the process, we do not have sufficient feedback to assess the reliability of a link. To do so, our protocol assigns each link a cost given by the following expression:

$$Metric(B, C) = K * e^{-age} * LQI_{BC} + \frac{1}{P_{hist BC}} \quad (2)$$

where LQI_{BC} denotes the link state indicator between nodes B and C and age corresponds to the delay since the LQI value has been recorded. The exponential function provides a decreasing function according to the age, which means that more recent values of LQI are considered as more significant. The $P_{hist BC}$ represents the probability of link success between nodes B and C. K is a constant used to weight the equation. These metrics are calculated periodically in the network to update the index assignments and make the protocol more robust to the environmental change. This update period depends on the packet error rate of the network: the faultier the network is, the more frequent the update occurs.

3) Algorithm description

Figure 1 represents a node B of rank N and its neighborhood. More particularly, it shows its N-1 neighbors C, D and E of index 1, 0, and 2, respectively. Let's notice that our protocol provides uniqueness of index to avoid collision problem: When different nodes have the same index, a random back off is added to the metric in order to have distinguish index. Once a packet with a Packet Identifier (PID) is received for the first time by a node of rank N from a node N+1, a transient context is created in its memory to manage this packet PID.

This context includes packet content and PID in order to allow possible retransmissions. If the node has index 0 relatively to the sender node, the context is considered as a 'Primary' one (P Ctxt) and the packet is immediately forwarded. Otherwise, the context is considered as 'Secondary' (S Ctxt) and the packet is cached waiting for a possible retransmission request.

To make it clearer, we will consider 3 scenarios shown in Figure 2. In the loss-free case (Figure 2/a), all the nodes C, D and E receive the packet. Node D (which has index 0 for B), creates a primary context for PID, while node C and E (which have an index greater than 0 for B) create a secondary context for packet PID. Because node D has

created a primary context for PID, it immediately forwards the packets towards its own neighbors. At this time, the node B also receives the packet forwarded by node D. There is an implicit acknowledgement for packet PID so the node B can release its primary context for PID. After a while, nodes C and E realize that node B didn't send any Explicit Retransmission Request (ERR) message. They can safely get rid of their secondary context for packet PID. In the loss-free case, this process goes on until the packet PID reaches the sink, without involving any waiting period in any of the forwarding nodes on the path to the sink.

Now, if we consider a case including packet loss (Figure 2/b), we can come back to the situation where node B has just received packet PID and has just created a primary context for this packet. Again, it forwards the packet to its neighbors, but we now assume that the node D doesn't receive the packet, while nodes C and E receive it. The node C has index 1 with respect to node B so it creates a secondary context for packet PID. Then the node C waits for a possible Explicit Retransmission Request (ERR) from node B with respect to packet PID. Note that this message is short, as it does not contain the data payload of packet PID.

When the node C receives the ERR message for packet PID, it immediately forwards this packet toward its neighbors, because its index for node B is 1. Once node B receives the implicit acknowledge from node C, it broadcasts an Explicit Retransmission Cancel message (ERC) with respect to packet PID. ERC is a short message similar to ERR. Once this message is received, the node E deletes this message and releases its secondary context for PID. We may now consider another case including packet loss (Figure 2/c). We come back to the same situation as before, but we now assume that among neighbors of rank N-1, only node E has received packet PID from node B. Once Node B detects the packet PID was not forwarded by D, it sends an ERR for PID as before to C, and E. However, node E does not immediately forwards packet PID (even if it is the only node which is able to retransmit the packet) because it is aware that its index for B is 2. That is why it waits for a delay T_{Delay} . Then, if no ERC message with respect to PID has been received from node B, it turns its secondary context for PID into a primary context, and it forwards packet PID to its own neighbors. The rule is that once a node having index n with respect to another node receives an ERR from this other node, it waits for a delay equal to $(n-1)$ times T_{Delay} for a possible ERC. If no ERC is received during this time, then retransmission occurs.

This process aims to avoid sending duplicate packets and consequently to reduce bandwidth consumption.

C. Protocol Analysis

It should be noted that in case of failure, this algorithm does not solicit the neighbor for which a failure was observed. Our protocol is particularly adapted to packet losses caused by a change in the channel state such as slow

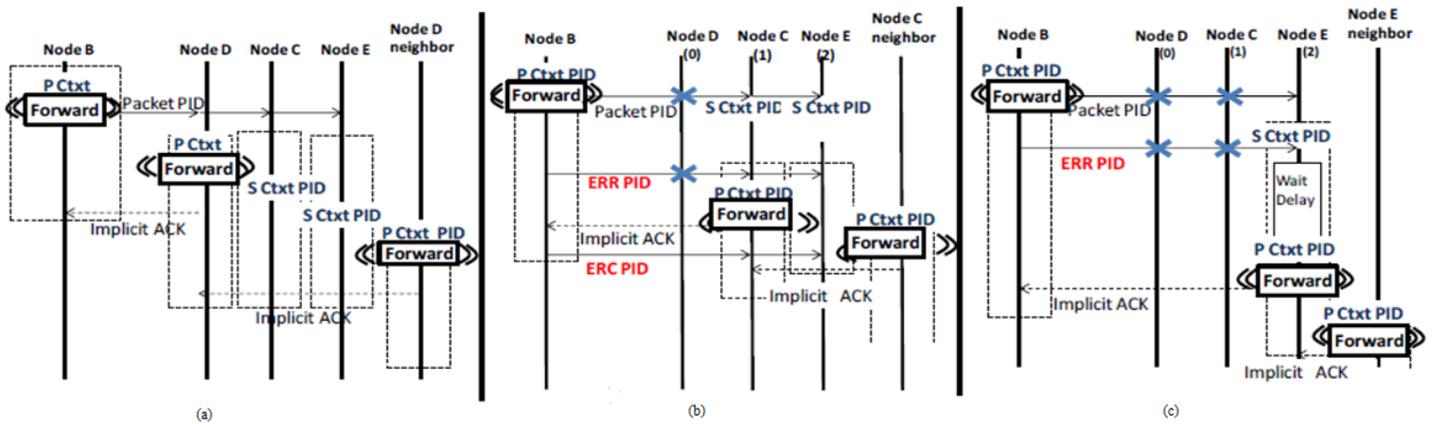


Figure 2. Protocol steps for data forwarding from Node B: (a)Loss free case, (b)Packet loss in Node D case, (c)Packet loss in Node D and Node C case

shadowing that affects the radio link. A retransmission occurring on the same radio link just after a failure is likely to bring a failure again while with the current proposed algorithm, another radio link is exploited. In the case of the set of neighbors of rank N-1 is limited to only one node (sparse network case), this principle doesn't apply, and repetitions should be carried out by the same neighbor.

Besides, packet losses on the implicit acknowledgement messages have not been considered in this paper. With the proposed algorithm, such losses may result in duplicate instances of a packet PID being forwarded up to the sink according to different paths. Given that paths are not disjoint, a node may receive the same packet twice. To alleviate this issue, it is recommended that nodes maintain a list of recently received packet PIDs and to drop packets once a duplicate is detected.

Moreover, the transmission of data from a node to its neighbor must be completed within a specified time. If the packet does not reach the next hop within this time limit, it is dropped and considered as it has been lost.

IV. CONCLUSION AND FUTURE WORK

In this paper, we proposed a lightweight protocol to tackle packet losses in WSNs. We provided a solution based on the implicit ACK mechanism and on an adaptive selection of the routing path based on the link quality evaluation. For future work, we intend to examine additional parameters that influence on the time-varying link reliability, and also plan to evaluate our protocol in comparison to other solutions of the state of the art using simulations.

REFERENCES

[1] Y. Xiao, X. Li, Y. Li, and S. Chen, "Evaluate reliability of wireless sensor networks with OBDD," In Proceeding of IEEE International Conference on Communications (ICC), Dresden, June 2009, pp. 1-5, doi:10.1109/ICC.2009.5199006.

[2] S. Qaisar and H. Radha, "Multipath distributed data reliability for wireless sensor networks," In Proceeding of IEEE International Conference on Communications (ICC), Dresden, June 2009, pp. 330-334.

[3] G. N. Lee and E. N. Huh, "Reliable data transfer using overhearing for implicit ack," ICROS-SICE, Fukuoka 2009, pp. 1976-1979

[4] B. Deb, S. Bhatnagar, and B. Nath, "ReInForM: Reliable information forwarding using multiple paths in sensor networks," Annual IEEE International Conference on Local Computer Networks. (LCN), Bonn, October 2003, vol. 0, pp. 406-415.

[5] D. Ganesan, R. Govindan, and S. Shenker, "Highly resilient, energy-efficient multipath routing in wireless sensor networks," ACM SIGMOBILE, October 2001, vol. 5, no. 4, pp. 11-25, doi:10.1145/509506.509514.

[6] I. Stojmenovic and A. Nayak, "Design guidelines for routing protocols in ad hoc and sensor networks with a realistic physical layer," IEEE Communications Magazine, 2005, pp. 101-106.

[7] <http://tools.ietf.org/pdf/draft-ietf-roll-trickle-mcast-04.pdf>, retrieved July 2013.

[8] J. Zhao, R. Govindan, and D. Estrin, "Computing Aggregates for Monitoring Wireless Sensor Networks," Proceedings of the First IEEE international Workshop on Sensor Network Protocols and Applications, Bonn 2003, pp. 139 - 148.

[9] H. She, Z. Lu, and A. Jantch, "Analytical evaluation of retransmission schemes in wireless sensor network," In Proceeding of IEEE Vehicular Technology Conference (VTC) Barcelona, April 2009, pp. 1-5.

[10] Campbell, Andrew T.; Krishnamurthy, and Lakshman, "Pump-slowly, fetch-quickly (PSFQ): a reliable transport protocol for sensor networks," IEEE Journal on selected area in communications, 2005, pp. 862 - 872.

[11] Miloš Blagojevi, Majid Nabi1, and Marc Geilen1, "A Probabilistic Acknowledgment Mechanism for Wireless Sensor Networks," IEEE international conference on networking, architecture and storage (NAS), Dalian, July 2011, pp. 63-72.

[12] Shinuk Woo and Hwangnam Kim, "Estimating Link Reliability in Wireless Networks," INFOCOM, San Diego, March 2010, pp. 1-5.

[13] Messina and Daniele, "Achieving Robustness through Caching and Retransmissions in IEEE 802.15.4-based WSNs," Proceedings of IEEE International Conference on Computer Communications and Networks (ICCCN), 2007 Honolulu, pp. 1117 - 1122.

[14] ZigBee Alliance, <http://www.zigbee.org>, retrieved July 2013.

Design for a UWB Down-Conversion Mixer for Multi-Band Applications

Wu-Shiung Feng, Chin-I Yeh, Prasenjit Chatterjee and Chen-Yu Hsu

Department of Electronic Engineering, Chang Gung University

Kwei-Shan, Tao-Yuan, Taiwan, R.O.C

e-mail: fengws@mail.cgu.edu.tw, yehci.aa@msa.hinet.net, prasenjit.80@gmail.com, real_killua77@hotmail.com

Abstract—In this paper, a CMOS down-conversion mixer for UWB applications is presented. The mixer circuit is designed using a VIS 0.25 μ m RF CMOS technology, working at the 0.9GHz-10.6GHz frequency range; it will be used in applications such as IEEE WiFi, GSM and WiMAX. The core of the mixer has been designed based on double-balanced cell architecture, and uses the current bleeding method to increase the linearity and improve the conversion gain. We put a resistor on the drain of the MOS between two RF inputs, which will improve the flatness of conversion gain. Regarding the arrangement of the mixer, the RF frequency is set at 0.9GHz-10.6GHz, the LO frequency 0.8GHz-10.5GHz and the IF of 100MHz. The simulated conversion gain of the mixer is 7 ± 1 dB. The 1dB compression point is higher than -11dBm at high frequency and -9dBm at low frequency. The RF input return loss is well below -11dB, and the LO input return loss is below -10dB. The noise figure is 12.93dB while IF is 100MHz, and the mixer core dissipates 9.8mW under a 1.8 V supply.

Keywords-UWB; Down Conversion; CMOS

I. INTRODUCTION

The requirements of wireless communication systems have been increasing rapidly over the last few decades. Higher speed transportation, lower power consumption, low cost and operating on multi-band have been very important considerations in wireless development [1]. The new design technique for ultra wide-band wireless communications must be carried out with future high-precision applications in mind. Wireless Fidelity (Wi-Fi), worldwide Interoperability for Microwave Access (WiMAX) [2-5] and Global Systems for Mobile Communications (GSM) are very common in modern life. A lot of mobile devices, like personal computers, mobile phones, and digital music players can be connected to Wi-Fi networks, but only within a limited area.

WiMAX is a form of wireless communication; its standard is known as IEEE 802.16x. The area of WiMAX is very large and can be used within entire residential districts, remote areas, or in SOHO internets. Its application will benefit the mobility of individuals, families and companies.

GSM is the most popular standard of mobile phones. By Q4 2012, over 3.2 billion people were using GSM technology [6] and the world market share of GSM is over 80% [7].

In this work, a 0.9GHz~10.6GHz ultra wide band (UWB) down conversion mixer was designed for use in a variety of applications, such as Wi-Fi, WiMAX and GSM. However, the linearity and noise figures of a mixer should be continuously improved for future high-sensitivity applications. In Section II, the circuit design aspect is

described. Section III includes post-simulation results, and conclusions are presented in Section IV.

II. CIRCUIT DESIGN

This circuit uses the process of standard complementary metal oxide semiconductor (CMOS) 0.25 μ m 1P5M technology, and we used Advanced Design System (ADS) requirements to simulate the circuit. The Radio Frequency (RF) input frequency was between 0.9GHz and 10.6GHz, and the Local Oscillator (LO) input frequency was set between 0.8GHz and 10.5GHz. The Intermediate Frequency (IF) output frequency was 100MHz. The Gilbert structure [8] was modified to complete this circuit, and improve it over a double-balanced mixer. Figure 1 shows the whole circuit diagram.

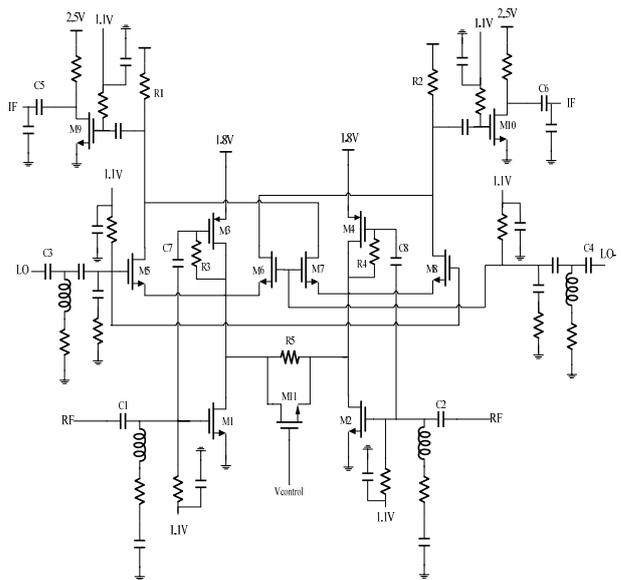


Figure 1. Down-Conversion Mixer

As shown in Figure 1, LO stages of transistors from M5 to M8 are designed using MOS switches. M1 and M2 are operated on the saturation region to act as conductors and to permit RF signals to go through the gate and become current signals. M3 and M4 act as the paths of current bleedings to make more current flow into M1 and M2, thereby increasing the conversion gain and reducing noise. R3 and R4 block the signals into M3 and M4, and can also act as the self-bias voltages of M3 and M4. C7 and C8 prevent the bias voltages and allow the signals to go through them to M3 and M4. If this is the case, M3 and M4 can not only provide the paths

for current bleeding but can also amplify the signals. Capacitors from C1 to C6 act as DC blocks, and a common source amplifier is implemented to be the IF output buffer.

The flatness of conversion gain is also very important to the UWB mixer, so a resistor is embedded between the drains of M1 and M2. Figure 2 shows that C7, R3 and R5 combine to form a high pass filter, as shown in Figure 3. This decreases the gain of low frequency to improve the flatness of conversion gain, as shown in Figure 4. The conversion gain can be sacrificed to improve flatness.

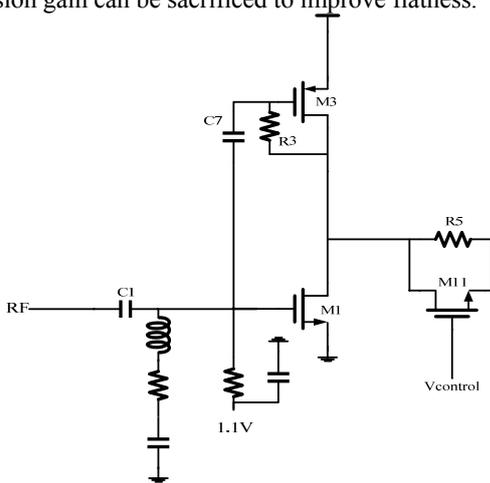


Figure 2. The Relationship among C7, R3, and R5

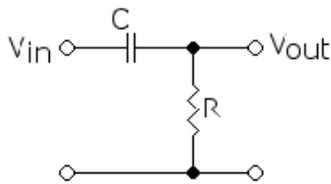


Figure 3: High-Pass Filter

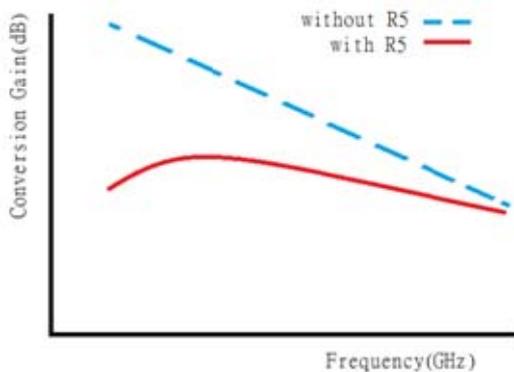


Figure 4. The Change in Flatness of Conversion Gain

III. SIMULATIONS AND RESULTS

Figures 5 to 11 depict the post-simulation result using ADS software. The supply voltage and power consumption

are 1.8V and 18.3mW, respectively. The maximum conversion gain of the mixer was simulated at 8.8 dB at a frequency of 2.9 GHz, while the minimum was 5.7 dB at 10.6 GHz, and an average of 7.25 dB over the band. A plot of the conversion gain versus frequency is shown in Figure 5. Figures 6 and 7 show RF and LO input matching, since it is a UWB mixer and the return loss is less than 8dB. Figure 8 represents P-1dB, with three frequencies chosen for simulation. The results range from about -8.5 to -11dBm.

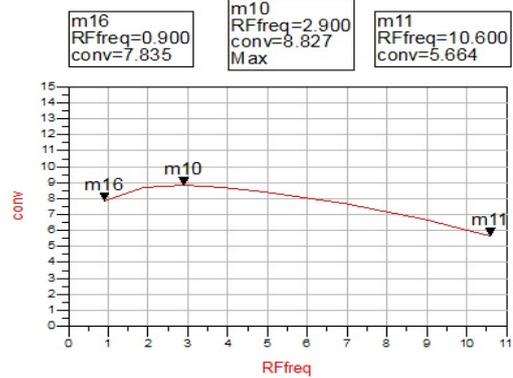


Figure 5. The Conversion Gain of 0.9~10.6GHz

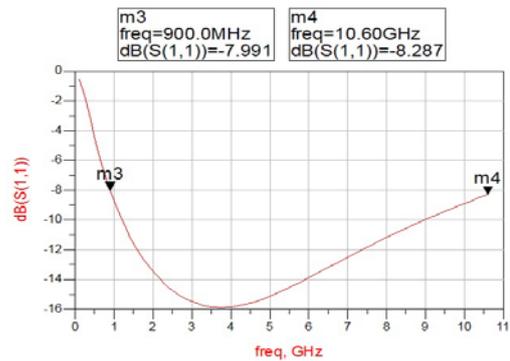


Figure 6. RF Input Return Loss

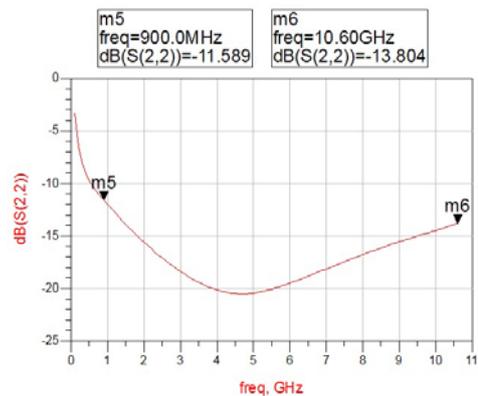
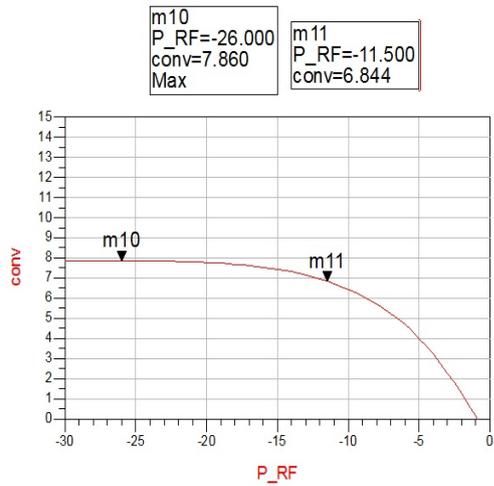
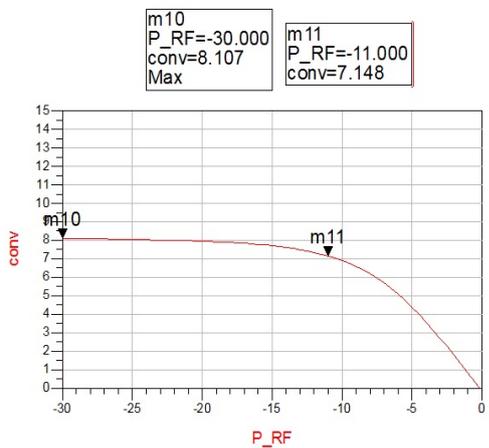


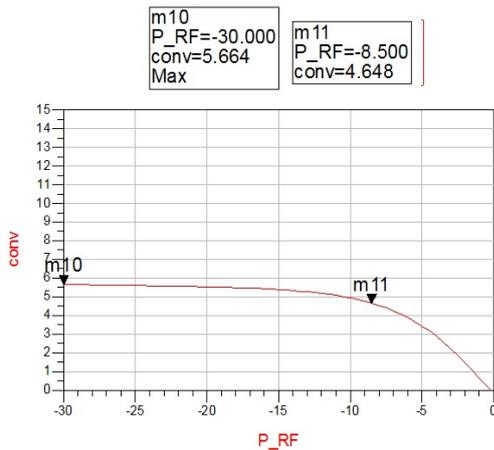
Figure 7. LO Input Return Loss



(a) 0.9GHz



(b) 5.8GHz



(c) 10.6GHz

Figure 8. 1dB Compression Point

Figure 9 shows the DSB noise figure of the mixer; the variation is very small, ranging between 0.9 and 10.6GHz.

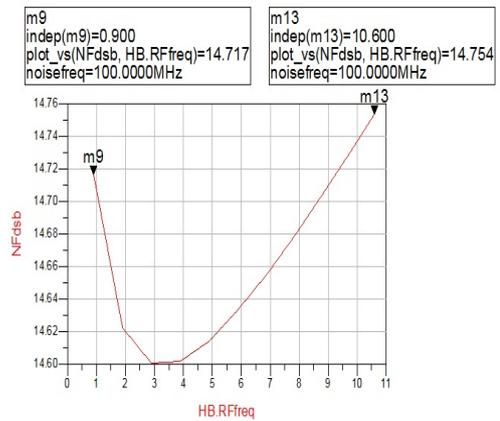
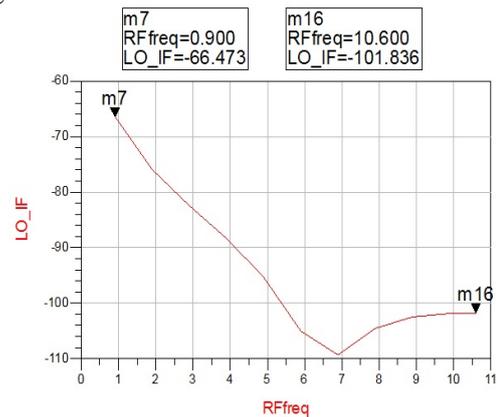
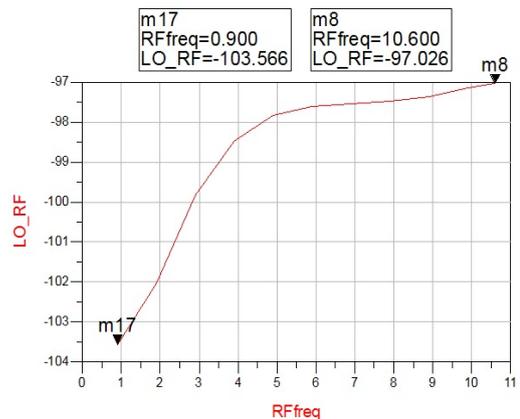


Figure 9. Noise Figure

Figure 10 shows the isolation of LO to IF and LO to RF. Since we used the double-balanced structure, isolation is very good.



(a)



(b)

Figure 10. Isolation (a) LO to IF and (b) LO to RF

Figure 11 shows the relationship of LO power and conversion gain. LO power was set at 0dBm in the simulation; we can see that the results are actually located at 0dBm and the support LO power range is down to -5dBm.

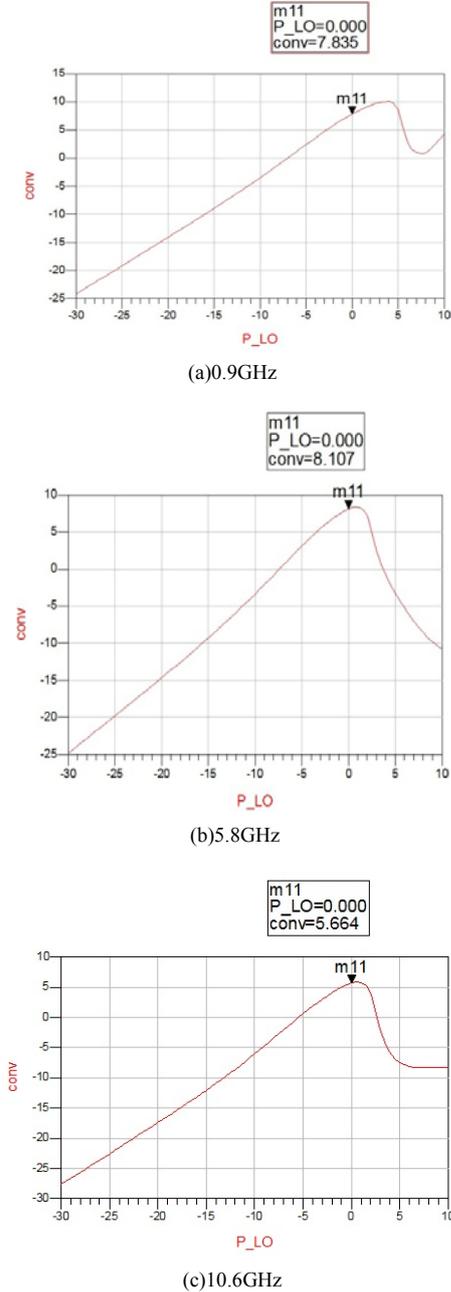


Figure 11. The Relationship of LO Power with Conversion Gain

Figure 12 shows the symmetrical waveforms from the mixer output at different operating frequencies. Using the simulated output signal of the proposed mixer, we are able to obtain the base band signal with a peak-to-peak value of approximately 50mV.

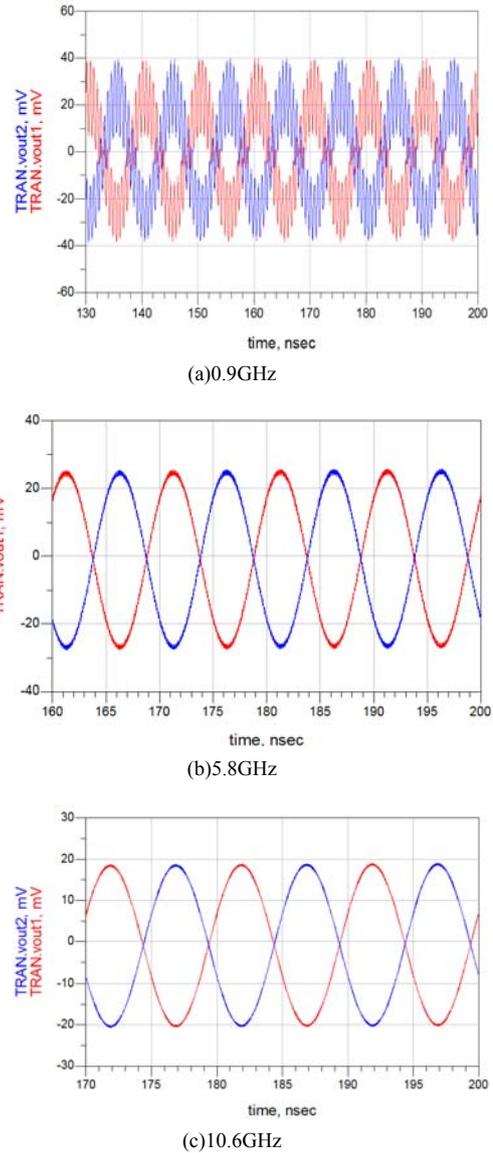


Figure 12. The Output Waveform Diagram

The layout of this circuit is shown in Figure 13, and will be fabricated in a VIS 0.25 μ m RF CMOS process. The total area of the chip is less than 1.33 mm². As observed, there are two pairs of on-chip spiral inductors in this design. Their values have been selected to be as small as possible, so that first-order parameters: inductance and resistance remain almost unchanged within the operating frequency.

The performance of the proposed wideband mixer and a comparison with other existing mixers [9-11] around the same frequency range are summarized in Table II. The advantage of the proposed mixer represents excellent properties of linearity and low noise figures.

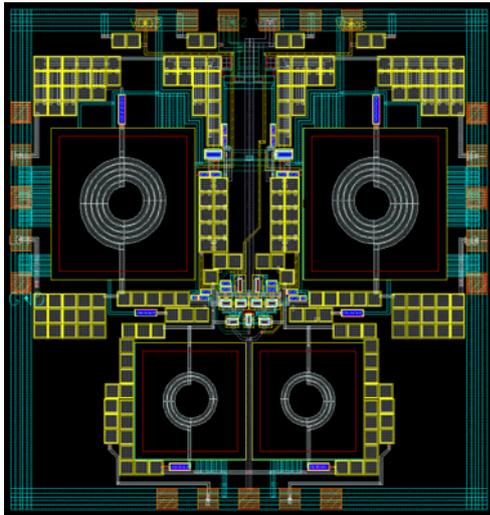


Figure 13. Mixer Layout

Table I lists the comparison of pre-simulation and post-simulation using 0.25 μ m CMOS technology..

TABLE I. COMPARISON OF PRE-SIM AND POST-SIM

Specifications	Pre-Sim.	Post-Sim.
Power Supply (V)	1/1.8/2.5	
Current (mA)	10.3	
Power Dissipation (mW)	18.54	
Frequency Range (GHz)	0.9~10.6	
Conversion Gain (dB)	6~8.7	5.7~8.8
DSB Noise Figure (dB)	14.57~14.75	14.6~14.76
RF Return Loss (dB)	<-7	<-7
LO Return Loss (dB)	<-11	<-11
LO-RF Isolation (dB)	<-60	<-60
LO-IF Isolation (dB)	<-97	<-97
P-1dB (dBm)	-10	-10
Chip Area (mm)	1.16*1.149	

TABLE II. COMPARISON WITH PREVIOUS WORK

	This work	[9]	[10]	[11]
Process (um)	0.25	0.18	0.13	0.13
Power Supply (V)	1.8	1.8	1.2	1.2
Frequency (GHz)	0.9~10.6 6	3.1~4. 8	3~7	0.7~6
Conversion Gain (dB)	6~8.7	9~11.5	5.3~8. 2	5~7

Noise Figure (dB)	14.57	10.8~1 3.2	9.6~13 .5	11~13. 2
Power Dissipation (mW)	18.54	11.3	5.8	5.8
Chip Area (mm)	1.16*1. 15	0.74*0 .4	0.36*0 .38	0.86*0. 75

IV. CONCLUSION AND FUTURE WORK

This UWB down-conversion mixer combines the advantages of the current-bleeding technique and a variable MOS-resistor with the VIS process of CMOS 0.25 μ m 1P5M. We used the technique of current bleeding and put a resistor between the drains of the RF input conductors to benefit the flatness of conversion gain. Since there are four different supply voltages, it was necessary to reduce the number of voltages to achieve lower cost. Combining these techniques with the use of a wideband low noise amplifier and voltage controlled oscillator to act as a SOC chip is our ultimate goal.

ACKNOWLEDGMENT

This work was supported by the National Science Council of Taiwan, ROC, under contracts of NSC101-2218-E-182-008, and NSC101-2221-E-182-078. The authors are grateful to Vanguard and CIC for technical support, and to NCHC for computer time under the assistance of the Green Technology Research Center, Chang Gung University.

REFERENCES

- [1] A. A. Abidi, "Direct-conversion Radio Transceivers for Digital Communications," IEEE Journal of Solid-State Circuits, vol. 30, no. 12, Dec. 1995, pp. 186-187.
- [2] D. M. Pozar, Microwave Engineering, Amherst, MA: Wiley, 1998.
- [3] C. S. Harsany, Principles of Microwave Technology, New Jersey: Prentice-Hall, 1997.
- [4] R. Mahmoud and K. Faitah, "A Low Power Consumption Gilbert-Cell Mixer in 65nm CMOS Technology," International Journal of Microwaves Applications, vol. 2, no. 2, pp. 93-96, Apr. 2013.
- [5] D. H. Na and T. W. Kim, "A 1.2V, 0.87-3.7GHz Wideband Low-Noise Mixer Using a Current Mirror for Multiband Application," IEEE Microwave and Wireless Components Letters, vol. 22, no. 2, pp. 91-93, Feb. 2012.
- [6] <http://www.gsma.com/newsroom/gsma-announces-new-global-research-that-highlights-significant-growth-opportunity-for-the-mobile-industry>, Jun. 2013.
- [7] http://www.gsacom.com/gsm_3g/market_update.php4, Jun. 2013.
- [8] B. Razavi, Design of Analog CMOS Integrated Circuits, Boston: McGraw Hill, 2001.
- [9] Y. Gao, F. Huang, J. Cheng, and L. Wu, "A 3.1-4.8 GHz Wideband Mixer for MB-OFDM UWB Receivers With Improved Noise Performance," IEEE Youth Conf. Information, Computing and Telecommunication in Beijing (YC-ICT), Sept. 2009, pp. 58-61.
- [10] K. Choi, D. H. Shin, and C. P. Vue, "A 1.2-V, 5.8mW, Ultra-Wideband Folded Mixer in 0.13- μ m CMOS," IEEE Radio Frequency Integrated Circuit Symposium in Honolulu (RFIC), Jun. 2007, pp. 489-492.
- [11] D. Zhao, F. Huang, X. Tang, and X. Sun, "A 0.7-6 GHz Low-Voltage Broadband Folded Mixer in 0.13- μ m CMOS," Microwave and Millimeter Wave Technology in Seoul (ICMMT), May 2012, pp. 1-4.

A Microwave Triplexer Based on Coupled Resonators Approach

Lidiane S. Araújo, Antonio J. Belfort de Oliveira

Dept. of Electronics and Systems

Federal University of Pernambuco

Recife, Brazil

e-mail: lidiane_sac@hotmail.com; belfort@ufpe.br

Abstract—This paper presents the design and computer simulation results of a triplexer based on coupled resonators approach. It comprises 6 resonators yielding three passbands which are centered at 1.800 GHz, 2.077 GHz and 2.400 GHz covering the GSM, UMTS and IEEE 802.11b wireless bands. The specification of each band is Chebyshev with maximum reflection loss of -20 dB (ripple of 0.04321 dB). A Gradient-based local optimization technique is here utilized to determine the coupling coefficients by minimizing a proposed Cost Function. The triplexer can be built regardless of the type of resonator. It is only required that they be able to provide the optimized coupling coefficients.

Keywords—coupling coefficient; triplexer; coupling optimization

I. INTRODUCTION

Multiplexers are devices commonly used in microwave front-end systems for channel separation. The need to operate with multiple services in multiple bands by using a single communication device, demands the input signal to be separated in various bands. This implies that the circuits of microwave transceiver systems have to be larger. In view of this, it has become imperative to build devices even smaller. Conventional diplexers and triplexers are usually built with a junction that splits the input power into two and three parts, respectively, each one being directed to a filter tuned at the desired frequency band. Normally, matching circuits need to be designed to match the filters to the transmission lines coming from the junction. This makes the resulting system bulkier. Multiplexers based on coupling resonators have proved to meet the demand of size reduction, especially because they avoid the use of energy distribution network to split the incoming signal. For this case, matching the circuit is part of its design. Very few triplexers based on coupling resonators have been reported. This article presents computer simulation results of a triplexer based on coupled resonators, without energy distribution network. The theory presented here is an extension of the one developed in [1], which is applied for the design of power dividers and diplexers. The triplexer proposed here comprises 6 resonators operating with three passbands centered at 1.800 GHz, 2.077 GHz and 2.400 GHz. The specification is Chebyshev with maximum reflection loss of -20 dB. A Gradient-based local optimization technique is used to minimize a proposed Cost Function to determine the coupling coefficients. The procedure applies to any type of resonator, provided that they give the coupling coefficients calculated in the design. This paper deals with the state of the art for triplexers, gives the

description of the triplexer treated here, shows the theory involved for its design and presents the simulation results for the reflection and transmission losses.

II. THE STATE OF THE ART

Triplexers using power divider junction have been reported in [2,3], where matching circuits need to be designed. Also very few triplexers based on coupled resonators have been described, but no optimization was considered [4,5]. This paper presents the design of a triplexer for the wireless band based on the coupling matrix optimization and without power junction. As a result, no matching circuit is needed. Following the procedure, the coupling coefficients are the only control variables to be optimized, since the external quality factors are normally calculated using the filter theory developed in [6].

III. THE TRIPLEXER DESCRIPTION AND DESIGN

A triplexer is a four-port device which is described in Figure 1. It receives a signal at its input port and splits it in three parts, each one tuned at a specific frequency.

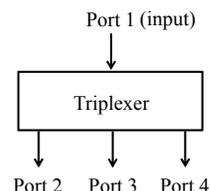


Figure 1. The triplexer diagram

The proposed triplexer is formed by 6 resonators with direct coupling between adjacent ones according to Figure 2. The lines linking the resonators represent the coupling between them.

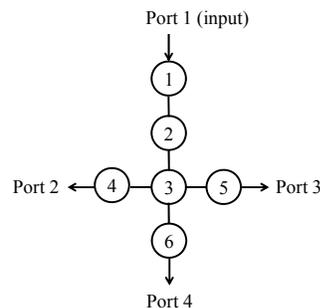


Figure 2. The triplexer formed by 6 coupled resonators

These resonators can be in microstrip, waveguide or other technology. The coupling coefficients between adjacent resonators are designed such that the mid passband is collected from resonator 6 (corresponds to port 4) and the two side passbands are collected from resonators 4 (port 3) and 5 (port 2). A central resonator distributes power to the three branches of the triplexer. The design theory of coupled resonator filters is well developed in [6] and extended in [1] for power dividers and diplexers. Figure 3 shows a circuit of coupled resonators, which exhibits electric and magnetic coupling through their capacitors and inductors and can be analyzed from the Kirchhoff's circuit laws.

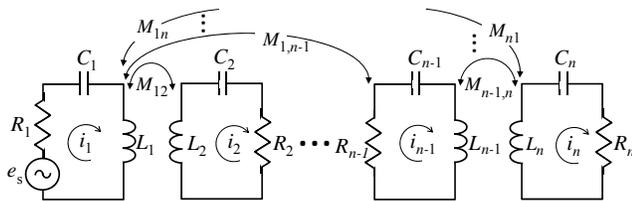


Figure 3. Multiport circuit of coupled resonators

In order to transform the circuit into a triplexer, the starting step is to consider suitable four ports, which are represented by resonators x , y and z whose resistors are R_x , R_y and R_z . Ports 2, 3 and 4 are associated to resonators x , y and z , respectively. Figure 4 shows the triplexer from the resulting four-port network described above.

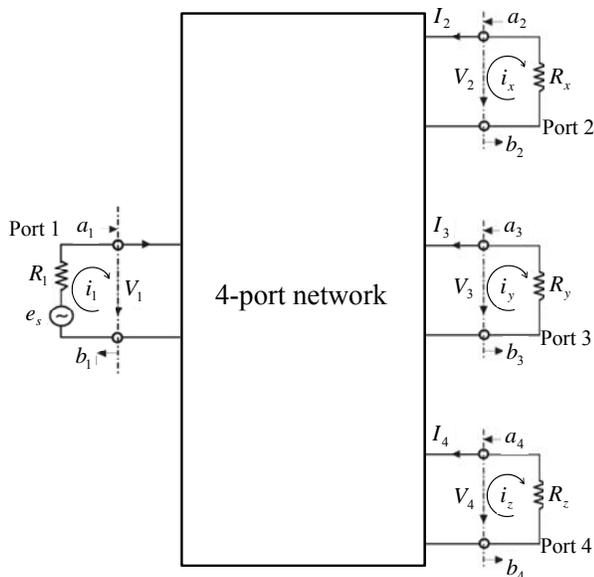


Figure 4. The triplexer resulting from the N-resonator network.

Referring to Figure 3, port 1 is the feeding port; port 2 is associated to resonator 4; port 3 associated to resonator 5 and port 4 to resonator 6. Manipulating the equations of Kirchhoff's circuit laws, the following expressions can be obtained for the scattering parameters:

$$S_{11}(z) = 1 - \frac{2 \text{cof}_{11}([B(z)])}{q_{e1} \Delta_B(z)} \quad (1)$$

$$S_{21}(z) = \frac{2 \text{cof}_{14}([B(z)])}{\sqrt{q_{e1} q_{e4}} \Delta_B(z)} \quad (2)$$

$$S_{31}(z) = \frac{2 \text{cof}_{15}([B(z)])}{\sqrt{q_{e1} q_{e5}} \Delta_B(z)} \quad (3)$$

$$S_{41}(z) = \frac{2 \text{cof}_{16}([B(z)])}{\sqrt{q_{e1} q_{e6}} \Delta_B(z)} \quad (4)$$

where q_{e1} , q_{e4} , q_{e5} and q_{e6} are the external quality factors at ports 1, 4, 5 and 6, and $\text{cof}_{ij}([B(z)])$ is the element ij of the cofactor of matrix $[B]$ calculated at $z = j2\pi f$. Also, $\Delta_B(z)$ is the determinant of matrix $[B]$ calculated at the same complex frequency. Matrix $[B]$ of the triplexer depicted in Figure 3 can be expressed by

$$\begin{bmatrix} \frac{1}{q_{e1}} + P & -jm_{12} & 0 & 0 & 0 & 0 \\ -jm_{12} & P & -jm_{23} & 0 & 0 & 0 \\ 0 & -jm_{23} & P & -jm_{34} & -jm_{35} & -jm_{36} \\ 0 & 0 & -jm_{34} & \frac{1}{q_{e4}} + P + jm_{44} & -jm_{45} & 0 \\ 0 & 0 & -jm_{35} & -jm_{45} & \frac{1}{q_{e5}} + P - jm_{44} & 0 \\ 0 & 0 & -jm_{36} & 0 & 0 & \frac{1}{q_{e6}} + P \end{bmatrix}$$

where m_{12} , m_{23} , m_{34} , m_{35} , m_{36} , m_{45} and m_{44} are the normalized coupling coefficients to be optimized. Since the four-port network is asynchronously tuned, normalized coupling coefficients m_{44} and m_{55} have to be introduced in the coupling matrix. In order to have symmetrical passbands in relation to f_0 , one has to set $m_{55} = -m_{44}$. Also, for the passbands to exhibit the same Chebyshev specifications, the relation $m_{34} = m_{35} = m_{36}$ must be established. Variable P is the prototype normalized frequency, which is related to the denormalized frequency by the expression

$$P = \frac{j}{FBW} \left(\frac{f}{f_0} - \frac{f_0}{f} \right) \quad (5)$$

where f is expressed in GHz and $FBW = (f_2 - f_1)/f_0$ is the fractional bandwidth. For the triplexer proposed here, $f_1 = 1.795 \text{ GHz}$, $f_2 = 2.405 \text{ GHz}$, and $f_0 = 2.077 \text{ GHz}$.

The Cost Function to be optimized is given by

$$CF = \sum_{i=1}^N \left| \Delta_B(z_{ri}) - \frac{2c\text{of}_{11}[B(z_{ri})]}{q_{e1}} \right|^2 + \sum_{j=1}^{N-3} \left| 1 - \frac{2c\text{of}_{11}[B(z_{pj})]}{q_{e1}\Delta_B(z_{pj})} - 10^{\frac{L_R}{20}} \right|^2 \quad (6)$$

where N is the number of resonators, z_{ri} is the i -th reflection zero and z_{pj} is the complex frequency where the j -th peak of the reflection loss is observed. Since the number of reflection zeros for each passband is 2, only one peak in the reflection loss response is observed. This means that 3 peaks are observed in the reflection loss response of the triplexer ($N-3$). The coupling coefficients and external quality factors form a set of parameters known as control variables. In this work, only the coupling coefficients are to be optimized, since the calculated external quality factors have proved to be very effective for good simulation results of the coupling coefficients. The first term in the Cost Function is derived from (1) and is minimum when calculated at z_{ri} . Likewise, the second one is related to the points where the peaks of $|s_{11}|$ occur. Theory shows that they occur for transmission loss given by $10^{\frac{L_R}{20}}$.

IV. PERFORMANCE OF THE TRIPLEXER

The design was initially done in the normalized frequency from the desired specifications in the high frequency range. A triplexer with Chebyshev characteristics [6] was simulated with maximum reflection loss of -20 dB (ripple 0.04321 dB) for each passband. The center frequency is 2.077 GHz in the denormalized frequency domain. A good guess for the external quality factors can be obtained if one considers the triplexer as a filter formed by 6 resonators. The filter normalized frequency prototype elements are then used to calculate the initial guess for the external quality factors at ports 1, 2, 3 and 4 in the computer simulations. A Gradient-based local optimization technique [7] was utilized to obtain the coupling coefficients m_{12} , m_{23} , m_{34} and m_{44} . The normalized external quality factors are calculated by employing expressions [6]

$$q_{e1} = g_0 g_1 \quad (7)$$

$$q_{e6} = g_6 g_7 \quad (8)$$

where g_0 , g_1 , g_6 and g_7 are the prototype normalized frequency elements of a Chebyshev filter of 6 elements. For the triplexer here simulated, the bandwidth of each passband was taken to be 0.125 Hz in the normalized frequency range. The calculated external quality factors were found to be $q_{e1} = 5.30$ and $q_{e4} = q_{e5} = q_{e6} = 15.9$. Having been calculated, rather than optimized, reduction in the number of

parameters to be optimized and in the computer time is achieved. Furthermore, very good results for the optimized coupling coefficients can also be attained. As for the coupling coefficients, the initial guess for the computer simulation was $m_{12} = m_{23} = m_{44} = 0.6$ and $m_{34} = 0.5$. Since the algorithm searches for a local minimum, the results are sensitive to the initial guess. After 90 iterations and a precision of 10^{-9} , the final optimized coupling coefficients values were $m_{12} = 0.6775$, $m_{23} = 0.6595$, $m_{34} = 0.1166$ and $m_{44} = 0.9213$. Figure 5 shows the response of the triplexer against frequency in the normalized frequency. The selectivity of the channels can be improved as the number of resonators increases, which makes the reflection loss greater in the rejection band.

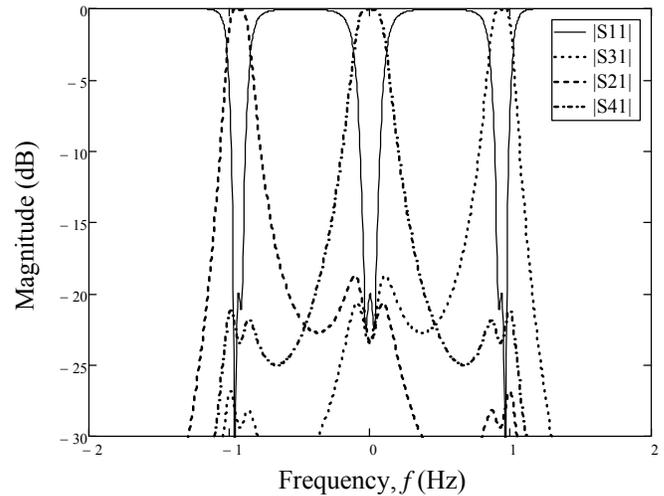


Figure 5. The triplexer response in the normalized frequency.

The denormalized frequency response of the triplexer is graphed in Figure 6. As can be seen, it is slightly compressed in the passband centered at 1.800 GHz when compared with the other side passband (centered at 2.447 GHz) due to the corresponding frequency transformation in (5). Even so, very similar responses were attained for the three output passbands. It was observed that the number of resonators is equal to the number of reflection zeros. Furthermore, in

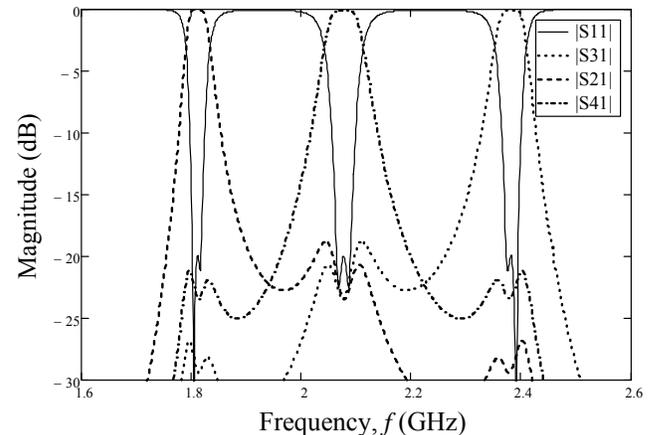


Figure 6. The triplexer response in the denormalized frequency.

order to have three outputs with similar Chebyshev specifications, the triplexer must have a number of resonators multiple of three, so that each one of the three passbands presents the same number of reflection zeros.

V. CONCLUSIONS AND FUTURE WORK

Design and simulation results of a triplexer based on coupled resonators approach was presented here. Six resonators were employed without using any energy distribution network. A Chebyshev response for the three passbands was achieved, with maximum reflection loss -20 dB (ripple 0.04321) and fractional bandwidths of 2.5%, 3.3% and 2.3% for the left, central and right passbands, respectively. A Gradient-based local optimization technique was used, which consisted in minimizing a proposed Cost Function whose control variables were the coupling coefficients. The external quality factors were excluded from the optimization procedure, since the calculated values yielded satisfactory optimized coupling coefficients. Very good simulation results were obtained for the reflection loss and transmission loss. As a future work, the triplexer will be implemented in microstrip with triangular resonators aiming to achieve size reduction.

ACKNOWLEDGMENT

The authors wish to acknowledge the support given to this research by the Brazilian funding agency, CAPES (Coordination for the Improvement of High Level Personnel).

REFERENCES

- [1] T. F. Skaik, M. J. Lancaster, and F. Huang, "Synthesis of multiple output coupled resonator circuits using coupling matrix optimization," *IET Microwave, Antennas and Propagation*, vol. 5, ed. 9, June, 2011, pp. 1081-1088, doi:10.1049/iet-map.2010.0447.
- [2] P. Deng, M. Lai, S. Jeng, and C. H. Chen, "Design of Matching Circuits for Microstrip Triplexers Based on Stepped-Impedance Resonators," *IEEE Trans. on Microwave Theory and Technique*, vol. 54, Dec. 2006, pp. 4185-4192, doi:10.1109/TMTT.2006.886161.
- [3] G. Macchiarella and S. Tamiazzo, "Design of triplexer combiners for base stations of mobile communications," *IEEE MTT-S International Microwave Symposium Digest*, Milan, Italy, May, 2010, pp. 429-432, doi:10.1109/MWSYM.2010.5517315.
- [4] C. Chen, T. Huang, T. Shen, and R. Wu, "A miniaturized microstrip common resonator triplexer without extra matching network," *Asia-Pacific Microwave Conference*, Yokohama, Japan, Dec. 2006, pp. 1439-1442, doi:10.1109/APMC.2006.4429677.
- [5] X. Zhang, M. He, and X. Zhao, "Compact design of high temperature superconducting duplexer and triplexer for satellite communications," *Supercond. Sci. Technol.*, vol. 25, July, 2012, pp. 1-5, doi:10.1088/0953-2048/25/10/105005.
- [6] J. S. Hong and M. J. Lancaster, *Microstrip filters for RF/microwave applications*. New York, Wiley, 2001.
- [7] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: theory and algorithms*. New Jersey, Wiley, 2006.

Spectral Resource Sharing for Uncoordinated Networks

Ayman Assra and Tricia Willink
Communications Research Centre
Ottawa, ON, Canada

E-mail: ayman.assra@gmail.com and tricia.willink@crc.gc.ca

Abstract—Spectrum resource sharing strategies for uncoordinated networks are investigated. The average capacity of 2-pair device-to-device (D2D) communication systems is derived, assuming a flat response of the frequency spectrum. Three different frequency allocation strategies are considered: (i) separate spectral allocation; (ii) full spectral allocation; and (iii) overlapped spectral allocation, considering deterministic and random locations of the communication devices. The analytical results, supported by simulations, show that at low to moderate received SNR the throughput assigned to each communication device can be enhanced by overlapping the frequency spectrum allocations.

Index Terms—Uncoordinated networks, frequency allocation strategies, D2D communication.

I. INTRODUCTION

Recently, device-to-device (D2D) communication has attracted more attention for its promising results in enhancing the system capacity [1]. With the increasing demand in establishing home networks, which connect a variety of communication devices including mobile phones, laptops, or other electronic appliances, D2D communication becomes a common paradigm for establishing these connections [2]. D2D communication schemes require low power consumption, cost, and human intervention [3]. Many studies have been performed to investigate the capacity of specific uncoordinated networks, i.e., cellular networks [4]. However, the results of these investigations cannot be generalized to other types of communication systems. Therefore, in this paper, we study the capacity of uncoordinated networks using physical parameters independent of the system specifications, such as the locations of communication devices, the number of devices which can co-exist in a certain region, and the environment.

We focus our study on 2-pair D2D communication systems considering three different spectral allocation strategies: separated, whole and overlapped spectral allocations. For tractability, we consider channels with a flat frequency response.

First, in Section II, we present the sum capacity of 2-pair D2D systems with deterministic locations. Then, in Section III, we derive the average sum capacity considering random locations of the communication devices. The results are introduced in Section IV. Finally, Section V concludes the paper.

II. SUM CAPACITY OF 2-PAIR D2D SYSTEMS WITH DETERMINISTIC LOCATIONS

In this section, we present the sum capacity of a circular cell in which two pairs of transmitters, (i.e., TX_1 and TX_2) and receivers (i.e., RX_1 and RX_2) are located in fixed positions inside the cell as shown in Fig. 1. In this figure, the solid lines represent the desired signals from the transmitters to the desired destinations (i.e., TX_1 - RX_1 and TX_2 - RX_2), while interfering signals (i.e., TX_1 - RX_2 and TX_2 - RX_1) are denoted by dotted lines.

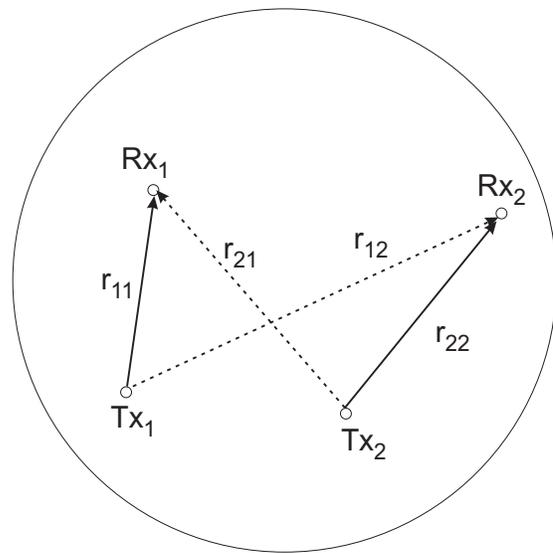


Fig. 1. Positions of communication devices inside a cell.

For simplicity, the channels of the desired and interfering communication links are modeled using the single-slope path loss, where the received power at a distance r from the transmitter is given by

$$P_r(r) \propto P_t r^{-\alpha} \quad (1)$$

where P_t is the transmit power spectral density, and α is the path loss exponent. We assume that P_t is the same for all links, and $\alpha = 4$. To simplify our analysis, we neglect the coefficient of proportionality and consider the effective transmit power density P_t such that (1) contains an equality.

Assuming that the frequency spectrum under investigation has a flat response over different frequencies, we use an

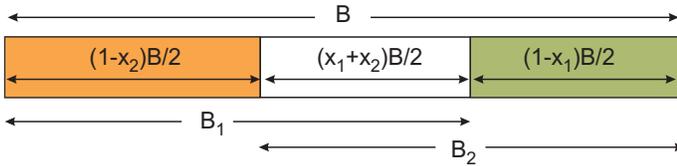


Fig. 2. Overlapped frequency spectrum allocations for 2-pair D2D communication systems.

overlapped spectral allocation strategy, where the bandwidth assigned to each transmitter is extended with a specific ratio over the neighboring frequency bands as shown in Fig. 2. The parameters x_1 and x_2 represent the spectral overlapping ratios of Tx_1 on Tx_2 and Tx_2 on Tx_1 , respectively.

In this case, the average capacity of the 2-pair D2D communication systems can be defined by

$$C_T = C_{F1} + C_{F2} + C_{O1} + C_{O2} \quad (2)$$

where C_{F1} and C_{F2} are the capacity of Tx_1 and Tx_2 in the region of the spectrum which is free of interference, respectively. The components C_{O1} and C_{O2} are the capacity of Tx_1 and Tx_2 in the overlapped spectral region between the two transmitters, respectively. The expressions corresponding to these capacities are

$$C_{F1} = (1 - x_2) \frac{B}{2} \log_2 \left(1 + \frac{P_t}{N} \frac{r_{11}^\alpha}{r_{11}^\alpha} \right) \quad (3)$$

$$C_{F2} = (1 - x_1) \frac{B}{2} \log_2 \left(1 + \frac{P_t}{N} \frac{r_{22}^\alpha}{r_{22}^\alpha} \right) \quad (4)$$

$$C_{O1} = (x_2 + x_1) \frac{B}{2} \log_2 \left(1 + \frac{P_t}{N} \frac{r_{11}^\alpha}{r_{21}^\alpha + N} \right) \quad (5)$$

$$C_{O2} = (x_2 + x_1) \frac{B}{2} \log_2 \left(1 + \frac{P_t}{N} \frac{r_{22}^\alpha}{r_{12}^\alpha + N} \right) \quad (6)$$

where B is the total frequency bandwidth, and N is the noise power spectral density. The distance $\{r_{ij}\}$, $i, j \in \{1, 2\}$, represents the length of the communication link between Tx_i and Rx_j . Also, $\{r_{ij}\}$ takes values between r_m and $2R$, where r_m is the minimum distance between two communication devices inside the cell, and R is the cell radius.

III. SUM CAPACITY OF 2-PAIR D2D SYSTEMS WITH RANDOM LOCATIONS

Here, we assume that Tx_1 , Tx_2 , Rx_1 and Rx_2 are located in random positions inside the cell. In this case, r_{11} , r_{21} , r_{12} , and r_{22} become variable parameters, which have independent identical distributions defined as follows. Let r represent the Euclidian distance between two points randomly located in a circle of radius R . Then, the probability distribution function of r is given by [5]

$$f(r) = \frac{2r}{R^2} \left(\frac{2}{\pi} \cos^{-1} \left(\frac{r}{2R} \right) - \frac{r}{\pi R} \sqrt{1 - \frac{r^2}{4R^2}} \right) \quad (7)$$

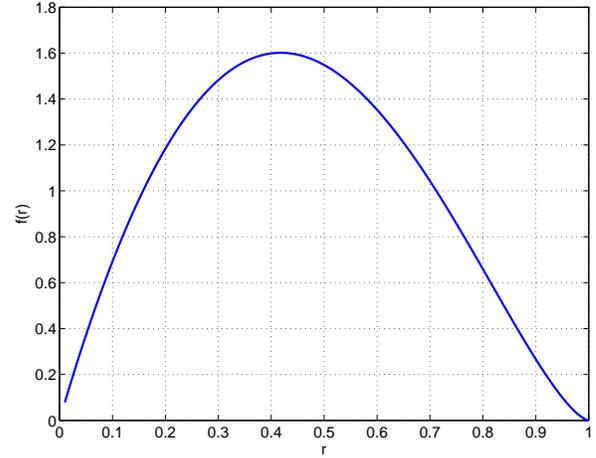


Fig. 3. Distribution of r inside a unit circle ($R = 0.5$).

which is plotted in Fig. 3. The validity of this probability distribution is investigated by Moltchanov in [5].

Using the closed-form expression of $f(r)$ in (7), we can estimate the average sum capacity of 2-pair D2D communications over a flat frequency response as follows

$$\bar{C}_T = \bar{C}_{F1} + \bar{C}_{O1} + \bar{C}_{O2} + \bar{C}_{F2} \quad (8)$$

where

$$\bar{C}_{F1} = \int_{r_m}^{2R} C_{F1}(r_{11}) f(r_{11}) dr_{11} \quad (9)$$

$$\bar{C}_{F2} = \int_{r_m}^{2R} C_{F2}(r_{22}) f(r_{22}) dr_{22} \quad (10)$$

$$\bar{C}_{O1} = \int_{r_m}^{2R} \int_{r_m}^{2R} C_{O1}(r_{11}, r_{21}) f(r_{11}, r_{21}) dr_{11} dr_{21} \quad (11)$$

$$\bar{C}_{O2} = \int_{r_m}^{2R} \int_{r_m}^{2R} C_{O2}(r_{12}, r_{22}) f(r_{12}, r_{22}) dr_{12} dr_{22}. \quad (12)$$

In (9) and (10), the closed-form expression of \bar{C}_{F1} and \bar{C}_{F2} can be obtained by deriving the average capacity of 1-pair D2D communication systems over bandwidths $(1 - x_2) \frac{B}{2}$ and $(1 - x_1) \frac{B}{2}$, respectively. The closed-form expressions of \bar{C}_{O1} and \bar{C}_{O2} can be obtained by deriving the average capacity of 1-pair D2D communication systems with one interferer over a bandwidth $(x_1 + x_2) \frac{B}{2}$. In the following subsections, we will show how the average capacity of the aforementioned cases can be estimated.

A. Average capacity of 1-pair D2D communications with no interference

This case is represented by (9) or (10), where C_{F1} is a function only of r_{11} , or C_{F2} is a function only of r_{22} . First, consider the derivation of \bar{C}_{F1} . Using (7), the average capacity of 1-pair D2D communication systems over a frequency bandwidth, $\tilde{B} = (1 - x_2) \frac{B}{2}$, is given by

$$\bar{C}_{F1} = \tilde{B} \left(\frac{4}{\pi R^2} \bar{c}_1 - \frac{2}{\pi R^3} \bar{c}_2 \right) \quad (13)$$

where

$$\bar{c}_1 = \int_{r_m}^{2R} \log_2(1 + \gamma_r) r_{11} \cos^{-1} \left(\frac{r_{11}}{2R} \right) dr_{11} \quad (14)$$

$$\bar{c}_2 = \int_{r_m}^{2R} \log_2(1 + \gamma_r) r_{11}^2 \sqrt{1 - \frac{r_{11}^2}{4R^2}} dr \quad (15)$$

where γ_r is the received signal-to-noise ratio (SNR) and given by $\gamma_r = \frac{\gamma_t}{r_{11}^\alpha}$, where $\gamma_t = \frac{P_t}{N}$. The mathematical derivations of the integrations in (14) and (15) are not tractable; therefore, we use the series expansions of the following mathematical functions to reduce the complexity of these integrations [6].

For $|\gamma_r| \leq 1$, the logarithmic function $\log_2(1 + \gamma_r)$ can be expressed as

$$\log_2(1 + \gamma_r) = \frac{1}{\ln(2)} \left(\gamma_r - \frac{\gamma_r^2}{2} + \frac{\gamma_r^3}{3} + \dots \right) \quad (16a)$$

whereas for $|\gamma_r| > 1$,

$$\log_2(1 + \gamma_r) = \frac{1}{\ln(2)} \left(\log(\gamma_r) + \frac{1}{\gamma_r} - \frac{2}{2\gamma_r^2} + \frac{1}{3\gamma_r^3} - \dots \right) \quad (16b)$$

where \ln is the natural logarithm. Also, for $|\frac{r}{2R} < 1|$,

$$\sqrt{1 - \frac{r^2}{4R^2}} = 1 - \frac{r^2}{8R^2} - \frac{r^4}{128R^4} - \dots \quad (17)$$

For $|\gamma_r| \leq 1$, we substitute for $\log_2(1 + \gamma_r)$ using the first three terms of the series expansion in (16a) since the values of the remaining terms are small and can be neglected. Accordingly, \bar{c}_1 can be approximated using

$$\bar{c}_1 \approx \frac{1}{\ln(2)} \left(\bar{c}_{a1} - \frac{\bar{c}_{a2}}{2} + \frac{\bar{c}_{a3}}{3} \right) \quad \forall |\gamma_r| \leq 1 \quad (18)$$

where

$$\bar{c}_{a1} = \int_{r_m}^{2R} \cos^{-1} \left(\frac{r_{11}}{2R} \right) \frac{\gamma_t}{r_{11}^{\alpha-1}} dr_{11} \quad (19)$$

$$\bar{c}_{a2} = \int_{r_m}^{2R} \cos^{-1} \left(\frac{r_{11}}{2R} \right) \frac{\gamma_t^2}{r_{11}^{2\alpha-1}} dr_{11} \quad (20)$$

$$\bar{c}_{a3} = \int_{r_m}^{2R} \cos^{-1} \left(\frac{r_{11}}{2R} \right) \frac{\gamma_t^3}{r_{11}^{3\alpha-1}} dr_{11}. \quad (21)$$

When $|\gamma_r| > 1$, we also include up to the third-order term, i.e., the first four terms of the series expansion in (16b), and consequently \bar{c}_1 can be given by

$$\bar{c}_1 \approx \frac{1}{\ln(2)} \left(\bar{c}_{b1} + \bar{c}_{b2} - \frac{\bar{c}_{b3}}{2} + \frac{\bar{c}_{b4}}{3} \right) \quad (22)$$

where

$$\bar{c}_{b1} = \int_{r_m}^{2R} r_{11} \cos^{-1} \left(\frac{r_{11}}{2R} \right) \log \left(\frac{\gamma_t}{r_{11}^\alpha} \right) dr_{11} \quad (23)$$

$$\bar{c}_{b2} = \int_{r_m}^{2R} \cos^{-1} \left(\frac{r_{11}}{2R} \right) \frac{r_{11}^{\alpha+1}}{\gamma_t} dr_{11} \quad (24)$$

$$\bar{c}_{b3} = \int_{r_m}^{2R} \cos^{-1} \left(\frac{r_{11}}{2R} \right) \frac{r_{11}^{2\alpha+1}}{2\gamma_t^2} dr_{11} \quad (25)$$

$$\bar{c}_{b4} = \int_{r_m}^{2R} \cos^{-1} \left(\frac{r_{11}}{2R} \right) \frac{r_{11}^{3\alpha+1}}{3\gamma_t^3} dr_{11}. \quad (26)$$

Finally, we estimate \bar{c}_2 using the series expansion in (17) as follows

$$\bar{c}_2 \approx \bar{c}_{e1} - \frac{1}{8R^2} \bar{c}_{e2} - \frac{1}{128R^4} \bar{c}_{e3} \quad (27)$$

where

$$\bar{c}_{e1} = \int_{r_m}^{2R} \log_2 \left(1 + \frac{\gamma_t}{r_{11}^\alpha} \right) r_{11}^2 dr_{11} \quad (28)$$

$$\bar{c}_{e2} = \int_{r_m}^{2R} \log_2 \left(1 + \frac{\gamma_t}{r_{11}^\alpha} \right) r_{11}^4 dr_{11} \quad (29)$$

$$\bar{c}_{e3} = \int_{r_m}^{2R} \log_2 \left(1 + \frac{\gamma_t}{r_{11}^\alpha} \right) r_{11}^6 dr_{11}. \quad (30)$$

The closed-form expression of the integrations in (19)-(21), (23)-(26) and (28)-(30) can be obtained from handbooks of integrations; however, due to the space limitation of the paper, we refer the reader to [6]. Similarly, we use the same mathematical derivations to get \bar{C}_{F2} over bandwidth $(1-x_1)\frac{B}{2}$ from (10).

B. Average capacity of one-pair D2D communications with one interferer

Now, we assume that there is an interfering signal to the desired communication link from a randomly located communication device. The average capacity of this case is presented in (11) and (12). In the following, we first solve the double integration in (11), and then we use the same procedure to evaluate the integrations in (12). Assuming that the locations of Tx_1 and Tx_2 inside the cell are statistically independent, i.e., $f(r_{11}, r_{21}) = f(r_{11})f(r_{21})$, (11) is reduced to

$$\bar{C}_{O1} = \tilde{B}_1 \int_{r_m}^{2R} \bar{\Phi}(r_{11}) f(r_{11}) dr_{11} \quad (31)$$

where $\tilde{B}_1 = (x_1 + x_2)\frac{B}{2}$, and

$$\begin{aligned} \bar{\Phi}(r_{11}) &= \int_{r_m}^{2R} \log_2 \left(1 + \frac{r_{21}^\alpha}{r_{11}^\alpha(1 + \gamma_t^{-1}r_{21}^\alpha)} \right) f(r_{21}) dr_{21} \\ &= \left(\frac{4}{\pi R^2} \bar{\Phi}_1(r_{11}) - \frac{2}{\pi R^3} \bar{\Phi}_2(r_{11}) \right). \end{aligned} \quad (32)$$

In (32),

$$\begin{aligned} \bar{\Phi}_1(r_{11}) &= \int_{r_m}^{2R} \log_2 \left(1 + \frac{r_{21}^\alpha}{r_{11}^\alpha(1 + \gamma_t^{-1}r_{21}^\alpha)} \right) r_{21} \\ &\quad \times \cos^{-1} \left(\frac{r_{21}}{2R} \right) dr_{21} \end{aligned} \quad (33)$$

$$\begin{aligned} \bar{\Phi}_2(r_{11}) &= \int_{r_m}^{2R} \log_2 \left(1 + \frac{r_{21}^\alpha}{r_{11}^\alpha(1 + \gamma_t^{-1}r_{21}^\alpha)} \right) r_{21}^2 \\ &\quad \times \sqrt{1 - \frac{r_{21}^2}{4R^2}} dr_{21}. \end{aligned} \quad (34)$$

In (33), the integration can be simplified by substituting for $\cos^{-1}(\theta)$ by the first two terms of its series expansion, which is given by

$$\cos^{-1}(\theta) = \frac{\pi}{2} - \theta - \dots \quad (35)$$

Also, we evaluate these integrations for $\alpha = 4$. Thus, the integrations in (33) can be approximated by

$$\bar{\Phi}_1(r_{11}) \approx \frac{\pi}{2} \bar{\Phi}_{11}(r_{11}) - \frac{1}{2R} \bar{\Phi}_{12}(r_{11}) \quad (36)$$

where

$$\begin{aligned} \bar{\Phi}_{11}(r_{11}) &= \int_{r_m}^{2R} r_{21} \log_2 \left(1 + \frac{r_{21}^4}{r_{11}^4 (1 + \gamma_t^{-1} r_{21}^4)} \right) dr_{21} \\ &= \left[\frac{r_{21}^2}{2} \log \left(\frac{r_{21}^4}{r_{11}^4 (\gamma_t^{-1} r_{21}^4 + 1)} + 1 \right) - \frac{r_{11}^2}{\sqrt{\gamma_t^{-1} r_{11}^4 + 1}} \right. \\ &\times \tan^{-1} \left(1 - \frac{\sqrt{2} r_{21} (\gamma_t^{-1} r_{11}^4 + 1)^{\frac{1}{4}}}{r_{11}} \right) - \frac{r_{11}^2}{\sqrt{\gamma_t^{-1} r_{11}^4 + 1}} \\ &\times \tan^{-1} \left(1 + \frac{\sqrt{2} r_{21} (\gamma_t^{-1} r_{11}^4 + 1)^{\frac{1}{4}}}{r_{11}} \right) + \frac{1}{\gamma_t^{-\frac{1}{2}}} \\ &\times \left. \left(\tan^{-1} \left(1 - \sqrt{2} \gamma_t^{-\frac{1}{4}} r_{21} \right) + \tan^{-1} \left(1 + \sqrt{2} \gamma_t^{-\frac{1}{4}} r_{21} \right) \right) \right]_{r_m}^{2R} \end{aligned} \quad (37)$$

and

$$\begin{aligned} \bar{\Phi}_{12}(r_{11}) &= \int_{r_m}^{2R} r_{21}^2 \log_2 \left(1 + \frac{r_{21}^4}{r_{11}^4 (1 + \gamma_t^{-1} r_{21}^4)} \right) dr_{21} \\ &= \frac{1}{6\sqrt{2}} \left[-\frac{\sqrt{2}}{\gamma_t^{-\frac{3}{4}}} \log \left(\gamma_t^{-\frac{1}{2}} r_{21}^2 - \sqrt{2} \gamma_t^{-\frac{1}{4}} r_{21} + 1 \right) + \frac{\sqrt{2}}{\gamma_t^{-\frac{3}{4}}} \right. \\ &\times \log \left(\gamma_t^{-\frac{1}{2}} r_{21}^2 + \sqrt{2} \gamma_t^{-\frac{1}{4}} r_{21} + 1 \right) + \frac{2\sqrt{2}}{\gamma_t^{-\frac{3}{4}}} \\ &\times \tan^{-1} \left(1 - \sqrt{2} \gamma_t^{-\frac{1}{4}} r_{21} \right) - \frac{2\sqrt{2}}{\gamma_t^{-\frac{3}{4}}} \tan^{-1} \left(1 + \sqrt{2} \gamma_t^{-\frac{1}{4}} r_{21} \right) \\ &+ 2r_{21}^3 \log \left(\frac{r_{21}^4}{(r_{11}^4 (\gamma_t^{-1} r_{21}^4 + 1))} + 1 \right) - \frac{2\sqrt{2} r_{11}^3}{(\gamma_t^{-1} r_{11}^4 + 1)^{\frac{3}{4}}} \\ &\times \tan^{-1} \left(1 - \frac{\sqrt{2} r_{21} (\gamma_t^{-1} r_{11}^4 + 1)^{\frac{1}{4}}}{r_{11}} \right) + \frac{2\sqrt{2} r_{11}^3}{(\gamma_t^{-1} r_{11}^4 + 1)^{3/4}} \\ &\times \tan^{-1} \left(1 + \frac{\sqrt{2} r_{21} (\gamma_t^{-1} r_{11}^4 + 1)^{\frac{1}{4}}}{r_{11}} \right) + \frac{\sqrt{2} r_{11}^3}{(\gamma_t^{-1} r_{11}^4 + 1)^{\frac{3}{4}}} \\ &\log \left(r_{21}^2 \sqrt{\gamma_t^{-1} r_{11}^4 + 1} - \sqrt{2} r_{11} r_{21} (\gamma_t^{-1} r_{11}^4 + 1)^{\frac{1}{4}} + r_{11}^2 \right) \\ &- \frac{\sqrt{2} r_{11}^3}{(\gamma_t^{-1} r_{11}^4 + 1)^{\frac{3}{4}}} \log \left(r_{21}^2 \sqrt{\gamma_t^{-1} r_{11}^4 + 1} + \sqrt{2} r_{11} r_{21} \right. \\ &\times \left. \left. \left(\gamma_t^{-1} r_{11}^4 + 1 \right)^{\frac{1}{4}} + r_{11}^2 \right) \right]_{r_m}^{2R}. \end{aligned} \quad (38)$$

The integrations of (37) and (38) can be evaluated with the aid of a handbook of integrations, e.g., [6]. Using the first two terms of the expansion in (17) and considering $\alpha = 4$, the integration in (34) can be approximated as

$$\bar{\Phi}_2(r_{11}) \approx \bar{\Phi}_{21}(r_{11}) - \frac{1}{8R^2} \bar{\Phi}_{22}(r_{11}) \quad (39)$$

where $\bar{\Phi}_{21}(r_{11}) = \bar{\Phi}_{12}(r_{11})$, given by (38) and

$$\begin{aligned} \bar{\Phi}_{22}(r_{11}) &= \int_{r_m}^{2R} r_{21}^4 \log_2 \left(1 + \frac{r_{21}^4}{r_{11}^4 (1 + \gamma_t^{-1} r_{21}^4)} \right) dr_{21} \\ &= \frac{1}{10 \log(2)} \left[-\frac{\sqrt{2}}{\gamma_t^{-\frac{5}{4}}} \log \left(\gamma_t^{-\frac{1}{2}} r_{21}^2 - \sqrt{2} \gamma_t^{-\frac{1}{4}} r_{21} + 1 \right) + \frac{\sqrt{2}}{\gamma_t^{-\frac{5}{4}}} \right. \\ &\times \log \left(\gamma_t^{-\frac{1}{2}} r_{21}^2 + \sqrt{2} \gamma_t^{-\frac{1}{4}} r_{21} + 1 \right) - \frac{2\sqrt{2}}{\gamma_t^{-\frac{5}{4}}} \tan^{-1} \left(1 - \sqrt{2} \gamma_t^{-1/4} \right. \\ &\times r_{21} \left. \right) + \frac{2\sqrt{2}}{\gamma_t^{-\frac{5}{4}}} \tan^{-1} \left(1 + \sqrt{2} \gamma_t^{-\frac{1}{4}} r_{21} \right) - \frac{8r_{21}}{\gamma_t^{-2} r_{11}^4 + \gamma_t^{-1}} \\ &+ 2r_{21}^5 \log \left(\frac{r_{21}^4}{r_{11}^4 (\gamma_t^{-1} r_{21}^4 + 1)} + 1 \right) + \frac{2^{\frac{3}{2}} r_{11}^5}{(\gamma_t^{-1} r_{11}^4 + 1)^{\frac{5}{4}}} \\ &\times \tan^{-1} \left(1 - \frac{\sqrt{2} r_{21} (\gamma_t^{-1} r_{11}^4 + 1)^{\frac{1}{4}}}{r_{11}} \right) - \frac{2^{\frac{3}{2}} r_{11}^5}{(\gamma_t^{-1} r_{11}^4 + 1)^{\frac{5}{4}}} \\ &\times \tan^{-1} \left(1 + \frac{\sqrt{2} r_{21} (\gamma_t^{-1} r_{11}^4 + 1)^{\frac{1}{4}}}{r_{11}} \right) + \frac{\sqrt{2} r_{11}^5}{(\gamma_t^{-1} r_{11}^4 + 1)^{\frac{5}{4}}} \\ &\times \log \left(r_{21}^2 \sqrt{\gamma_t^{-1} r_{11}^4 + 1} - \sqrt{2} r_{21} r_{11} (\gamma_t^{-1} r_{11}^4 + 1)^{\frac{1}{4}} + r_{11}^2 \right) \\ &- \frac{\sqrt{2} r_{11}^5}{(\gamma_t^{-1} r_{11}^4 + 1)^{\frac{5}{4}}} \log \left(r_{21}^2 \sqrt{\gamma_t^{-1} r_{11}^4 + 1} + \sqrt{2} r_{21} r_{11} \right. \\ &\times \left. \left. \left(\gamma_t^{-1} r_{11}^4 + 1 \right)^{\frac{1}{4}} + r_{11}^2 \right) \right]_{r_m}^{2R}. \end{aligned} \quad (40)$$

The integration in (40) is obtained using [6]. Substituting (36) and (39) in (32), we derive the closed-form expression of $\bar{\Phi}(r_{11})$, and subsequently, we can perform the second integration in (31) with respect to r_{11} . However, due to the high complexity of the resultant integrations, we obtain the final estimate of \bar{C}_{O1} using numerical integration techniques, specifically, we used the adaptive Gauss-Kronrod quadrature method [7].

IV. SIMULATION RESULTS

In this section, we examine the accuracy of the derived expression of the average sum capacity in (8) of 2-pair D2D systems considering different spectral allocation strategies. For illustrative purposes, we perform our simulations using design parameters similar to those of GSM systems: carrier frequency $f_c = 2\text{GHz}$, bandwidth $B = 200\text{kHz}$ and cell with radius $R = 5\text{km}$. The value of r_m , which is the minimum distance between two devices inside the cell to initiate transmission, is taken as 1% of the cell diameter, i.e., $r_m = 0.02R$. In other words, the two devices are not allowed to communicate with each other when $r < r_m$. As noted previously, the channel between different communication devices is modeled using the single-slope path loss with $\alpha = 4$. In our results, we present the average capacity versus the received SNR at the median distance r_e between Tx_i and Rx_i , which is denoted as γ_m . The value of r_e can be estimated from (7) or can be detected from Fig. 3. With this value of γ_m , the transmit power density,

P_t , can be determined from

$$P_t = Nr_e^\alpha \gamma_m. \quad (41)$$

In the following results, the average sum capacity \bar{C}_T is estimated by averaging (2) over 10^5 realizations of $\{r_{ij}\}$, $i, j \in \{1, 2\}$.

The average capacity of 1-pair D2D systems is shown in Fig. 4 for two cases: (i) no interferer (0-Interferer); and (ii) 1-Interferer. The analytical results, supported by the simulations, show that the average capacity of 1-pair systems can be enhanced by increasing P_t ; however, this enhancement becomes limited when an interferer is present. The figure also shows the accuracy of the analytical derivation of the scenario of 0-Interferer. Even for the 1-Interferer case, the error in the derivations at high values of γ_m is less than 23%, which can be reduced by adding more terms from the series expansions in (17) and (35), but this requires additional mathematical computations.

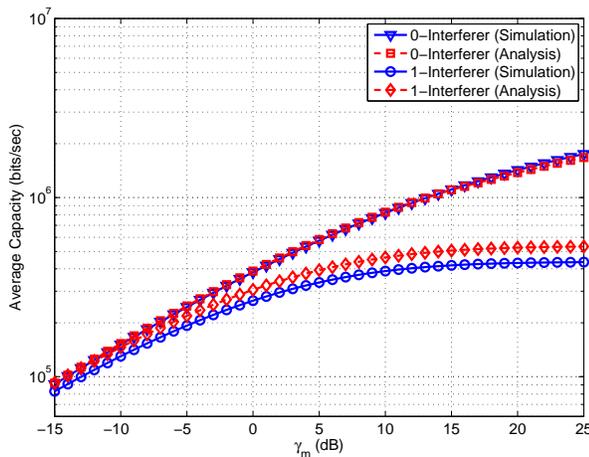


Fig. 4. Average capacity of 1-pair D2D systems with (i) 0-Interferer, (ii) 1-Interferer ($R=5\text{km}$ and $B=200\text{kHz}$).

The average sum capacity, i.e., \bar{C}_T , of different frequency allocation strategies are shown in Fig. 5. The values of the overlapping ratios, i.e., x_1 and x_2 , vary from 0 to 1. The case of $x_1 = x_2 = 0$ represents the separated spectral allocation, while the values $x_1 = x_2 = 1$ refer to the fully overlapped spectral allocation. The results show that at low to moderate values of γ_m , \bar{C}_T can be enhanced by overlapping the frequency bands assigned to each pair of communication devices, which means that a fully-overlapped spectral allocation can be the preferred strategy at low received SNR values. At high values of γ_m , the interference power increases and thus limits the capacity enhancement provided by the overlapped spectral allocation.

V. CONCLUSION

We derived the average capacity of 1-pair and 2-pair D2D communication systems assuming deterministic and random locations of the transmitting and receiving devices. We found

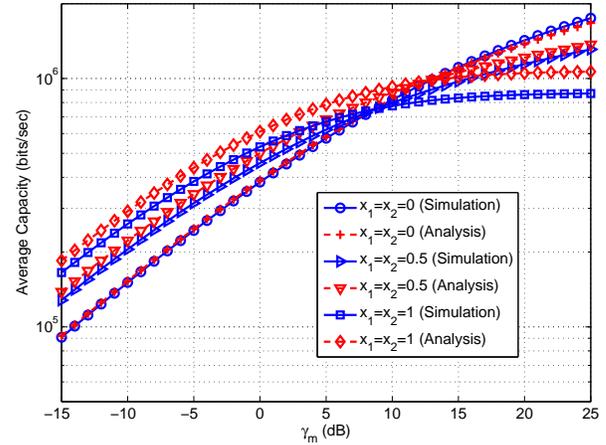


Fig. 5. Average capacity of 2-pair D2D systems ($R=5\text{km}$ and $B=200\text{kHz}$).

that at low and moderate received SNR values, the average system capacity can be enhanced by allowing an overlapping between the frequency spectrum allocations. As future work, it would be interesting to find the optimal overlapping ratios, x_1^{opt} and x_2^{opt} , which maximize the average throughput assigned to each user considering the limits on the transmit power and the number of users which can co-exist in the area under investigation.

VI. ACKNOWLEDGEMENT

This work was performed while A. Assra was an NSERC Visiting Fellow at the Communications Research Centre. This work was supported by Defence Research and Development Canada

REFERENCES

- [1] K.-D. Lee, S. Kim, and B. Yi, "Throughput comparison of random access methods for M2M service over LTE networks," in *Proc. of IEEE GLOBECOM Workshops*, Dec. 2011, pp. 373–377.
- [2] S. Whitehead, "Adopting wireless machine-to-machine technology," *IEE Computing and Control Engineering*, vol. 15, no. 5, pp. 40–46, Oct. 2004.
- [3] C. Inhyok, Y. Shah, A. Schmidt, A. Leicher, and M. Meyerstein, "Trust in M2M communication," *IEEE Vehicular Technology Magazine*, vol. 4, no. 3, pp. 69–75, Sep. 2009.
- [4] H. Min, J. Lee, S. Park, and D. Hong, "Capacity enhancement using an interference limited area for device-to-device uplink underlying cellular networks," *IEEE Trans. Wireless Communications*, vol. 11, no. 12, pp. 3995–4000, Dec. 2011.
- [5] D. Moltchanov, "Distance distributions in random networks," *Elsevier:Ad Hoc Networks*, vol. 10, no. 6, pp. 1146–1166, Aug. 2012.
- [6] I. S. Gradshteyn and I. Ryzhik, *Table of Integrals, Series and Products*, 6th ed. New York: Academic, 2000.
- [7] L. Shampine, "Vectorized adaptive quadrature in MATLAB," *Journal of Computational and Applied Mathematics*, vol. 211, no. 2, pp. 131–140, Feb. 2008.

Performance Analysis of Coordinated Multi-Point with Scheduling and Precoding schemes in the Heterogeneous Network

Bora Kim¹, Saransh Malik¹, Sangmi Moon¹, Daejin Kim¹, Youngil Kim², Kunmin Yeo² and Intae Hwang¹

1) Department of Electronics and Computer Engineering,
Chonnam National University,
Gwangju, Republic of Korea,

bora54321@naver.com, saranshmk@gmail.com, msm0804@naver.com, djinkim@chonnam.ac.kr and hit@chonnam.ac.kr

2) Electronics and Telecommunications Research Institute,
Daejeon, Republic of Korea,
yikim@etri.re.kr and kunmin@etri.re.kr

Abstract—Coordinated Multi-Point (CoMP) is considered as a technology in the 3rd Generation Partnership Project (3GPP) Long Term Evolution-Advanced (LTE-A) system. In this paper, we design and analyze the performance of the Coordinated Scheduling/Beamforming (CS/CB) technique, which is one major category of CoMP. We perform Monte Carlo simulations with a Heterogeneous Network (HetNet) in LTE-A. Simulation results show that the proposed CoMP can improve Signal to Interference plus Noise Ratio (SINR), and the spectrum efficiency of macrocell and picocell users through a graph of the Cumulative Distribution Function (CDF). From these results, we also show significant performance gain when we apply various techniques of scheduling and precoding with the CoMP.

Keywords- CoMP; HetNet; LTE-A; Precoding; Scheduling

I. INTRODUCTION

Mobile communication technology has been constantly evolving to satisfy the communication market, which requires high-speed support for high-capacity and high-quality. The next generation of wireless communication requires a data rate of 100Mbps class for dynamic user equipment (UE), and 1Gbps class for static UE. It aims to enable high-speed network access using information devices anytime, anywhere, through integration of wired and wireless communications and broadcasting [1-2].

Recently, through the prevalence of smart phones, the needs of users for quality anytime, anywhere data services have increased rapidly. Techniques that can support high data rates to users located within cell edges, as well as cell centers, came into demand. A cell center can increase data transmission speed, simply by using the support of additional antenna ports for each cell. But in the case of a cell edge, it is difficult to increase the data rate up to any limitations without cooperation, because this location receives too much interference from adjacent cells. Also frequency reuse techniques being deployed using small cells, such as femto cell or pico cell, within the macro cell area can provide high speed data services to dense user areas. Accordingly, the need for efficient methods to control the interference between transfer points is increasing [3].

Current standards and academic issues, such as how to control the interference from point-to-point transmission naming, identify Coordinated Multi-Point (CoMP), which was selected as a work item for Long Term Evolution-Advanced (LTE-A) Release 11.

In the paper, we will discuss the basic techniques of CoMP, and the effect of CoMP combined with several techniques, through simulation results. As an issue in fourth generation mobile communication, the CoMP environment in this paper is based on LTE-A systems.

The rest of the paper is organized as follows. Section II defines the CoMP basic system model. Section III briefly presents the various scheduling and precoding schemes. After evaluating the performance of the proposed schemes in Section IV, we conclude the paper in Section V.

II. SYSTEM MODEL

The Heterogeneous Network (HetNet) is one of the LTE-A system networks. HetNet complying with the LTE-A Release 11 Scenario 4 is discussed as follows [4].

As shown in Figure 1 (a), it is assumed that a macro cell is divided into three sectors. Three adjacent sectors consist of one cell site, and communication is coordinated. In Figure 1 (b), pico cells exist in a macro cell, and communication is coordinated with the macro cell. Users are uniformly and randomly distributed in each macro cell and pico cell [4-5].

In the conventional cellular system, users receive a signal from each anchor cell with interference signals from adjacent cells. We assume that user M_j is located randomly in C_j and receives signals from three cells (denoted as C_1 , C_2 and C_3) [6]. Assume H_{ij} is the channel gain from C_i to M_j . The received signal Y_1 at M_1 can be expressed as

$$Y_1 = H_{11}W_1X_1 + H_{21}W_2X_2 + H_{31}W_3X_3 + N \quad (1)$$

where X_i is the signal transmitted at C_i , W_i is the precoding matrix at C_i , and N is the additive white Gaussian noise at M_1 .

As shown in the expression, Inter-Cell Interference (ICI) occurs, and the channel capacity is limited, according to the signal from an adjacent cell being considered as an interference signal. The following expression for C_1 is

located within the user's Signal to Interference plus Noise Ratio (SINR).

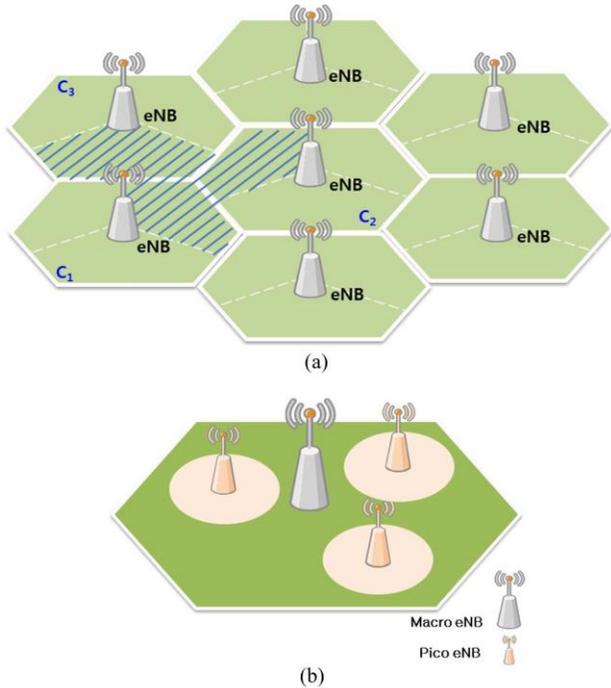


Figure 1. Structure of the HetNet complying with the LTE-A Release 11 Scenario 4: (a) Macro only, (b) Macro and Pico.

If one user receives a signal from its own serving cell, the signals from the other cells affect the user as interference, and then the SINR of M_1 can be expressed as

$$SINR = \frac{||H_{11}W_1||^2 P_1}{||H_{21}W_2||^2 P_2 + ||H_{31}W_3||^2 P_3 + N} \quad (2)$$

CoMP technology can control the interference among neighboring cells through cooperative communication technology, which is a promising 4G cellular standard being developed by the 3rd Generation Partnership Project (3GPP) standardization group. CoMP can improve the coverage, cell-edge capacity and/or system efficiency. For downlink CoMP, two different approaches are being considered, namely, Joint Processing (JP) and Coordinated Scheduling/Beamforming (CS/CB) in 3GPP LTE-A standard [4].

The SINR for a user that is located within C_1 in the CoMP JP scheme can be formulated as

$$SINR = \frac{||H_{11}W_1\sqrt{P_1} + H_{21}W_2\sqrt{P_2} + H_{31}W_3\sqrt{P_3}||^2}{N} \quad (3)$$

From the above expression, you can see that the interference signal from the neighboring cells also provides a useful signal through the cooperation among multiple cells, where only noise interferes with the signal.

The received signal at the user within each cell in the CoMP CS/CB scheme can be expressed as

$$Y_1 = H_{11}W_1X_1 + H_{21}W_2X_2 + H_{31}W_3X_3 + Z_1 \quad (4)$$

$$Y_2 = H_{12}W_1X_1 + H_{22}W_2X_2 + H_{32}W_3X_3 + Z_2 \quad (5)$$

$$Y_3 = H_{13}W_1X_1 + H_{23}W_2X_2 + H_{33}W_3X_3 + Z_3 \quad (6)$$

We measure the SINR for each cell by the above expressions, by the precoding matrix that has the highest SINR applied to each cell. As a result, the interference coming from the surrounding cell is minimized, and the received signal strength can be maximized. In this case, the precoding matrix is expressed as follows.

$$W_1' = \frac{||H_{11}W_1||^2 P_1}{||H_{21}W_2||^2 P_2 + ||H_{31}W_3||^2 P_3 + N} \quad (7)$$

$$W_2' = \frac{||H_{22}W_2||^2 P_2}{||H_{12}W_1||^2 P_1 + ||H_{32}W_3||^2 P_3 + N} \quad (8)$$

$$W_3' = \frac{||H_{33}W_3||^2 P_3}{||H_{13}W_1||^2 P_1 + ||H_{23}W_2||^2 P_2 + N} \quad (9)$$

We perform simulations using CoMP CS/CB, which shows a tradeoff between performance and complexity, and which is the best parameter, compared with other schemes.

III. SCHEDULING AND PRECODING TECHNIQUES

In this section, scheduling and precoding techniques are proposed as a way to improve the performance of CoMP. Each technique is described below.

A. Scheduling

We suggest a scheduling scheme to improve the CoMP performance through using scheduling, as Flexible Frequency Allocation Plan (FFAP) and Coordinated Scheduling (CS) schemes are used.

The FFAP scheme suggests that the whole frequency band be divided into two parts: a CoMP frequency zone for the cell-edge user (CEU)'s transmission, and a single sector frequency band for the cell-center user (CCU)'s transmission, as shown in Figure 2.

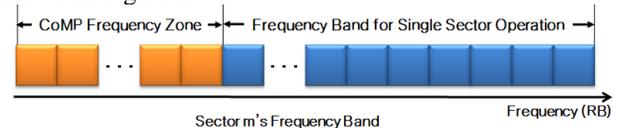


Figure 2. Structure of the CoMP FFAP scheme.

The differentiation between CEUs and CCUs can be made based on the received SINR at the UE.

$$SINR \leq \gamma \quad (10)$$

where γ is a predetermined threshold in dB.

The CS scheme is a way in which all users that exist in one cell can be distinguished into CCU or CEU and then, the sum of highest priority of CCU for each sector in one cell and the highest priority of CEU in one cell are compared. The priority is based on the received SINR.

If the CEU's priority is higher, the system will be calculated based on the CoMP mode. Otherwise, CCUs are calculated based on the Non-CoMP mode. In this mode, we send a signal to users using the common transmission

method. The flow chart for the CS scheme is shown in Figure 3 [7].

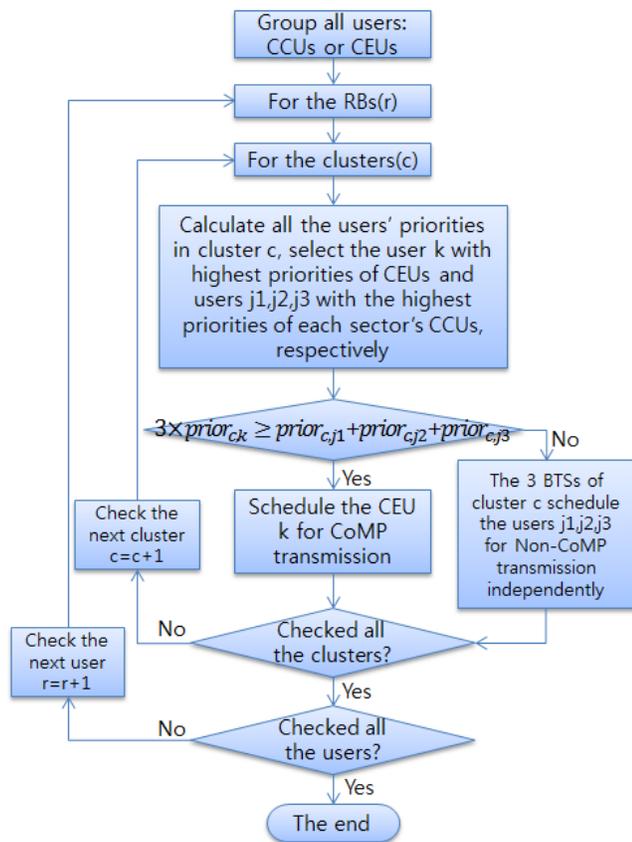


Figure 3. Flow Chart of the CoMP CS scheme.

B. Precoding

A precoding scheme is another way to improve the performance of CoMP. Precoding is a technique to increase the SINR and spectrum efficiency of a signal, by multiplying a specific matrix with the channel in the transmitter. Singular Value Decomposition (SVD), Polar Decomposition (PD), Tomlinson Harashima Precoding (THP) and QR Decomposition (QRD) schemes are used as precoding techniques.

Using the SVD scheme, the channel is separated in parallel, by multiplying an orthogonal matrix U and V . The basic equation for the channel matrix H and the precoding matrix W are expressed as

$$\begin{aligned} H &= U \Sigma V^H \\ W &= V \end{aligned} \quad (11)$$

Using the PD scheme, the channel is separated, based on the SVD scheme. The channel matrix H and the precoding matrix W are expressed as follows.

$$\begin{aligned} H &= Q S P^H = A U^H \\ A &= Q S Q^H \\ W &= U = P Q^H \end{aligned} \quad (12)$$

The THP scheme is a nonlinear precoding based on Costa's "writing on dirty paper result" information theory [8].

In other words, if the transmitter knows the interference signal beforehand, it can get the same channel capacity as the non-interference channel condition. THP precoding additionally uses modulo operation, which is symmetric nonlinear operation based on Costa's precoding. The modulo operation is given as

$$\text{mod}_A(x) = x - A \left\lfloor \left(x - \frac{A}{2} \right) / A \right\rfloor \quad (13)$$

A block diagram of this technique is shown in Figure 4.

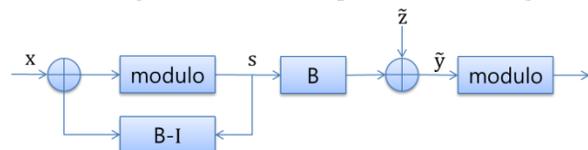


Figure 4. Block Diagram of THP precoding scheme.

QRD scheme creates the precoding matrix through QR decomposition, which factorizes the channel matrix. The basic equations for the channel matrix H and the precoding matrix W are given as

$$\begin{aligned} H &= R Q^H \\ W &= Q F \end{aligned} \quad (14)$$

IV. SIMULATION PARAMETERS AND RESULTS

We confirm the performance gain according to the use of CoMP technique in heterogeneous network scenario in LTE-A system. In addition, the performance gain for the scheduling and precoding with CoMP techniques will be confirmed. We use Cumulative Distribution Function (CDF) to analyze the SINR and spectrum efficiency.

A. Simulation Environments

The simulation environment follows the 3GPP LTE-A standard. We perform system-level simulations using Matlab, based on the parameters of Table 1 [9]. We assume that three pico cells exist within one macro cell, and are located in the edge site of the macro cell.

TABLE I. SIMULATION PARAMETERS

Parameter	Value
Carrier Frequency	2 GHz
Bandwidth	20 MHz
Cellular Structure	Hexagonal grid, 2-tiers, 7 cell sites, 3 sectors per site, 3 pico cells per site
No. of MS per cell	100 MSs
Antenna Configuration	BS: 2, MS: 2
BS Max TX Power	49 dBm – 20 MHz Carrier
Cell Radius (R) =	$1732/\sqrt{3} \approx 1000\text{m} \rightarrow$ Macro cell
Path Loss Model	Macro cell: $L = 128.1 + 37.6\log_{10}(R)$, R in km Pico cell: $L = 140.7 + 37.6\log_{10}(R)$, R in km
Shadow Std. Deviation	8 dB
MS Noise Level	174 dBm/Hz
UE Noise Figure	9 dB
NodeB Noise Figure	5 dB
Correlation distance of Shadowing	50 m
Shadowing correlation between cells/sectors	0.5 / 1.0
Minimum distance	Macro-Pico: >75m, Pico-Pico: >40m, Macro-UE: >35m, Pico-UE: >10m
BS antenna gain plus cable loss	14 dBi for micro, macro cell case
Antenna Pattern	70 degree sectored beam \approx 70 degree, $A_m=20\text{dB}$
Scheduling scheme	FFAP, CS
Precoding scheme	SVD, PD, THP, QRD

B. Simulation Results

Using Non-CoMP scheme as a baseline for evaluating the performance of the proposed scheme, we send a signal to users.

1) CoMP simulation with Scheduling

The CDF graphs of the SINR are shown when we apply two kinds of scheduling techniques to CoMP in Figures 5 and 6, respectively. Figure 5 is for a Macro UE, which is for a macro cell when communication is coordinated among macro cells. Figure 6 is for a Pico UE, which is for a pico cell when communication is coordinated among macro cell and pico cell.

From these figures, we can find that using the scheduling (FFAP or CS) with CoMP brings a performance gain, compared with the non-CoMP. Especially the CS with CoMP can be seen to provide the highest contribution towards improving the performance.

The CDF graphs of spectrum efficiency are shown when we apply two kinds of scheduling techniques to CoMP as in Figures 7 and 8, respectively. Figure 7 is for Macro UE, when communication is coordinated among macro cells.

Figure 8 is for Pico UE when communication is coordinated among macro cell and pico cell.

From these figures, we can find that using the scheduling with CoMP brings a performance gain, compared to with non-CoMP. CS with CoMP can be seen to offer the highest contribution to improve the performance.

Also, in the case of Pico UE, the overall performance is better than in the case of Macro UE, because the pico cell is located at the edge of the macro cell.

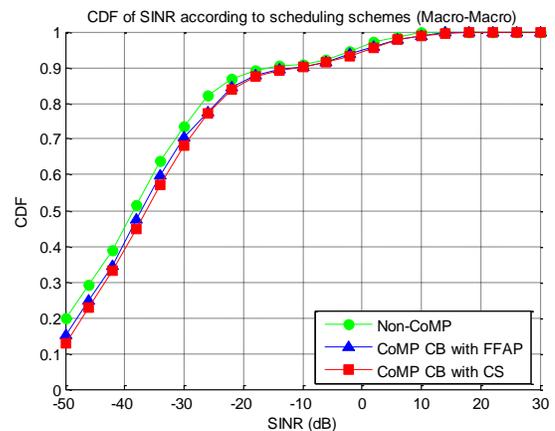


Figure 5. CDF of Macro UE SINR according to scheduling schemes.

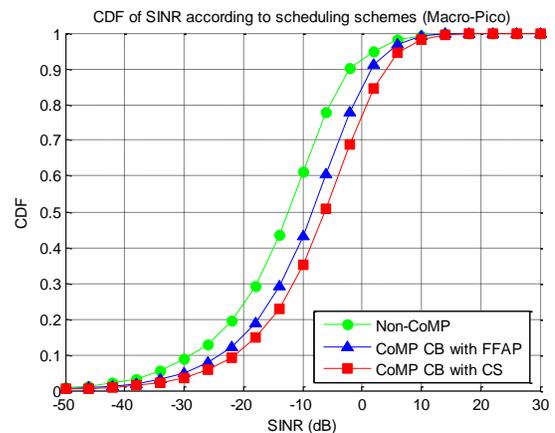


Figure 6. CDF of Pico UE SINR according to scheduling schemes.

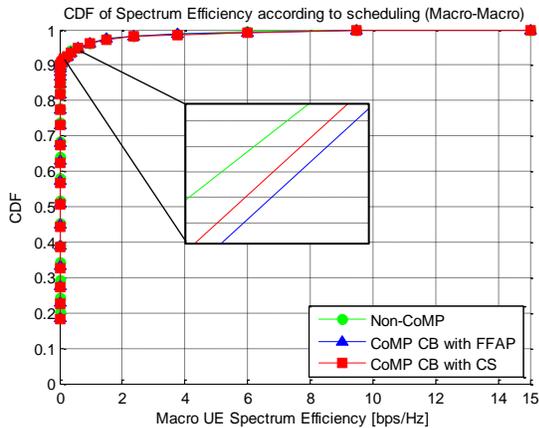


Figure 7. CDF of Macro UE Spectrum Efficiency according to scheduling schemes.

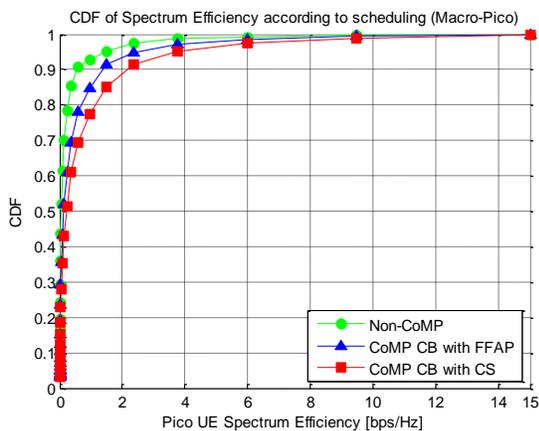


Figure 8. CDF of Pico UE Spectrum Efficiency according to scheduling schemes.

2) CoMP simulation with Precoding

Here, the CDF graphs of SINR are shown when we apply various precoding techniques to CoMP, in Figures 9 and 10, respectively. Figure 9 is for Macro UE, when communication is coordinated among macro cells. Figure 10 is for Pico UE, when communication is coordinated among macro cell and pico cell.

From these figures, we can find that using SVD precoding with CoMP brings a performance gain, compared to PD, THP and QRD schemes.

The CDF graphs of spectrum efficiency are shown when we apply various precoding techniques to CoMP, in Figures 11 and 12, respectively. Figure 11 is for Macro UE, when communication is coordinated among macro cells. Figure 12 is for Pico UE when communication is coordinated among macro cell and pico cell.

From these figures, we can find that using SVD precoding with CoMP brings a performance gain, compared to PD, THP and QRD schemes.

Also, the performance of the Pico UE is better than the Macro UE case.

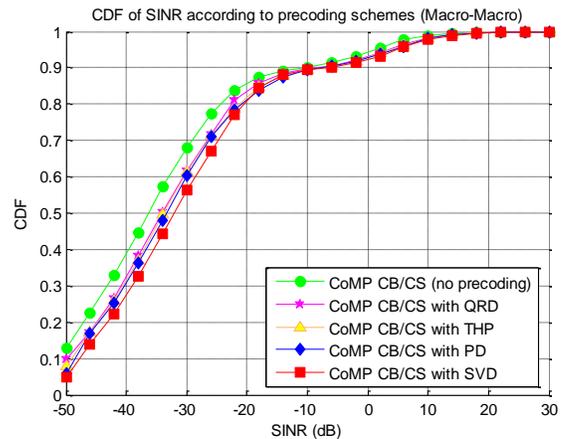


Figure 9. CDF of Macro UE SINR according to precoding schemes.

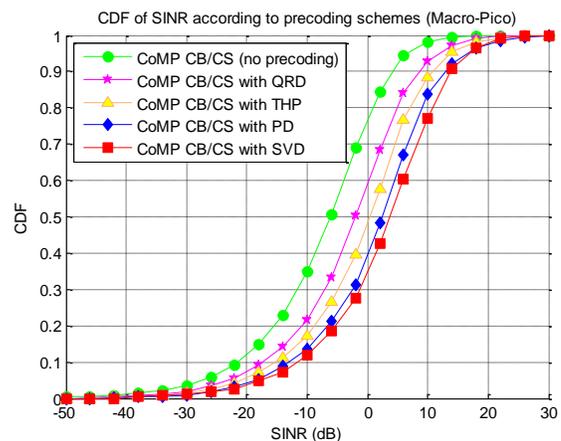


Figure 10. CDF of Pico UE SINR according to precoding schemes.

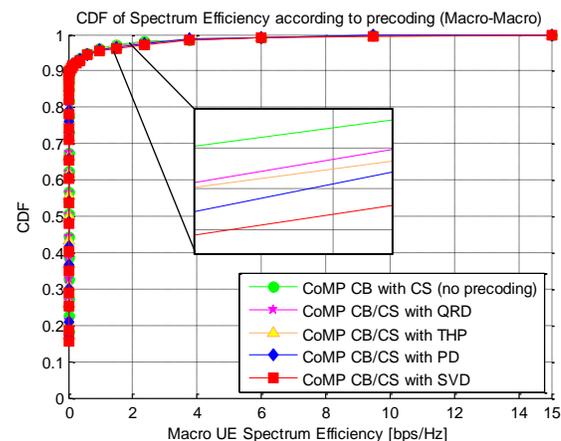


Figure 11. CDF of Macro UE Spectrum Efficiency according to precoding schemes.

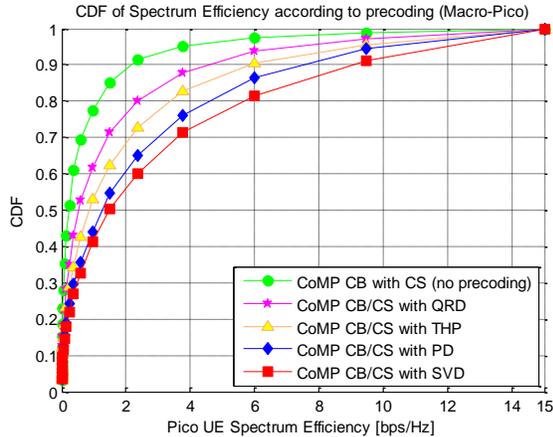


Figure 12. CDF of Pico UE Spectrum Efficiency according to precoding schemes.

V. CONCLUSIONS

We demonstrate a performance analysis of CoMP technique, for multi-point cooperation among users for next-generation cellular systems. System-level simulation results are based on HetNet in the LTE-A system.

We compare and analyze the performance according to the CoMP technique applied, in two cases; one is among only macro cells, the other is among macro cell and pico cell. In both cases, we can get an improved performance when using CoMP technique. It is also shown that CoMP techniques combined with scheduling and precoding achieved further improvement of performance.

As a result, we confirm that CoMP technology can be applied to HetNet, which has pico cells within a macro cell in the LTE-A system.

ACKNOWLEDGMENT

This research was supported by the MSIP (Ministry of Science, ICT & Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (NIPA-2013-H0301-13-3005) supervised by the NIPA (National IT Industry Promotion Agency). This study was financially supported by Chonnam National University, 2012. This research is supported by Korea Research Council for Industrial Science and Technology under B551179-12-07-00.

REFERENCES

- [1] 3GPP TR 36.913, "Requirements for Further Advancements for E-UTRA (LTE-Advanced)", 3rd Generation Partnership Project.
- [2] 3GPP, "IMT Advanced Technical Requirements—An India Perspective", REV- 080050, Shenzhen, China, April 7-8, 2008, CEWiT.
- [3] Samsung, "Standard Trends and Performance Analysis of Cooperative Communication on a Point-to-Point Transmission based on LTE-A", TTA Journal, Vol.139, pp.94-99, 2012.
- [4] Yong-Ping Zhang, "Joint Transmission for LTE-Advanced Systems with Non-Full Buffer Traffic", Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th, pp.1-6, May 2012.
- [5] 3GPP TR 36.819, "Coordinated multi-point operation for LTE physical layer aspects (Release 11)", 3rd Generation Partnership Project, Sept. 2011.
- [6] Young-Han Nam, "Cooperative Communication Technologies for LTE-Advanced", ICASSP, 2010 IEEE, pp.5610-5613, March 2010.
- [7] Jing LIU, "A Novel Transmission Scheme and Scheduling Algorithm for CoMP-SU-MIMO in LTE-A System", 2010 IEEE 71st, pp.1-5, May 2010.
- [8] M.H.M. Costa, "Writing on dirty paper", IEEE Trans. Info. Theory, vol. 29, no. 3, pp.439-441, May 1983.
- [9] 3GPP TR 36.814, "Further advancements for E-UTRA physical layer aspects(Release 9)", 3rd Generation Partnership Project.

A New Approach for Multistage Analog Complex Filter Design

Daniele Grasso, Giuseppe Avellone

Automotive Product Group - Radio Frequency Competence Center

STMicroelectronics

Catania, Italy

daniele.grasso@st.com, giuseppe.avellone@st.com.

Abstract— In this paper, a design approach for g_m -C complex filter for intermediate frequency (IF) is formalized. It is based on decoupled first order g_m -C sections, each one centered at a different frequency with respect to the others that can be stacked in series to get the final band pass filter response, with order equal to the number of stages. A simple case with two stages is presented to show selectivity improvement when difference between the two center-bands increase, with ripple increase as drawback. An optimal setup with good selectivity increment and still zero ripple (flat pass-band) is also shown. Then, the approach is extended to third order. The approach has been used for IF filtering in STMicroelectronics GNSS receivers, but it is applicable to other wireless receivers.

Keywords- g_m -C filter, complex filtering, low-IF receivers.

I. INTRODUCTION

The complex intermediate frequency (IF) filters have been proposed for radio frequency (RF) front-end (especially for the low-IF ones) for their characteristic of rejecting both out-of-band and image signals, due to their asymmetrical transferring function [1]. Several implementations have been disclosed in technical literature, such as [1]-[8], mainly based on active circuits, while related impairments has been also analyzed in [9] and the references therein. Amongst these implementations, it has been chosen the one based on decoupled first order stacked stages with operational trans-conductor amplifier (OTA) and frequency shift of low-pass prototype. This choice presents lower current consumption, good response at high frequency and simple, reconfigurable, modular design [10].

In the state-of-the-art implementation of g_m -C multistage complex IF filter, each stage is centered on the same frequency with the same bandwidth, as in [1], [4] and [15] and the references therein. In this contribution, we formalize an approach considering different center frequency for each stage (and eventually different bandwidth). These center frequencies will be placed symmetrically around the center frequency of the final filter. This approach achieves better performances, especially in terms of out-of-band attenuation, image rejection and flatness in group delay response with lower current consumption (as shown in [10]). Moreover, it allows more freedom in shaping the filter frequency response for a given filter order. We adopted it in designing an IF filter for Global Navigation Satellite Systems (GNSS) receivers described in [10]-[12].

Section II describes the complex IF filter architecture, as in [10]; Section III presents the proposed approach considering the second order complex filter case, providing equations for filter design, and comparing the behavior between coincident center and non-coincident center cases. Section IV gives a brief extension to third order, and the main conclusions are drawn in Section V.

II. COMPLEX FILTER ARCHITECTURE

Filter architecture is based on decoupled first order sections stacked in series to get the final band-pass filter order. Each single stage band-pass filter response is a frequency shifted version of a low-pass one, designed using active components with trans-conductors and capacitors (g_m -C or OTA-C). Fig. 1 shows the architecture used in [10]-[12] for the single-stage, where the first order low-pass frequency response is set by the OTA g_{m1} and capacitor C values, with bandwidth given by: $f_{LP} = (g_{m1}/2\pi C)$. The structure with the g_{m2} OTA is a gyrator that creates a feedback between in-phase (I) and quadrature (Q) branches performing the frequency shift of the low-pass prototype response, obtaining a band-pass centered at a frequency defined through g_{m2} and C values according to the formula: $f_{center} = (g_{m2}/2\pi C)$. The transfer function can be obtained just putting the translation

$$j\omega \rightarrow j\omega - j\omega_{center} \quad (1)$$

in the transfer function of the low pass version, [1],[13]-[14]:

$$H_{LP}(j\omega) = \frac{G}{1 + j\omega/\omega_{LP}} \quad (2)$$

resulting in

$$H_{BP}(j\omega) = \frac{G}{1 + j(\omega - \omega_{center})/\omega_{LP}} \quad (3)$$

with $G = (g_{m3}/g_{m1})$, where the third OTA, g_{m3} , is used both to decouple each stage with the previous one and to provide a gain (as a free parameter).

In [10], we performed several schematic level simulations for defining design parameters values as a better trade-off amongst requirements of bandwidth, in-band group-delay variation, image and out-of-band rejection. We considered a 3rd or 4th order filter comparing in detail two cases:

- stages with same center frequency and bandwidth;
- stages with different center frequencies, same bandwidth.

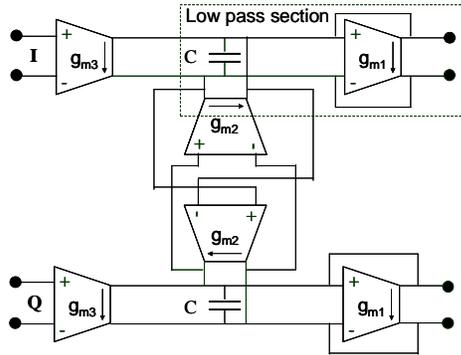


Figure 1. Complex filter single-stage basic architecture.

It was easily recognizable as the second case presents better performances, especially in terms of out-of-band attenuation, image rejection and flatness in group delay response. In fact, in [10], we have observed that, when all stages share the same center frequency, if wider bandwidth are needed with low-order filters, the properties of both out-of-band and image rejection rapidly becomes not acceptable. On the contrary, the different center frequency configuration can achieve wider bandwidth while maintaining low-order IF filter with good rejection for both out-of-band and image signals. The drawback of this latter approach is that the implementation needs different gyrators, one for each stage, instead of only one type for all the stages. Once fixed the stages main parameters (f_{center} , f_{LP} and G), then g_{m1} , g_{m2} and g_{m3} are fixed from the previous formulas and all the transconductors can be designed. For the details of the circuit design and measurement results, see [10]-[12].

III. COMPLEX FILTER FORMALIZATION

To formalize what has been observed both by simulation and by measurements on the implemented IF filters, it is better to take into account a simplified version with only two stages, obtaining a second order filter (but the approach and the conclusions can be straightforwardly extended to higher order complex analog filters). In this case, the transfer function for two equal stages is

$$H_C(j\omega) = \frac{G^2}{\left(1 + j \frac{\omega - \omega_{center}}{\omega_{LP}}\right)^2}. \quad (4)$$

Meanwhile, for two stages with centers respectively in $\omega_{center1}$ and $\omega_{center2}$ is

$$\begin{aligned} H_{NC}(j\omega) &= H_{BP1}(j\omega) \cdot H_{BP2}(j\omega) = \\ &= \frac{G}{1 + j \frac{\omega - \omega_{center1}}{\omega_{LP}}} \cdot \frac{G}{1 + j \frac{\omega - \omega_{center2}}{\omega_{LP}}}. \end{aligned} \quad (5)$$

For a final pass-band filter centered at ω_{center} and band equal to $2\omega_B$, $\omega_{center1}$ and $\omega_{center2}$ must be chosen symmetrical to ω_{center} , that is $\omega_{center} = (\omega_{center1} + \omega_{center2})/2$. Normalizing all the frequencies with respect to ω_B , (4) and (5) become:

$$H_C(j\omega_n) = \frac{G^2}{\left(1 + j \frac{\omega_n - \omega_c}{\omega_o}\right)^2} \quad (6)$$

$$H_{NC}(j\omega_n) = \frac{G^2}{\left(1 + j \frac{\omega_n - \omega_{c1}}{\omega_o}\right) \cdot \left(1 + j \frac{\omega_n - \omega_{c2}}{\omega_o}\right)} \quad (7)$$

where $\omega_n = \omega/\omega_B$, $\omega_c = \omega_{center}/\omega_B$, $\omega_o = \omega_{LP}/\omega_B$, $\omega_{c1} = \omega_{center1}/\omega_B$, $\omega_{c2} = \omega_{center2}/\omega_B$, and the normalized filter bandwidth is 2. In the same way the upper 3dB corner of the band is $\omega_c + 1$ and the lower is $\omega_c - 1$.

A. Coincident Center Frequency Case

Without losing general validity, we set $|H_{coinc}(j\omega_c)| = 1$, this lead to $G=1$. Now, we want calculate the single stage low-pass normalized band ω_b needed for a final normalized bandwidth equal to 2 of the two stages filter. This happens when the magnitude at the normalized upper band corner $\omega_c + 1$ is 3dB lower than the center-band one, that is:

$$\frac{|H_C[j(\omega_c + 1)]|}{|H_C[j(\omega_c)]|} = \frac{1}{\sqrt{2}}; \quad (8)$$

considering (6), condition (8) is verified for $\omega_b = 1.554$.

This means that in order to have a two coincident stages band-pass filter with bandwidth of $2\omega_B$ we need a low pass bandwidth of $1.554 \cdot \omega_b$ for each single stage.

B. Different Center Frequency Case

Setting $\omega_d = (\omega_{c2} - \omega_{c1})/2$, then $\omega_{c1} = \omega_c - \omega_d$ and $\omega_{c2} = \omega_c + \omega_d$, and equation (7) may be written as:

$$\begin{aligned} |H_{NC}(j\omega_n)|^2 &= \\ &= \frac{G^4 \omega_o^4}{(\omega_n - \omega_c)^4 + 2(\omega_o^2 - \omega_d^2)(\omega_n - \omega_c)^2 + (\omega_d^2 + \omega_o^2)^2} \end{aligned} \quad (9)$$

Analyzing this function it is possible to distinguish two cases:

- if $\omega_d < \omega_o$ then the filter frequency response has only one maximum (MAX) at $\omega_n = \omega_c$ equal to

$$MAX = |H_{NC}(j\omega_c)| = G^2 \omega_o^2 / (\omega_d^2 + \omega_o^2) \quad (10)$$

- if $\omega_d > \omega_o$, then the filter frequency response has one minimum (min) at $\omega_n = \omega_c$ equal to

$$\min = |H_{NC}(j\omega_c)| = G^2 \omega_o^2 / (\omega_d^2 + \omega_o^2) \quad (11)$$

and two equal maximum values at

$$\omega_n = \omega_{MAX} = \omega_c \pm \sqrt{\omega_d^2 - \omega_o^2} \quad (12)$$

both with value

$$\text{MAX} = \left| H_C \left[j(\omega_c \pm \sqrt{\omega_d^2 - \omega_o^2}) \right] \right| = G^2 \omega_o / 2\omega_d . \quad (13)$$

Considering the first case, we want to calculate, also for non-coincident center, the single stage low-pass normalized band needed for a final normalized bandwidth equal to 2 for the two-stage filter. In general, we obtain a different value respect the one calculated in section A; for this reason, let put ω'_o instead of ω_o to distinguish the new variable. Further, we set $|H_{NC}(j\omega_c)| = 1$ obtaining from (9)

$$G^2 = (\omega_d^2 + \omega_o'^2) / \omega_o'^2 \quad (14)$$

and

$$\begin{aligned} |H_{NC}(j\omega_n)| &= \\ &= \frac{\omega_d^2 + \omega_o'^2}{\sqrt{(\omega_n - \omega_c)^4 + 2(\omega_o'^2 - \omega_d^2)(\omega_n - \omega_c)^2 + (\omega_d^2 + \omega_o'^2)^2}} . \end{aligned} \quad (15)$$

Then, setting (8) for 3dB cut off at $\omega_c + 1$ we obtain:

$$\omega_o'^2 = 1 - \omega_d^2 \pm \sqrt{\omega_d^4 - 4\omega_d^2 + 2} . \quad (16)$$

These are valid when $0 \leq \omega_d \leq \sqrt{2 - \sqrt{2}}$; and $\omega_d \geq \sqrt{2 + \sqrt{2}}$; the plot of (16), '+' case, results as in Fig. 2: all the negative values are not acceptable, so we can discard the values for $\omega_d \geq \sqrt{2 + \sqrt{2}}$. Furthermore, we must discard the solutions not compliant with $\omega_d < \omega'_o$; plotting the square root of (16), it is easy to verify that only the solutions for $0 \leq \omega_d \leq 0.7596$ can be considered. Other solutions of (16) can be found in the '-' case, but the condition $\omega_d < \omega'_o$ is never satisfied and cannot be considered. So, final solution in $\omega_d < \omega'_o$ case is

$$\omega'_o = \sqrt{1 - \omega_d^2 + \sqrt{\omega_d^4 - 4\omega_d^2 + 2}} \quad (17)$$

true for $0 \leq \omega_d \leq 0.7596$. Fig. 3 shows the needed low-pass bandwidth of each stage versus frequency distance of the responses of the two stages. Observing this plot, for $\omega_d = 0$ we should have the previous coincident center case, and in fact we found the previous value $\omega'_o = \omega_o = 1.554$; when ω_d increases then ω'_o decreases: this means that each stage can have a narrower bandwidth for obtaining the desired resulting bandwidth of the whole two-stage filter, that is a normalized value equal to 2. This lead to higher slope of the out-of-band response, that is higher out-of-band attenuation with the same bandwidth, so it is possible to obtain a more selective filter. Just to make an example, choosing $\omega_d = 0.7$, we obtain $\omega'_o = 1.0194$, less than $\omega_o = 1.554$ in the coincident center case also satisfying condition: $\omega_d < \omega'_o$.

In order to evaluate the selectivity improvement, the out-of-band attenuation can be calculated anywhere for both cases, e.g. in $\omega_c + 10$. For not coincident case from (15), substituting $\omega_d = 0.7$ and $\omega'_o = 1.0194$, we have

$$A_{NC}(\omega_c + 10) = 36.4 \text{dB} \quad (18)$$

for coincident center case, from (6), (8) and $G=1$, we have

$$A_C(\omega_c + 10) = 32.5 \text{dB} \quad (19)$$

with 3.9dB of more attenuation in the not coincident case, resulting in more selectivity as both have same normalized bandwidth equal to 2.

Choosing another value of out-of-band frequency, nearer the band of the filter, for example $\omega_c + 3$, we have

$$A_{NC}(3) - A_C(3) = 16.0 - 13.5 = 2.5 \text{dB} \quad (20)$$

and that confirms the selectivity enhancement.

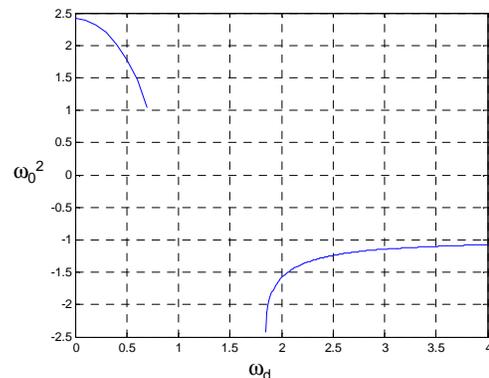
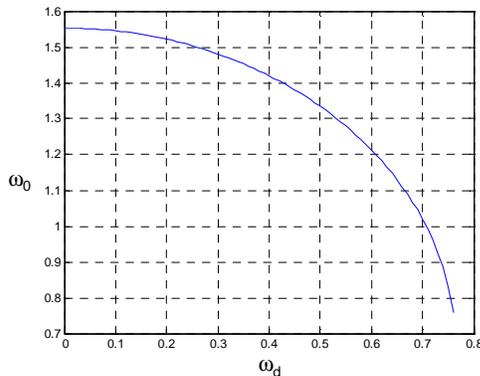


Figure 2. Plot of Eq. 16, plus case, for the allowed values of ω_d .


 Figure 3. Plot of Eq. 17, for the allowed values of ω_d .

Now, let's consider $\omega_d > \omega'_o$, still with not-coincident center, and set the values of the two peaks equal to MAX=1 without losing general validity, so from (13) is

$$G^2 = 2\omega_d/\omega'_o \quad (21),$$

and (9) becomes

$$\begin{aligned} |H_{NC}(j\omega_n)|^2 &= \\ &= \frac{4\omega_d^2\omega'^2_o}{(\omega_n - \omega_c)^4 + 2(\omega'^2_o - \omega_d^2)(\omega_n - \omega_c)^2 + (\omega_d^2 + \omega'^2_o)^2} \end{aligned} \quad (22)$$

Then, once again, we impose a normalized bandwidth of the resulting filter equal to 2; this means that in $\omega_c \pm 1$ the response must be $1/\sqrt{2}$ times the maximum. From (22) imposing the condition (8) we obtain:

$$\omega'_o = \sqrt{3\omega_d^2 - 1 \pm 2\omega_d\sqrt{2\omega_d^2 - 1}} \quad (23)$$

valid for $\omega_d \geq 1/\sqrt{2}$. The plot of (23) results as in Fig. 4: as easily recognizable the '+' case (dashed line) does not match the condition $\omega_d > \omega'_o$, so can be discarded. Observing the '-' case (continuous line) is possible to split the curve in two parts: the first one, on which we focus our interest, with decreasing value of ω'_o for $1/\sqrt{2} \leq \omega_d < 1$, up to $\omega_d = 1$, that means when the center frequency separation of the two stage is equal to the bandwidth of the whole filter. In this point the ω'_o required is zero, so it is only a theoretical limit. The second part for $\omega_d > 1$, that is center frequency separation of the two stages greater than the bandwidth of the whole filter, lacks of interest because the response of each stage is completely separated from the other, so not useful to obtain an appropriate overall pass-band response.

Returning to the range $1/\sqrt{2} \leq \omega_d < 1$ in (23), as noted before, the decreasing characteristic of ω'_o suggests more

selective behavior for higher ω_d ; further, the presence of two peaks, as formalized in (11), (12) and (13), means presence of in-band ripple: when ω_d increases, the minimum in ω_c becomes lower and the ripple increases. It is important to check when ripple is too big.

If the difference between maximum and minimum becomes greater than 3dB then we can consider the response completely separated into two distinct lobes, non-acceptable as band-pass characteristic. The limit of the center frequencies separation ω_d that still gives us a unique filter band is given for ripple equal to 3dB, that is:

$$F(\omega_d) = \frac{|H_{NC}[j(\omega_c)]|}{|H_{NC}[j(\omega_{MAX})]|} = \frac{1}{\sqrt{2}} \quad (24)$$

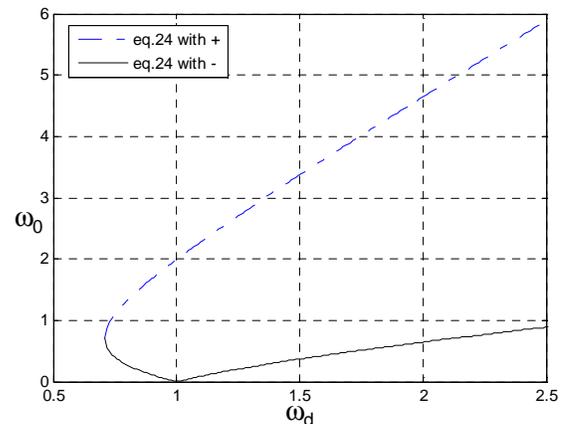
Considering the (21) and the (22) then is

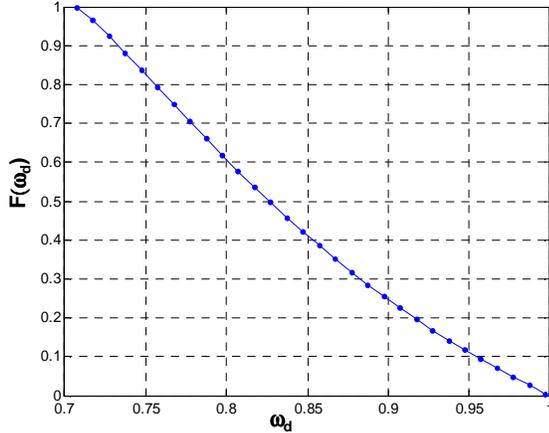
$$F(\omega_d) = 2\omega_d\omega'_o/(\omega'^2_o + \omega_d^2) \quad (25)$$

where ω'_o has the value for a final band equal to 2, given by (23) for $1/\sqrt{2} \leq \omega_d < 1$; substituting it we have:

$$F(\omega_d) = \frac{2\omega_d(\omega_d - \sqrt{2\omega_d^2 - 1})}{4\omega_d^2 - 2\omega_d\sqrt{2\omega_d^2 - 1} - 1} \quad (26)$$

Plotting (26) (see Fig. 5) the condition in (24) is satisfied for $\omega_d = 0.77679$; this value gives the maximum acceptable ripple. Observing the same plot is fine to note that for $\omega_d = 1/\sqrt{2}$ we found $F(\omega_d)=1$. This is the case with no ripple, because the minimum is equal to the two maximum values, and coincides with the case previously studied for $\omega_d < \omega'_o$ having only one maximum, where we observed that higher ω_d brings to higher selective filter.


 Figure 4. Plot of Eq. 23, for the allowed values of ω_d .


 Figure 5. Plot of Eq. 26, for the allowed values of ω_d .

So, choosing $\omega_d = 1/\sqrt{2}$ gives the resulting filter with higher selectivity and no ripple; if we can accept ripple, we can choose a greater value of ω_d in the range $1/\sqrt{2} \leq \omega_d \leq 0.77679$. In particular we can choose values nearer to 0.77679 for better selectivity despite of worse ripple. If for example we choose $\omega_d = 0.77$ from (23), ' case, we have $\omega'_0 = 0.3389$, very low compared with the one of the best values we can have for $\omega_d < \omega'_0$, for example for $\omega_d = 0.7$ from (16) we already found $\omega'_0 = 1.0194 \gg 0.3389$; in this way, with so low ω'_0 , we have much more selectivity. If we compare it with the coincident center case, the difference is even bigger (from (9) it was $1.554 \gg 0.3389$). Evaluating the attenuation in the same out-of-band frequencies used in the examples in all the previous cases we obtain the values in Table I.

Another interesting example can be considered for the "optimum" setup that gives maximum out-of-band attenuation with still zero ripple. As already discussed, it is for $\omega_d = 1/\sqrt{2}$ in $\omega_d \geq \omega'_0$ case, and for about $\omega_d = 0.7596$ in $\omega_d < \omega'_0$ case. The values of ω'_0 are $1/\sqrt{2}$ and 0.7622 respectively, while the related attenuations for ω_c+3 and ω_c+10 are in Table II. In both cases, the maximum attenuation with zero ripple is when the two-stages frequency separation, ω_d , is equal to the single stage bandwidth ω'_0 .

TABLE I. COMPARISON OF OUT-OF-BAND ATTENUATIONS AMONGST NON COINCIDENT FREQUENCIES CASES AND COINCIDENT FREQUENCIES

	Ripple $\omega_d=0.77$	No ripple $\omega_d=0.7$	Equal Freq.
ω_c+3	$A_{NC} = 24.3$ dB	$A_{NC} = 16.0$ dB	$A_C = 13.5$ dB
ω_c+10	$A_{NC} = 45.6$ dB	$A_{NC} = 36.4$ dB	$A_C = 32.5$ dB

TABLE II. MAXIMUM OUT-OF-BAND ATTENUATION WITH NO RIPPLE

	No ripple $\omega_d \geq \omega'_0$	No ripple $\omega_d < \omega'_0$
ω_c+3	$A_{NC} = 19.1$ dB	$A_{NC} = 17.9$ dB
ω_c+10	$A_{NC} = 40.0$ dB	$A_{NC} = 38.7$ dB

Lastly, maximum selectivity, with 3dB ripple, is for $\omega_d = 0.77679$, obtaining $A_{NC}(\omega_c+3) = 24.6$ dB and $A_{NC}(\omega_c+10) = 46.0$ dB.

IV. THIRD ORDER CASE

In this section, we will give a brief look to the third order case. For three stages with normalized centers, respectively, in ω_{c1} , ω_{c2} and ω_c the transferring function is

$$H_{NC}(j\omega_n) = \frac{K}{\left(1 + j \frac{\omega_n - \omega_{c1}}{\omega_o}\right) \cdot \left(1 + j \frac{\omega_n - \omega_{c2}}{\omega_o}\right) \left(1 + j \frac{\omega_n - \omega_c}{\omega_o}\right)} \quad (27)$$

where $K = G_1 G_2 G_3$ is the filter gain, $\omega_n = \omega/\omega_B$, $\omega_c = \omega_{center}/\omega_B$, $\omega_{c1} = \omega_{center1}/\omega_B$, $\omega_{c2} = \omega_{center2}/\omega_B$, $\omega_o = \omega_{LP}/\omega_B$ and the normalized filter bandwidth is 2. Moreover, ω_{c1} and ω_{c2} must be chosen symmetrical to ω_c , that is $\omega_c = (\omega_{c1} + \omega_{c2})/2$. Setting $\omega_d = (\omega_{c2} - \omega_{c1})/2$, then $\omega_{c1} = \omega_c - \omega_d$ and $\omega_{c2} = \omega_c + \omega_d$, equation (27) may be written as

$$H_{NC}(j\omega_n) = \frac{K}{\left(1 + j \frac{\omega_n - \omega_c + \omega_d}{\omega_o}\right) \cdot \left(1 + j \frac{\omega_n - \omega_c - \omega_d}{\omega_o}\right) \left(1 + j \frac{\omega_n - \omega_c}{\omega_o}\right)} \quad (28)$$

Setting $x = (\omega_n - \omega_c)$ and developing denominator in (28) we obtain:

$$H_{NC}(j\omega_n) = \frac{K\omega_o^3}{[\omega_o(\omega_o^2 - 3x^2 + \omega_d^2) + jx(3\omega_o^2 - x^2 + \omega_d^2)]} \quad (29)$$

The square modulus becomes:

$$|H_{NC}(j\omega_n)|^2 = \frac{K^2\omega_o^6}{[\omega_o^2(\omega_o^2 - 3x^2 + \omega_d^2)^2 + x^2(3\omega_o^2 - x^2 + \omega_d^2)^2]} \quad (30)$$

Maximum and minimum values will be related to minimum and maximum values of denominator in (30). So, let us consider the derivate of the denominator with respect to x equating it to zero (defining for which x there could be maximum and minimum values):

$$6x^5 + 4x^3(3\omega_o^2 - 2\omega_d^2) + 2x(3\omega_o^2 + \omega_d^2) = 0 \quad (31)$$

The first solution is $x=0 \Rightarrow \omega_n = \omega_c$; other solutions from:

$$x^4 + \frac{2}{3}x^2(3\omega_o^2 - 2\omega_d^2) + \frac{1}{3}(3\omega_o^2 + \omega_d^2) = 0 \quad (32)$$

setting $x^2=y$, solution of (32) has the following expressions:

$$y_{1,2} = -\frac{1}{3}(3\omega_o^2 - 2\omega_d^2) \pm \frac{1}{3}\omega_d\sqrt{\omega_d^2 - 12\omega_o^2} \quad (33)$$

Solutions (33) are real for $\omega_d^2 \geq 12\omega_o^2$. If this last holds, these are acceptable solutions for x if and only if $y_{1,2} \geq 0$. In this latter case, both solutions are non-negative; so there are two couples of frequencies to be considered:

$$\begin{aligned}\omega_{n1} &= \omega_c \pm \sqrt{y_1} \\ \omega_{n2} &= \omega_c \pm \sqrt{y_2}\end{aligned}\quad (34)$$

On the contrary, if $y_{1,2} < 0$, there will be only one real solution for (33) and for $\omega_i = \omega_c$ the transfer function present only a maximum and no ripple:

$$|H_{NC}(j\omega_c)|^2 = \frac{K^2 \omega_0^4}{(\omega_0^2 + \omega_d^2)^2}. \quad (35)$$

Considering $y_{1,2} \geq 0$ for the frequencies in (34), the transfer function presents more solutions. In this case, we can have three maxima (one always from (35)) and two minima, and ripple will be present as in Fig. 6.

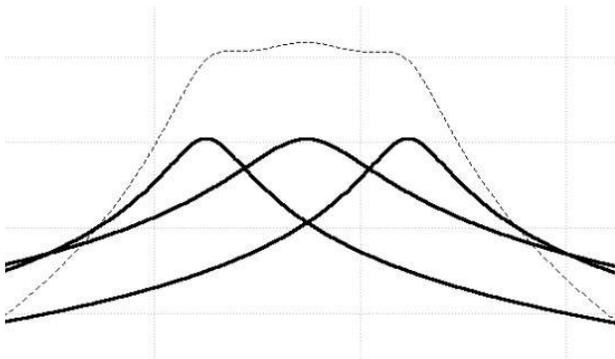


Figure 6. Third order complex filter transfer function (dot) and transfer function of each first order stage.

For this case, we have to verify both that $\omega_i = \omega_c + 1$ is cut-off frequency and the in-band ripple; imposing the maximum value for this last parameter in order to verify if there are acceptable solutions for ω_0 and ω_d . For instance, for $\omega_d^2 = 12\omega_0^2$ the ripple is not acceptable. Moreover, it can be observed that the selectivity increases as a function of ω_d .

V. CONCLUSIONS AND FUTURE WORK

We have described an approach for multistage complex IF filter design able to reach a good tradeoff between in-band ripple and rejection of both out-of-band and image frequency using different center frequencies in each stage, also choosing properly their bandwidths. Formulas are also provided for the filter design in the two stages case, obtaining the desired selectivity and ripple just choosing the bandwidth and the center frequency of each stage. We have also briefly described the third order case and extension to higher order is quite straightforward. We have adopted this approach for IF filter design in STMicroelectronics GNSS receiver, [10]-[12].

Future work could be the extension of the proposed approach to design filter where each stage shows different order, bandwidth and center frequency asymmetrically placed respect to the center of the passband.

ACKNOWLEDGMENT

The authors are grateful to Angela Bruno for support, discussions and helpful advices.

REFERENCES

- [1] K.W. Martin, "Complex signal processing is not complex", IEEE Trans. on Circ. and Syst. I: Regular Papers, vol.51, no.9, pp.1823,1836, Sept. 2004
- [2] P. Khumsat and A. Worapishet, "5th-order Chebyshev Active-RC Complex Filter Employing Current Amplifiers", Proc. IEEE Int. Symp. on Integrated Circuit, Sept. 26-28, 2007, pp. 281-284, Singapore.
- [3] V. Della Torre, M. Conta, R. Chokkalingam, G. Cusmai, P. Rossi, and F. Svelto, "A 20 mW 3.24 mm² Fully Integrated GPS Radio for Location Based Services", IEEE J. of Solid-State Circuits, vol. 42, no. 3, March 2007, pp. 602-612.
- [4] L. Laccisaglia, "Progetto di un filtro passa banda per segnali I-Q in un ricevitore a low-IF per applicazione GPS in tecnologia CMOS" Master Thesis, Rome University "La Sapienza", 2004.
- [5] F. Behbahani, H. Firouzkouhi, R. Chokkalingam, S. Delshadpour, A. Kheirkhahi, M. Nariman, et al., "A 27mW GPS Radio in 0.35 μ m CMOS", Proc. IEEE Int. Solid-State Circ. Conf., vol.1, Feb. 3-7, 2002, pp. 398-476, San Francisco.
- [6] C-W Chen, Y-C Chung, O. T-C. Chen, and R.Y.J. Tsen, "A Concurrent L1 and L2-band GPS Receiver Using an Integrated Low-IF and Weaver Architecture", Proc. IEEE Int. Symp. on Micro-NanoMechatronics and Human Science, Dec. 27-30, 2003, vol.3, pp. 1163-1166.
- [7] A. Abuelma'atti and M.T. Abuelma'atti, "A new active polyphase filter for image rejection using second generation current conveyors", Proc. 9th WSEAS Int. Conf. on Circuits, July 11-13, 2005, Athens, Greece.
- [8] H. Kondo, M. Sawada, N. Murakami, S. Masui, M. Fujioka, K. Maruyama, M. Naganuma, and H. Fujiyama, "Complex BPF for Low-IF receivers with an Automatic Digital Tuning Circuit", RFIT2007-IEEE Int. Workshop on Radio-Frequency Integration Technology, Dec. 9-11, 2007, pp.74-77, Singapore.
- [9] J. Mahattanakul, "The effect of I/Q imbalance and complex filter component mismatch in Low-IF receivers", IEEE Trans. on Circ. and Syst. I: Regular Papers, vol. 53, no. 2, Feb. 2006, pp. 247-253.
- [10] D. Grasso, G. Avellone, G. Rivela, G. Calí, and S. Scaccianoce, "An Adaptive Low-Power Complex gm-C IF Filter for L1/E1 GPS/Galileo Signals", IEEJ Int. Analog VLSI Workshop, AVLSIWS 2010, Sept. 8-10, 2010, Pavia, Italy.
- [11] G. Rivela, P. Scavini, D. Grasso, A. Calcagno, M. Castro, G. Di Chiara, G. Avellone, G. Calí, and S. Scaccianoce, "A 65 nm CMOS Low Power RF Front-End for L1/E1 GPS/Galileo Signals", Great Lake Symposium on VLSI, GLSVLSI, May 2-4, 2011, Lausanne, Switzerland.
- [12] G. Rivela, P. Scavini, D. Grasso, M. Castro, A. Calcagno, G. Avellone, A. Di Mauro, G. Calí, and S. Scaccianoce, "A low power RF front-end for L1/E1 GPS/Galileo and GLONASS signals in CMOS 65nm technology," Localization and GNSS Int. Conf. on ICL-GNSS, June 29-30, 2011, pp.7-12.
- [13] C. Cuypers, N.Y. Voo, M. Teplechuk, and J.I. Sewell, "General synthesis of complex analogue filters", IEE Proc. Circuits, Dev. and Syst., vol. 152, no. 1, Feb. 2005, pp.7-15.
- [14] S. Hintea, G. Csipkes, C. Rus, D. Csipkes, and H. Fernandez-Canque, "On the Design of a Reconfigurable OTA-C Filter for Software Radio", IEEE 2nd NASA/ESA Conf. on Adaptive Hardware and Systems (AHS 2007), Aug. 5-8, 2007, pp. 541-546, Edinburgh.
- [15] H. Liu, Z. Fu, F. Lin, "A Low Power Gm-C Complex Filter for ZigBee Receiver", IEEE Microwave and Millimeter Wave Technology (ICMMT 2012) Int. Conf. on, May 5-8, 2012, pp.1-4, Shenzhen, China.

QoS-aware Traffic Offloading in 4G/WiFi Multi-RAT Heterogeneous Networks: Opnet-based Simulations and Real Prototyping Implementation

Ahmed S. Ibrahim*, Ahmed M. Darwish, Shady O. Elbassiouny, and Hani Elgebalay

Intel Corporation, Wireless Communications Lab, Middle East Mobile Innovation Center (MEMIC), Cairo, Egypt

* Electronics and Electrical Communications Department, School of Engineering, Cairo University, Egypt
{ahmed.s.ibrahim, ahmedx.m.darwish, shadyx.o.elbassiouny, hani.elgebalay}@intel.com

Abstract— Multiple radio access technologies (Multi-RAT) Heterogeneous Networks (HetNets) are considered as one of the recent fundamental ideas in next generation mobile networks. Multi-RAT HetNets aim to increase the network capacity needed to meet the high data demand of mobile users. In Multi-RAT HetNets, data flow can be offloaded to unlicensed bands (e.g., WiFi) to free some of the resources of the 4G licensed band (e.g. LTE or WiMAX). However, it needs to be guaranteed that the WiFi can provide the needed Quality of Service (QoS). In this paper, we propose multiple QoS-aware 4G/WiFi offloading schemes that try to maximize the utilization of the WiFi air interface, taking into consideration the network loading conditions and maintaining the required QoS. The proposed offloading schemes are evaluated in two 4G/WiFi Multi-RAT testbeds. The first testbed is based on the Opnet network simulator, while the second testbed is based on a real small-cell prototype. The performance evaluation of these offloading schemes is shown in a time-cost tradeoff graph.

Keywords- HetNets; LTE; Multi-RAT; Opnet; Prototype; WiFi; WiMAX.

I. INTRODUCTION

Recently, there has been increased interest in Heterogeneous Networks (HetNets) as a means to meet the huge data demand needed for the current and future mobile applications. For example, various HetNets topics are actively investigated in the 3GPP-Long Term Evolution (LTE) [1] and Worldwide Interoperability for Microwave Access (WiMAX) IEEE 802.16 [2] standards. Research in HetNets can be divided into two main topics, namely, multi-tier and multiple radio access technologies (multi-RAT). The multi-tier HetNets comprise a hierarchical deployment of low power, smaller footprint Femto/Pico/Relay stations within a Macro-cell coverage area, aiming to increase the network capacity and provide reliable indoor coverage [3]. Interference mitigation (IM) is one of the challenging topics in multi-tier HetNets, and we have previously proposed multiple IM techniques that span the time, frequency, space, and power dimensions [4], [5].

The multi-tier HetNets utilizes one air interface only (e.g. LTE). However, multiple air interfaces can be utilized too, which is the case of multi-RAT HetNets [6], [7], [8], [9], [10], [11], [12], [13]. Multi-RAT HetNets allow the support of mobile users with more than one air interface, such as the combination of the licensed 3GPP LTE along with the unlicensed IEEE 802.11 Wireless Local Area Network (WLAN) technologies in order to increase the overall network capacity with reduced cost. It is known that the 4G technology (LTE or WiMAX) utilizes license bands with more charging cost

compared to utilizing the unlicensed WiFi air interface. Hence, it is cost-effective to be utilizing the WiFi interface, as long as it is providing the required service quality.

In the literature, there has been a number of works considering 3G-based Multi-RAT systems such as [6], [7], [8]. A feasibility study on multiple radio access (MRA) networking was introduced in [7] and an implementable MRA was proposed in [8]. Beyond 3G, Multi-RAT systems were analyzed with the objective of network capacity maximization in [9], [10], [11], [12]. Other than capacity-maximization, we have focused on minimizing the transmission power in [13], and we have proposed multiple minimum-power LTE/WiFi offloading schemes that minimize the transmission power, while guaranteeing the required QoS.

In this paper, we aim to propose multiple QoS-aware 4G/WiFi offloading schemes that try to maximize the utilization of the WiFi interface, while maintaining the required QoS. Maximizing the WiFi utilization, compared to the 4G utilization, corresponds to lowering the transmission cost. In order to propose these offloading schemes, we develop two independent testbeds. The first testbed is based on the Opnet network simulator [14], in which the available 4G technology is WiMAX, while the second testbed is based on a real prototype, in which the available 4G technology is LTE. We point out that considering more than one 4G technology (LTE or WiMAX) in more than one testbed (simulations or real prototype) strengthens this paper and generalizes the presented ideas.

The main contributions of this paper are as follows. First, we develop an Opnet-based WiMAX/WiFi multi-RAT HetNet simulation environment, which consists mainly of a multi-RAT base station (BS) or small cell and a *newly* created multi-RAT UE. Second, we propose a QoS-aware offloading scheme, which assigns the users to either the WiMAX or the WiFi air interface, depending on the network loading condition. The proposed scheme utilizes the WiFi interface unless it cannot guarantee the required QoS. Third, we build a real LTE/WiFi Multi-RAT small-cell prototype and explain how this prototype can be utilized in sending a video clip over the two air interfaces (LTE and WiFi). Finally, we propose two QoS-aware offloading schemes, implement them on the small-cell prototype, and evaluate their time-cost performance tradeoff graph.

The rest of the paper is organized as follows. In the next section, we focus on the Opnet-based Multi-RAT simulation environment, clarify the problem, and evaluate the proposed

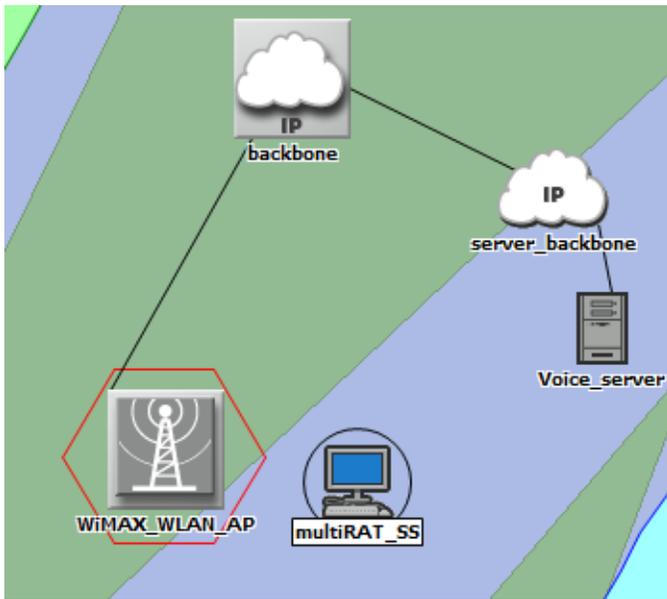


Fig. 1. Opnet-based Multi-RAT HetNet.

offloading scheme. The real LTE/WiFi small-cell prototype is introduced in Section III, after which two additional offloading schemes are proposed. Finally, Section IV concludes the paper.

II. OPNET-BASED QOS-AWARE WiMAX/WiFi OFFLOADING SCHEME

In this section, first we introduce our developed Opnet-based WiMAX/WiFi Multi-RAT HetNet environment. Second, we explain our motivation to take into consideration the QoS while designing the offloading schemes. Third, we pose the problem statement. Finally, we propose a QoS-aware WiMAX/WiFi offloading scheme.

A. Opnet-based Multi-RAT Heterogeneous Network Model

Fig. 1 depicts a multi-RAT HetNet, which consists of multi-RAT small cell (or BS), multi-RAT UE, VoIP server, and core network (server backbone and IP backbone). The multi-RAT BS along with the VoIP server and the core network are already built in the Opnet library (v14.5). However, there was no Multi-RAT UE that can communicate with both WiFi and WiMAX, which represented our first challenge.

Our first contribution is to create a novel multi-RAT (WiMAX/WiFi) UE that can receive the data over either the WiFi or the WiMAX air interfaces. Fig. 2 depicts the node model of the proposed Multi-RAT UE in details. As shown, the set of all the upper layers (e.g. IP) remains untouched as in the single-RAT node model. On the other hand, the lower layers, namely, Address Resolution Protocol (ARP) and Media Access Control (MAC) are duplicated for each interface. The MAC module is the one differentiating WiMAX from WiFi. We note that each UE will have one specific IP address, through which it can be addressed in the network. This IP address will not be changed if the UE is utilizing either the WiMAX or the WiFi

TABLE I
WiMAX AND WiFi AIR INTERFACE CONFIGURATION
PARAMETERS.

Parameter	Value
WiMAX PHY profile	OFDMA
WiMAX Bandwidth	20MHz
WiFi PHY profile	Extended rate PHY (802.11g)
WiFi Data rate	11 Mbps
Multipath channel model	ITU Pedestrian A
Pathloss model	Free space

air interface. Moreover, either the small cell or the UE may decide either to activate the WiMAX or the WiFi air interfaces.

In order to test the performance of the developed HetNet environment, we assume the UE initiates a VoIP call with the VoIP server. We note that we focus on real time application (mainly VoIP) and evaluate its QoS, measured in terms of Mean Opinion Score (MOS). The VoIP data will be sent over WiFi or WiMAX to the small cell, which in turn will be delivered through the core network to the VoIP server. The VoIP packets from the VoIP server to the UE follow the same path in reversible order. Table I shows the major configuration parameters for both WiMAX and WiFi air interfaces.

Fig. 3 depicts the performance of the VoIP application, measured in MOS, for both the WiMAX and WiFi cases. In addition, it shows the loading condition on both air interfaces, to indicate which interface is activated. First, we consider the case of having WiMAX as the active interface, which is depicted in Fig. 3 (a). As shown, the load on the WiMAX network is non-zero, while it is zero for the WiFi case. The MOS of the VoIP application is around 3.8, which gives great user experience. Second, the WiFi is activated and the resulting performance is shown in Fig. 3 (b). Similar to the WiMAX, we find that the WiFi also achieves the same MOS value.

From Fig. 3, we conclude that if the VoIP application is the only traffic in the network, offloading to WiFi does not degrade the MOS below that achieved by the WiMAX. In such scenario, traffic offloading from WiMAX to WiFi should happen immediately once WiFi is available. However, this might not be necessarily true in case there is other additional traffic in the network. This case will be discussed in the next sub-section.

B. Motivation and Problem Statement

In general, we note that there is a major difference between the WiMAX (IEEE 802.16e) and the WiFi (IEEE 802.11g), which is the consideration of the QoS classes. Particularly, in WiMAX each type of application is assigned a specific class, which corresponds to the needed QoS. Traffic types with high QoS class (e.g. VoIP) are assigned resource blocks with high priority, compared to traffic types with lower QoS classes (e.g. best effort as in web browsing). On the contrary, in WiFi all the traffic types are treated equally in the best effort class. This fact follows mainly from the utilized multiple access scheme,

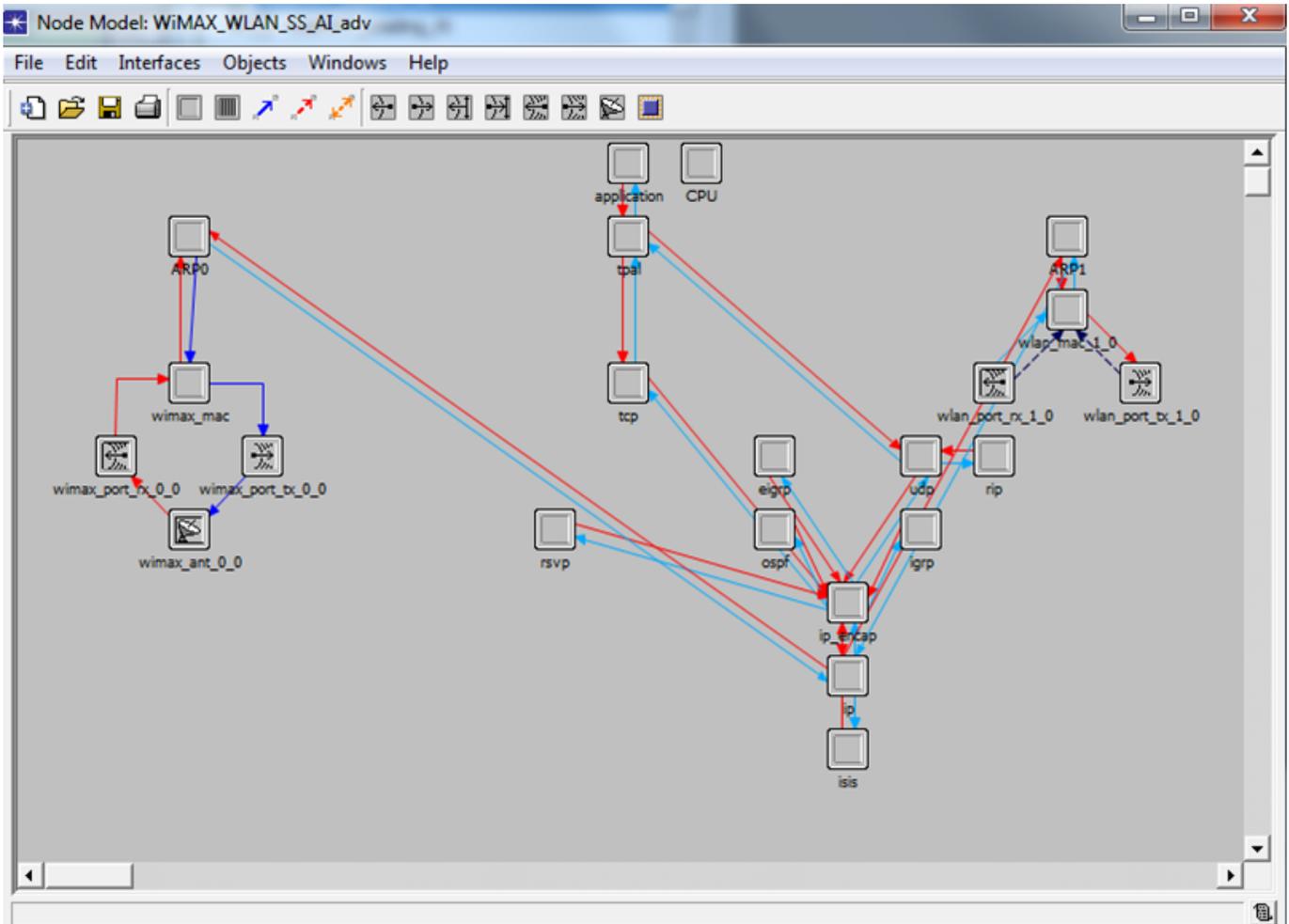


Fig. 2. Newly created Opnet Multi-RAT UE.

namely, Carrier Sense Multiple Access/Collision Avoidance (CSMA/CA). Hence, in WiFi there is no distinction between the real time applications and non real time ones. Such issue may degrade the performance of the VoIP applications sent over WiFi, once the network gets congested.

In order to investigate this issue, we revisit the multi-RAT scenario shown in Fig. 1 and apply a best effort background data traffic of rate 2 Mbps, in addition to the existing VoIP application. This background data traffic may represent the accumulation of many UEs accessing the same air interface, or it may represent data traffic to/generated from the same UE. Fig. 4 shows the network performance in the case of WiMAX and WiFi interfaces for 2 Mbps background best-effort data traffic.

As shown in Fig. 4 (a), the WiMAX air interface maintains the high value of the VoIP MOS. This high performance is a result of assigning high QoS class to the VoIP application. Hence, the WiMAX resource blocks needed for the VoIP application are always allocated irrespective of the other background traffic. On the contrary, we find in Fig. 4 (b) that the WiFi air interface dramatically degrades the MOS of the VoIP application. Such performance degradation is

because the best effort background traffic is competing with the VoIP application, which results in lower chances for the VoIP application in accessing the WiFi air interface.

From Fig. 4, we conclude that offloading from WiMAX (or 4G in general) to WiFi should not happen automatically once WiFi access is available. Other factors (e.g. background traffic) need to be considered to make such offloading decision. This finding represents our motivation to develop a QoS-aware offloading scheme that takes into consideration the loading condition of the network. Consequently, our problem can be formulated as follows: Given a multi-RAT (4G/WiFi) system, find an offloading scheme that utilizes the WiFi interface the most (i.e. minimizes the charging cost), while guaranteeing the required QoS. The solution to this problem will be proposed in the next sub-section.

C. QoS-aware Traffic Offloading Scheme in WiMAX/WiFi Multi-RAT HetNets

Fig. 5 depicts the flow chart of the proposed QoS-aware offloading scheme, and it can be described as follows. Initially, the 2 RATs are deactivated. Then, the UE senses the total data rate (loading condition) on each interface. If the loading

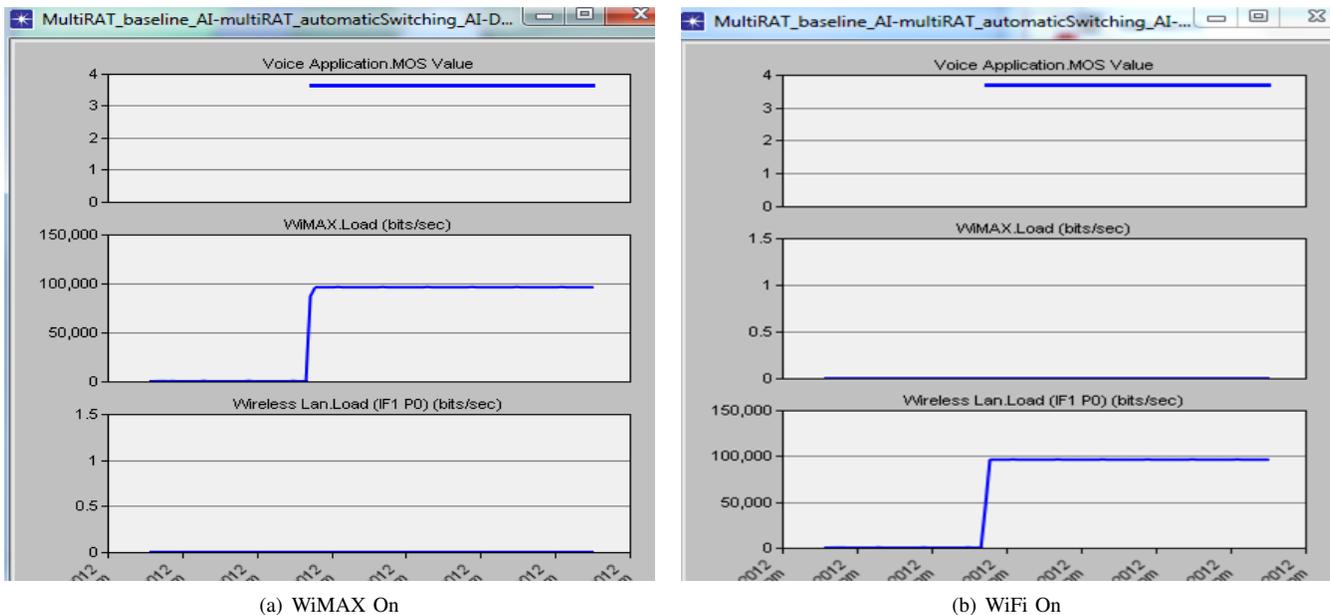


Fig. 3. Performance of WiMAX and WiFi interfaces for VoIP-only network.

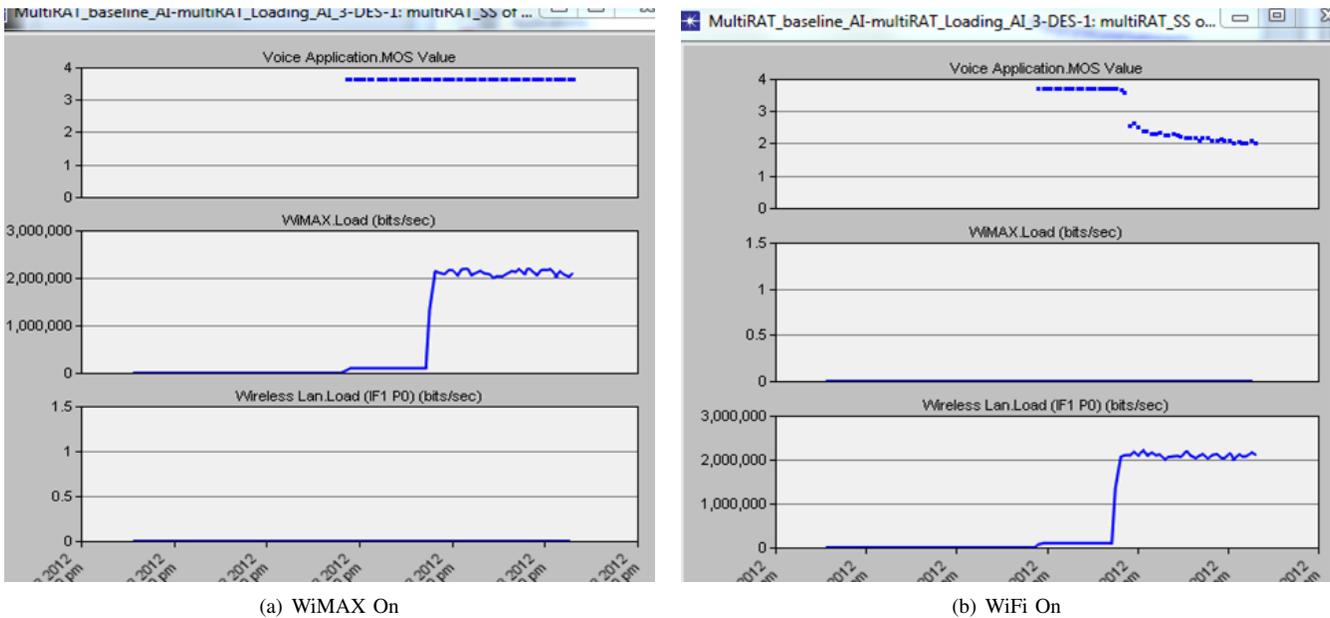
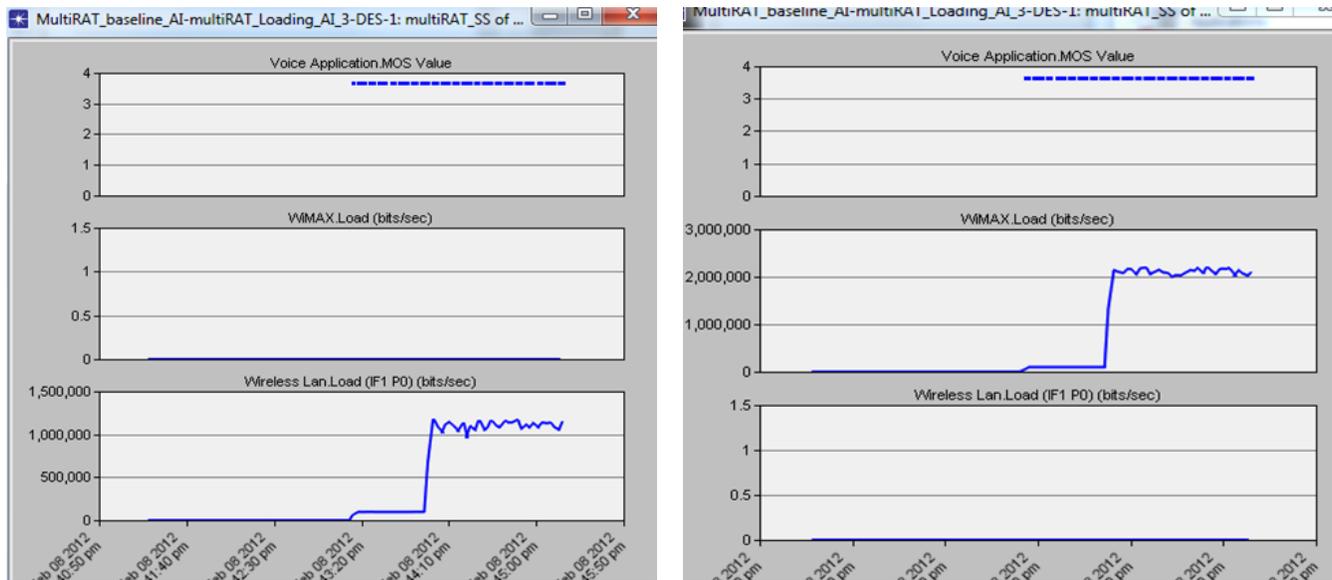


Fig. 4. Network performance with 2Mbps background best-effort traffic.

rate is higher than a specific threshold, then it activates the WiMAX interface. In this case, the WiFi will not be able to guarantee the required QoS of the real-time application. Otherwise, it offloads the traffic to the WiFi interface. Utilizing our developed Opnet-based HetNet environment, we have investigated the threshold after which the WiFi breaks down and starts downgrading the VoIP MOS, and found that to be 1.5 Mbps. In other words, for background traffic higher than 1.5 Mbps, the VoIP MOS degrades as was shown in Fig. 4 (b).

We have implemented the proposed offloading scheme in the Opnet-based Multi-RAT environment at the IP process

module of the Multi-RAT UE node, shown in Fig. 2. We have tested the offloading scheme by applying different background traffic rates, which are 1 Mbps and 2 Mbps. Fig. 6 shows the performance of the offloading scheme at these two scenarios. As shown, for 1 Mbps loading scenario in Fig. 6 (a), the WiFi interface is activated and the MOS is maintained at high value. In this case, offloading to WiFi guarantees the required QoS. As for the higher loading rate (2 Mbps in Fig. 6 (b)), the WiMAX is activated and the high MOS is maintained. Hence, our proposed QoS-aware offloading scheme achieves the required QoS with minimum charging cost. More precisely, it offloads the traffic to the unlicensed minimum-cost WiFi



(a) WiFi On at 1 Mbps loading

(b) WiMAX On at 2Mbps loading

Fig. 6. Network performance with QoS-aware offloading scheme at different network loading conditions.

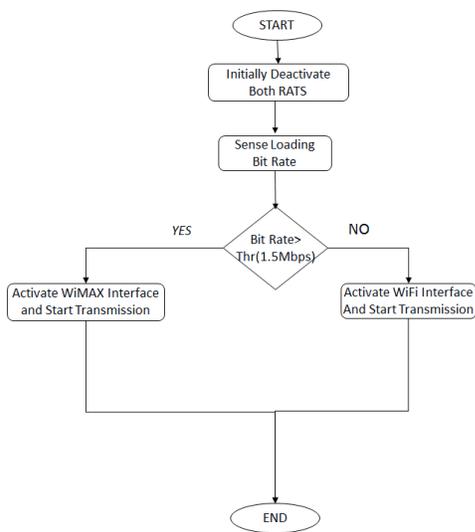


Fig. 5. Flow chart of the QoS-aware traffic offloading scheme.

only if the required QoS is guaranteed.

III. REAL PROTOTYPE-BASED QOS-AWARE LTE/WIFI OFFLOADING SCHEMES

In the previous section, we have focused on Opnet-based multi-RAT simulation environment. In this section, we present a real prototype that represents the implementation of LTE/WiFi small cell. Such prototype is extremely necessary to validate the concept of multi-RAT system and to augment our simulation testbed with a real one. In this prototype, we utilize Intel-proprietary LTE software stack, consisting of all the LTE layers (RRC/PDCP/RLC/MAC). As for the WiFi, we

utilize the commercial Intel 6205 WiFi card, which can be set to work in the access point mode. Hence, we create an integrated LTE/WiFi small cell that has both WiFi and LTE software stacks.

A. Real LTE/WiFi Multi-RAT Prototype

Fig. 7 depicts the system architecture of the implemented multi-RAT prototype. The prototype consists of two terminals: the first terminal has an integrated LTE/WiFi small cell, and the second terminal represents the UE side. The LTE data between the two terminals is physically sent over the Ethernet cable, which mimics the LTE PHY layer, while the WiFi data is physically sent over the WiFi air interface. At the small cell side, the LTE down-link path is represented by the (PDCP-TX/RLC-TX/MAC-TX) path, and the LTE up-link path is represented by the (MAC-RX/RLC-RX/PDCP-RX) path. The WiFi path is represented by the (NDIS/MAC/PHY) path. At the UE side, we create a simple software client that can send/receive data to/from the Ethernet and WiFi interfaces.

A photo of the real prototype is shown in Fig. 8. The small-cell and the UE terminals are shown. In addition, we have created a WiFi loading environment, which consists of two terminals having file transfer between each other. Such file transfer causes huge degradation on the WiFi link, which results in huge WiFi congestion.

In order to test this prototype, we have created a demonstration experiment in which a video clip is sent from the small-cell terminal to the UE terminal via both LTE and WiFi software stacks. The Ethernet link is used to transfer the LTE data, while the WiFi is sent over the conventional WiFi air interface. Initially, a part of a video clip (a number of packets) is sent over the LTE. Once the WiFi coverage is available, the transmission is offloaded to the WiFi. Once the WiFi

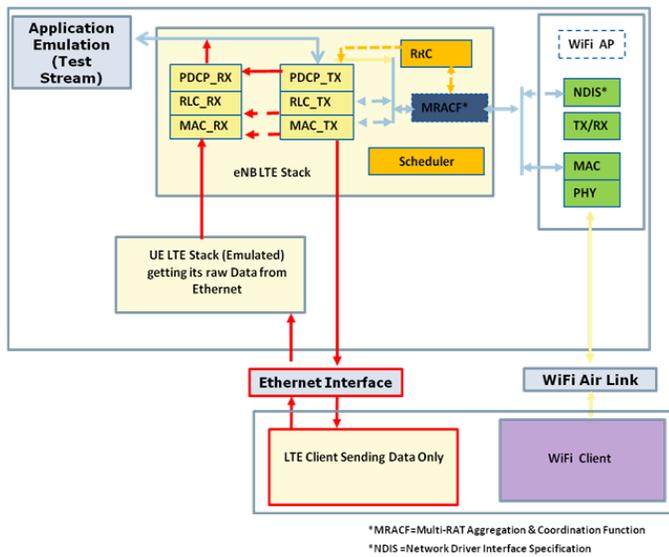


Fig. 7. System architecture of the LTE/WiFi Multi-RAT prototype.



Fig. 8. Real setup of the LTE/WiFi Multi-RAT prototype.

network is congested (due to loading the WiFi link with the file transfer), the rest of the video transmission is sent over LTE. Such offloading scheme was implemented in the multi-RAT aggregation and coordination function (MRACF) at the small cell side. At the receiver side, the received packets over the two interfaces are assembled, and the video is displayed.

The small-cell prototype described above is utilized in evaluating the performance of our proposed QoS-aware offloading schemes. Particularly, we consider two performance criteria to evaluate these schemes, namely, the time and the cost. First, the time metric represents the total number of transmission slots needed to successfully receive the whole video file. Second, the cost metric represents the number of cost units

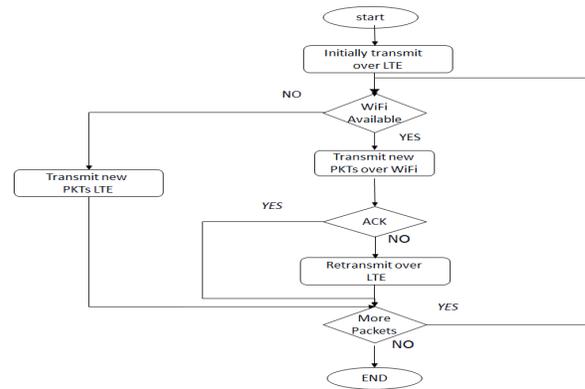


Fig. 9. Flow chart of the “LTE/WiFi Offloading Retransmit-Once” offloading scheme.

needed to successfully receive the video file. We assume that the video file is divided into a number of packets. Each packet requires 1 transmission slot and costs 1 unit over LTE and zero units over WiFi. These cost values reflect the fact that WiFi is unlicensed band, while LTE is a licensed one. The proposed offloading schemes are presented in the next sub-section.

B. QoS-aware LTE/WiFi Offloading Schemes

We propose 2 QoS-aware LTE/WiFi offloading schemes. The first scheme is entitled “LTE/WiFi Offloading Retransmit-Once” and its flowchart is depicted in Fig. 9. In this scheme, the packets are initially sent over LTE. Once the WiFi coverage is available, the data is offloaded to the WiFi air interface. Due to the created congestion on the WiFi, we have developed an acknowledgment mechanism that allows the UE to transmit back to the small cell either a positive acknowledgment (ACK) or a negative acknowledgment (NACK) for each received packet. A NACK is declared if a particular packet index is missing. Once a NACK is declared, the erroneous packet is retransmitted over the LTE interface. The LTE is utilized to guarantee a congestion-free retransmission. As for the next new packet, the WiFi is activated again and it carries the upcoming packet. Hence, the description in the name “Retransmit-Once.” This procedure continues until the video is successfully received.

The second QoS-aware offloading scheme, namely “LTE/WiFi Offloading Retransmit-Rest”, is similar to the first one in many aspects except in the response following the retransmission of the erroneous packet. More precisely, once an erroneous packet is retransmitted over the LTE, the LTE air interface is kept active and transmits the rest of the video packets. Hence, the second scheme utilizes the LTE interface more and hence it costs more. This is shown in the time-cost tradeoff curve shown in Fig. 10.

Fig. 10 depicts the different offloading schemes results where the x-axis represents the cost of transmission (0 for WiFi packet and 1 for LTE packet) and the y-axis represents the total time (in transmission slots) needed to transmit one video clip of size 23 packets. The “LTE-only” point corresponds to having LTE transmission only and results in the minimum

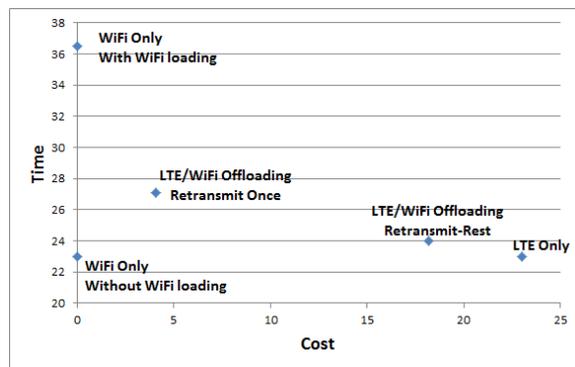


Fig. 10. Time-Cost tradeoff curve for the LTE/WiFi Multi-RAT offloading schemes.

time of 23 units, but with the maximum cost of 23 units. In order to reduce the cost, we have proposed the 2 QoS-aware offloading schemes described earlier. We find that the “LTE/WiFi Offloading Retransmit-Rest” slightly reduces the cost to 18, while slightly increasing the time to 24 units. The extra time units are due to the retransmission time slots happening because of congestion and packet drop over the WiFi link.

On the other hand, the “LTE/WiFi Offloading Retransmit-Once” reduces the cost much more to 4 units while slightly increasing the time to 27. Compared to “LTE-only” case, we find that the “LTE/WiFi Offloading Retransmit-Once” achieves cost saving of 83% while increasing the time by 15%. For reference, we have also plotted the points corresponding to the WiFi-only transmission. More specifically, we observe “WiFi-only without WiFi loading,” which results in both minimum time and zero cost. If loading is injected, the required time is increased to 37 units, which is shown in the “WiFi-only with WiFi loading.”

IV. CONCLUSION AND FUTURE WORK

In this paper, we have developed 2 4G/WiFi Multi-RAT testbeds and proposed various QoS-aware offloading schemes. In particular, we have developed an Opnet-based WiMAX/WiFi multi-RAT simulation environment. We have shown, via this simulation platform, that the WiFi degrades the VoIP quality as the background traffic rate exceeds 1.5 Mbps. On the other hand, the WiMAX preserves high quality for the VoIP application irrespective of the background traffic, as it guarantees the needed QoS. Consequently, we have proposed a QoS-aware offloading scheme that decides either to offload the traffic to the WiFi or to utilize the WiMAX, based on the network loading data rate. Such offloading scheme achieves the required rate with minimum charging cost.

In addition to developing the Opnet-based simulation environment, we have also developed a real LTE/WiFi prototype that can be used to transmit any file over the 2 air interfaces. Finally, we have proposed and implemented 2 QoS-aware LTE/WiFi offloading schemes and have evaluated the time-cost performance tradeoff. We have shown that by considering our proposed QoS-aware offloading we can achieve cost saving

of 83% while increasing the time by 15% only, compared to “LTE-only” case.

In the future, we will investigate which of the different LTE layers (PDCP,RLC,MAC) is the best one at which we can offload the data to the WiFi. We will utilize our developed real LTE/WiFi prototype as a testbed to evaluate the pros and cons of the different offloading points.

V. ACKNOWLEDGMENT

The authors would like to thank Dr. Nageen Himayat, Dr. Kirk Skeba, Dr. Shilpa Talwar, Mohamed Fadel, Wafaa Taie, Mohamed Yousef, and Hossam Samy of Intel Wireless Communications labs, for their constructive discussions and cooperation throughout this work.

REFERENCES

- [1] 3GPP TS36.300 V9.0.0, “Evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (E-UTRAN),” 2009.
- [2] IEEE 802.16e, “IEEE standard for local and metropolitan area networks part 16: Air interface for fixed and mobile broadband wireless access systems,” 2009.
- [3] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, “Femtocell networks: A survey,” in <http://arxiv.org/ftp/arxiv/papers/0803/0803.0952.pdf>, 2008.
- [4] A. S. Ibrahim, A. Darwish, E. Makled, and H. Elgebal, “Adaptive interference mitigation techniques for femtocells,” *Proc. IEEE Wireless Communication and Networking Conference (WCNC'12)*, Apr. 2012.
- [5] E. Makled, A. S. Ibrahim, A. Darwish, and H. Elgebal, “Non-unanimous power inter-cell interference coordination in heterogeneous networks,” *Proc. the IEEE 75th Vehicular Technology Conference Spring (VTC'12-Spring)*, May 2012.
- [6] K. Dimou, R. Agero, M. Bortnik, R. Karimi, G. Koudouridis, S. Kaminski, H. Lederer, and J. Sachs, “Generic link layer: a solution for multi-radio transmission diversity in communication networks beyond 3g,” in *Proc. IEEE Veh. Techn. Conf. (VTC)*, Sept. 2005, vol. 3, pp. 1672–1676.
- [7] G. P. Koudouridis, R. Agero, E. Alexandri, M. Berg, A. Bria, J. Gebert, L. Jorgueski, H. R. Karimi, I. Karla, P. Karlsson, J. Lundsja, P. Magnusson, F. Meago, M. Prytz, and J. Sachs, “Feasibility studies and architecture for multi-radio access in ambient networks,” in *Wireless World Research Forum (WWRF) Meeting*, Dec. 2005.
- [8] E. Gustafsson and A. Jonsson, “Always best connected,” *Wireless communication, IEEE Trans.*, vol. 10, no. 1, pp. 49–55, Feb. 2003.
- [9] Y. Choi, H. Kim, S. Han, and Y. Han, “Joint resource allocation for parallel multi-radio access in heterogeneous wireless networks,” vol. 9, no. 11, pp. 3324–3329, Nov. 2010.
- [10] L. Liu, Y. Wei, Y. Bi, and J. Song, “Cooperative transmission and data scheduling mechanism of multiple interfaces in multi-radio access environment,” in *Proc. Global Mobile Congress (GMC)*, Oct. 2009.
- [11] S. Ghamari, M. Schellmann, M. Dillinger, and E. Schulz, “An approach for automated spectrum reforming for multiple radio access technologies,” in *Proc. Technical Symposium at Telecom World (ITU WT)*, Oct. 2011.
- [12] W. Song and W. Zhuang, “Multi-service load sharing for resource management in the cellular/wlan integrated network,” vol. 8, no. 2, pp. 725–735, Feb. 2009.
- [13] M. Fadel, A. S. Ibrahim, and H. Elgebal, “QoS-aware multi-RAT resource allocation with minimum transmit power in multiuser OFDM system,” *Proc. the IEEE Globecom International Workshop on Emerging Technologies for LTE-Advanced and Beyond-4G (GC'12 Workshop - LTE-B4G)*, Dec. 2012.
- [14] “<http://www.opnet.com/>,” v 14.5.

A Top-Down Heuristic for TCFA Problem in WAN

Roza Goscién, Iwona Pozniak-Koszalka, Leszek Koszalka, and Andrzej Kasprzak

Department of Systems and Computer Networks
Wrocław University of Technology
Wrocław, Poland

roza.goscién@gmail.com, {iwona.pozniak-koszalka, leszek.koszalka, andrzej.kasprzak}@pwr.wroc.pl

Abstract—This paper presents the results of investigation focused on a new topology design problem – a very important issue affecting nowadays Wide Area Networks (WANs). We formulate a TCFA (Topology, Capacity and Flow Assignment) problem for WANs and propose a novel heuristic algorithm to solve it. Moreover, we present findings of computational experiments, carried out to compare the properties of the created algorithm with other TCFA methods (both exact and heuristic) and also to determine dependences between processing time and dimensions of problems for TCFA tasks. The obtained results confirm that the proposed Top Down heuristic algorithm is promising.

Keywords- algorithm; network design; optimization; WAN; TCFA

I. INTRODUCTION

The most important thing about Wide Area Network (WANs) is its necessity to connect Local Area Networks (LANs), PCs or terminals over large distances and to fulfill the users' requirements. Evidently, realization of abovementioned tasks is connected with some costs. Abovementioned requirements and bounded budget make modeling WAN so important and increasingly popular nowadays [1].

Modeling WAN includes two main groups of problems: (i) optimization of existing topologies, which are not efficient enough, (ii) new topologies design. These problems can be divided into three main groups: FA (Flow Assignment), CFA (Capacity and Flow Assignment), TCFA (Topology, Capacity and Flow Assignment) defined in detail in [1], and [2]. This paper is focused on TCFA problems.

In this work, the defined TCFA problem is considered. We propose our own algorithm to solve this problem. We present the results of the designed computational experiments, carried out in order to compare the proposed algorithm with other algorithms solving TCFA (both exact and heuristic) and to determine dependences between processing time and dimensions of problems to be solved as well.

Optimization problems like FA, CFA and TCFA were a subject of many papers. TCFA problems are the most complex. Different forms of TCFA problems were discussed and examined in literature. Walkowiak [3] considered the problem with no network topology given a priori – placement of networks' nodes was a part of the optimization task. The problem objective was system's overall costs. In the same paper, various methods solving problem were presented – for example, algorithm based

on Lagrangian relaxation. Other forms of TCFA problem were presented by M. Gerla and L. Kleinrock [4], where the authors considered several propositions for problem objective and constraints formulation. The notation of TCFA problem used in this paper was presented by A. Kasprzak [5], where a heuristic algorithm to solve the problem was presented. We used this algorithm to evaluate our own algorithm.

The paper is organized as follows. In Section II, we formulate the formal model of the considered problem. Section III presents our novel *Top-Down* algorithm. Findings from computational experiments are presented in Section IV. The final remarks and conclusion appear in Section V.

II. OPTIMIZATION PROBLEM

A. The model of a computer network

The model of a computer network can be created using graph theory. Graph's vertices $V=\{v_1, v_2, \dots, v_n\}$ correspond to network's nodes and graph's edges $E=\{e_1, e_2, \dots, e_m\}$ correspond to links. Numbers $c(e)$ assigned to edges determine links' capacities (in, e.g., bits per second) [6]. We considered only directive graphs in modeling.

In computer memory, a model can be saved as a weighing matrix N or adjacency matrix N' and vector of capacity c . There is an example of computer network model presented in Fig. 1 and described as below.

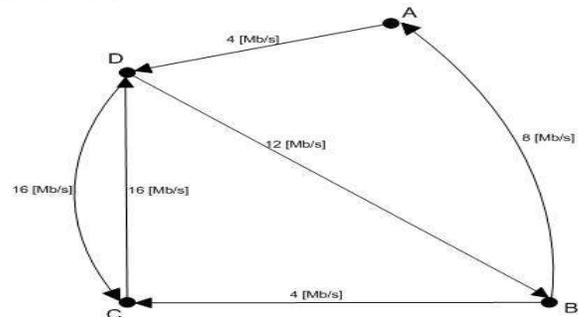


Figure 1. An example of computer network model.

$$N = \begin{bmatrix} 0 & 0 & 0 & 4 \\ 8 & 0 & 4 & 0 \\ 0 & 0 & 0 & 16 \\ 0 & 12 & 16 & 0 \end{bmatrix}, N' = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}, c = [4 \quad 8 \quad 4 \quad 16 \quad 12 \quad 16]$$

Weighting matrix $N=[n_{ij}]_{n \times n}$ contains capacities of all existing links. n_{ij} is equal to 0 if the link between nodes i and j does not exist. Otherwise, there is a capacity value of this link. Adjacency matrix $N'=[n'_{ij}]_{n \times n}$ informs only whether a specific link exists. n'_{ij} is equal to 0 if the link between nodes i and j does not exist. Otherwise, there is a value 1. Links' capacities are saved sequentially in vector c .

B. Multicommodity flows

Multicommodity flow in WAN is defined as the average flow of information in a particular slot of time. Flow commodity is a set of packets with the same i -th source and j -th destination node.

Let r_{ij} denotes the average packet rate transmitted from node i to node j and $R=[r_{ij}]_{n \times n}$ is a demand matrix, defined in [5] and [7]. Note that d -th flow commodity is connected with a pair of nodes: source s_d and termination t_d . The value $h_d=r_{ij}$, where $i=s_d$ and $j=t_d$, is known as volume of d -th commodity.

Let $x=e^-$ and $y=e^+$ be a source and destination node of link e , respectively.

Mathematically, multicommodity flow is a function $f^k: E \rightarrow R^+ \cup 0$; $k=1,2,\dots,q$, which assigns to networks links $e \in E$ values $f^k(e)$ [b/s] satisfying constraints (1) – (3) (see [2] and [8]).

- for $v \in V$ and $k=1,2,\dots,q$:

$$\sum_{e^+=v} f^k(e) - \sum_{e^-=v} f^k(e) = \begin{cases} r_k, v = s_k \\ -r_k, v = t_k \\ 0, v \neq s_k \wedge v \neq t_k \end{cases} \quad (1)$$

- for $e \in E$ and $k=1,2,\dots,q$:

$$f^k(e) \geq 0 \quad (2)$$

$$f(e) = \sum_{k=1}^q f^k(e) \leq c(e) \quad (3)$$

The value $f(e)$ [b/s] is known as an entire link's flow. Abovementioned formulation is based on node-link notation of the flow. We considered bifurcated flows as defined in [8],[9].

C. Optimization criteria

In optimization tasks, we used two mathematical dependences: cost of network and average packet delay.

The cost of a network is a sum of leasing all networks' links with specified capacities. Formally, it is described by formula (4). This goal function was used only as an additional constraint in TCFA problems.

$$d(\mathbf{e}) = \sum_{e=(x,y) \in E} k_{e,c} \quad (4)$$

where $k_{e,c}$ is a cost of leasing link e with capacity $c=c(e)$.

The goal function in optimization problems was the average packet delay – nonlinear function of flow, described by formula (5).

$$T(\mathbf{f}) = \frac{1}{\gamma} \sum_{e \in E} \frac{f(e)}{c(e) - f(e)} \quad (5)$$

where $\gamma = \sum_{k=1}^q r_k$ represents the average packet rate transmitted in the network per second.

D. TCFA formulation

We follow the formulation of TCFA problem presented e.g., in [9].

Given:

- Location of network's nodes,
- Demand matrix R ,
- Set of feasible links,
- Sets of candidate links' capacities and corresponding cost of leasing – matrixes C and D respectively,
- *costBound* – as a maximum acceptable cost of topology,
- average packet size in the network in b/s - *pSize*.

Minimize:

- Average packet delay.

Over:

- Bifurcated flows,
- Links' capacities,
- Network's structure.

Subject to constraints:

- Realization of demand matrix,
- For each link e : link's flow does not exceed link's capacity,
- Sum of leasing all selected links does not exceed *costBound*.

III. TOP DOWN ALGORITHM

In this section, we present the *Top-Down* algorithm for TCFA problem in WAN. The flow chart of this method is presented in Fig. 2.

The idea of the algorithm is based on starting analyzing the problem with fully adjacency matrix $N'(i=0)$ and gradually removing useless links or least used links, until the topology satisfies constraint *costBound* and enables to fulfill all traffic demands.

The choice of links for removing from topology is made using criterion $\delta(e)$ formulated as follows:

$$\delta(e) = \frac{f(e)}{c(e) - f(e)} \quad (6)$$

Where $c(e)$ is a capacity of link e and $f(e)$ is the flow allocated to this link.

Criterion $\delta(e)$ is a ratio of flow allocated to link e to free (still available) capacity of this link. The smaller the criterion $\delta(e)$, the higher the probability that link e will be removed from the topology.

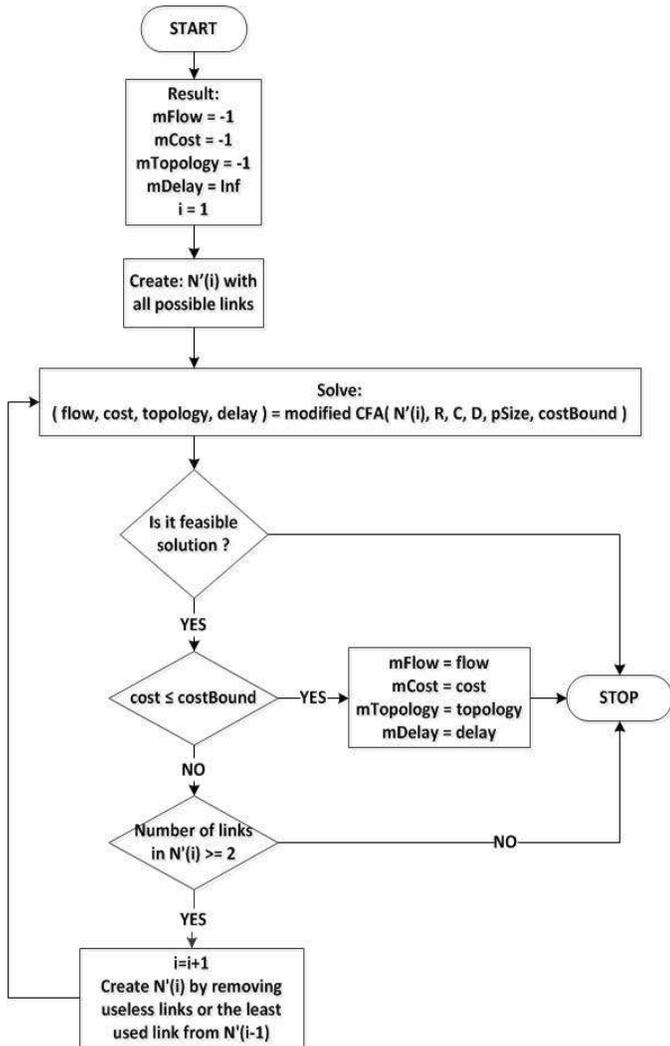


Figure 2. The flow chart of *Top-Down* algorithm.

In each iteration, the *modified-CFA* method runs. This method finds a feasible solution that satisfies all current problems' constraints or solution that exceeds *costBound* of the smallest value.

The potential final solution is marked with "m" prefix, e.g., *mCost* is a cost of final topology.

IV. INVESTIGATION

In this section, we present and discuss the results of computational experiments. The goal of experiments was twofold. First, we tried to determine the dependence between processing time of TCFA algorithms and the dimension of the input data (size of optimization problem).

Second, we wanted to evaluate the proposed *Top-Down* algorithm, based on comparison with other methods. We considered the following comparison criteria: processing time, number of returned optimal and feasible solutions.

Top-Down algorithm was compared with two different TCFA methods: exact algorithm based on complete search and heuristic proposed by M. Gola [9], called as *Modified topology* in this paper. Abovementioned methods can run with different CFA algorithms. In our work we used exact CFA algorithm based on complete search and heuristic *Bottom-Up*, proposed by R. Goscién [6]. Thus, we examined five different TCFA algorithms.

The input data for algorithms were sets of random, solvable topologies, regardless of *costBound*. All presented results are averages of 4 – 10 measurements. We used solutions returned by exact TCFA algorithm as a reference, optimal solution.

Because of unreasonable processing time of exact algorithm, we compared only results for the number of 2 and 3 network nodes. Thus, the presented investigation is only a first step in evaluating our *Top-Down* algorithm and its results should help to decide if further work with greater and real topologies (e.g., 10 nodes) is promising.

To solve partial FA problems, a nonlinear programming technique was used.

A. Impact of nodes' number on processing time

We examined topologies with 2 flow commodities and 2 candidate links' capacities. We assumed *costBound* as a 2/3 of sum of the most expensive candidate capacities for all links.

We compared only results for the number of 2 and 3 network nodes. The results are presented in Table I.

TABLE I. PROCESSING TIME AS A FUNCTION OF NODES' NUMBER (SECONDS)

TCFA method + CFA method	Number of nodes	
	2	3
Exact TCFA + exact CFA	2,003	271,901
Exact TCFA + Bottom-Up	0,374	26,458
Modified topology + exact CFA	0,807	902,476
Modified topology + Bottom-Up	0,282	1,107
Top-Down	0,118	31,706

The processing time of all methods increases with the number of nodes. The explanation is based on some proportions. The number of nodes in the network determines the maximum possible number of links and flow commodities in topology. Moreover, dimensions of constraints matrixes used in nonlinear programming techniques are proportional to the number of nodes. The explanation of this fact was discussed in [6].

In this case we wanted to emphasize that the number of nodes in the topology significantly influences processing time, particularly in case of exact method, and to show that even for relatively small topologies processing time of exact method can increase to unacceptable value.

B. Impact of candidate capacities' number on processing time

We examined topologies with 3 nodes and 2 flow commodities. We assumed *costBound* as a 2/3 of sum of the most expensive candidate capacities for all links. The results are presented in Table II.

TABLE II. PROCESSING TIME AS A FUNTION OF CANDIDATE CAPACITIES' NUMBER (SECONDS)

TCFA method + CFA method	candidate capacities' number	
	2	3
Exact TCFA + exact CFA	131,662	1847,504
Exact TCFA + Bottom-Up	17,903	19,635
Modified topology + exact CFA	309,178	2696,440
Modified topology + Bottom-Up	0,924	1,002
Top-Down	2,976	27,211

According to our findings, we can say that processing time of TCFA methods increases with candidate capacities' number. Moreover, this influence is very significant for both – exact and heuristic algorithms.

C. Impact of flow commodities' number on processing time

In this case, we used input data with 3 nodes, and 2 candidate capacities and *costBound* assumed as in previous point.

TABLE III. PROCESSING TIME AS A FUNTION OF FLOW COMMODITIES' NUMBER (SECONDS)

TCFA method + CFA method	Flow commodities' number			
	1	2	3	4
Exact TCFA + exact CFA	53,382	124,106	178,364	176,292
Exact TCFA + Bottom-Up	7,774	37,405	43,837	45,033
Modified topology + exact CFA	63,730	467,398	564,145	479,912
Modified topology+Bottom-Up	0,287	3,704	3,795	11,577
Top-Down	0,158	2,943	2,923	64,190

The results proved that the number of flow commodities influences processing time of TCFA algorithms. Moreover, according to the findings presented in Tables 1-3, we can say that *Top-Down* is the third best method in terms of processing time.

D. Impact of costBound constraint on processing time

In this part of our work, we discuss changes of processing time as a repercussion of different value of constraint *costBound*.

The results are presented for a topology with 3 nodes, 3 flow commodities and 2 candidate links' capacities.

As we can see in Fig. 3, methods based on exact CFA solve problems in the longest time. Furthermore, processing time of this algorithms increases very fast with constraint *costBound* – this is a consequence of enlarging space of topologies that we can create and analyse in CFA tasks.

After the cost of the most expensive topology is reached (about 710 euro) the processing time is approximately steady in relation to *costBound*.

Both figures (Fig. 3 and Fig. 4) also present one more important issue. The *Top-Down* algorithm is a method that solves problems in relatively short and acceptable time.

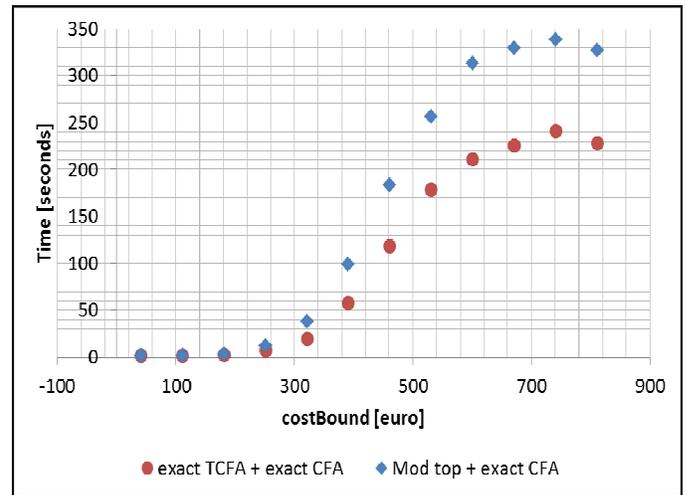


Figure 3. Processing time in relation to *costBound* – methods based on exact CFA.

In Fig.4, the results for the fastest algorithms are presented. The connection of exact TCFA and CFA heuristic *Bottom-Up* solves problems in the longest time, regardless of *costBound*. *Top-Down* method works faster.

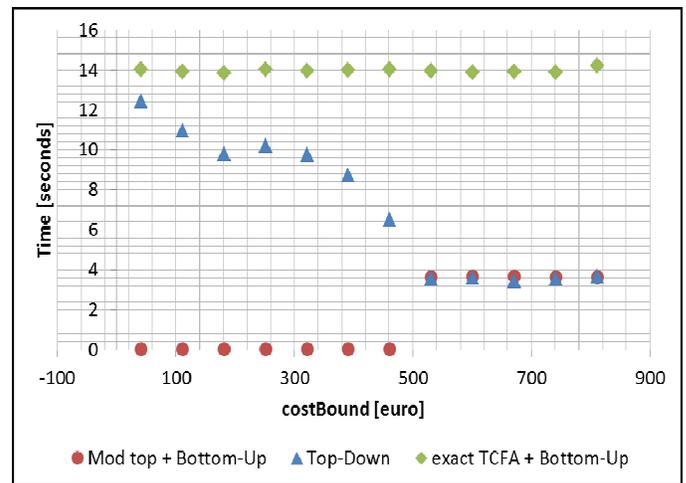


Figure 4. Processing time in relation to *costBound* – heuristic methods.

The processing time of *Top-Down* is the longest at the beginning, when *costBound* is small and we iteratively try to find a topology that is cheaper than our budget limit. Processing time of *Top-Down* algorithm decreases with increasing *costBound*, until the cost of one of the most expensive topologies is reached. Since changes of *costBound* do not influence on processing time – the space of potential solutions does not enlarge significantly.

When analysing dependences between dimensions of problems and processing time of solving methods, it is important to notice that:

- Values of all input data influence on processing time of TCFA algorithms,
- Processing time of exact algorithm is always longer than processing time of heuristic algorithm for the same set of input data,
- Exact TCFA algorithm is not always able to solve problem in reasonable time.

E. Comparison of algorithms

Overall, we compared 5 different TCFA algorithms, both exact and heuristic.

In terms of processing time, the most attractive are methods based on heuristic CFA and the proposed *Top-Down* method.

The percentage results of comparison according to number of returned optimal and feasible solutions are presented in Fig. 5 and Fig. 6. The solutions returned by connection of exact TCFA and exact CFA methods were used as reference.

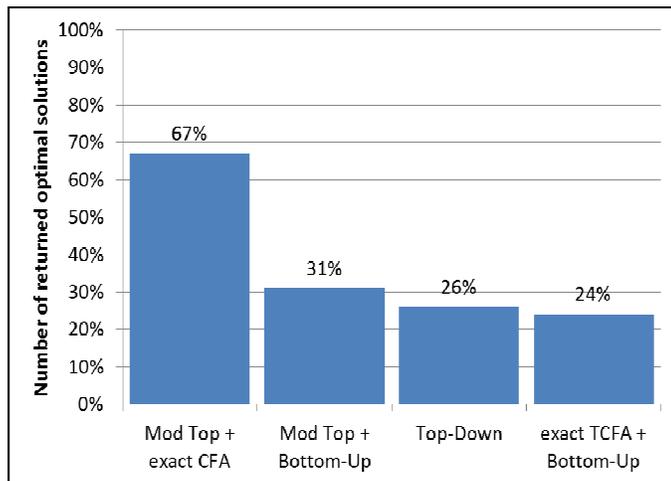


Figure 5. Number of returned optimal solutions.

The efficiency of TCFA methods that can run with different CFA methods (e.g. presented exact TCFA and *Modified Topology*) depends on the efficiency of the used CFA algorithm.

In terms of the number of the returned optimal solutions, the best was the method based on exact CFA – this method returned optimal solution in 67% of all experiments. The probability of reaching optimal solution by *Top-Down* algorithm is smaller.

The result achieved by *Top-Down* was of 26%, but it is important to notice that processing time of *Top-Down* is shorter than processing time of methods based on exact CFA.

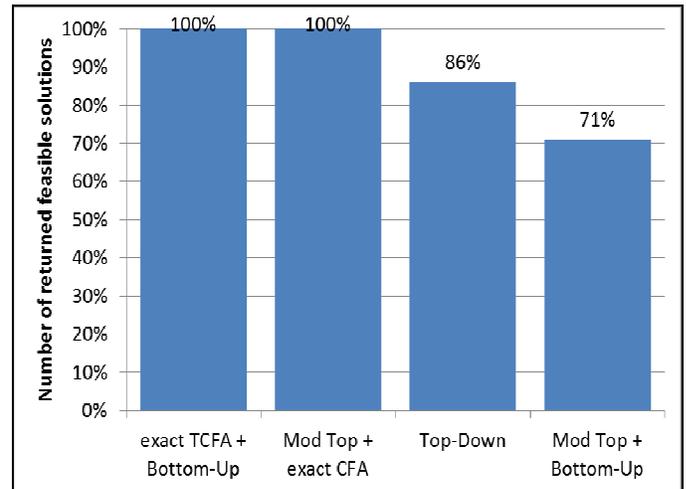


Figure 6. Number of returned feasible solutions.

The number of returned feasible solutions was equal to 100% for methods based on exact CFA or exact TCFA approach. For typically heuristic algorithms (*Top-Down*, connection of *Modified topology* and *Bottom-Up*), results were fewer. This is very important to underline in this comparison, that *Top-Down* returned more feasible solutions for solvable topologies than second heuristic method.

To summarize the evaluation of the *Top-Down* algorithm, it is important to notice that

- It is a method with a high number of returned feasible solutions and short, acceptable processing time, especially in comparison with other methods.
- If optimal solution is a crucial issue – the exact method has to be used.
- When some deviation from optimal solution is acceptable but there is restrictive time constraint then the heuristic method should be applied and in that case *Top-Down* algorithm is a good, candidate tool.

V. CONCLUSION

In the summary of all computational experiments, it is important to emphasize, that TCFA optimization problems are tasks with high computational and memory complexity (especially when the goal function is nonlinear), even for relatively small computer networks.

Moreover, complexity of these problems increases with increasing number of dimensions of input data. The growth of the problem complexity is connected with greater demands for memory and time, which are necessary to find optimal solution by exact methods.

Because of time and technical constraints, there is a necessity to find ways to solve optimization tasks using fewer resources.

This is the main reason and purpose of inventing heuristic methods – algorithms, which can find feasible solutions using fewer resources (e.g., time, memory) than exact method.

The selection of a suitable algorithm is a compromise between the accuracy of the solution and the required resources to solve the problem. Depending on specified requirements, another algorithm may be optimal tool. To choose the best one, some factors to be considered are: computer/technical equipment (environment), time constraint, allowed deviation from optimal solution.

In the further research in the area, the authors are planning to consider the algorithms based partially on the evolutionary approaches, e.g., presented in [10].

There are also several interesting issues that might be considered in the future work on this project. The most important include experiments with: more exact/heuristic algorithms, path-link notation of flow, larger topologies (number of nodes greater than 5) and detailed analysis of computational complexity of algorithms and memory usage following the ideas of multistage experiment design presented in [11].

ACKNOWLEDGMENT

This work was supported by the statutory funds of the Department of Systems and Computer Networks, Faculty of Electronics, Wrocław University of Technology, No S20010W4.

REFERENCES

- [1] T. Miksa, L. Kozzalka, and A. Kasprzak, "Comparison of heuristic methods applied to optimization of computer networks", Proceedings of the 11-th International Conference on Networks, ICN 2012, IARIA, cop. 2012., pp. 34-38.
- [2] K. Walkowiak, "Ant algorithm for flow assignment in connection-oriented networks", *Int. J. Appl. Math. Comput. Sci.*, vol. 15, No. 2, 2005, pp. 205–220.
- [3] B. Gavish, "Topological design of computer communication networks", Proceedings of the 22th Annual Hawaii International Conference on System Sciences, vol. III, 1989, pp. 770-779
- [4] M. Gerla and L. Kleinrock, "On the topological design of distributed computer networks", *IEEE Transactions on Communications*, vol. COM-25, 1977, pp. 48-60
- [5] A. Kasprzak, *Wide Area Networks*, OWPW Publishing House, Wrocław, 2001 /in Polish//.
- [6] R. Goscién, W. Kmiecik, P. Ryba, I. Pozniak-Koszalka, and A. Kasprzak, "Bottom-Up heuristic for CFA problem in WAN", Proceedings of International Conference on Systems Engineering 2012, pp. 85-90
- [7] M. Pioro and D. Medhi, *Routing, Flow and Capacity Design in Communication and Computer Networks*, Morgan Kaufman Publishers, San Francisco, 2004.
- [8] K. Walkowiak, "Heuristic algorithms for assignment of non-bifurcated multicommodity flows", Proceedings of the XXVth International Autumn Colloquium, MARQ, 2003., pp. 243-248.
- [9] M. Gola and A. Kasprzak, "Topology design problem with combined cost criterion and time varying traffic in wild area networks", Proceedings to 17-th IMACS World Congress, Ecole Centrale de Lille, vol. 5, 2005.
- [10] D. Ohia, L. Kozzalka, and A. Kasprzak, "Evolutionary algorithm for solving congestion problem in computer network", *Lecture Notes in Computer Science*, Springer, vol. 5711, 2009, pp. 112-121.
- [11] L. Kozzalka, D. Lisowski, and I. Pozniak-Koszalka, "Comparison of allocation algorithms for mesh structured networks using multistage simulation", *Lecture Notes in Computer Science*, Springer, vol. 3984, 2006, pp.58-67.

Frequency and Time Resource Allocation for Enhanced Interference Management in a Heterogeneous Network based on the LTE-Advanced

Sangmi Moon¹, Bora Kim¹, Saransh Malik¹, Cheolsung Kim¹, Youngil Kim², Kunmin Yeo² and Intae Hwang¹

1) Department of Electronics and Computer Engineering,
Chonnam National University,
Gwangju, Republic of Korea,

msm0804@naver.com, bora54321@naver.com, saranshmk@gmail.com, chskim@chonnam.ac.kr and hit@chonnam.ac.kr

2) Electronics and Telecommunications Research Institute,
Daejeon, Republic of Korea,
yikim@etri.re.kr and kunmin@etri.re.kr

Abstract— In this paper, we propose a resource allocation scheme in the frequency and time domain, to reduce interference in the Heterogeneous Network (HetNet) scenario. In the frequency domain, the macrocell allocates frequency band by using Soft Frequency Reuse (SFR), and the picocell chooses sub-bands that are not used in the macrocell sector, to avoid interference. In addition, to allocate the limited frequency resource is difficult. Therefore, we can manage the cross-tier interference using Almost Blank Subframe (ABS) in the time domain. Simulation results show that the proposed scheme can improve Signal to Interference plus Noise Ratio (SINR), and the spectrum efficiency of macrocell and picocell users. Eventually, the proposed scheme can improve overall cell performance.

Keywords- ABS; HetNet; Interference; Picocell; SFR

I. INTRODUCTION

Recently, small cells have been developed to gain cell subdivision, according to trends of the increase of data rate and bandwidth. Also, cells of hot zone form, such as the picocell, developed in specific areas, are becoming common for offloading explosion data traffic.

Picocells are low-power operator-installed cell towers, with the same backhaul and access features as macrocells. They are usually deployed in a centralized way, serving few users within a radio range of 300m or less, with a typical transmit power range of 23 to 30 dBm. Picocells are mainly utilized for capacity, and also for outdoor/indoor coverage infill with insufficient macro penetration (e.g. office buildings).

A network consisting of small cells, such as the low-cost, low-power picocell over the macrocell, is called a HetNet.

Interference between neighboring picocells and between picocell and macrocell occurs under the HetNet scenario. In general, the two types of interference are as follows [1]:

(a) Co-tier interference: This type of interference occurs among network elements that belong to the same tier in the network. In the case of a picocell network, co-tier interference occurs between neighboring picocells.

(b) Cross-tier interference: This type of interference occurs among network elements that belong to different tiers of the network, i.e. interference between picocells and macrocells as depicted in Figure 1. For example, when a

terminal accessed picocell is located in an edge area of a picocell, access from the picocell can be disconnected, due to interference from the macrocell. This means that the picocell coverage is less than the expectation of the operators [2].

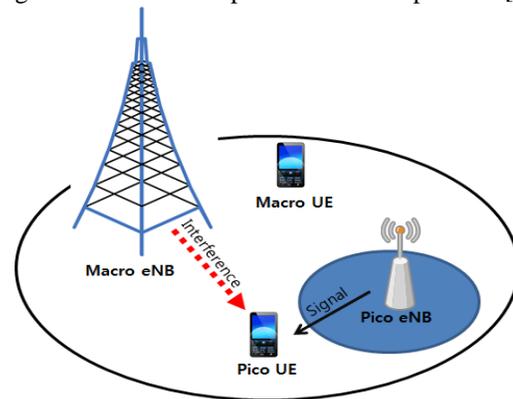


Figure 1. Cross-tier interference in HetNet.

Conventional radio resource management techniques are not suitable for HetNet. Thus, it is essential to adopt an effective and robust interference management scheme that could mitigate the co-tier interference, and reduce the cross-tier interference considerably, in order to enhance the throughput of the overall network.

In this paper, we propose not only co-tier interference management, but also cross-tier interference management for the HetNet scenario. Resource allocation is done by SFR and ABS, to manage the interference in the frequency and time domain, respectively.

The rest of the paper is organized as follows: conventional related work is explained in Section II, Section III shows the proposed scheme, and Section IV shows a performance analysis of the proposed scheme through simulation. Finally, we conclude the results in Section V.

II. CONVENTIONAL RELATED WORK

Long Term Evolution (LTE) - Advanced has developed a low-power, low-cost picocell to reduce the traffic load of the base station. However, there are many problems, such as the

transmission rate, and transmission quality degradation due to interference between macrocell and picocell using the same frequency, in a HetNet scenario deploying picocells over the macrocell.

The 3rd Generation Partnership Project (3GPP) has discussed Interference control technology in an LTE-Advanced system. 3GPP with its REL-8/9 supports Fractional Frequency Reuse (FFR) and Relative Narrowband Transmission Power (RNTP) for interference control. Both of these methods are to reduce the interference level, by adjusting the level of signal that is transmitted through a certain frequency resource. However, picocells are deployed over the existing macrocell network, and share the same frequency spectrum. Due to spectrum scarcity, to eliminate interference is limited in the frequency domain. Therefore, REL-10/11 has introduced the Cooperative Silencing scheme, using ABS in the time domain for interference control [3]. Interference management has been actively discussed in the time domain. Time-Domain Resource Partitioning was proposed [4]. This determines users protected by ABS, and then finds the optimal amount of ABS, by evaluating the overall system utility.

Also, Enhanced Dynamic Spectrum Sharing (EDSS) was proposed for interference management in the picocell. This combines the DSS with FFR for the HetNet scenario. According to the principle of DSS, and the fact that the coverage area of the macro center area (which can be efficiently controlled, by adjusting the transmitting power of the macro system) has no overlap with the coverage area of cell-edge Pico cells. The cell-edge Pico cells can use not only the whole spectrum in the macro cell uplink environment, but also the spectrum in the macro center area downlink environment, while central Pico cells can just use the spectrum in the macro cell uplink environment in TDD mode [5].

FFR allocates resource in the frequency domain for interference management. In addition, SFR is one of the solutions to reduce inter-cell interference in the Orthogonal Frequency Division Multiple Access (OFDMA) based macro system. In FFR, each cell turns off the transmission power of the subband using a Frequency Reuse Factor of 3, but in the case of SFR, this factor can increase the overall system efficiency, by transmitting with a reduced transmission power. Under this condition, cross-tier interference that occurs due to the development of the picocell can be minimized. Additionally, ABS can be applied, for interference management in the time domain. Therefore, we focus on the interference management between the macrocell and the picocell, by allocating resource in the frequency and time domains.

III. THE PROPOSED SCHEME

In this section, we propose a new scheme that allocates resource in the frequency and time domain. In the frequency domain, resource is allocated based on SFR. Additionally, resource is allocated based on ABS in the time domain.

A. Frequency Domain

We allocate the frequency band into macrocell and picocell, based on SFR for intercell interference management, as depicted in Figure 2 (a). The macrocell coverage is divided into a center zone and edge region, each of which include three sectors, denoted by C1, C2, and C3, and E1, E2, and E3. The whole frequency band is partitioned into three parts, denoted by A, B, and C.

For the macrocell, a different frequency sub-band is allocated to each macrocell sub-area, according to the SFR. A reuse factor of one is applied in the center zone, while the edge region adopts a factor of three. The entire frequency bands are used in the center zone (C1, C2 and C3), and sub-bands A, B and C are applied in the E1, E2 and E3 regions, respectively. Also, the overall system efficiency is increased, by reducing the transmission power in the center zone of the macrocell, as illustrated in Figure 2 (b).

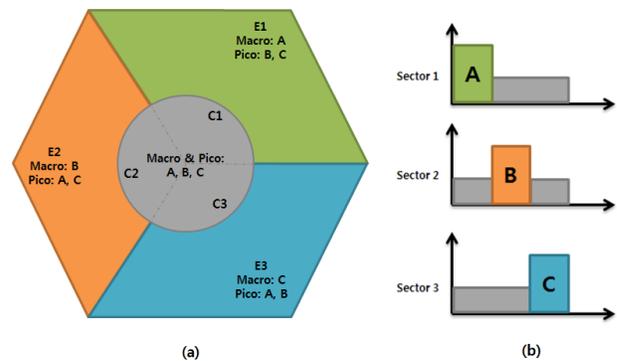


Figure 2. The proposed interference management scheme using SFR in the frequency domain: (a) frequency planning, and (b) power allocation.

Under this circumstance, the picocell chooses sub-bands that are not used in the macrocell sub-area. For example, when a picocell is located in the region E1, it uses sub-bands B and C, while the macrocell uses sub-band A. However, when a picocell is located in the center zone, it uses the entire frequency band, because when a macrocell is located in the center zone, it uses the entire frequency band.

B. Time Domain

When a picocell is located in the center zone of a macrocell, it uses the entire frequency band. The interference problem between the macrocell and picocell still exists. Therefore, we solve the cross-tier interference, by applying the ABS in the time domain.

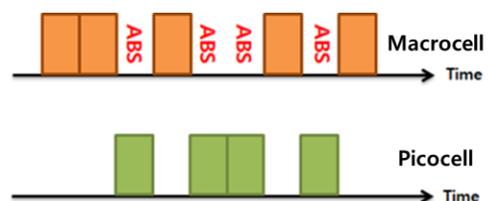


Figure 3. The proposed interference management scheme, using ABS in the time domain.

The aggressor macrocell uses ABS, which doesn't transmit a signal during some subframe for the victim pico user, as depicted in Figure 3. Therefore, the picocell can avoid interference, by transmitting a signal during the ABS of the macrocell.

IV. SIMULATION RESULTS AND PERFORMANCE ANALYSIS

A. Simulation Model and Simulation Parameters

The simulation results are based on 3GPP LTE-Advanced system level simulation parameters [6]. The overall network is composed of 7 macrocells, and picocells are randomly deployed over the macrocells. The number of picocells is 3 in one macrocell coverage. The main simulation parameters are listed in Table I.

TABLE I. SIMULATION PARAMETERS.

Parameter	Value	
	Macrocell	Picocell
Number of Cells	7	3/Macro
Cell Coverage	Radius 1km (ISD = 1,732m)	Radius 250m
Channel Bandwidth	10MHz	
BS Transmit Power	46dBm	30dBm
Number of Users	30	6
Size of Center zone	0.63 of macrocell coverage	
White Noise Power Density	-174 dBm/Hz	
Path Loss	L = 128.1 + 37.6logR (R in km)	L = 140.7 + 37.6logR (R in km)

B. Problem Formulations

To demonstrate the improvement of Quality of Service (QoS) for users, it is necessary to obtain the Signal to Interference plus Noise Ratio (SINR) of these users. In our simulation, the received SINR of a macro user m on sub-carrier k can be expressed by the following definition:

$$SINR_{m,k} = \frac{P_{M,k}G_{m,M,k}}{N_0 + \sum_{M'}(P_{M',k}G_{m,M',k}) + \sum_P(P_{P,k}G_{m,P,k})} \quad (1)$$

where, $P_{M,k}$ and $P_{M',k}$ are the transmit powers of serving macrocell M and neighbor macrocell M' on sub-carrier k , respectively. $G_{m,M,k}$ is the channel gain between macro user m and serving macrocell M on sub-carrier k . Channel gains from neighboring macrocells are denoted by $G_{m,M',k}$. Similarly, $P_{P,k}$ is the transmit power of the neighboring picocell P on sub-carrier k . $G_{m,P,k}$ is the channel gain between macro user m and the neighboring picocell P on sub-carrier k . N_0 is the white noise power spectral density.

In the case of a picocell user, it is interfered from macrocells and the adjacent picocell. The received SINR of a pico user p on sub-carrier k can be similarly given by

$$SINR_{p,k} = \frac{P_{P,k}G_{p,P,k}}{N_0 + \sum_{M'}(P_{M,k}G_{p,M,k}) + \sum_{P'}(P_{P',k}G_{p,P',k})} \quad (2)$$

where, $P_{P,k}$ and $P_{P',k}$ are the transmit powers of the serving picocell P and neighboring picocell P' on sub-carrier k , respectively. $G_{p,P,k}$ is the channel gain between pico user m and serving picocell P on sub-carrier k . The channel gains from neighboring picocells are denoted by $G_{p,P',k}$. Similarly, $P_{M,k}$ is the transmit power of neighboring macrocell M on sub-carrier k . $G_{p,M,k}$ is the channel gain between pico user p and neighboring macrocell M on sub-carrier k . N_0 is the white noise power spectral density.

Also, capacity of the MUE m in macrocell M and the PUE p in picocell P can be expressed by equation (3) and (4), respectively.

$$C_{m,M} = \sum_{k=0}^{N_{RB}} \frac{BW}{N_{RB}} \log_2 (1 + SINR_{m,k}) \quad (3)$$

$$C_{p,P} = \sum_{k=0}^{N_{RB}} \frac{BW}{N_{RB}} \log_2 (1 + SINR_{p,k}) \quad (4)$$

where, BW is the system bandwidth, and N_{RB} is the number of the resource block that assigned the user.

C. Performance Analysis of Proposed Scheme

The proposed scheme is compared with several comparison schemes, as follows. In the No ICIC (Inter-cell Interference Coordination) scheme, both macro and pico users are randomly assigned the resource in time and frequency domain. Also, SFR is the proposed scheme that assigns the resource using only SFR in the frequency domain, and ABS with SFR is the proposed scheme that assigns the resource using SFR and ABS in the frequency and time domain, respectively.

1) Performance Analysis of SINR

Figure 4 shows the Cumulative Distribution Function (CDF) of macro user SINR. In the case of the proposed scheme applying only SFR, co-tier interference between the macrocells is reduced. Also, cross-tier interference between the macrocell and picocell is reduced, because users of picocell located in the macrocell edge use a sub-carrier that is not used by a macrocell in that area. Therefore, the reduction of interference of the cell edge brings improved SINR performance.

The picocell transmits signal with low power. So, a macrocell is less affected by interference. If the center of the macrocell is applied with ABS as well as SFR, the proposed ABS with SFR scheme brings only a slight performance improvement over the scheme applied with only SFR, because the user of the picocell located in the center uses ABS time, which is not used by a macrocell in that area. Otherwise, the sub-carriers are randomly assigned to the macro and pico users. If the same sub-carrier can be used by an edge user in a different cell, co-tier interference occurs. Also, cross-tier interference between the macro user and pico users occurs, when the same sub-carrier can be used by the macro user and pico users who are very close to each other.

Therefore, the No ICIC scheme does not bring SINR performance improvement, because not only co-tier interference, but also cross-tier interference is higher than the proposed scheme.

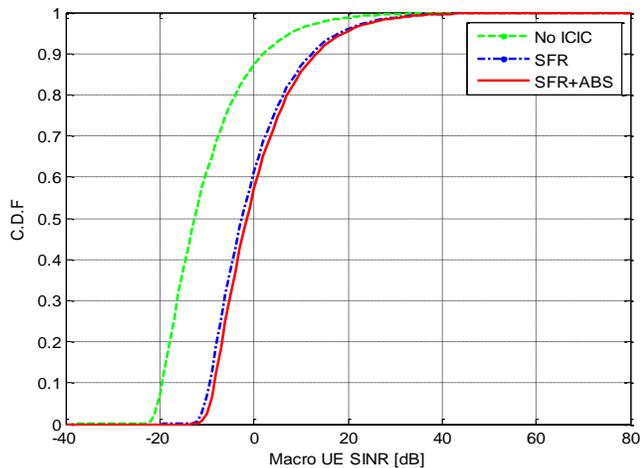


Figure 4. CDF of the Macro User SINR.

Figure 6 shows the CDF of the total user combined macro user and pico user. Simulation results show that the proposed scheme improves the total user SINR.

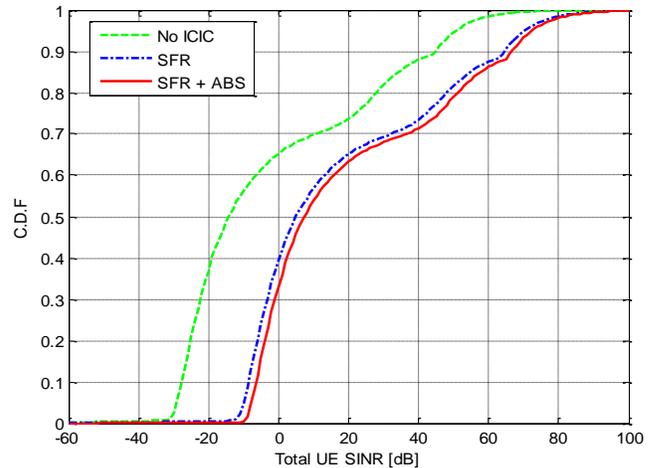


Figure 6. CDF of the Total User SINR.

Figure 5 shows the CDF of the pico user SINR. It can be seen that the SINR performance of the pico user is similar to the SINR performance of the macro user, and is improved in the following order: No ICIC, SFR, and ABS with SFR.

But, the proposed ABS with SFR scheme brings significant performance improvement, compared with the macro user. The reason can be found from Section III. For SFR, performance of the cell edge is improved. Also, with the reduction of cross-tier interference from the macrocell using ABS, SINR performance is improved correspondingly in the center area.

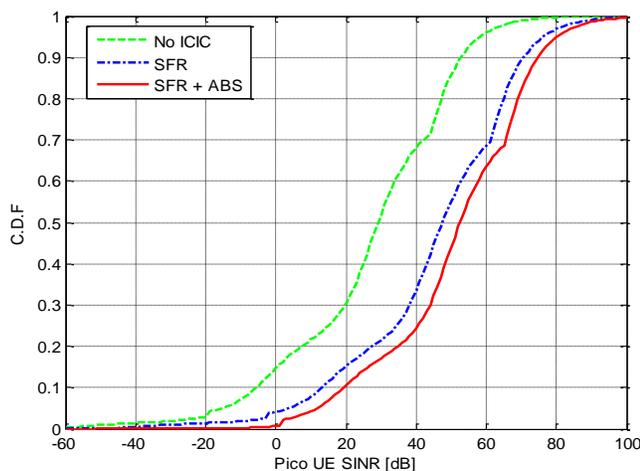


Figure 5. CDF of the Pico User SINR.

2) Performance Analysis of the Spectrum Efficiency

We also evaluated the proposed scheme in terms of spectrum efficiency. Figure 7 and Figure 8 show the CDF of the macro user and pico user spectrum efficiency, respectively, and Figure 9 shows the total user spectrum efficiency.

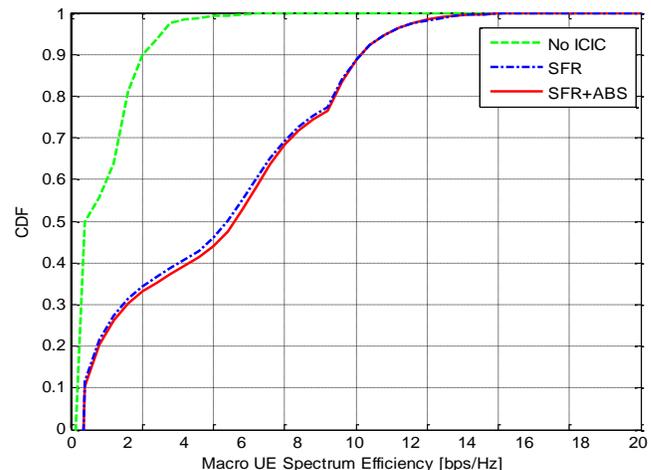


Figure 7. CDF of the Macro User Spectrum Efficiency.

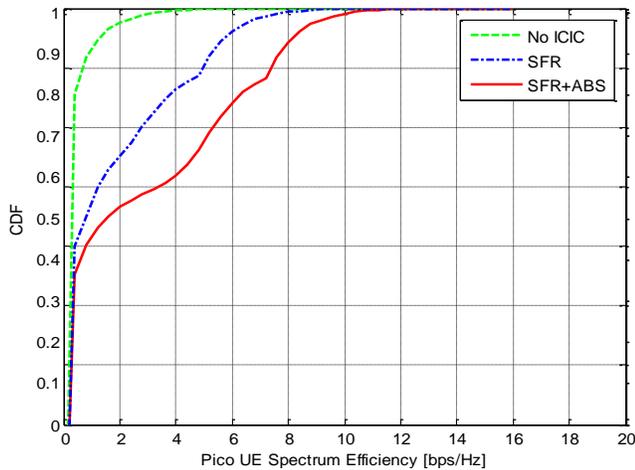


Figure 8. CDF of the Pico User Spectrum Efficiency.

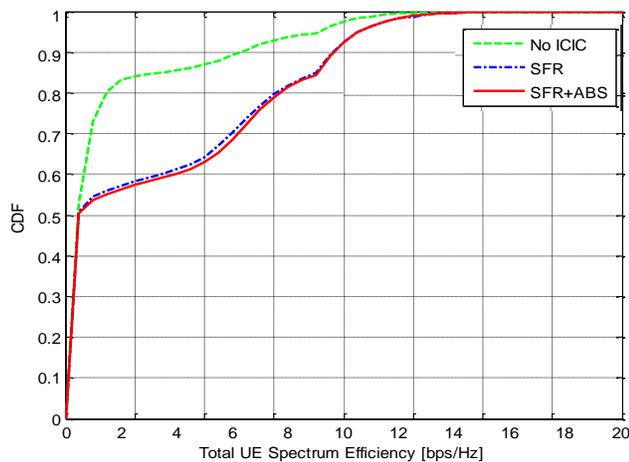


Figure 9. CDF of the Total User Spectrum Efficiency.

It can be seen that the spectrum efficiency is improved in the following order: No ICIC, SFR, and ABS with SFR.

V. CONCLUSIONS

We propose an interference management scheme for the HetNet scenario with coexisting macrocell and picocell, by allocating resource in the frequency and time domains. In the frequency domain, under the macrocell allocating frequency band by the SFR, the picocell chooses sub-bands that are not

used in the macrocell sector, to avoid interference. In addition, to allocate the limited frequency resource is difficult. Therefore, we can manage the cross-tier interference, using ABS in the time domain. Simulation results demonstrate that the proposed scheme can improve the performance, in terms of SINR and spectrum efficiency.

We expect that the proposed scheme configures a more efficient cellular environment based on LTE-Advanced, due to the improvement of performance. Also, this improves the overall cell performance in the HetNet scenario, for the next generation wireless communication environment.

ACKNOWLEDGMENT

This research was supported by the MSIP(Ministry of Science, ICT&Future Planning), Korea, under the ITRC(Information Technology Research Center) support program (NIPA-2013-H0301-13-3005) supervised by the NIPA(National IT Industry Promotion Agency). This study was financially supported by Chonnam National University, 2012. This research is supported by Korea Research Council for Industrial Science and Technology under B551179-12-07-00.

REFERENCES

- [1] N. Saquib, E. Hossain, L. B. Le, and D. I. Kim, "Interference Management in OFDMA Femtocell Networks: Issues and Approaches", *IEEE Wireless Communications*, Vol. 19, June 2012, pp. 86-95.
- [2] Y. W. Yoon, "LTE-Advanced Standards and Technology (REL-10 Trends and REL-11 prospect)", *Journal of the Korea Telecom (Information and Communication)*, Vol. 28, No. 6, May 2011, pp. 61-83.
- [3] R1-104968, Summary of the description of candidate eICIC solutions, 3GPP Std., Madrid, Spain, Aug. 2010.
- [4] J. Pang, J. Wang, D. Wang, G. Shen, Q. Jiang, and J. Liu, "Optimized Time-Domain Resource Partitioning for Enhanced Inter-Cell Interference Coordination in Heterogeneous Networks", *IEEE WCNC*, April 2012, pp. 1613-1617
- [5] M. Li, J. Li, W. Gao, A. Li, Z. Fei, and J. Kuang, "Enhanced dynamic spectrum sharing for multi-cell heterogeneous networks", *International conference on Wireless Communications and Signal Processing (WCSP)*, Nov. 2011, pp. 1-5.
- [6] 3GPP, TR 36.814 v9.0.0, "Further advancements for EUTRA physical layer aspects".

A New No Reference Metric for Estimation Video Quality Based on Blur Effect

Jean Nunes Ribeiro Araujo and Claudio de Castro Monteiro

e-mail: jean.to@gmail.com, ccm@ifto.edu.br

Network Computer Group - GREDES

IFTO Federal Institute of Education, Science and Technology of Tocantins
Palmas, Brazil

Abstract—The measurement and evaluation of video quality is a great challenge for real-time communication. Among those algorithms that evaluate the quality of video, many of them usually make use of reference the original video to compare pixel by pixel with the transmitted video. This fact makes the application of these algorithms complicated in real-time environments. Moreover, these algorithms based on structural similarity do not take into account human visual system information. This paper presents an algorithm that aims to assess the quality of a video sequence using the blurring effect based on human visual system information.

Keywords—video; quality; metric; blur

I. INTRODUCTION

Streaming multimedia content is very sensitive to network conditions, especially in real-time transmissions. The noises introduced into an image in many instances are not perceived by the Human Visual System (HVS). Thus, packet loss and delays may not influence decisively the level of video quality perceived by the viewer. Human perception tends to tolerate more visual distortions since the images have an acceptable level of comprehension. Therefore, the concept of Quality of Experience (QoE) is essential for the development of models for measuring video quality [8], [9].

There are two different ways to calculate the quality of video transmitted over a network. The subjective tests are based directly on visual perception of the viewer, while the objective tests attempt to estimate the quality of a video without the direct intervention of the viewer, simulating the Human Visual System.

The subjective test is known as Mean Opinion Score (MOS) and has higher cost of implementation because it requires adequate space, specific technological resources, planning time and viewers in good eye health. However, the subjective test is presented as more reliable technique to human perception of quality. The recommendation BT. 500-11 ITU-R has specific procedures to perform the subjective tests [6].

Another way is through objective tests. The greatest interest of a metric objective is to be able to replace subjective tests, which are more expensive. Recent surveys show that the objective metrics that consider widely the Human Visual System have better results.

Most objective metrics are Full Reference (FR) or Reduced Reference (RR), only some are Null Reference (NR) [10], [11].

FR metrics determine the quality by comparing the images pixel by pixel of video transmitted over the corresponding images of the original video. RR metric selects some information from both video, and compares to obtain the quality threshold. NR metrics measure quality based on the video itself without receiving any information of the original video.

One of the best known metrics for objective evaluation is the Peak Signal Noise Ratio (PSNR). PSNR is the ratio between the input and output of a lossy compression process, which assesses how much compression added noises in the original image. Therefore, to perform the calculation, one needs to compare the original video and the video transmitted, featuring then a FR metric [1].

The metric PSNR has been widely used and has proven useful in many papers. However, it is not entirely reliable, especially because it takes into account the Human Visual System [10], [12].

In real-time environments it is clearly not feasible to have the original video as a parameter for measuring quality, thus justified the importance of NR metrics.

This paper aims to propose a NR metric for estimating video quality by using the artifact blurring in transmissions in real time [3].

Section 2 presents the description of the metric estimation of video quality using the blurring artifact and the conception of the algorithm. Section 3 illustrates the environment and details the methodology of the experiments. Section 4 describes the analysis of the results obtained in the experiments, highlighting the impact of QoS metrics have on the blur metric, and demonstrate the performance of this metric over the visual quality of the frames. Prediction models are defined in Section 5; the NR blur metric is taking as input the values of the variables of QoS. Section 6 presents the conclusion of our work and final remarks.

II. DESCRIPTION OF BLUR METRIC

Blur effect is caused by a loss of the high frequency content and can occur when we have video sequences, which characterizes well the video streams in a network environment.

In this sense, the estimate of the degree of blurring suffered by a sequence of frames can greatly influence the video quality and be able to demonstrate the current state of a network through the relationship with the QoS metrics. Therefore, the reverse is true if there is a coherent correlation between the metric and QoS metrics jitter, delay and loss.

To find this correlation we must first understand how the blur effect is perceived by people. When comparing a sharp image with this same image blurred, human perception detects a significant difference in terms of loss of detail between the first and second image. However, if we compare an image already blurred and the same image further blurred the difference between the two can be small. This situation is depicted in Fig. 1.



Figure 1. From left to right: original frame, frame transmitted with limited 100Kbps, frame transmitted with limited 50Kbps

An important detail is that, when observing an image containing a small part blurred over a homogeneous area, human perception identifies that the image is blurred, even if only a small part is blurred. For this reason, the analysis of the variation of neighboring pixels, the blurring metric considers only the pixels that have changed after the process that causes the uncomfortable image in [3]. In this sense, the main idea is to determine the degree of blurring suffered by frame, compared to a previous frame.

Considering the phenomena explained, it is then possible to quantify the discomfort caused by the blur effect in a video frame. The first step consists of calculating the metric in determining the variations in intensity between neighboring pixels of an initial frame. Then, calculate the variations between neighboring pixels of the next frame and compare the variations of intensity of the initial frame and the next frame, allowing to evaluate the nuisance blur effect between the two frames.

Thus, a greater variation between the initial frame and the next frame means that the initial frame was clear. However, if there is less variation between the original frame and the next frame means that the initial frame was already blurred. This description is summarized in Fig. 2.

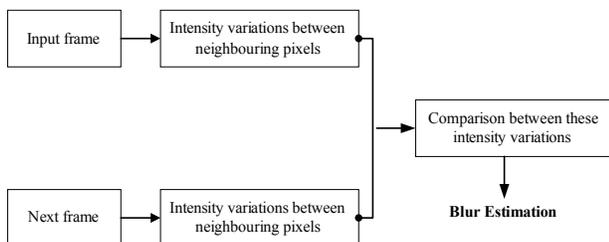


Figure 2. Simplified flow-chart of the blur estimation principle

Importantly, the metric uses frames that were transmitted from the server to the client, not requiring the use of the original video to analyze video quality, unlike the PSNR metric that needs the original frames of reference. Therefore, this estimation of blurring is characterized as a method of gauging the video quality without reference.

This feature eliminates the inconsistency problem that can occur using the PSNR, which may be caused by losses in the transmission frames, which in many cases causes a loss of reference for calculating the quality of certain frames [3].

It is, then, necessary to analyze the variation of neighboring pixels. If this variation between the two frames is high, the initial frame has a higher degree of sharpness. Therefore, in order to represent the degree of blurring, the result is normalized within the defined range between 0 and 1, which are respectively the best and worst quality in terms of perceived blurring. Based on the concepts and equations described in [3] was possible to adapt the metric in the form of algorithm running within an environment of streaming video on demand, returning the value of blurring each frame, allowing the achievement of threshold quality of the video as a whole.

We used a function written in GNU Octave [13], developed for mathematical computation. Initially, we used an implementation of the blurring of perception developed in MATLAB described in [3] and adapted their execution in Octave, in order to insert a loop for process all frames in video sequence.

The original function returns the value of blurring only comparing the original image with a blurred image. As in our model we want to perform video transmission, it was necessary to adapt the metric to function within a repeating structure that calls the function for each frame received by the client.

After transmission of the video in the network, we need to transform the mp4 video received by the user into png images and then execute the program that calculates the blurring.

With these frames stored in a folder, we can change the algorithm of perceived blurring to obtain the quality of each frame. Thus we call the function that calculates the blurring comparing the current frame to the next frame, for later analysis.

III. METHODOLOGY

The structure of the test environment was built with the goal of providing adequate conditions for implementation of multimedia streams, in order to facility the assessment of that the network quality of services causes in the video quality reception.

The physical layout and functionality of computers followed the recommendation G. 1082 ITU-T described in [7], which defines well most environments used in research of this nature. We used four netbooks prepared as described

in Fig. 3; each netbook played certain roles within the proposed scheme.

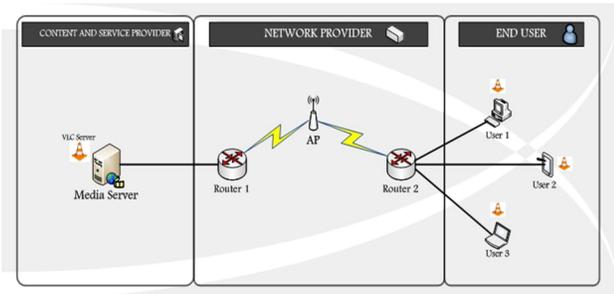


Figure 3. Environment execution of experiments

In the environment we have a media server that delivers content on demand to customers via VLC server. The core network consists of two routers that communicate with each other through a WLAN 802.11g. The video streams stored in the Content Provider and Service Provider are sent via the Network and viewed by End-User. In the user's host is deployed a testbed for optimization and automation of media requests and data collection regarding network transmission and video.

The hosts running the Linux operating system Ubuntu 11.04. The other programs used in the experiments are all based on free software. The videos were encoded in H264/MPEG4 using 300Kbps bitrate and GOP size 12 obtained in [2]. The characteristics of videos used are described in Table I.

TABLE I. DESCRIPTION OF VIDEO SEQUENCES USED IN TRANSMISSIONS

Video	Info	Image
<i>elephants_dream</i>	3500 frames 120 seconds 16:9	
<i>highway</i>	2000 frames 120 seconds 4:3	

The experiments were composed of 4 test groups categorized by the band width limitation imposed on the network for multimedia streams, as shown in Table II.

TABLE II. GROUP OF EXPERIMENTS DEFINED FROM THE BANDWIDTH LIMITATION IMPOSED ON THE NETWORK

Group of experiment	Multimedia traffic limited to:
Group 01	50 Kbps
Group 02	100 Kbps
Group 03	300 Kbps
Group 04	500 Kbps

To determine the sample size needed for population results presented 95% confidence, a sample was defined model of 10 broadcasts each video stream, in order to assess statistical data from the results. We use a bandwidth

limitation for multimedia transmission of 50Kbps, characterized as the worst in our proposal.

Thus, we found the mean and average standard deviation of each population parameter used. Analyzing the results, the delay showed the values of standard deviation higher, so the estimate of the ideal size of population was based on this metric.

Thus, in each group, we executed the 3 video streams described, 135 transmissions for each stream, totaling 270 transmissions/group and 1080 overall transmissions. For each group of experiments, we observed variable values of delay, jitter, and packet loss during transmission of each video stream in order to identify and analyze the impacts of these parameters on video quality NR blur metric.

IV. RESULTS

Based on these procedures, we analyzed the data obtained in the transmission, where it was possible to find the boundary values for each parameter raised to a video that can be delivered with quality customer service. The information collected in each test group used to calculate statistically the influence that each has on QoS metrics to measure objectively without reference BLUR. This determination makes it possible to predict video quality based on network conditions, by analyzing correlations between QoS and variable QoV BLUR.

Based on what was described in the previous section, it was determined that the samples used in each assay were 135 repeats for each video sequence mounted on the environment. To the test groups, the average values of delay, jitter, packet loss, and BLUR were computed for transmission. After the 135 transmissions, the average values for the variables were calculated, establishing a confidence interval for these values, with a significance level of 95%, achieving a margin of error depending on the average standard deviation of each variable examined and the sample size population under consideration.

The blur metric values are shown through Figures 4, 5, 6, 7, 8, 9, 10, and 11. The details of the results of QoS metrics and blur metric are described in Tables III, IV, V, VI, VII, VIII, IX, and X.

A. Group 1 (50 Kbps)

In Figures 4 and 5, we can observe the values of the metric blur in the 135 transmissions performed for both video sequences, subject to the limitation 50Kbps. In this scenario, we can say that the network is in a situation poorly suited for performing multimedia streams, which reflects negatively on the values of delay, jitter, and loss shown in Tables III and IV. The average delay measured exceeded 2000ms for both sequences.

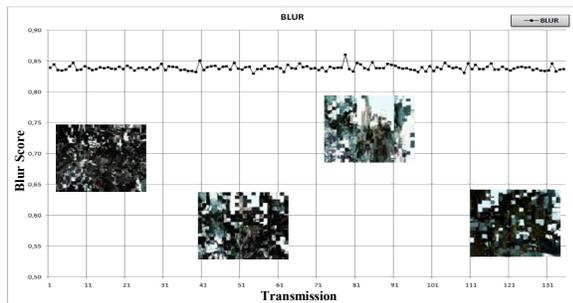


Figure 4. Elephants_dream Blur Score

TABLE III. AVERAGE, MAXIMUM, MINIMUM AND ASSURANCE INTERVAL MEASURED FOR DELAY, JITTER, LOSS AND, BLUR FOR GROUP 01 FOR VIDEO ELEPHANTS_DREAM

	Delay (ms)	Jitter (ms)	Loss	Blur
Avg	2308.50	122.91	15.90	0.8387
Max	2388.28	144.46	24	0.8601
Min	2056.07	92.04	7	0.8291
DP	83.38	11.17	3.61	0.0042
Error	14.06	1.88	0.61	0.0007
Assurance Interval	2294.43 – 2322.56	121.03 – 124.80	15.30 – 16.51	0.8380 – 0.8394

The values of packet loss showed variation due to the characteristics of the video. The first sequence has a larger number of frames, which often change during transmission. So, the average loss was larger than the other sequence.

This fact is observed by the four frames extracted, reflected in the averages of the metric without reference BLUR. By analyzing the blur metric, we conclude that the quality of streaming videos is unsatisfactory for any viewer. However, the sequence *elephants_dream* showed a worse state, as described by the blur metric value, which reached a

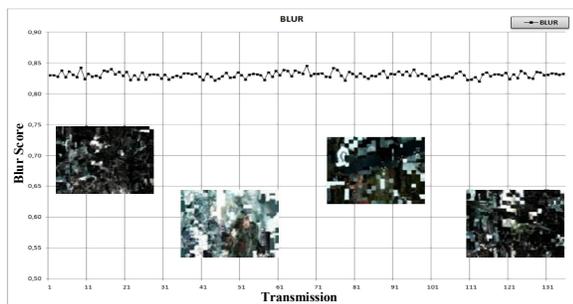


Figure 6. Elephants_dream Blur Score

TABLE V. AVERAGE, MAXIMUM, MINIMUM AND ASSURANCE INTERVAL MEASURED FOR DELAY, JITTER, LOSS, AND BLUR FOR GROUP 02 FOR VIDEO ELEPHANTS_DREAM

	Delay (ms)	Jitter (ms)	Loss	Blur
Avg	1160.99	60.11	9.526	0.8308
Max	1189.97	74.26	20	0.8453
Min	1010.88	46.06	2	0.8200
DP	39.51	5.19	3.032	0.0048
Error	6.66	0.87	0.51	0.0008
Assurance Interval	1154.33 – 1167.66	59.24 – 60.99	9.02 – 10.04	0.8300 – 0.8316

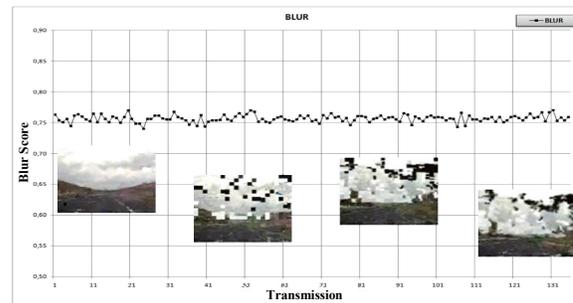


Figure 5. Highway Blur Score

TABLE IV. AVERAGE, MAXIMUM, MINIMUM AND ASSURANCE INTERVAL MEASURED FOR DELAY, JITTER, LOSS, AND BLUR FOR GROUP 01 FOR VIDEO HIGHWAY

	Delay (ms)	Jitter (ms)	Loss	Blur
Avg	2272.78	153.56	10.97	0.7564
Max	2394.24	201.43	21	0.7701
Min	1878.42	103.50	4	0.7399
DP	150.03	16.81	3.01	0.0058
Error	25.3078	2.8361	0.5073	0.0010
Assurance Interval	2247.47 – 2298.09	150.72 – 156.39	10.46 – 11.48	0.7554 – 0.7574

maximum value of 0.8601 and an average of 0.8387, values nearly 10% higher than the other sequence.

B. Group 2 (100 Kbps)

In Figures 6 and 7, we can observe the values of the metric blur in the 135 transmissions performed for both video sequences, subject to the limitation 100Kbps. In this scenario, there were improvements in the conditions of the network, but not enough to assert that the state was suitable for performing multimedia streams. The average delay measured exceeded 1000ms for both sequences.

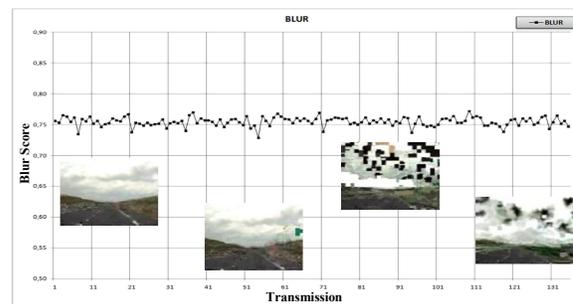


Figure 7. Highway Blur Score

TABLE VI. AVERAGE, MAXIMUM, MINIMUM AND ASSURANCE INTERVAL MEASURED FOR DELAY, JITTER, LOSS, AND BLUR FOR GROUP 02 FOR GROUP 01 FOR VIDEO HIGHWAY

	Delay (ms)	Jitter (ms)	Loss	Blur
Avg	1147.01	77.12	6.50	0.7549
Max	1199.37	99.44	13	0.7720
Min	951.55	53.64	1	0.7287
DP	59.02	8.52	2.29	0.0073
Error	9.96	1.44	0.39	0.0012
Assurance Interval	1137.06 – 1156.97	75.68 – 78.55	6.12 – 6.89	0.7537 – 0.7561

Again, the values of packet loss were higher in the sequence *elephants_dream*. The improved values of QoS metrics is not sufficient to promote a significant increase in the quality of the videos, which was adequately demonstrated by the metric blur.

C. Group 3 (300 Kbps)

In Figures 8 and 9, we can observe the values of the metric blur in the 135 transmissions performed for both video sequences, subject to the limitation 300Kbps. In this scenario, the packet loss in the network reduces to a value close to zero, since the video was encoded with bitrate

300Kbits and therefore found better conditions for transmission.

Thus, with the loss, the other variables network also decrease. We can see more clearly in the frames extracted from sequences, demonstrated in a value of blur metric, which showed lower averages compared to the first two groups. The highway sequence adapted better network conditions by having fewer frames and little variation between the images, showing an average of 0.6787 blur metric.

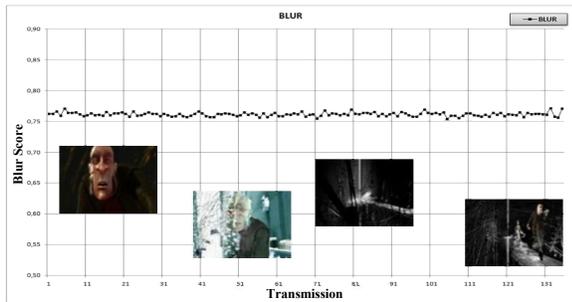


Figure 8. Elephants_dream Blur Score

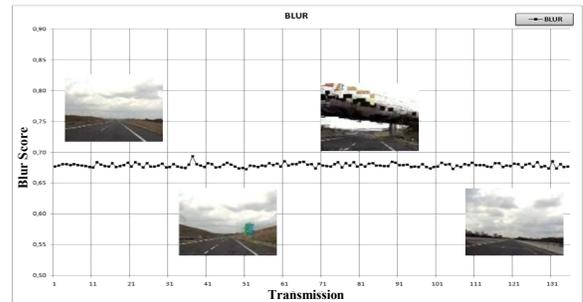


Figure 9. Highway Blur Score

TABLE VII. AVERAGE, MAXIMUM, MINIMUM AND ASSURANCE INTERVAL MEASURED FOR DELAY, JITTER, LOSS, AND BLUR FOR GROUP 03 FOR VIDEO ELEPHANTS_DREAM

	Delay (ms)	Jitter (ms)	Loss	Blur
Avg	241.27	56.41	1.56	0.7615
Max	247.46	61.26	5	0.7712
Min	234.28	46.91	0	0.7537
DP	3.11	2.58	1.22	0.0031
Error	0.52	0.44	0.21	0.0005
Assurance Interval	240.75 – 241.80	55.97 – 56.84	1.36 – 1.77	0.7610 – 0.7620

TABLE VIII. AVERAGE, MAXIMUM, MINIMUM AND ASSURANCE INTERVAL MEASURED FOR DELAY, JITTER, LOSS, AND BLUR FOR GROUP 03 FOR GROUP 01 FOR VIDEO HIGHWAY

	Delay (ms)	Jitter (ms)	Loss	Blur
Avg	313.84	52.98	1.12	0.6787
Max	341.85	69.16	4	0.6931
Min	285.35	46.72	0	0.6723
DP	9.75	3.91	0.94	0.0031
Error	1.64	0.66	0.16	0.0005
Assurance Interval	312.20 – 315.49	52.32 – 53.64	0.96 – 1.28	0.6781 – 0.6792

D. Group 4 (500 Kbps)

In Figures 10 and 11, we can observe the values of the metric blur in the 135 transmissions performed for both video sequences, subject to the limitation 500Kbps. This scenario presents conditions much more suitable for video traffic on the network, significantly reducing the variable values of QoS. With the average delay and jitter less than 15ms, packet loss reduces to a level very close to zero,

reflecting as expected the blur metric and sharpness of the images taken.

The *highway* sequence in this scenario presents values blur allowing qualify frames of videos as well sharp to the viewer. However, the *elephants_dream* sequence has higher average blur, demonstrating that the specific features of each video can cause variations in QoV metrics, even under identical conditions of traffic.

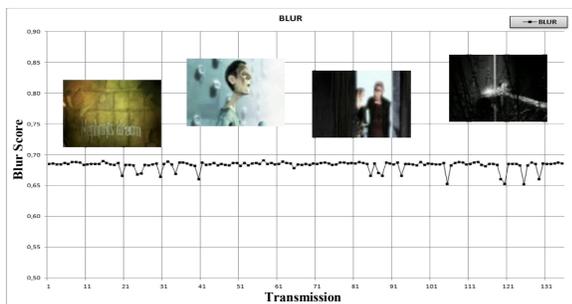


Figure 10. Elephants_dream Blur Score

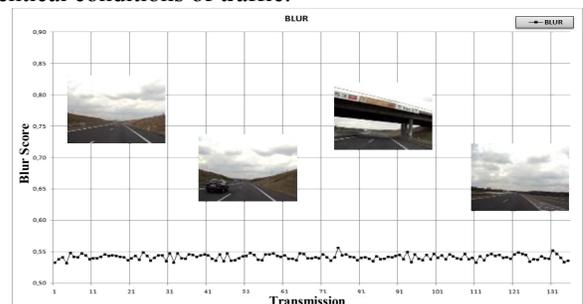


Figure 11. Highway Blur Score

TABLE IX. AVERAGE, MAXIMUM, MINIMUM AND ASSURANCE INTERVAL MEASURED FOR DELAY, JITTER, LOSS AND, BLUR FOR GROUP 04 FOR VIDEO ELEPHANTS_DREAM

	Delay (ms)	Jitter (ms)	Loss	Blur
Avg	17.41	13.53	0.25	0.6832
Max	24.67	18.63	1	0.6912
Min	10.87	7.22	0	0.6525
DP	2.76	2.51	0.44	0.0076
Error	0.47	0.42	0.07	0.0013
Assurance Interval	16.94 – 17.87	13.10 – 13.94	0.18 – 0.32	0.6820 – 0.6845

TABLE XI. ASSURANCE INTERVALS (AI) OF POPULATION PARAMETERS OF VIDEO ELEPHANTS_DREAM

Video: <i>elephants dream</i>				
	AI – Delay	AI – Jitter	AI – Loss	AI – Blur
Group 01 (50Kbps)	2294.43 – 2322.56	121.03 – 124.80	15.30 – 16.51	0.8380 – 0.8394
Group 02 (100Kbps)	1154.33 – 1167.66	59.24 – 60.99	9.02 – 10.04	0.8300 – 0.8316
Group 03 (300Kbps)	240.75 – 241.80	55.97 – 56.84	1.36 – 1.77	0.7610 – 0.7620
Group 04 (500Kbps)	16.94 – 17.87	13.10 – 13.94	0.18 – 0.32	0.6820 – 0.6845

In order to check the impacts of network conditions on each variable used, Tables XI and XII show the assurance intervals for each parameter of the sample population.

As can be observed, the variables delay, jitter, and loss exhibits values that vary depending of terms of limitation and traffic in network. The intervals of confidence show with clarity the differences between the Averages of each parameter for each video transmitted.

We also conclude that the metrics without reference BLUR translates well the quality of a video transmitted in a network based on state of middle of transmission, visa that the same if showed sensitive to variations at the limitations and in incidence of traffic competitor ally to character specific of each sequence of video.

V. PREDICTION MODELS OF BLUR METRIC

Considering a linear relationship between the variables of QoS (delay, jitter, and packet loss) and the Blur metric, we consider a linear model of blur prediction expressed by equation (1).

$$Y = b_0 + b_1 \cdot x_j + b_2 \cdot x_a + b_3 \cdot x_p \quad (1)$$

where, Y = Blur Estimated;
 b_0 = Coefficient of adjustment linear;
 b_1 = Coefficient of jitter;
 x_j = Value average of jitter;
 b_2 = Coefficient of delay;
 x_a = Value average of delay;
 b_3 = Coefficient of loss;
 x_p = Value average of loss;

With the goal of checking the degree of influence that each variable of QoS exercises about the blur metric,

TABLE X. AVERAGE, MAXIMUM, MINIMUM AND ASSURANCE INTERVAL MEASURED FOR DELAY, JITTER, LOSS AND, BLUR FOR GROUP 01 FOR VIDEO HIGHWAY

	Delay (ms)	Jitter (ms)	Loss	Blur
Avg	12.29	12.80	0.09	0.5420
Max	106.96	56.85	2	0.5566
Min	5.74	4.44	0	0.5322
DP	10.71	6.49	0.31	0.0044
Error	1.81	1.09	0.05	0.0007
Assurance Interval	10.48 – 14.10	11.70 – 13.89	0.04 – 0.14	0.5413 – 0.5427

TABLE XII. ASSURANCE INTERVALS OF POPULATION PARAMETERS OF VIDEO HIGHWAY

Video: <i>highway</i>				
	AI – Delay	AI – Jitter	AI – Loss	AI – Blur
Group 01 (50Kbps)	2247.47 – 2298.09	150.72 – 156.39	10.46 – 11.48	0.7554 – 0.7574
Group 02 (100Kbps)	1137.06 – 1156.97	75.68 – 78.55	6.12 – 6.89	0.7537 – 0.7561
Group 03 (300Kbps)	312.20 – 315.49	52.32 – 53.64	0.96 – 1.28	0.6781 – 0.6792
Group 04 (500Kbps)	10.48 – 14.10	11.70 – 13.89	0.04 – 0.14	0.5413 – 0.5427

mathematical models were developed using multiple linear regression. As output, we obtain the estimate of blur in function of variables delay, jitter, and loss, using as input the values collected in the groups of experiments described in the previous session. Within this context, we consider that the variables of QoS act together and identify the state of the network, thus the thresholds of blur are dependent on these metrics. This is expressed through a linear system of equations into (2).

$$\begin{bmatrix} n \\ \sum_{i=1}^n x_{ji} \\ \sum_{i=1}^n x_{ai} \\ \sum_{i=1}^n x_{pi} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n x_{ji} \sum_{i=1}^n x_{ai} \sum_{i=1}^n x_{pi} \\ \sum_{i=1}^n x_{ji}^2 \sum_{i=1}^n x_{ai} x_{1i} \sum_{i=1}^n x_{3i} x_{ji} \\ \sum_{i=1}^n x_{ai} \sum_{i=1}^n x_{ji} x_{ai} \sum_{i=1}^n x_{ai}^2 \sum_{i=1}^n x_{3i} x_{ai} \\ \sum_{i=1}^n x_{pi} \sum_{i=1}^n x_{ji} x_{pi} \sum_{i=1}^n x_{2i} x_{pi} \sum_{i=1}^n x_{pi}^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{ji} y_i \\ \sum_{i=1}^n x_{ai} y_i \\ \sum_{i=1}^n x_{pi} y_i \end{bmatrix} \quad (2)$$

The coefficients b_0 , b_1 , b_2 , and b_3 represent the impact that each variable of QoS exercises on result of blur metric and were obtained through the Gauss-Seidel iterative method [14] for solutions of linear equations, using $k = 8$ iterations for each group of experiments.

After the 8 iterations, 8 values for the coefficients were found; they were used in each of 4 groups, to estimate the values of blur. This form the 8 values of blur estimated for each group, were compared with those obtained experimentally, in order check whether the 8 sets of

coefficients obtained have more coefficient of determination.

Thus, the coefficients that presented more coefficient of determination were the ones chosen to complete the model of prediction of blur.

Using these models, it was possible to also determine equations of prediction generic independent of band limitation and network traffic for each video. For either, we obtained the averages of jitter, delay, loss, and blur for each group of experiments, and we apply over the results, the equations of prediction. Table XIII demonstrates the generic prediction equations for each video.

TABLE XIII. BLUR GENERIC PREDICTION EQUATIONS FOR EACH VIDEO

Video	Prediction Equation	Coefficient of determination
elephants dream	$Y = 0,00021780x_j + 0,00001082x_a + 0,00002303x_p + 0,73916$	0,9963362342
highway	$Y = 0,00014226x_j + 0,00003491x_a + 0,00079661x_p + 0,65460$	0,9994591652

Therefore, according to the values of the QoS metrics, it is possible, by means of these general equations, to determine the degree of blurring the video sequence suffered during transmission.

VI. CONCLUSION AND FUTURE WORK

The analysis of network quality is done using QoS metrics that describes, in practice, the current state of a computer network. These metrics (delay, jitter, and loss) are known well enough and used for purposes of determination of network overload. However, the values of QoS network variables don't shows a clear relationship with the quality of sequences multimedia transmitted over a computer network.

The ideal method is the subjective test, that sets the quality of agreement with the opinion of spectators. However, to implement this model of analysis, it requires a high cost in terms of technological and human resources. The PSNR objective metrics to determine the video quality still have particularities that generate some inconsistency; therefore, they are not fully reliable metrics.

In this context, we present a new prediction model for video quality, based on a no reference metrics, using blur estimate. We adapt and implement the metrics in a network

environment, transmitting video sequences, in order to do several experiments that allowed to get information about the performance of metrics and analyze accordingly the data obtained.

We can conclude, based on results described in this work, that the metrics no reference of blur determination, it is demonstrated be an efficient method of video quality prediction, presenting correlation indexes very good in relationship to QoS variables. Beside this, we notice that the metrics has a very low computational cost, allowing its use in network selection solutions without incremental computational cost to mobile devices.

REFERENCES

- [1] Q. Huynh-Thu and M. Ghanbari, "The accuracy of PSNR in predicting video quality for different video scenes and frame rates," *Telecommunication Systems*, pp. 35-48, 2012.
- [2] Xiph, 2012. <http://media.xiph.org/video/derf/>, [retrieved: May, 2013]
- [3] F. Crete, T. Dolmiere, P. Ladret, and M. Nicolas, "The BLUR effect: perception and estimation with a new no-reference perceptual blur metric," *Human Vision and Electronic Imaging XII*, 64920I, 2007.
- [4] H. Wu, K. Rao, and A. Kassim, "Digital video image quality and perceptual coding," *Journal of Electronic Imaging*, 2007.
- [5] A. Eden, "No-reference estimation of the coding PSNR for H. 264-coded sequences," *Consumer Electronics, IEEE Transactions*, pp. 667-674, 2007.
- [6] ITU-R Recommendation BT.500-11, "Methodology for the subjective assessment of the quality of television pictures," In *International Telecommunication Union*, 2002.
- [7] ITU-T, Recommendation G. 1082, "Measurement-based methods for improving the robustness of IPTV performance," In *International Telecommunication Union*, 2009.
- [8] C. Mingardi and M. Brunner, "IPTV Quality of Service Management in Home," *IEEE Communications Society*, 2009.
- [9] D. Migliorini, E. Mingozi, and C. Vallati, "Performance evaluation of H.264/SVC video streaming over mobile WiMAX," *Computer Networks – Vol. 55*, pp. 3578–3591, 2011.
- [10] C. Farias and A. Rodrigues, "Métricas objetivas para avaliação da qualidade de vídeo," *Simpósio Brasileiro de Sistemas Multimídia e Web – Webmedia*, 2010.
- [11] J. Cardoso, A. Mariano, C. Regis, and M. Alencar, "Comparação das Métricas Objetivas de Qualidade de Vídeos Baseadas na Similaridade Estrutural e na Sensibilidade ao Erro," *Revista de tecnologia da informação e comunicação*, vol. 1, no. 2, pp. 33-40, 2012.
- [12] C. Monteiro, "Um ambiente para apoio a integração de redes sem fio heterogêneas," *Tese de doutorado, Universidade de Brasília*, 2012.
- [13] GNU OCTAVE, 2013. <http://www.gnu.org/software/octave/>, [retrieved: June, 2013]
- [14] R. Larson and D. Falco, "Elementary Linear Algebra," *Pennsylvania State University*, Ed. 7, 2012.

Traffic Steering Framework for Mobile-Assisted Resource Management in Heterogeneous Networks

Anton Dogadaev*, Aleksandra Checko*[†], Andrijana Popovska Avramova*, Anna Zakrzewska*, Ying Yan*, Sarah Ruepp*, Michael S. Berger*, Lars Dittmann* and Henrik Christiansen*

*Department of Photonics Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark

[†]MTI Radiocomp, Hillerød, Denmark

Email: {akdo, amch, apop, azak, yiya, sruru, msbe, ladit, hlch}@fotonik.dtu.dk

Abstract—With the expected growth of mobile data traffic it is essential to manage the network resources efficiently. In order to undertake this challenge, we propose a framework for network-centric, mobile-assisted resource management, which facilitates traffic offloading from mobile network to Wi-Fi or open access small cells. A provision of desired quality of experience to the end-user is carried out by an operator-configurable monitoring application that is running on a mobile device. A potential to enhance network-centric resource management is provided by delegating traffic steering authorities to the network backbone. What is more, we give an overview of existing standardization activities on offloading the mobile traffic through Wi-Fi.

Keywords—Heterogenous Networks; Network Centric; Resource Management; Offloading

I. INTRODUCTION

Cisco forecasts that mobile data traffic will grow 13-fold from 2012 to 2017. As the 4G networks are generating 19 times more traffic than their predecessors [1], it is essential to maintain the network in a way that ensures certain Quality of Experience (QoE) level. In Long Term Evolution (LTE) networks, the load on the network has a crucial impact on achievable user data rates. Improving the network capacity by introducing additional cells challenges interference management. The exponential growth in mobile data traffic also puts additional load on the backhaul network, which becomes congested and requires considerable investments. Therefore, mobile operators would like to use, e.g., the Wi-Fi networks, or open small cells whenever possible to offload the traffic through them.

In this work, we describe a network-centric resource management framework where a mobile facilitates the networks' decisions by gathering feedback and requesting its demands. However, it is the network that makes the final decisions in order to ensure the optimum network operation, in terms of network offloading and power saving in the first place. For the mobile operators, who are interested in reducing their investments on the licensed spectrum and backhaul upgrading, our solution will enable that by facilitating the traffic offload. Users benefit from this approach, as the mechanisms for the network selection are enhanced, therefore connecting over a less loaded cellular network will improve their QoE and reduce the price of connection. Reduction of the traffic cost is especially important, because while data connections in a home country are relatively cheap, in roaming the price becomes considerable.

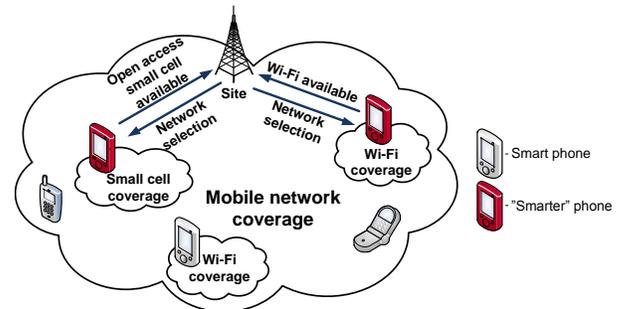


Fig. 1. Mobile network offloading.

In the sections below, the phone, working according to our framework, will be called as a mobile device with "smarter" application on it. Such a "smarter" phone will seamlessly use, e.g., Wi-Fi or open access small cells for offload as shown in Figure 1.

What is more, our platform will enhance possibility of ubiquitous Internet access, complementing services offering global connectivity from one account, like the iPass [2].

The remainder of this paper is organized as follows. In Section II, we state the challenges of resource management and load balancing in heterogeneous networks. Section III introduces related standardization and research work towards enhancement of handover across the standards, traffic offloading, interworking and network selection. In Section IV, we provide a definition of our framework from the network-centric point of view. Section V gives detailed overview of the proposed framework architecture and covers its extension possibilities by employing IEEE 802.11u and IEEE 802.21. Section VI concludes the paper.

II. RESOURCE MANAGEMENT AND TRAFFIC STEERING ISSUES IN HETEROGENEOUS NETWORKS

LTE has been deployed around the world since the end of 2009. For example in the big European cities, where coverage is assured, a lot of people are subscribing to the mobile access through LTE. It is enabled by a growing popularity of the user devices supporting the standard. In the same time the QoE is decreasing as more users share the capacity of the network.

A Wi-Fi connection is a cost efficient way to gain access to the Internet, thanks to its widely spread coverage and availability of the devices that support it. Mobile operators

could use the ubiquitous Wi-Fi coverage to ensure the mobile service accessibility, especially indoors. The opposite solution is also present in the market, where the Wi-Fi routers are connected to the Internet through LTE. It is a clear evidence of the convergence of those two standards, however taking into account the additional load on the mobile infrastructure that it imposes, it should be used carefully and only when needed.

Users are also getting used to the always-on connectivity over flat rate subscription and relying on their smartphones. When they are traveling abroad, the roaming data rates are still considerable. Seamless offloading the traffic to local Wi-Fi hotspots is a convenient solution to this problem. That is the reason for the popularity of the services like iPass, which cover a wide range of Wi-Fi hotspots. However, their coverage could be improved; they also require user to get and configure the client.

Nowadays widely-employed 3G/4G cellular networks are lacking the possibility of traffic offloading and interconnecting with other access networks that may exist in the area. In particular, Wi-Fi networks might be employed for interworking with standards for high-speed mobile communications like LTE/LTE-Advanced (LTE-A). Deployment of this functionality will sufficiently improve traffic load balancing. However, currently existing Wi-Fi networks lack the possibility of global access and authentication. In order to settle this problems ideas of Global Wi-Fi networks and their interworking with the cellular networks received increasing interest in scientific and standardization community.

In today's mobile devices, a lot of control is still left for the user - he/she needs to know how to configure a smartphone or any other network device in order to use Wi-Fi instead of cellular connection whenever possible. What is more, users very often do not have the possibility to configure certain operations to be scheduled when they are connected through Wi-Fi, e.g., to update the installed applications, to backup data from the phone, etc. The mobile devices should require as little as possible interaction from the users, while on the other hand be able to provide (assist) the network with sufficient information which will ensure that the user's expectations of service quality are met.

Mobile network coverage is deployed using macro cells. In densely populated areas, where high peak data rates are expected, small cells can boost the overall network capacity by introducing the additional resources. Small cells can be of micro-, pico- or femtocell size. Some of them can be for enterprise use, but also an open access small cells can be introduced. In both cases offloading the traffic from macro- to small cell will equalize the average cell load, improving the peak achievable data rates.

III. OVERVIEW OF STANDARDIZATION EFFORTS AND RELATED WORK

A lot of standardization efforts have been put in preparing the standards that ensure a seamless vertical handover between wireless technologies. In the tight coupling architecture among heterogeneous networks, a seamless handover can be achieved. The enhanced Generic Access Network (eGAN) [3], [4], a Third Generation Partnership Project (3GPP) standard, is an

example of a tight coupling approach. In eGAN, the mechanisms for exchanging inter-Radio Access Technology (RAT) measurements are defined which enable well timed inter-RAT handover. Due to the fact that tight coupling approach is expensive, complex and requires modification in the terminals and access networks, a loose coupling approach is preferred. The Interworking WLAN (I-WLAN) [5] is a 3GPP standard for loose coupling interworking between WLAN and 3GPP networks. The Mobile IP is used as the mobility protocol, but there are no mechanisms defined for timely triggering of the vertical handover.

The Media Independent Handover (MIH) [6], an IEEE 802.21 standard, facilitates vertical handover using a loose coupling approach. It provides mechanisms for the make-before-break handover by defining initialization and preparation phases prior to the execution of a handover. In the initialization phase the current link is observed and new access networks are detected. In the preparation phase, a decision for handover is made, query regarding the available resources is performed and information, required for mobile IP execution, is retrieved. The IEEE 802.21a amendment to the MIH covers mechanisms for latency reduction during the authentication, as well as authorization and data protection for the MIH services. The main issue that is yet not defined in the standard is how to perform a reservation of the resource in order to ensure the required QoS and methods for translation of the Quality of Service (QoS) context definitions between different technologies and operators [7].

From a user's point of view, maintaining seamless connection to Wi-Fi networks is not as straightforward as connecting to cellular networks. Often users do not recognize the available Wi-Fi networks, and, even more, they get discouraged to use Wi-Fi when credentials are necessary to connect to open or public networks. In order to provide a cellular-like experience and the Global Wi-Fi implementation, the Wi-Fi Alliance Hotspot 2.0 Specification, which references the IEEE 802.11u amendment, has been presented [8]. The IEEE 802.11u [9] aims to provide an overall end-to-end solution for interworking with external networks. It defines Layer 2 transport for a query-response protocol which can be the IEEE 802.21 protocol or the Access Network Query Protocol (ANQP). This allows users to effectively query the network for the information relevant to the network selection prior to performing the authentication procedures.

Nowadays mobile devices are capable of multihoming, allowing simultaneous connection to different networks, receiving and sending data on multiple interfaces at the same time. Seamless mobility thus becomes not just mobility per user, but more per service (flow). The IP Flow Mobility (IFOM) [10] provides the means to select and offload a single flow to a complementary access network. A shortcoming of IFOM is that it is limited to one active 3GPP and one active non-3GPP connections, while all IP flows are intended to one Packet Data Network (PDN) connection [11]. In [12], the missing interface towards the Policy and Charging Control (PCC) system is identified as a limitation for dynamic IFOM. This interface is needed in order to inform the gateways (PDN, signalling gateway) of the routing policies.

In the IEEE 802.21, a MIH Information Server (IS) is defined to maintain information of available networks. On

the other hand, 3GPP has defined the Access Network Detection and Selection Function (ANDSF) to maintain a map of coverage in a form of a static database containing the available access networks in a certain location (e.g., a cell). The mobile nodes can provide ANDSF with its capabilities (IFOM, Multiaccess PDN Connectivity or Seamless Wi-Fi) and location, so that ANDSF can limit the information that is valid for the particular mobile node. However, the standard does not include generation procedures for the policies and the dependence of the requested service, the QoS required for the service, the available signal strength levels, load balancing etc. The issue of a database design and maintenance has not been covered as well. As indicated in [13] and [14] the ANDSF does not provide sufficient tools to the operators for control of network access.

The Fixed/Mobile Convergence initiated the alignment of fixed and mobile management requirements between the 3GPP and the Broadband Forum (BBF). The Technical Report (TR) [15] defines architectural framework for interworking between Service Provider (SP) that offer 3GPP and/or fixed access. The TR considers several functionalities among which is Wi-Fi offloading and IFOM. The Broadband Policy Control Framework (BPCF), described in [16], defines a policy based activation of broadband services as well as policy management and control. The main logical entities are policy decision, policy enforcement points and admission control function. The policies are applied at an IP session, IP flow or at an aggregate level. The motivation and the architectures for interworking between 3GPP system and a fixed broadband access network using WLAN or femto access is elaborated in [17]. The authors indicate the different methods for policy control as a main problem for 3GPP-BBF interworking and discuss the procedures for initial attachment and policy and charging control session establishment.

The research community has also been focusing on suggesting and evaluating different frameworks for traffic steering. Important research directions such as: minimization of delay for the vertical handover, load balancing among networks, traffic offloading etc., have been covered to high extent. Different schemes with respect to user's QoE and optimization of network resources have been proposed in the research literature. For example, a possible procedure for vertical handoff between cellular and Wi-Fi networks were suggested in [18]. However, this paper considers architecture for handover decision management taking into account only users and applications requirements. In [19] the decision algorithm relies on application profile handling, and manages a fleet of mobile nodes. The target of the framework is to either reduce the cost of communication from user point of view or to achieve the required QoS, while considering the energy consumption. Again the operator point of view is not taken into account, nor resource availability at the cellular or Wi-Fi networks. In [20], IEEE 802.21 is evaluated for handover of data services, where the decision is made within a time interval during which the data rates over two networks are compared. In [21], a context aware mobility framework is presented, where decisions are made in a cross-layer and interactive approach. The mobile node is responsible for determining the optimal point of attachment, while the network is responsible for optimizing the network resources.

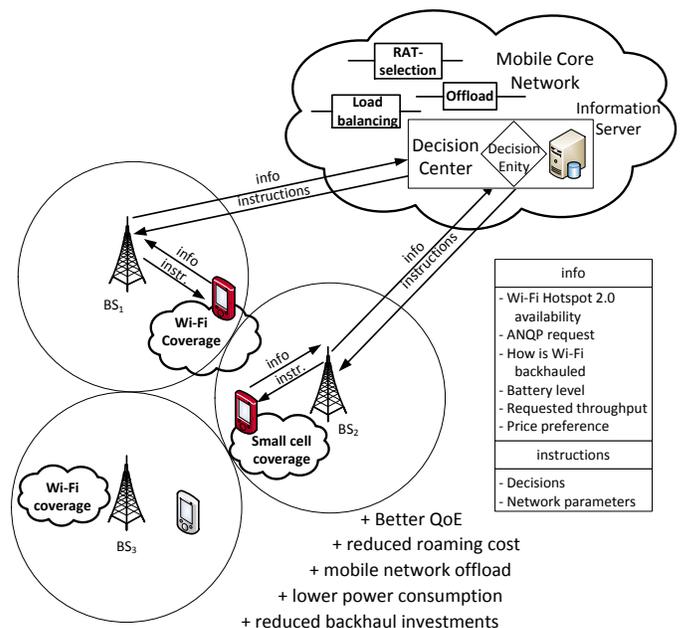


Fig. 2. A framework for mobile-assisted, network-centric resource management and its benefits

In our previous work [22], we have introduced an approach to the network resource management, where the system functionalities are classified according to the time needed to perform particular control action. The framework is universal and applicable to a wide variety of network deployments characterized by multiple RATs, Heterogeneous Networks and different architectures of Base Stations (BSs) such as Distributed BS and Cloud Radio Access Network. In this work, we focus on traffic steering as one functional part of resource management. Here, we propose a framework where both users' profiles per application/preferred operator and network mode are considered in the decision algorithm. More over, we argue that the network centric approach is more appropriate in achieving improved network operation and utilization, while the user is ensured with good connection and service anytime, anywhere.

IV. NETWORK-CENTRIC FRAMEWORK DEFINITION

The proposed resource management framework in this paper adopts the network-centric approach, where assistance from mobile nodes is required. As the final decision is made at the network side, this framework allows the operator to be in control of the network load and resources in general. The mobile nodes are responsible to provide the network with various information such that the decision will improve users' QoE.

As represented by Figure 2, the resource management framework consists of two parts: a network part or decision center and a mobile part or "smarter" application residing on the side of a Mobile Node (MN). The proposed decision center contains two (logical) elements: the Decision Entity and the IS. The Decision Entity is responsible for analyzing data and steering the traffic demands. It has several main operations that can indicate to the MN to perform: selection of a suitable access network from a list of hotspots/small cells, handover to a small cell or a Wi-Fi hotspot, traffic offloading

to a small cell/Wi-Fi, etc. The goals of these operations are to always ensure best connectivity and service, optimize the network operation, balance the load among networks that belong to the operator. The IS represents a database of the up-to-date information gathered from the mobile devices, available eNodeBs and access points. The "smarter" application is responsible to send sufficient information that will aid the network to reach a decision. The users are encouraged to install this application as it will enhance the ubiquitous access to the Internet and improve the experienced QoS. In the following subsections, two approaches are presented: the basic and the enhanced framework.

A. Basic Framework

In the basic framework, the Decision Entity considers the data received from the MN, the network load and operation. Given that the mobile node is the only one able to sense alternative access networks, it is therefore responsible to detect and report a list of available networks. The "smarter" application is responsible to communicate with the decision center and send information on certain time intervals. The time intervals can be set by the decision center. The information gathered about Wi-Fi hotspot may not be just limited to the Service Set Identification (SSID). In case of support of IEEE 802.11u and ANQP, the MN can retrieve more information about the capabilities of the networks such as presented in [8]: Network Authentication Type information, Roaming consortium list, IP Address Type Availability, Hotspot Operator Friendly Name, Operating Class etc. Additionally in case of the support of IEEE 802.21, the MN can retrieve QoS parameters such as throughput, packet error rate and/or different classes of service [6].

All this information can be collected by the IS and made available to the Decision Entity for analysis. In this case the task of the Decision Center is to provide a decision on the MNs that are best suited to be offloaded to alternative access networks and possibly forward the required security credentials for the new access network. The decision is made so that the operator's network is relieved from load, and the users that remain are ensured with the required level of quality.

B. Enhanced Framework

In the enhanced framework, the IS maintains a database (a Profile DB) of available mobile Network Profiles (NPs) and User Profiles (UPs) that indicate their operation and possibility to provide certain QoE. An example of the network profile can be presented as follows:

- *Guaranteed QoS parameters:* The available network is able to ensure a certain level of QoS parameters.
- *Non Guaranteed QoS:* The available network might be unknown to the operator, but can offer free access to the Internet without information on any QoS parameters.
- *Limited Resources:* The available network is capable to ensure a connectivity under certain limitations. For example the throughput per user is limited to a certain threshold, the number of users connected to

the network is limited and/or only certain users are allowed to connect.

In order to assure its delegated functionality, a Profile DB is allocated in the backbone. The Profile DB can be constructed dynamically based on the information from the MNs and/or agreements among network operators. In the second case, the networks need to update their profile in case there is a change. For example, if the network becomes overloaded and the available resources are limited. Additionally, in case of Wi-Fi hotspots, the credentials for access could be available at the Decision Center and therefore simplify the access of MNs to the available Wi-Fi.

The "smarter" application is enhanced as well. It is responsible for a profile activation depending on the type of the service that is requested by the user. The profiles can be based on internal parameters of the MN, that may include, as an example, desired throughput requirements, residual battery lifetime and activated subscription. The information on the active profiles per user is collected at the side of IS as well. The profiles can be defined according to the parameters that have higher priority, for example:

- *Required QoE parameters* profile determines the key quality parameters for the requested service. Here different classes of service can be defined depending of the type of application. For example it can be a high demand on bandwidth for streaming service, low latency for voice services, low packet error rate for file synchronization etc.
- *Minimum price* profile controls traffic consumption of the MN and its applications, minimizing the overall cost of data transmission. The application will look for Wi-Fi networks, checking their quality. Known Wi-Fi networks and security credentials could be stored and marked as the preferred ones. Application will also create UPs in open access Wi-Fi networks on behalf of the user, if needed for connecting, to facilitate the connection.
- *Maximum battery lifetime* profile controls different parameters of the MN, maximizing total time that MN can work on its battery power.

In the enhanced framework, the Decision Entity reaches decisions based on the data received from the MN, the user profiles and network conditions that are reflected through NPs. The proposed architecture and the communication among different entities in the enhanced framework is elaborated in the next section.

V. PROPOSED ARCHITECTURE FOR DATA OFFLOADING

We build our approach on the assumption that MN with "smarter" application offloads traffic to the Wi-Fi network, which is superior in terms of provided QoS parameters and most closely fits to the UP that is currently active on the MN. In the basic approach, the Decision Entity may require MN to offload all traffic each time when Wi-Fi is available for connection. However, in this paper, we propose an advanced technique that potentially can provide an optimal load balancing solution for network operators, still keeping the user of MN satisfied

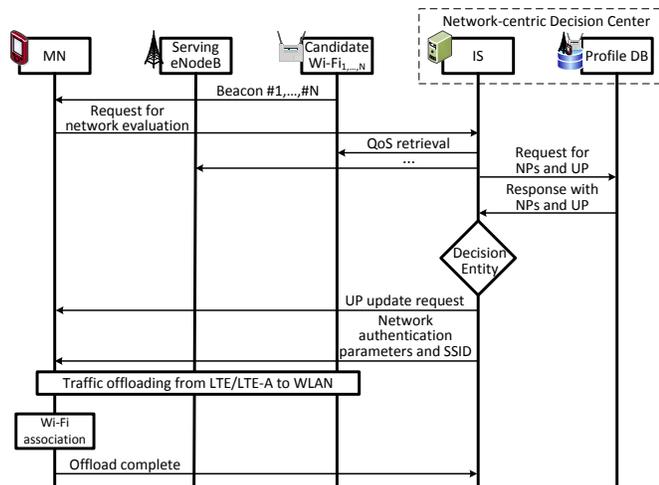


Fig. 3. Proposed sequence diagram for the profile-based traffic offloading from LTE/LTE-A cellular networks towards Wi-Fi.

by the acquired network resources. As represented by Figure 3 a traffic offloading scenario begins on the side of MN by performing network discovery procedure in a passive low power consumption scanning mode by retrieving information about new Wi-Fi hotspots from broadcasted beacon frames. Since the MN may not recognize all received SSIDs and cannot retrieve their QoS parameters as well as an active NP, it transfers the acquired data to the network-centric Decision Center together with the request to evaluate recently discovered Wi-Fi.

Upon the retrieval of network QoS and NPs together with UP of MN that is analyzed for offloading options, the IS executes the algorithm of decision making as depicted by Figure 4. After performing mapping procedures of received SSIDs to NPs from the Profile DB, the Decision Center analyses if the discovered networks meet the requirements and are reasonable for data offloading for a given UP, by executing the following compliance test:

- Requirement 1: NP should support connection of MN that has an active UP.
- Requirement 2: Provided QoS parameters of the network should be equal or greater than those requested by UP.

If both requirements are met, the IS puts SSIDs corresponding to current NP into the set C that defines available candidates for offloading. If after searching procedures it reveals that $C = \emptyset$, the IS sends an information message to a "smarter" application that MN should stay in the serving cellular network. In this case, IS also sends notification that "smarter" application declares too high requirements for the networks that are currently available for offloading. As soon as the final list of the SSIDs that meet requirements of UP is created, the final decision may be derived. We define the task of optimal NP selection that is performed by the IS as an optimization problem that can be formulated as follows:

$$NP^* = \max_{NP_i \in C} f(NP_i, UP), \quad i = 1, \dots, |C|, \quad (1)$$

where the objective function f takes the set C and the UP as initial parameters in order to calculate NP^* that corresponds

to the optimal load balancing solution and maximizes the satisfaction of the end-user by provided network resources. However, the overall solution for the introduced optimization problem highly depends on the final implementation of this traffic steering framework and lies beyond the scope of the current paper.

After transmission of security credentials and the SSID that corresponds to the profile NP^* , offloading procedures are initiated between the MN and selected network. In this case, the MN may simply switch connection to the designated network or employ advanced protocol for traffic offloading and vertical handover, such as the MIH. If necessary, the IS also provides security parameters for authorization and indication of roaming partners, so the MN correctly identifies recently selected access network and is able to proceed with association.

Interworking solutions based on the MIH, like the IEEE 802.21, may also facilitate the task of the vertical handover realization between LTE/LTE-A and Wi-Fi networks as well as provide feasible opportunity for a possible functionality extension of the proposed framework. The MIH IS may provide information about discovered networks and indicate roaming partners for a MN. However, some procedures that are usually performed on higher layers, like pre-association and user authorization, may still require additional interaction with 3GPP and IEEE 802.11 technologies. Information exchange can be carried out by employing the ANQP from the IEEE 802.11u amendment. It can potentially simplify procedures of network identification, providing comprehensive access to information about network operators, type of the service (public, private, paid-for, etc) and connection parameters like cost of the traffic, for example. In order to have a common and up-to-date information about Wi-Fi hotspots and eNodeBs that are available in the area, their operators are supposed to negotiate a joint agreement to share this information with each other. However, we foresee that this requirement may be neglected in the nearest future after overall deployment of Wi-Fi access points with support of the Hotspot 2.0 Technical Specification that includes the IEEE 802.11u amendment. The IEEE 802.11u gives an opportunity to retrieve the information about network specifications through user generated probe requests and responses. The access network information retrieval is conducted by the ANQP, which is transported by the Generic Advertisements Service public action frames,

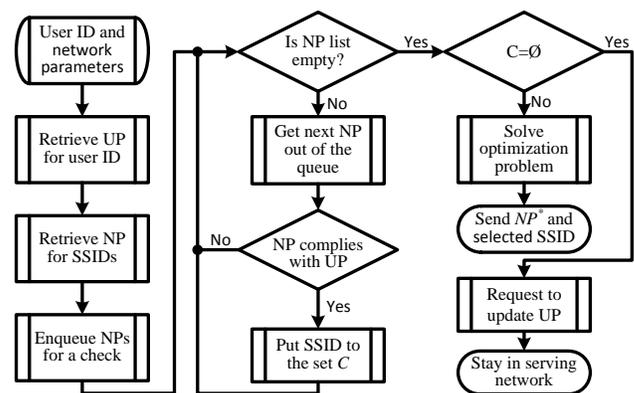


Fig. 4. The state diagram of decision making procedures in the IS.

carrying out transportation frames of higher layer advertisements between Wi-Fi hotspots (or IS) and the MN. The information repository for this type of a protocol, in terms of this framework definition, is provided by the Decision Center. According to the fact that location of this server lies outside the scope of the IEEE 802.11u specification we bind the physical location of the IS and the Profile DB to a backbone network.

VI. CONCLUSION AND FUTURE WORK

In a scope of broad deployment of the 3G/4G access technologies, this paper introduces a novel network-centric framework for offloading decision making supported by the mobile-assisted discovery of candidate-networks.

The fast and reliable approach is provided by allocating the decision center in the backbone, which facilitates optimization of overall network resources and implementation of an efficient load balancing scheme. At the same time the quality of end-user experience is controlled at an appropriate level without noticeable increase in the power consumption on its side due to the passive network discovery. In order to facilitate and accelerate adjustments according to the changes in available network resources and the Quality of Experience demands of the user, a profile-based control is presented for the user and network entities.

As another advantage, the proposed framework also fits currently available IEEE standards for the Media Independent Handover, which facilitates its implementation and retrieval of network parameters. Together with the profile approach, this resource management framework allows a network operator to determine and control the Quality of Experience of the end-user from the network-centric perspective. All the benefits listed above give potential for the presented framework to be successfully implemented and deployed in future mobile communication networks.

Performance evaluation of the proposed solution is the subject of our future work.

ACKNOWLEDGMENT

This work was partially sponsored by the Danish Advanced Technology Foundation through the research project SAIRS and by the Seventh Framework Programme for Research of the European Commission HARP project under grant number 318489.

REFERENCES

- [1] "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2012-2017," White Paper, Cisco, Feb. 2013.
- [2] Enterprise Mobility Services and Global Wi-Fi Network — iPass. [Online]. Available: <http://www.ipass.com/>
- [3] 3GPP, "Generic Access Network (GAN); Stage 2," 3rd Generation Partnership Project (3GPP), TS 43.318 version 10.1.0 Release 10, Apr. 2011.
- [4] 3GPP, "Enhanced Generic Access Networks (EGAN) study," 3rd Generation Partnership Project (3GPP), TR 43.902 version 11.0.0 Release 11, Oct. 2012.
- [5] 3GPP, "3GPP system to Wireless Local Area Network (WLAN) interworking; System description," 3rd Generation Partnership Project (3GPP), TS 23.234 version 11.0.0 Release 11, Sep. 2012.
- [6] IEEE Standard for Local and metropolitan area networks - Part 21: Media Independent Handover Services, IEEE Std. 802.21-2008, Jan. 2009.
- [7] G. Lampropoulos, A. Salkintzis, and N. Passas, "Media-independent handover for seamless service provision in heterogeneous networks," in *Communications Magazine*, IEEE, vol. 46, no. 1, 2008, pp. 64–71.
- [8] "Wi-Fi Certified Passpoint Architecture for Public Access," White Paper, Aruba Networks, 2011.
- [9] 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 9: Interworking with External Networks, IEEE Std. 802.11u-2011, Feb. 2011.
- [10] 3GPP, "IP flow mobility and seamless Wireless Local Area Network (WLAN) offload; Stage 2," 3rd Generation Partnership Project (3GPP), TS 23.261 version 11.0.0 Release 11, Sep. 2012.
- [11] T. Ahmed, S. Antoine, S. Dong, and D. Barankanira, "Multi Access Data Network Connectivity and IP Flow Mobility in Evolved Packet System (EPS)," in *Wireless Communications and Networking Conference (WCNC)*, 2010 IEEE, 2010, pp. 1–6.
- [12] P. Loureiro, M. Liebsch, and S. Schmid, "Policy routing architecture for IP flow mobility in 3GPP's Evolved Packet Core," in *GLOBECOM Workshops (GC Wkshps)*, 2010 IEEE, 2010, pp. 2000–2005.
- [13] M. Corici, A. Diez, D. Vingarzan, T. Magedanz, C. Pampu, and Z. Qing, "Enhanced access network discovery and selection in 3GPP Evolved Packet Core," in *Local Computer Networks*, 2009. LCN 2009. IEEE 34th Conference on, 2009, pp. 677–682.
- [14] M. Corici, J. Fiedler, T. Magedanz, and D. Vingarzan, "Access Network Discovery and Selection in the Future Wireless Communication," in *Mob. Netw. Appl.*, vol. 16, no. 3. Hingham, MA, USA: Kluwer Academic Publishers, Jun. 2011, pp. 337–349. [Online]. Available: <http://dx.doi.org/10.1007/s11036-011-0309-3>
- [15] Broadband Forum, "Interworking between Next Generation Fixed and 3GPP Wireless Networks," Broadband Forum, TR 203, Aug. 2012.
- [16] Broadband Forum, "Broadband Policy Control Framework (BPCF)," Broadband Forum, TR 134, Jul. 2012.
- [17] H. Kim, L. Kim, and A. Kunz, "Enhanced 3gpp system for interworking with fixed broadband access network," *Communications Magazine*, IEEE, vol. 51, no. 3, pp. 88–95, 2013.
- [18] F. Hadji, F. Zarai, and A. Kamoun, "Architecture of mobile node in heterogeneous networks," in *Communications and Information Technology (ICCIT)*, 2012 International Conference on, 2012, pp. 260–264.
- [19] J.-M. Bonnin, I. Lassoued, and Z. B. Hamouda, "Automatic multi-interface management through profile handling," in *Mob. Netw. Appl.*, vol. 14, no. 1. Hingham, MA, USA: Kluwer Academic Publishers, Feb. 2009, pp. 4–17. [Online]. Available: <http://dx.doi.org/10.1007/s11036-008-0117-6>
- [20] S. J. Bae, M. Y. Chung, and J. So, "Handover triggering mechanism based on IEEE 802.21 in heterogeneous networks with LTE and WLAN," in *Information Networking (ICOIN)*, 2011 International Conference on, 2011, pp. 399–403.
- [21] S. Fernandes and A. Karmouch, "Design and analysis of an IEEE 802.21-based mobility management architecture: a context-aware approach," in *Wirel. Netw.*, vol. 19, no. 2. Secaucus, NJ, USA: Springer-Verlag New York, Inc., Feb. 2013, pp. 187–205. [Online]. Available: <http://dx.doi.org/10.1007/s11276-012-0459-7>
- [22] A. Zakrzewska et al., "A Framework for Joint Optical-Wireless Resource Management in Multi-RAT, Heterogeneous Mobile Networks," in *IEEE International Conference on Communications 2013: IEEE ICC'13 - Workshop on Optical-Wireless Integrated Technology for Systems and Networks 2013*, June 2013, unpublished.

A Proposal of Prediction of Peak-Signal-to-Noise-Ratio based on QoS Networks

Carlos Soares Noleto Junior, Claudio de Castro Monteiro
 Network Computer Group - GREDES
 IFTO Federal Institute of Education, Science and Technology of Tocantins
 Palmas, Brazil
 carlosnoleto@gmail.com, ccm @ifto.edu.br

Abstract— Video quality perceived by the human eye is an area of research that is gaining space in today's scientific circles. Nowadays, a lot of media content is made available in the form of videos. The relationship between network quality metrics and quality perceived for these videos assumes great importance, showing that if the requirements of network quality are not achieved the outcome will be impaired. This work presents a proposal to evaluate what the quality of a video would be based only on the quality of the network metrics (jitter, loss and delay), thus attempting to predict what the end-user would perceive using a widely used metric of objective quality (the PSNR) to evaluate the quality of multimedia content.

Keywords- PSNR; multimedia; QoS; QoV;

I. INTRODUCTION

Many researches [2][3][5] have been completed in the area of video quality analysis, due to the enlarging bandwidth offered through the latest mobile network technologies. Among those networks, we cite the standards: Wireless Local Area Network (WLAN) [1], 3G [1] and WiMax [3].

With more bandwidth availability, new frontiers may be opened, increasing user satisfaction with the services that are available. When we talk about evaluating user satisfaction, we refer to the experience that is provided by the use of services offered, from which emerges a term that is becoming popular: Quality of Experience (QoE) [4].

This work intends to analyze the possibility of predicting an objective metric's value for verification of video quality, using QoS information from the network. To do this, we need to evaluate objectively the quality of the video broadcasts, and the impact that the variables of Quality of Service (QoS) have on this value. The metric that has been chosen for video quality analysis was the Peak Signal-to-Noise Ratio (PSNR) [3], due to being commonly used in studies of this nature and for being proven valid for the proposed scenario.

In Section 2, we comment on some works related to Quality of Video (QoV) analysis and QoE in IP networks. In Section 3, we present the scenario and the resources used, as well as the methodology applied on the experiments of this research. In Section 4, the tests accomplished will be described with a summary of the results. In section 5, the results will be analyzed and discussed. Section 6 completes

the report with the conclusion and presentation of possible future works.

II. RELATED WORKS

The growth of network capacity, the standardization of transport protocols, the rising demand for multimedia content, and the standardization of video codecs are among the factors that propel the scientific community to continue research about how to improve the quality of this content, as well as a rational use of modern operating networks. The studies presented below are a few examples of research that contribute to improving the experience of the users when they access multimedia content.

Gomathi et al. [5] proposed the implementation of a transport layer protocol to make available an improvement of the QoS metrics for multimedia content in wireless ad-hoc networks. Furthermore, he proposes the optimization of parameters from the Media Access Control (MAC) layer to meet the objectives of the study, and as a result of the proposal presented, the author got a network delay reduction as well as an increase in the PSNR value.

In [3] the evaluation of the protocol H.264/SVC with multimedia content through WiMax networks is performed, the evaluation using the objective metric (PSNR) and the subjective metric Mean Opinion Score (MOS), making use of a formula that relates the PSNR metric with the MOS metric, and identifying possible factors that influence the quality perceived by the user.

Monteiro [1] worked with mobility management, aiming for the best user experience possible. He also presents a metric that was called Network Quality Metric (NQN) and this metric will be used to analyze network quality in this study.

Forchhammer et al. [6] presented an analysis and processing proposal without references, evaluating video quality without the need of accessing the original video, and as a result managing to obtain the value of PSNR with an error of about 0.3db. For this the coefficients of measured Discrete Cosine Transform (DCT) were used.

In order to compare video quality metrics, D. Z. Rodriguez and G. Bressan [7] compare the quality of information of the metrics PSNR, Structural SIMilarity (SSIM), and Video Quality Metric (VQM) in an environment of stream over IP and digital TV according to the ISDB-T.

As a result, it may be that the quality measured is related to the content and characteristics of the video.

C. C. Monteiro and P. R. L. Gondim [8] presented a proposal for the selection of the network based on characteristics of QoS and QoE for mobile environments, thus defining what would be the best moment for the execution of soft handover between 3G and Wireless Local Area Network (WLAN) networks to improve user experience.

III. METHODOLOGY

Aiming to achieve the objective of this study, a scenario was created containing all the requirements to meet the project's demands. A series of procedures were necessary for the execution process and this section deals with the presentation of this scenario, the resources and methodology used.

The scenario devised for the tests was composed of a multimedia content server that made videos available for the client, and a client that processed the multimedia content provided by the server. All material available on the server could only be visualized on demand, thus giving more flexibility to a client that might use the services at more convenient times.

One of the premises for the development of this study was that all the tools used should be based on free software. Following this concept, the server and the client were computers with Ubuntu 11.04 operating system installed. For content availability and its subsequent acquisition through network transmissions, VLC software [9] was used. VLC is a multimedia content player capable of publishing content for other users in the network.

A great way of simulating congestion on the network is to use bandwidth limiting. This process allows you not to have to generate unnecessary background traffic and guarantees the velocity that will be available for the user. To accomplish this task the Class-based queuing (CBQ) shaper was used. Besides the already mentioned software, a little script written in python was necessary for sampling the state of the network, the program ffmpeg, necessary for the conversion of the videos, and lastly the file psnr.exe together with the program wine for the calculation of the PSNR.

For research development the first step was the creation of a pilot test, having as its objective to make statistical calculations from the data to define the size of the ideal sample for the study. Thus the data used for the definition of the sample size were the jitter data that presented the greatest variation in pattern bypass during the pilot test, becoming the worst case. Below we have the values used for the calculations.

$$\sigma = 12,14ms$$

$$\rho = 95\%$$

$$\theta = 3,0ms$$

where,

σ – Average bypass pattern.

ρ – Level of trust.

θ – Maximum error allowed.

The level of trust for 95% of confidence level in normal distribution is 1,96.

With these values at hand, it was possible to determine the size of the sample using the following formula:

$$\theta = \frac{\sigma * \rho}{\sqrt{\omega}} \quad (1)$$

Changing the values we found

$$3,0 = \frac{12,14 * 1,96}{\sqrt{\omega}} \quad (2)$$

$$\omega \cong 63 \quad (3)$$

We then arrived at the value of 65 repetitions for each evaluated video for a confidence level of 95%.

Fifteen sequences of video were selected for the execution of the tests, all with spatial resolutions of 352 x 288 pixels. All videos acquired in the format .yuv were converted to .mp4 to make possible their transmission through the network. The sequences are presented in Table I.

TABLE I. SEQUENCES OF VIDEO

Name	Video	Frames
Sequence 1	Bridge (Close)	2001
Sequence 2	Bridge (Far)	2101
Sequence 3	Coastguard	200
Sequence 4	Container	300
Sequence 5	Flower	250
Sequence 6	Foreman	300
Sequence 7	Hall Monitor	300
Sequence 8	Highway	2000
Sequence 9	Mobile	300
Sequence 10	Mother and Daughter	300
Sequence 11	News	300
Sequence 12	Silent	300
Sequence 13	Stefan	90
Sequence 14	Tempete	260
Sequence 15	Waterfall	260

After reception, the videos must be recodified in the format .yuv to make calculations of the PSNR possible. The data should be then compared with the PSNR data previously obtained, aiming to identify the real losses caused in this metric during the transmission process.

The simulation of congestion in the network was accomplished via cbq shaper. The videos were codified in the format H264/MPEG4 with a bitrate of 300k. Three distinct speeds were selected for the making of the tests: 200k forcing the lost of packages due to network limitations, 300k allowing the video to be played inside the limits of bandwidth requested, and 400k allowing the transmissions to happen without limitation.

IV. RESULTS

After identifying the quantity of repetitions needed, the quantity of sequences to be evaluated, and the velocities that should be tested, we see what needs to be done: two thousand nine hundred and five tests. In each test we should

verify the information of QoS at the moment of transmission, complete the calculations of the QoV values, and, in the end, record all data in an organized way.

Considering that this process would take a long time to be performed manually, and that little mistakes during its execution could harm the test results, we realized that the creation of a mechanism to automate these proceedings would be of great value, and in this way a testbed was developed.

The testbed developed has the functions of performing the acquisition of the multimedia content of the server, storing the content in a client directory, acquiring and keeping the collected information about the network state during the transmissions, completing the process of modification of the video received from the .mp4 format to .yuv format, performing the PSNR calculations, comparing the original video previously saved in the client’s machine with the newly converted video, and finally consolidating the received data to facilitate the process of tabulation and analysis for the researcher.

As this work does not deal with the presentation of the developed testbed, as it is just a tool used to facilitate the activities performed during the research, the codes and algorithms used for the development of the testbed will not be shown. We will only present its components as well as the functions of each.

According to what is presented in the Table II, the testbed shows only four components. All components were developed in a way that their reuse may be possible in other environments with few modifications.

TABLE II. TESTBED COMPONENTS

Component	Function
cliente.sh	Main component of the framework where the algorithms are implemented. This component has a series of dependences which if not attended do not permit its execution, making the process more secure and ensuring the functioning of all algorithms in a correct form.
Qos.py	Script in python developed for the collection of information about the network. Using the information contained in ping it can return data of QoS from the network like delay, jitter and loss.
Psnr.exe	Program responsible for the realization of comparison of the videos and calculation of PSNR, used in Linux with the help of the program wine.

With the help of the framework the data collection process has become simpler, only one call being needed from the system to perform all the collection and calculation tasks the command line are “cliente.sh 65 300K”.

During the system call, we show that we need to make 65 tests at a speed of 300K. As the quantity of videos has already been defined as 15, 975 tests will be executed, and when finishing all the tests we will have 60 files containing the data to be analyzed, that being four files for each video, one for delay, one for jitter, one for loss and finally one for PSNR. In Tables III, IV and V, we present the averages and standard deviation of data for every transmission rate.

TABLE III. AVERAGE AND SD OF DATA IN 200K

Name	Average Jitter	SD Jitter	Average Delay	SD Delay	Average Loss	SD Loss	Average PSNR	SD PSNR
Sequence 1	30,774	1,849	509,813	7,151	0,585	0,051	22,308	1,409
Sequence 2	31,389	2,216	524,284	7,664	1,708	1,331	25,165	0,044
Sequence 3	94,707	7,033	477,109	7,076	0,462	0,502	17,357	0,194
Sequence 4	90,645	4,827	467,672	10,443	0,215	0,414	22,770	0,140
Sequence 5	110,516	9,113	465,371	9,176	0,185	0,391	10,805	0,064
Sequence 6	102,169	9,154	474,679	9,975	0,662	0,713	15,460	0,228
Sequence 7	89,203	10,912	471,867	46,788	0,231	0,552	22,964	0,172
Sequence 8	32,849	2,414	534,996	6,243	3,138	1,694	18,860	0,499
Sequence 9	96,741	8,108	463,774	7,678	0,569	0,529	11,926	0,103
Sequence 10	99,957	9,108	474,404	8,884	0,462	0,811	25,913	0,283
Sequence 11	100,754	7,360	476,421	10,275	0,215	0,483	21,111	0,437
Sequence 12	97,670	6,353	470,264	11,086	0,015	0,124	23,443	0,258
Sequence 13	200,737	35,328	389,353	80,768	0,077	0,268	15,332	0,118
Sequence 14	76,178	5,335	489,374	33,787	0,277	0,515	13,659	0,139
Sequence 15	72,476	3,395	492,667	10,229	0,169	0,377	21,811	0,216

TABLE IV. AVERAGE AND SD OF DATA IN 300K

Name	Average Jitter	SD Jitter	Average Delay	SD Delay	Average Loss	SD Loss	Average PSNR	SD PSNR
Sequence 1	36,869	1,409	276,815	4,218	0,215	0,414	28,824	0,298
Sequence 2	35,733	1,558	287,778	4,416	0,308	0,497	34,771	0,288
Sequence 3	71,473	4,902	271,467	8,219	0,323	0,471	18,205	0,630
Sequence 4	65,249	3,916	257,472	8,173	0,031	0,174	23,953	0,222
Sequence 5	74,211	17,538	278,522	17,523	0,092	0,384	11,006	0,070
Sequence 6	66,654	4,815	283,352	9,203	0,015	0,124	16,528	0,255
Sequence 7	66,905	2,666	275,372	7,622	0,031	0,174	23,585	0,379
Sequence 8	34,256	2,083	270,862	4,487	0,631	0,719	25,608	0,264
Sequence 9	76,690	8,175	255,558	8,227	0,169	0,377	12,064	0,058
Sequence 10	68,912	3,105	271,854	11,668	0,385	0,490	27,966	0,412
Sequence 11	70,968	3,815	270,132	7,021	0	0	23,106	0,259
Sequence 12	66,501	6,110	259,593	9,947	0	0	24,441	0,333
Sequence 13	164,494	38,250	263,176	24,457	0	0	15,202	0,042
Sequence 14	57,623	5,488	290,460	9,139	0,246	0,434	14,222	0,191
Sequence 15	54,703	4,506	284,247	10,511	0,031	0,174	22,913	0,283

TABLE V. AVERAGE AND SD OF DATA IN 400K

Name	Average Jitter	SD Jitter	Average Delay	SD Delay	Average Loss	SD Loss	Average PSNR	SD PSNR
Sequence 1	7,783	1,280	18,500	15,515	0,015	0,124	29,905	0,418
Sequence 2	7,737	0,856	17,502	1,070	0,015	0,124	36,309	0,043
Sequence 3	52,915	4,812	69,179	5,997	0	0	19,924	0,088
Sequence 4	56,846	11,187	145,048	8,401	0,031	0,174	29,256	0,367
Sequence 5	67,617	8,107	82,099	9,387	0	0	13,605	0,049
Sequence 6	59,477	10,409	113,302	7,093	0	0	19,782	0,342
Sequence 7	54,778	4,937	141,948	7,138	0,015	0,124	30,090	0,363
Sequence 8	7,071	0,615	11,164	1,115	0,277	0,450	26,484	0,040
Sequence 9	48,253	2,732	32,099	5,996	0	0	13,229	0,045
Sequence 10	53,232	4,264	141,930	7,768	0	0	29,758	0,356
Sequence 11	59,768	7,976	151,949	9,116	0,231	0,424	27,080	0,448
Sequence 12	54,748	6,364	142,035	7,409	0	0	26,820	0,348
Sequence 13	134,811	40,128	166,845	30,067	0,123	0,331	17,937	0,460
Sequence 14	58,334	5,839	61,318	7,389	0	0	16,332	0,093
Sequence 15	69,268	14,547	146,231	7,234	0,231	0,424	26,613	0,297

After all the transmissions have been completed and all the files generated for each velocity, we will have a great quantity of data to be analyzed. As this work has as an objective to find any relation, if it exists, among the variables of QoS with the video quality calculated by PSNR, it is necessary to do multiple linear regression. This way we can verify how each variable influences PSNR, and if any variable can.

V. DATA ANALYSIS

Now, possessing all data in an organized form, tabulated in a spreadsheet, we can perform the next step to try to identify some relation among the variables involved. To do this we use the process of normal linear regression, aiming to establish a relation between the variable PSNR and the network metrics collected, jitter, delay and loss, that were evaluated as independent variables as can be seen below.

$$\gamma = \beta_0 + \beta_1 x_j + \beta_2 x_a + \beta_3 x_p \quad (4)$$

where,
 γ – Estimated PSNR
 β_0 – Coefficient of linear adjust.

β_1 – Coefficient of jitter.
 x_j – Average value of jitter
 β_2 – Coefficient of Delay.
 x_a – Average Delay value
 β_3 – Coefficient of loss.
 x_p – Average Loss value.

The values of $\beta_0, \beta_1, \beta_2$ and β_3 represent the values of the angular variables of each independent variable, and influence directly the result of the estimated PSNR. To find these values we use the interactive regression method of Gauss-Seidel using K=8 interactions for each broadcasted video at each velocity. At the end of each group of interactions we obtained 8 values of PSNR that were then compared to the values obtained during the tests in order to identify which one had the best coefficient of correlation with the real data that was used in the following formula:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - Y_i)^2}{\sum_{i=1}^n y_i^2 - \frac{1}{n} (\sum_{i=1}^n y_i)^2} \quad (5)$$

where,
 R^2 – Coefficient of correlation

Y – Value of PSNR obtained on the tests

45 functions, defined by which function presented the best relevance through the coefficient of correlation, were obtained, derived from the 15 sequences of videos selected and the 3 velocities of transmission. The functions are presented below accompanied by the coefficient of correlation for each one, the values used having a precision of 5 decimal places.

TABLE VI. FUNCTIONS OF SEQUENCE 1

Sequence 1		
Rate	Function	R ²
200K	$\gamma = 22,30275 - 0,00008\beta_{1_{xj}} - 0,00001\beta_{2_{xa}} - 0,01629_{xp}$	0,99994
300K	$\gamma = 28,82447 + 0,00000\beta_{1_{xj}} - 0,00000\beta_{2_{xa}} + 0,12414_{xp}$	0,99975
400K	$\gamma = 29,90543 - 0,00488\beta_{1_{xj}} - 0,00968\beta_{2_{xa}} - 1,88240_{xp}$	0,99467

TABLE VII. FUNCTIONS OF SEQUENCE 2

Sequence 2		
Rate	Function	R ²
200K	$\gamma = 25,19843 + 0,00110\beta_{1_{xj}} + 0,00006\beta_{2_{xa}} + 0,01208_{xp}$	0,97232
300K	$\gamma = 34,70796 - 0,00193\beta_{1_{xj}} - 0,00021\beta_{2_{xa}} + 0,00806_{xp}$	0,95519
400K	$\gamma = 36,31333 + 0,00048\beta_{1_{xj}} + 0,00032\beta_{2_{xa}} - 0,12876_{xp}$	0,99994

TABLE VII. FUNCTIOS OF SEQUENCE 3

Sequence 3		
Rate	Function	R ²
200K	$\gamma = 17,39035 + 0,00034\beta_{1_{xj}} + 0,00007\beta_{2_{xa}} - 0,07372_{xp}$	0,98514
300K	$\gamma = 18,61859 + 0,00551\beta_{1_{xj}} + 0,00158\beta_{2_{xa}} - 0,02556_{xp}$	0,99509
400K	$\gamma = 19,93211 - 0,00005\beta_{1_{xj}} + 0,00009\beta_{2_{xa}} - 0,12041_{xp}$	0,99916

TABLE IX. FUNCTIOS OF SEQUENCE 4

Sequence 4		
Rate	Function	R ²
200K	$\gamma = 22,78018 + 0,00010\beta_{1_{xj}} + 0,00002\beta_{2_{xa}} - 0,00433_{xp}$	0,99920
300K	$\gamma = 23,97115 + 0,00028\beta_{1_{xj}} + 0,00005\beta_{2_{xa}} - 0,14057_{xp}$	0,99899
400K	$\gamma = 29,21160 - 0,00084\beta_{1_{xj}} - 0,00030\beta_{2_{xa}} - 0,37113_{xp}$	0,99994

TABLE X. FUNCTIONS OF SEQUENCE 5

Sequence 5		
rate	Function	R ²
200K	$\gamma = 10,81638 + 0,00009\beta_{1_{xj}} + 0,00002\beta_{2_{xa}} - 0,02618_{xp}$	0,99999
300K	$\gamma = 10,98440 - 0,00024\beta_{1_{xj}} - 0,00007\beta_{2_{xa}} - 0,06850_{xp}$	0,99356
400K	$\gamma = 13,60414 - 0,00002\beta_{1_{xj}} + 0,00001\beta_{2_{xa}} - 0,00133_{xp}$	0,99999

TABLE XI. FUNCTIONS OF SEQUENCE 6

Sequence 6		
Rate	Function	R ²
200K	$\gamma = 15,496300 + 0,00031\beta_{1_{xj}} + 0,00007\beta_{2_{xa}} + 0,00482_{xp}$	0,99600
300K	$\gamma = 16,54088 + 0,00081\beta_{1_{xj}} + 0,00001\beta_{2_{xa}} - 0,26808_{xp}$	0,99996
400K	$\gamma = 19,81150 - 0,00010\beta_{1_{xj}} + 0,00040\beta_{2_{xa}} - 0,11325_{xp}$	0,98254

TABLE XII. FUNCTIONS OF SEQUENCE 7

Sequence 7		
Rate	Function	R ²
200K	$\gamma = 22,96395 + 0,00002\beta_{1_{xj}} - 0,00000\beta_{2_{xa}} - 0,03407_{xp}$	0,99999
300K	$\gamma = 23,60513 - 0,00001\beta_{1_{xj}} + 0,00009\beta_{2_{xa}} + 0,06614_{xp}$	0,99793
400K	$\gamma = 30,10198 + 0,00016\beta_{1_{xj}} + 0,00011\beta_{2_{xa}} + 0,02540_{xp}$	0,99474

TABLE XIII. FUNCTIONS OF SEQUENCE 8

Sequence 8		
Rate	Function	R ²
200K	$\gamma = 18,86019 + 0,00007\beta_{1_{xj}} - 0,00000\beta_{2_{xa}} + 0,00747_{xp}$	0,99995
300K	$\gamma = 25,90413 + 0,00810\beta_{1_{xj}} + 0,00110\beta_{2_{xa}} + 0,20133_{xp}$	0,97307
400K	$\gamma = 26,48370 - 0,00011\beta_{1_{xj}} - 0,00001\beta_{2_{xa}} + 0,02009_{xp}$	0,99999

TABLE XIV. FUNCTIONS OF SEQUENCE 9

Sequence 9		
Rate	Function	R ²
200K	$\gamma = 11,92616 - 0,00004\beta_{1_{xj}} + 0,00000\beta_{2_{xa}} - 0,05556_{xp}$	0,99997
300K	$\gamma = 12,03541 - 0,00031\beta_{1_{xj}} - 0,00011\beta_{2_{xa}} - 0,06790_{xp}$	0,99862
400K	$\gamma = 13,22607 - 0,00004\beta_{1_{xj}} + 0,00008\beta_{2_{xa}} - 0,02882_{xp}$	0,99900

TABLE XV. FUNCTIONS OS SEQUENCE 10

Sequence 10		
Rate	Function	R ²
200K	$\gamma = 25,58458 - 0,00369\beta_{1_{xj}} - 0,00064\beta_{2_{xa}} - 0,39505_{xp}$	0,99015
300K	$\gamma = 27,99559 + 0,00001\beta_{1_{xj}} + 0,00003\beta_{2_{xa}} - 0,34296_{xp}$	0,99922
400K	$\gamma = 29,78410 + 0,00069\beta_{1_{xj}} + 0,00014\beta_{2_{xa}} + 0,10573_{xp}$	0,99995

TABLE XVI. FUNCTIONS OF SEQUENCE 11

Sequence 11		
Rate	Function	R ²
200K	$\gamma = 21,06641 - 0,00061\beta_{1_{xj}} - 0,00009\beta_{2_{xa}} + 0,02737_{xp}$	0,99969
300K	$\gamma = 23,11265 + 0,00000\beta_{1_{xj}} + 0,00003\beta_{2_{xa}} - 0,40068_{xp}$	0,99939
400K	$\gamma = 27,06671 - 0,00047\beta_{1_{xj}} - 0,00004\beta_{2_{xa}} + 0,06318_{xp}$	0,99992

TABLE XVII. FUNCTIONS OF SEQUENCE 12

Sequence 12		
Rate	Function	R ²
200K	$\gamma = 23,41065 - 0,00019\beta_{1xj} - 0,00007\beta_{2xa} + 0,17841_{xp}$	0,99624
300K	$\gamma = 24,30987 - 0,00207\beta_{1xj} - 0,00051\beta_{2xa} - 0,01020_{xp}$	0,99593
400K	$\gamma = 26,88529 + 0,00005\beta_{1xj} + 0,00053\beta_{2xa} - 0,45183_{xp}$	0,99998

TABLE XVIII. FUNCTIONS OF SEQUENCE 13

Sequence 13		
Rate	Function	R ²
200K	$\gamma = 15,32524 - 0,00003\beta_{1xj} - 0,00002\beta_{2xa} - 0,05070_{xp}$	0,99246
300K	$\gamma = 15,22704 + 0,00026\beta_{1xj} + 0,00004\beta_{2xa} + 0,01078_{xp}$	0,99488
400K	$\gamma = 17,84301 - 0,00088\beta_{1xj} - 0,00029\beta_{2xa} + 0,15210_{xp}$	0,99996

TABLE XIX. FUNCTIONS OF SEQUENCE 14

Sequence 14		
Rate	Function	R ²
200K	$\gamma = 13,65806 - 0,00013\beta_{1xj} - 0,00000\beta_{2xa} - 0,02347_{xp}$	0,99986
300K	$\gamma = 14,32875 + 0,00227\beta_{1xj} + 0,00032\beta_{2xa} - 0,02272_{xp}$	0,99746
400K	$\gamma = 16,33751 + 0,00015\beta_{1xj} + 0,00037\beta_{2xa} + 0,01849_{xp}$	0,99996

TABLE XX. FUNCTIONS OF SEQUENCE 15

Sequence 15		
Rate	Function	R ²
200K	$\gamma = 21,72162 - 0,00113\beta_{1xj} - 0,00019\beta_{2xa} + 0,08800_{xp}$	0,99999
300K	$\gamma = 22,91307 - 0,00005\beta_{1xj} + 0,00001\beta_{2xa} - 0,39961_{xp}$	0,98264
400K	$\gamma = 26,57632 - 0,00008\beta_{1xj} - 0,00034\beta_{2xa} + 0,15489_{xp}$	0,99728

The functions listed in the Tables VI-XX show that it is possible to find a relation between the metrics of the network and the quality of videos transmitted, taking into consideration the values of the Coefficient R².

VI. CONCLUSION AND FUTURE WORKS

This article proposed a model for the prediction of video quality based on an objective metric (PSNR), through the use of the QoS parameters of the network.

With the data analysis collected during the experiments and the use of linear regression on the data, we could detect a relation between the metrics of objective quality evaluated (PSNR) and the metrics of QoS of the network. The functions found demonstrate this relation, as the PSNR suffers changes depending on the characteristics of the videos. Because of this, we have great differences in the generated formulas.

The study demonstrated the possibility of the development of applications that use the parameters of a network's QoS to predict the values of PSNR of a

predetermined content in a way that the user will always have the best experience possible accessing multimedia content. For example, we may cite an application that uses this model to define what would be the best moment to make a change in the access network, or what video resolution should be made available in a predetermined environment.

For future studies, we intend to use other objective metrics to verify if these aspects are confirmed. We intend to compare subjective metrics, objectives and the parameters of the network in order to consolidate the results of this research.

REFERENCES

[1] C. C. Monteiro, "Um ambiente para apoio à integração de redes sem fio heterogêneas". Tese (Doutorado em Engenharia Elétrica) – Universidade de Brasília, Brasília, 2012.

[2] Q. Huynh-Thu and M. Ghanbari, "The accuracy of PSNR in predicting video quality for different video scenes and frame rates" *Telecommunication Systems* – no. 49 pp. 35-48, 2012.

[3] D. Migliorini, E. Mingozzi, and C. Vallati, "Performance evaluation of H.264/SVC video streaming over mobile WiMAX". *Computer Networks* no. 55 pp. 3578–3591, 2011.

[4] A. Vishwanath, P. Gupta, P. DUTTA, S. Kalyanaraman, M. Chetlur, and A. Ghosh, "Perspectives on Quality of Experience for Video Streaming over WiMAX ACM SIGMOBILE Mobile" *Computing and Communications Review* no. 13 Issue 4 pp. 15-25 , 2009.

[5] N. Gomathi, P. Seethalakshmi, and A.Govardhan, "A cross layer Design to Enhance Throughput for Multimedia Streaming over Mobile Ad hoc Networks" *International Journal on Computer Science and Engineering* no. 3 No. 1 pp. 114-126 , 2011.

[6] S. Forchhammer, H. Li, and J. D. Andersen, "No-reference analysis of decoded MPEG images for PSNR estimation and post-processing" *Journal of Visual Communication and Image Representation*, no. 22, n. 4, pp. 313-324, 2011.

[7] D. Z. Rodriguez and G. Bressan, "Video Quality Assessments on Digital TV and Video Streaming services using Objective Metrics" *Latin America Transactions, IEEE (Revista IEEE America Latina)*, no. 10, n. 1, pp. 1184-1189, 2012.

[8] C. C. Monteiro and P. R. L. Gondim "Video quality guarantee for mobile users across WLAN/3G networks" *Advanced Communication Technology (ICACT), 13th International Conference, IEEE*. Pp. 1075-1079, 2011.

[9] VLC media player. <http://www.videolan.org/vlc/>, 2013.

Effective ARP Based Call Admission Control (CAC) Scheme for LTE Systems

Abhishek Nayak, Saurabh Verma, Anurag Jain

HCL Technologies Ltd.

Noida, India

abhishek_n@hcl.com, saurabh_verma@hcl.com, anuragjn@hcl.com

Abstract-- In wireless communication technologies like GSM, WCDMA, LTE, etc. Radio is stressed by the various traffic conditions, in terms of radio resources to provide the admission of new calls and required QoS. There are situations and scenarios, where the load takes unprecedented conditions and as part of services the requirements may change, like to improve on non-blocking calls though recessing on Quality of Service (QoS). Though such situations and scenarios are intermittent and last for limited periods of time, the importance of services becomes operator's credentials. We propose a top-up algorithm, which does not interfere with existing admission control mechanism in direct way, but provide them with effective Allocation and Retention Priority (E-ARP) for each call. This algorithm is based on "inferences and intentions based combinational logic". It works on the system specific inferences of certain parameters and intentional classification of existing and new demands based on weightage-based combinational logic, which results in modification of ARPs into effective ARPs. For evaluation of proposed Admission Control (AC) scheme blocking and outage probability of calls is used as performance metrics.

Keywords – LTE, RRM, Admission Control, Effective ARP, SON, Self organizing networks, scheduling algorithms, spectral efficiency, next generation networks, NGN, radio resource management, combinational logic, inferences and intentions.

I. INTRODUCTION

The task of radio admission control is to admit or reject a new connection request referred to as a call, depending on whether the required QoS of the new call request will be fulfilled. Each call request has its own ARP value, which describes its relative importance for admission and retention. In the proposed scheme we deduce effective ARP (EARP) for each call based on our top-up algorithm. The algorithm modifies the relative importance of the calls based on their inferred impact on the system and intentional bias for the services.

In industry perspective, admission control mechanisms are provided through vendor specific algorithms. There are various kinds of scheduler algorithms [4][5], available in LTE systems, like Proportional Fair, Modified Largest Weighted Delay First, and Exponential Proportional Fair etc. These algorithms take best use of system and resources for the specific load conditions. With the advancement of radio technologies like LTE, the spectrum efficiency of the system is of much concern from the vendors perspectives; that is to say, to provide maximum throughput with the limited resources within a system. The operators are also compelled for it, as for the sake of maximum Return on

Investment (ROI). Such algorithms or scheme for admission controls become efficient in best possible scenarios. Though from a service provider perspective, the scenario changes with respect to many factors like multiplicity of applications and devices, geo-position, time, whether, social events, natural events, power conditions and so on. An operator need to plan his services as per, to retain his credentials. Some time may have to be intentionally biased, like compromise on QoS, service outage, blocking of calls, etc.

For LTE networks, where the resources are allocated through shared channels and various adaptations like link adaptation etc. The efficiency of schedulers could be improved through statistical measures of various parameters. These parameters, with respect to each call, could be made available at single point to take an inference for system specific impact of the call. We have taken into consideration QoS Class Identifier (QCI) [7], Allocation and Retention Priority (ARP) [7], Guaranteed Bit Rate (GBR) [7], Aggregate Maximum Bit Rate (AMBR) [7], Channel Quality Indicator (CQI) [1] and Buffer Status Report (BSR) [1], as parameters for inferences.

With the help of EARP based admission control we would be able to create intentional models, for the service credentials. As for evaluation here we use an intentional model, to reduce the blocking and outage probability, in simulated LTE set up.

The rest of the paper is organized as follows. Section II provides the inference specifics of system parameters used for the proposed algorithm. Section III gives the description of combinational logic for deduction of effective ARP. Section IV contains the advantages of algorithm. Section V of the paper evaluates the proposed algorithm, for the specific intentions like blocking probability, outage and cell throughput. Section VI concludes the paper.

II. INFERENCE SPECIFICS

There are various algorithms [4][5] as part of Radio Resource Management (RRM) functionality. These algorithms provide admission control and resources, based on the availability and demands. Though such algorithms take care of best and efficient use of resources and system, but the traffic load on the system is not always under the controlled variations.

This algorithm approach helps to create intentional bias for exiting scheduler and to cater unprecedented traffic conditions.

This is an inferences-based and intension-based algorithm. This algorithm will be effective into the systems, where the inferences for the given parameters could be rated fairly on a fixed linear scale. These inferences are basically system specific impacts, with respect to parameters taken for the inference, for particular call.

The specifics, to take inferences from the given parameters in LTE networks, are being expressed in the following way.

A. Allocation and Retention Priority (ARP)

The priority level defines the relative importance of a resource request; values 1 to 15 are defined, with value 1 as the highest level of priority.

B. QoS Class Identifier (QCI)

The QCI label for a bearer determines how it is handled in the eNodeB. It is defining the packet forwarding requirements, through packet delay budget, and acceptable packet loss rate. This parameter could be inferred as real time traffic requirements. Table 1 provides standardized QoS in LTE networks, as given below.

Table 1. Standard QoS Identifier for LTE [5]

QCI	Bearer Type	Priority	Packet Delay	Packet Loss	Example
1	GBR	2	100	10^{-2}	VoIP Call
2		4	150	10^{-3}	Video Call
3		3	50		Online Gaming (Real Time)
4		5	300	Video Streaming	
5	Non-GBR	1	100	10^{-6}	IMS Signaling
6		6	300	Video, TCP based services e.g., email, chat, ftp, etc.	
7		7	100	10^{-3}	Voice Video Interactive gaming
8		8	300	10^{-6}	Video, TCP based services e.g., email, chat, ftp, etc.
9	9				

C. Aggregate Maximum Bit Rate (AMBR)

Each call would be having AMBR. The demand for the radio resource, for that call, would be inferred based on the existing load for AMBR. For higher Load the individual call impact for radio resource would be less.

D. Guaranteed Bit Rate (GBR)

Each call with specified GBR could be inferred as reservation for radio resources. Call with higher GBR value will require more resources in comparison with the calls with lower GBR.

E. Channel Quality Indicator (CQI)

CQI reports, the channel conditions. The value ranges from 1 to 15, where 15 indicate the most efficient. We can infer it like more data can be transferred using less resource blocks for a UE, which reporting higher CQI. The eNodeB can use higher Modulation and Coding Scheme (MCS) to send the data. Spectral efficiency increases with higher MCS as shown through Figure 1, which provides standard data with respect to SNR.

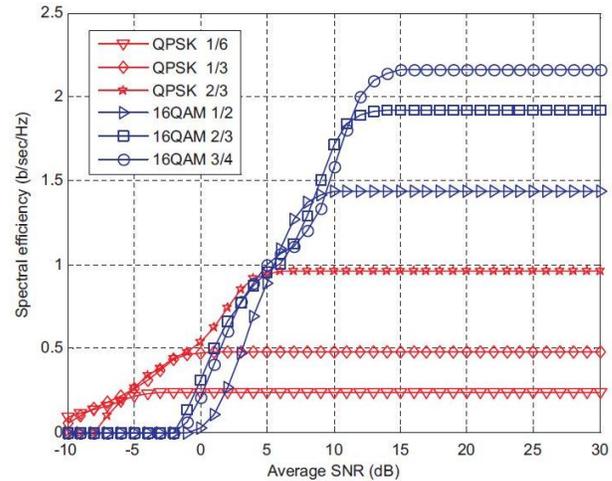


Figure 1. Spectral efficiency vs. SINR for different MCS [8].

F. Buffer Status Report (BSR)

This is a kind of MAC layer control element, carrying the information, on how much data is in UE buffer to be sent out. The index varies from 0 to 63, where a UE reporting Index 0 indicates no buffered data bytes are available at UE buffer and 63 indicates that more than 150000 buffered data bytes are available at the UE buffer. This can be inferred as UE traffic durability.

III. COMBINATIONAL LOGIC AND DEDUCTION OF EARP

The concept of EARP is based on the combinational logic. This logic depends on the system specific inferences for the parameters like QCI, ARP, GBR, AMBR, CQI and BSR. The deduction of EARP is done through combinational logic and classification based virtual scheduler, which is biased to intentions given in terms of weightage to parameters inferences.

The inferences could be direct or indirect (conditional, i.e., taking other parameter in consideration under specific conditions) based on the inference function’s algorithmic decision.

For example, in some cases, impact on system resources by a call, could be inferred based on the combination of CQI and BSR values reported. So, for User Equipment (UE) reporting better channel condition would require less radio resource blocks for the call, in comparison to one reporting bad channel condition. But, if the buffer (inferred by BSR)

that needs to be transmitted for second UE is very small when compared to the buffer of previous UE, in that case, impact on resources has to be the combination of CQI and BSR.

Explicitly stating the above example, it would be that even since low CQI can be inferred as poor impact on system, with intention for spectral efficiency, but could be tolerated to the extent BSR value is low.

That also means to say, that this combinational logic basically isolates the ‘extremes’ in terms of weightage of impact of individual call on the system. This weightage factor could be again magnified with intention factor, and this eventually helps in classification and deduction of Effective ARP or E-ARP through a virtual scheduler.

So, inferences of any parameter could be a combination of others. This is provided in the form of inferences function, implemented by inferences algorithms, as given below.

$$U_{ij} = f_j (J_{ARP}, J_{QCI}, J_{GBR}, J_{AMBR}, J_{CQI}, J_{BSR}) \quad (1)$$

Function “*f*” takes all the parameters, for the call “*i*”, with actual values and provides inferred value “*U_{ij}*” for each parameter “*j*”.

The inferred value of each parameter will be weighted with specified intentions with respect to the same parameter. The combined value with respect to all parameters will be used for classification, and that eventually used for the virtual scheduler. This combinational logic could be canonically represented in diagram below.

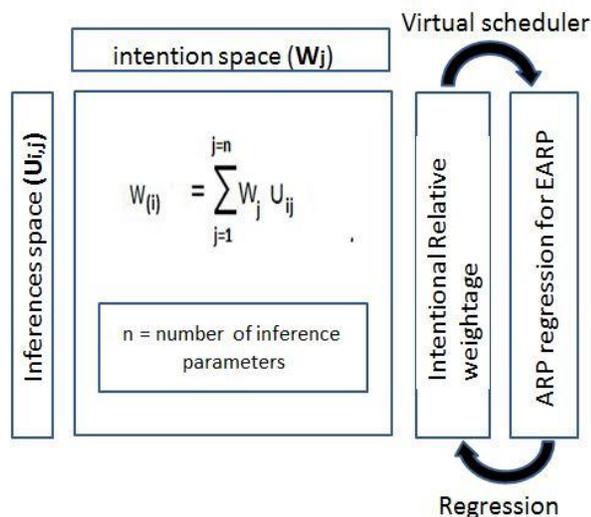


Figure 2. Combinational logic for deduction of E-ARP.

$W_{(i)}$ total weightage of a particular call “*i*” in the system.

W_j corresponds to the intentional weightage of each parameter “*j*”, e.g. ARP, QCI, GBR, AMBR, CQI and BSR.

U_{ij} inferred value of parameters reported for inference space.

A virtual scheduler implements the logic to provide the EARP values. It works on the resulted weightage (W_i) and existing ARP and deduces EARP through classification based regression.

IV. FORESEEN ADVANTAGE OF THIS MECHANISM

1. Various intentional model could be prepared for various traffic conditions in the system
2. Statistical regression could be performed for the EARP
3. This mechanism is independent of number of parameter and could be used in uplink downlink separately.
4. This mechanism is top-up mechanism, which does not interfere with existing mechanisms but to provide theirs intentional efficiency
5. This may need not to be in the existing RRM but may be in Self Organizing Networks (SON) servers etc.

V. EVALUATION

This top-up algorithm is evaluated over the network simulator compliant with 3GPP LTE standards [1][2][3] and provide real radio environment as well. This is tested against a particular vendor specific RRM and load is generated with varying system specific conditions. We intend to improve on blocking and outage probability.

Blocking Probability (P_b) is defined as the ratio of the number of blocked users to the total number of users requesting admission and retention. Outage Probability (P_o) is calculated as the ratio of the number of users not fulfilling their GBR requirement.

In the given simulation, when the peak load is reaching above 40 calls, with various types of traffic, there is sharp increase in blocking rate. After implementing with this top-up algorithm, with traffic specific intentions, a significant drop in blocking call is observed as show in Figure 3.

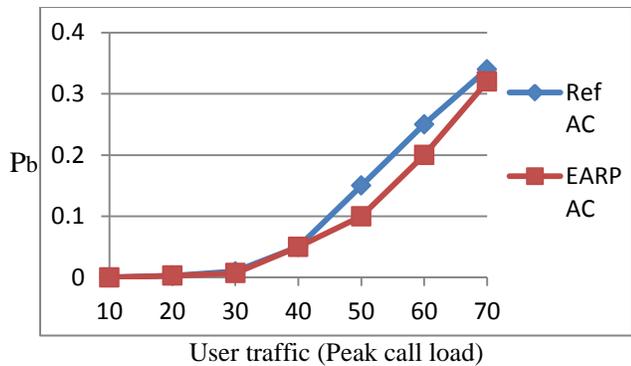


Figure 3. Traffic vs. Blocking Probability Graph.

Similarly, in the given simulation, when the peak load is reaching above 40 calls, there is sharp increase in outage under specific traffic load. After implementing this top-up algorithm, with traffic specific intentions, a significant drop in outage rate is observed as show in Figure 4.

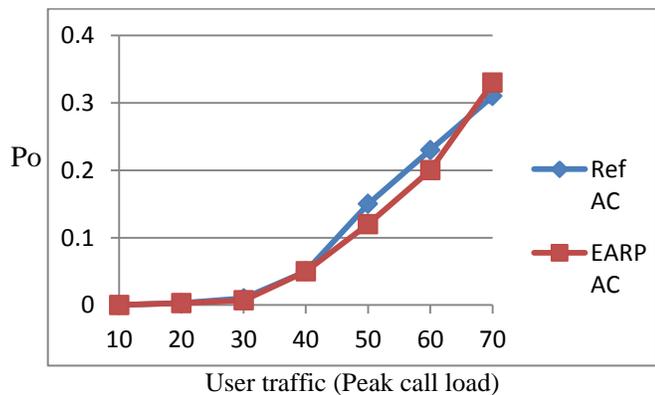


Figure 4. Traffic vs. Outage Probability Graph.

After implementing this top-up algorithm, there is significant improvement found in cell throughput, in almost all the traffic specific load, as shown in Figure 5.

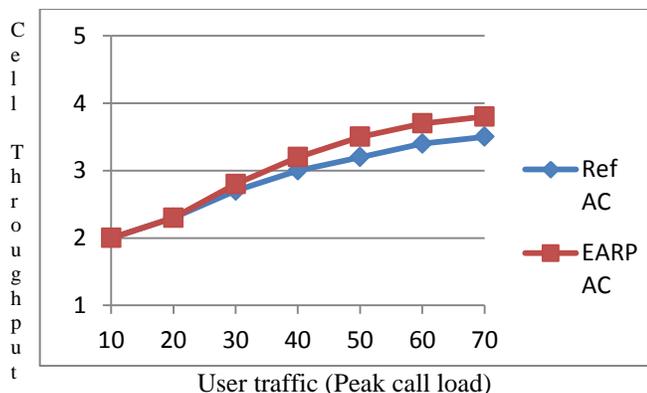


Figure 5. Traffic vs. Average Cell Throughput [Mbps].

The results showcase the significance of this top-up algorithm, in the extreme conditions. As it help to create various intentional models to provide control over unprecedented load conditions.

VI. CONCLUSION AND FUTURE WORK

The results show that the given EARP based AC, unlike reference AC, is robust and automatically adjusts to the various traffic scenarios, cell load, and user channel conditions. Various models (intentions) can be prepared for different kind of unprecedented traffic statistically. It can also be incorporated in the SON techniques as part of top-up intentional modification and control of existing AC algorithms. Future studies will address to addition of more parameters in the inference space. Such top-up mechanism are being studied for SON perspective and would be developed for traffic load control and RAN selections.

The traffic classifications and traffic conditions are being studied at various levels in the LTE systems and such mechanism could be of help for effective forwarding and prioritizations of classified traffics under various traffic condition scenarios.

REFERENCES

- [1] 3rd Generation Partnership Project; Technical Specification Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) protocol specification; 3GPP TS36.321 version 10.7.0, Jan. 2013.
- [2] 3rd Generation Partnership Project; Technical Specification Evolved Universal Terrestrial Radio Access (E-UTRA); Requirements for support of radio resource management; 3GPP TS 36.133 version 10.8.1, Oct. 2012.
- [3] 3rd Generation Partnership Project; Technical Specification Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); 3GPP TS 36.300 version 10.9.0; Jan. 2013.
- [4] H. Holma and A. Toskala, "LTE for UMTS: Evolution to LTE-Advanced", 2nd Edition, Apr. 2011.
- [5] S. Sesia, I. Toufik, and Matthew Baker, "LTE – The UMTS Long Term Evolution: From Theory to Practice", 2009.
- [6] Y. Fang and Y. Zhang, "Call Admission Control Schemes and Performance Analysis in Wireless Mobile Networks" IEEE Trans. Veh. Tech. vol. 51, no. 2, Mar. 2002
- [7] 3rd Generation Partnership Project; Technical Specification 23.203 version 10.8.0 Release 10.
- [8] M. Anas, PhD-Thesis; "Uplink Radio Resource Management for QoS Provisioning in Long Term Evolution", Aalborg University, Denmark, Jan. 2009.
- [9] K. Spaey, B. Sas, and C. Blondia, "Self Optimizing Call Admission Control for LTE Downlink", 'FP7 ICT-SOCRATES' workshop, Feb. 2010.

Performance Limitation of HSDPA MIMO by Pre-Coding Induced Phase Distortion

Wolfgang Aichmann
 Nokia Siemens Networks
 D-89081 Ulm, Germany
 Email: wolfgang.aichmann@nsn.com

Abstract—In multiple-input multiple-output (MIMO) antenna systems both strength of the data layers as well as separation of them is achieved by pre-coding the transmitted signal with different factors for different layers and different antennas. In W-CDMA systems, the selection of these factors is decided anew every transmission time interval, which causes a phase distortion with every change. Due to raised cosine filtering this phase distortion is broadened even if the physical channel is flat in frequency, which distorts the affected symbols and thus limits the maximum bit rate to be transmitted in such cases if no counter measures are taken. If the pre-coding weights of adjacent time intervals are known by the receiver, the distortion can be removed by the equalizer; otherwise high code rates should be avoided by the scheduler.

Keywords—W-CDMA; HSDPA; MIMO; pre-coding; raised cosine filter;

I. INTRODUCTION

Multiple-input multiple-output is a well known method to increase the capacity of radio systems on the air interface [1]–[5] and was introduced in 3G high speed downlink packet access (HSDPA) with Release 7 of 3GPP TS 25.214 [6] in 2007 and in high speed uplink packet access (HSUPA) with Release 11 [7] in 2012. Two modes of operation are defined, namely single and dual layer transmission. The distribution of data to the transmitting antennas is controlled by so called pre-coding vectors. The effect of this pre-coding is twofold: In case of single data layer transmission, the pre-coding vector is adjusted such as to optimize the power at the receiving antennas. This is possible as shifting the phase difference between the signals fed to the transmitting antennas changes the interference behaviour at the receiving antennas, and constructive interference results in higher power.

If two layers are transmitted, orthogonal pre-coding vectors are applied, which allow to separate the layers at the receiving side, typically by a linear minimum squared equalizer in space and time as it is, e.g., described in [8]. Non-linear receiver algorithms such as interference cancellation improve the de-coding performance on receiver side but are quite sensitive to very good channel knowledge [9], [10].

As the interference conditions are changing with time (fast fading), especially for moving user equipments (UE), one tries to always find and use the best pre-coding vector out of the set of allowed ones, and to change it if necessary. Switching between single and dual stream as well as between

MIMO and non-MIMO can be regarded as a generalization of switching the pre-coding vector: In non-MIMO mode, the weights for the second transmitting antenna are 0 and single layer transmission can formally be regarded as dual layer transmission with the same data and the same pre-coding for both layers.

Any change in the pre-coding vector however causes a superposition of signals coded with the old and ones coded with the new pre-coding vector in the range of channel delay due to multi-path propagation including raised cosine filtering. This has some impact on the detection of the affected bits and might even prohibit at all the correct block detection for high coding rates. In real networks switching of MIMO mode as well as switching the pre-coding vector within a MIMO mode is a quite frequent operation, which leads to a significant restriction of the achievable maximum data rate even in very good radio conditions. This could be one of the reasons why HSDPA MIMO isn't present in commercial networks so far besides other reasons such as the chicken or egg (here better base station or UE) dilemma and the need for a second pilot signal reducing thus the power available for data transmission.

The remainder of the paper is organized as follows: In Section II, we investigate the impact of pre-coding induced phase distortions on the received symbols. We show then in Section III that a smooth transition of equalizer weights at the boundaries of transmission time intervals (TTI) pre-coded with different vectors is able to remove this distortion. Unfortunately, the pre-coding vector of adjacent TTIs is known on receiving side only if they bear data for the same UE. Therefore, it is useful to take counter measures on transmitter side if TTIs are dedicated to different UEs. Some of them are introduced in Section IV followed by some concluding remarks in Section V.

II. BASIC CONSIDERATIONS

In order to understand the principles of MIMO pre-coding, the basic concept of MIMO transmission in HSDPA is displayed in Fig. 1.

The signal s_λ of each data layer $\lambda \in \{1, 2\}$ is fed to the $m \in \{1, 2\}$ transmitting antennas with weights $w_1 = 1/\sqrt{2}$ and $w_2 \in W_2 \equiv \{(1+i)/2, (1-i)/2, (-1+i)/2, (-1-i)/2\}$ [6]; w_1 and w_2 build the pre-coding vector

and the pre-coding index (PCI) defines which component w_2 out of the code book W_2 is taken.

If only one data layer is transmitted, w_2 typically is chosen such as to achieve the maximum channel power at receiving side as this is related to the highest possible throughput if we ignore here that also the post-equalizer interference power may depend on the applied pre-coding vector. In case of two layers being transmitted the two pre-coding vectors have to be orthogonal in order to enable the receiver to separate the layers and hence $w_2|_{\lambda=2} = -w_2|_{\lambda=1}$. As exchanging the PCI values within two layers transmitted in parallel has no impact on the aggregated throughput this means that only two choices remain for w_2 , namely either PCI 1 and 4 or PCI 2 and 3, and the aggregated possible throughput is used to decide which of these two possibilities is taken.

The signal r_j received at antenna j and time k counted in chip length ($0 \leq k < n$) consists of contributions of all transmitted data layers and can be expressed as

$$r_j(k) = \underbrace{\sum_{\delta=0}^{n-1} [w_1 h_{1j}(\delta) + w_2 h_{2j}(\delta)] s_1(k - \delta)}_{\text{layer } \lambda=1} + \underbrace{\sum_{\delta=0}^{n-1} [w_1 h_{1j}(\delta) - w_2 h_{2j}(\delta)] s_2(k - \delta)}_{\text{layer } \lambda=2} \quad (1)$$

To achieve high data rates it is necessary the coherence time of the channel impulse response to be large compared to the delay spread in the presence of multi-path propagation, and large compared to the averaging length required for appropriate channel estimation. Practical experience shows that code rates of 80% or higher require a coherence time of at least one slot, which is achieved for UEs moving not faster than about $3km/h$.

However, this constraint is violated if the pre-coding vector is changed: The effective channel impulse response becomes unsteady at the TTI boundary and hence the coherence time approaches zero, at least in the time range of the delay spread. It has to be emphasized that the raised cosine (RC) filter has to be included explicitly in this considerations as on transmitting side, the pre-coding usually is executed before the root raised cosine (RRC) filtering and on receiving side the equalizer algorithm containing the inverse pre-coding is fed with data, which have already passed the filter.

To demonstrate the impact of RC filtering we have investigated the positions of symbols in the phase space diagram, modulated with 64QAM and passing a one-tap channel with additional RC filter, i.e. a RRC filter was applied on both transmitting and receiving side, but without equalizing. Four different scenarios have been taken into account, namely keeping the pre-coding vector constant at PCI=1, as well as switching it in the middle of a data block with 2560 chips to another pre-coding vector out of W_2 . The distance of the

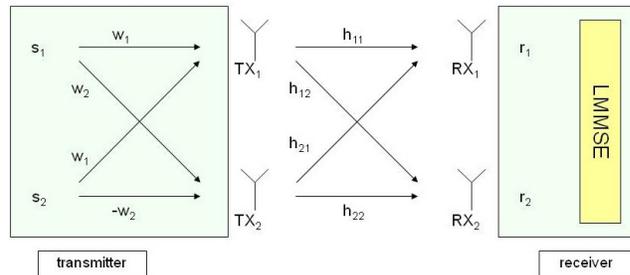


Figure 1. MIMO dual layer transmission, schematic overview.

positions of the symbols in the phase space diagram from the position they would have if the complete data block had been sent with the new pre-coding vector is then measured. The result is shown in Fig. 2 displaying the symbol shifts for an RRC filter with oversampling factor 4 and tail length¹ of 20 chips.

Fig. 2a shows the constellation of the symbols in the phase space. The circles mark the symbol positions for constant pre-coding whereas crosses are placed at the effective symbol positions. Symbols beyond the RRC coherence length are drawn in blue and those within the coherence length around the switching time are red colored. Fig. 2b displays the distance of shift of the symbols with respect to the position without shift, normalized to the average symbol distance from the origin. Please note the reduced scale on ordinate for the case of no switching. It is obvious that symbols within the tail length are affected heavily by switching the pre-coding vector, whereas symbols beyond that range relax quickly to the equilibrium position.

Additional equalizing turns and stretches the symbols such as to have them at the same positions as on the transmitting side, but it does *not* remove the distortion induced by switching the pre-coding vector, i.e., the relative positions of the red crosses to the blue circles.

Two major effects are worth to be noted, namely that the symbol shift increases with increasing length of the RRC filter and that especially for larger tail lengths a phase shift of π is worse than a shift of $\pi/2$. Whether the latter effect also results in higher bit error rates is not clear so far as a longer distance from the ideal position has no more influence on the bit error rate when the bit is detected wrongly anyway.

Furthermore it has to be noted that on transmitting side the pre-coding filter usually is applied before the RRC filter, which causes the disturbance by switching the filter to be twice as long as the length of the RRC filter. If the order of filtering and pre-coding is changed, the length of disturbance is halved.

Link level simulations with a Pedestrian A channel at $3km/h$ for MIMO single stream (see Fig. 3a) show how

¹Tail length denotes the length of RRC filter, i.e., the length from middle position of the filter to the edge.

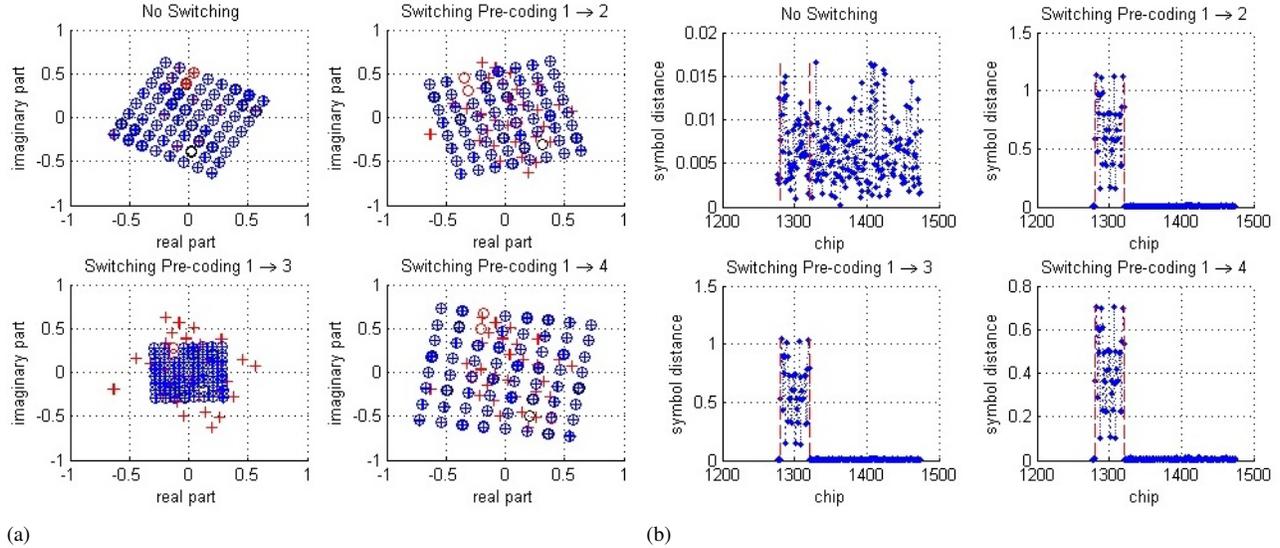


Figure 2. Shifting of 64QAM symbol position due to switching pre-coding vector for a one tap channel and RRC filter with tail length 20. In (a), the expected (o) and measured (+) symbol positions in the complex phase space are shown. In a range around the pre-coding change (red), determined by the maximum channel delay including RC filter, the symbols are distorted. The absolute value of the dislocations as a function of chip index is shown in (b).

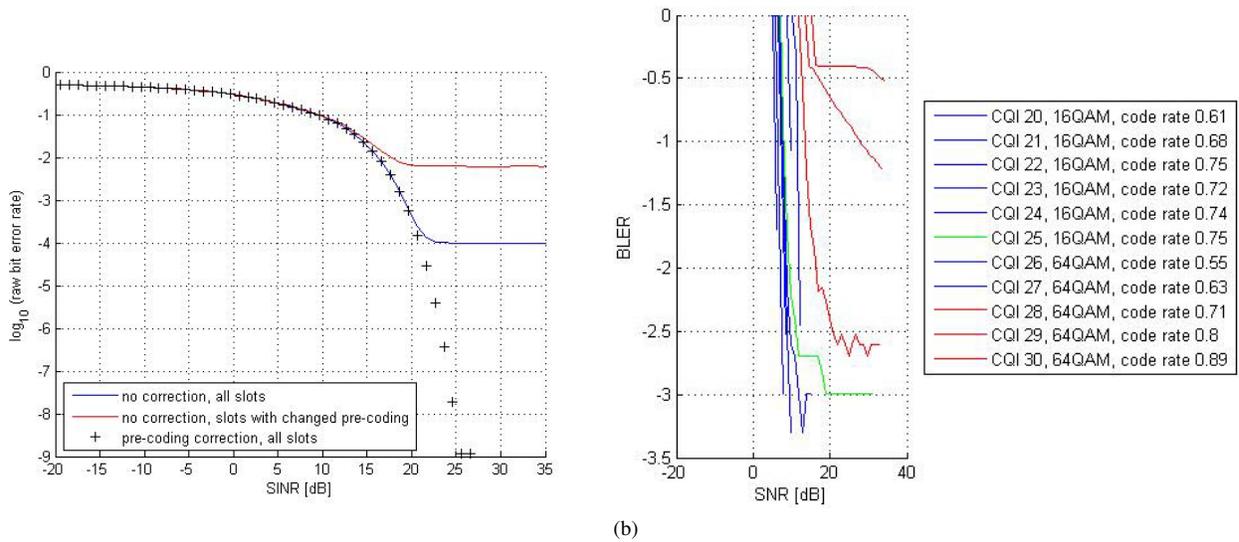


Figure 3. Link level simulations for a Pedestrian A channel at 3km/h, UE category20 and CQI values according table G in [7] (a) Raw bit error rate for CQI 30. Slots with changed pre-coding vector show a significantly higher error floor compared to all slots. Correct handling of these pre-coding changes (see III-B) removes the error floor completely. (b) High code rates show an error floor in the block error rate (BLER) which can even be above the working point of 10% BLER.

switching the pre-coding vector causes an error floor of the raw bit error rate of about 0.6% in affected slots. These bit errors directly cause erroneous decoding of the blocks they belong to in case of coding rates higher than about 0.75 as is demonstrated in Fig. 3b.

III. EQUALIZER WITH SMOOTH WEIGHT TRANSITION

A. Calculation of Symbol Distortion

To see how the distortion induced by switching the pre-coding vector in the presence of channel impulse response h including the RC filtering can be properly taken into account

on receiver side, let us start with (1) but restricted to one data layer only, namely:

$$r_j(k) = \sum_{\delta=0}^{n-1} [w_1(k-\delta)h_{1j}(\delta) + w_2(k-\delta)h_{2j}(\delta)] s_1(k-\delta) \quad (2)$$

We have thus generalized the pre-coding vector to depend on time via $k-\delta$. Let us now introduce a single switching of the pre-coding vector at time 0, i.e., $s_1(k < 0)$ is pre-coded with \tilde{w} and $s_1(k \geq 0)$ with w :

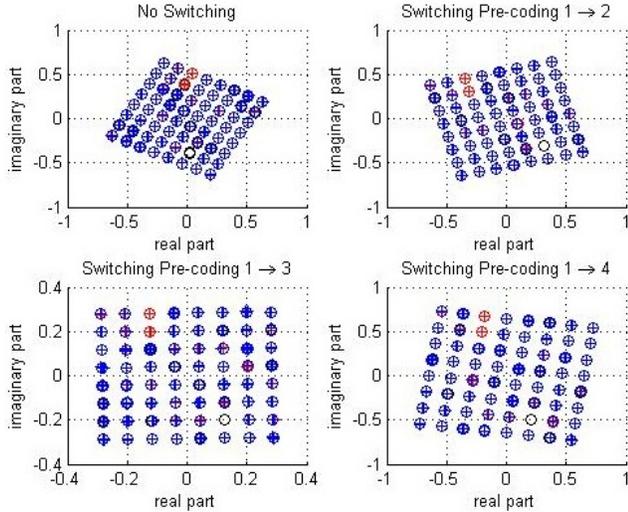


Figure 4. Corrected positions of 64QAM symbols, which are distorted by switching the pre-coding vector in presence of a one tap channel and RRC filter with filter length 20 and OSF 4.

$$\begin{aligned}
 r_j(k \geq 0) &= \sum_{\delta=0}^k [w_1 h_{1j}(\delta) + w_2 h_{2j}(\delta)] s_1(k - \delta) \\
 &+ \sum_{\delta=k+1}^{n-1} [\tilde{w}_1 h_{1j}(\delta) + \tilde{w}_2 h_{2j}(\delta)] s_1(k - \delta) \\
 &= \bar{r}_j + \Delta r_j \quad (3)
 \end{aligned}$$

$$\bar{r}_j(k) \equiv \sum_{\delta=0}^{n-1} [w_1 h_{1j}(\delta) + w_2 h_{2j}(\delta)] s_1(k - \delta) \quad (4)$$

$$\Delta r_j(k) \equiv \sum_{\delta=k+1}^{n-1} [\tilde{w}_2 - w_2] h_{2j}(\delta) s_1(k - \delta) \quad (5)$$

The last step in (3) holds because the same phase factor is applied to the first transmitting antenna for all pre-coding vectors, i.e., $w_1 \equiv \tilde{w}_1$. One can see that received symbols up to index $k = n - 2$ are affected from the switch in the pre-coding vector. Hence we can limit the further investigations to $0 \geq k \geq n - 2$. It remains $\Delta r_j(k)$ as error correction to the 'static' calculation $\bar{r}_j(k)$ of $r_j(k)$.

One can hence use (4) to calculate \bar{r}_j assuming ideal knowledge of the transmitted symbols $s_1(k < 0)$ also on receiver side, i.e., symbols of the data block preceding the switch of the pre-coding vector, as is shown in Fig. 4.

The assumption of ideal symbol knowledge demonstrates the impact of switching the pre-coding index, but it doesn't help to remove the distortion in real scenarios as of course the symbols aren't known a priori except of pilot measurements, which is beyond the current scope. We must therefore find another solution of removing the symbol distortion.

B. Smooth Weight Transition

In a linear minimum mean square equalizer (LMMSE) the weights are typically re-calculated every slot or every TTI but always with the pre-coding weights as they are relevant for the current data block. If the pre-coding vector is changed on transmitting side also the equalizer weight vector obtained with the old pre-coding weights is replaced as a whole by the new one now calculated also with the new pre-coding weights. This implies that there is a transition interval, in which either data originating from the 'old' data block are equalized with weights obtained for 'new' pre-coding or data already originating from the 'new' data block are still equalized with weights obtained for 'old' pre-coding or a mixture of both depending on when exactly the replacement of the weight vector takes place.

Heuristically, this error can be avoided if the weight vector isn't replaced as a whole but element by element. This means that as soon as the new data block enters the equalization procedure the corresponding element of the 'old' equalizer vector (either the first or the last one, depending on the ordering convention) is replaced by the element of the 'new' weight vector. Then, the elements are exchanged one by one with every chip until the weight vector consists completely of elements calculated with the 'new' pre-coding.

In order to motivate the validity of this approach we remember the calculation of space time LMMSE equalizer weights as, e.g., elaborated in [8]. The weights are proportional to a generalized matrix of channel impulse response, Γ_E , but with Γ_E appearing in a quadratic form in the denominator². If all data layers are transmitted with the same power as is the case in HSDPA MIMO, this quadratic form becomes independent of the applied pre-coding. The latter persists only in the numerator and makes the elements of the weight vector proportional to the respective pre-coding weights, leading directly to the heuristic approach described above.

Applying this procedure to full link level simulations a significant gain can be obtained at the high data rate edge as can be seen in Fig. 5. Although the working point is only slightly shifted to lower SNR, the error floor present without any correction is avoided completely. Compared to the simulation with filtering the pre-coding switching with the goal to reduce the number of transitions, see Fig. 6, the averaged working point is nearly 1dB better in the scenario investigated here.

The formalism above can easily be extended to switching between MIMO single and dual stream: Single stream transmission is formally the same as dual stream transmission with identical pre-coding vectors (and of course identical data). To obtain single layer transition weights from dual layer ones the latter have to be summed up whereas

²We use here the term denominator for an inverted matrix in order to simplify the verbal description.

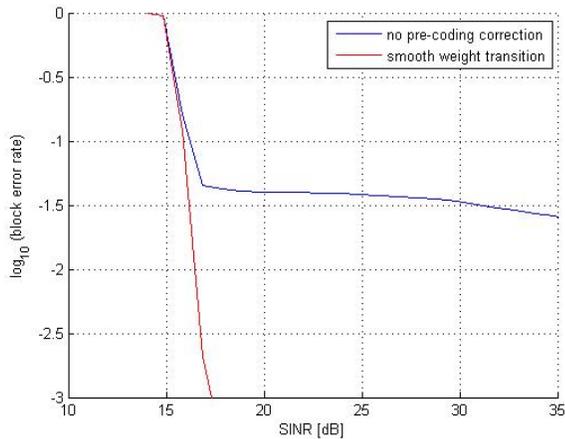


Figure 5. Performance improvement of MIMO Single Stream by smooth transition of equalizer weights. The simulation was executed for a Pedestrian A channel at 3km/h, UE category 20 and CQI 30 according table G in [7]

dual layer transition weights can be obtained from single layer ones by dividing them into two equal parts, of course always with proper scaling such as to conserve the overall transmit power.

As also non-MIMO data transfer is affected if switching between this mode and any of the MIMO modes occurs it has also to be included by defining it to be a single stream transmission with constant pre-coding vectors and vanishing pre-coding weight for the second transmitting antenna.

IV. OTHER MEANS OF AVOIDING SYSTEMATIC BLOCK ERRORS

There are many methods, which can help to avoid the occurrence of systematic block errors induced by changing the pre-coding weights in case the receiver is not able to handle the corresponding symbol distortion properly. Some of them, e.g., the introduction of a cyclic prefix as in long term evolution (LTE), would require to change the specification which would make it impossible for legacy UEs to be operated. They are beyond the scope of the current paper. Others can be implemented either on base station or UE side within the valid specification framework. Such proposals are discussed now.

A. Attenuation of Switching Pre-Coding Vector

One quite simple strategy to decrease the negative impact of switching the pre-coding vector is to attenuate the switching itself. We have investigated a simple filter algorithm on receiver side for single layer transmission, which is working in two steps. This approach is easily extensible to the dual layer case. It is based on an ordinary calculation of channel

power for all possible pre-coding vectors, i.e.,

$$P_s^{(k)} = \sum_{n=1}^{N_{Rx}} [H_{kn} * H_{kn}^\dagger]$$

$$H_{kn} = \sum_{m=1}^{N_{Tx}} \left[w_m^{(k)} \sum_{\delta=1}^L h_{mn}(\delta) \right]$$

where $w^{(k)}$ denotes the k^{th} element out of the code book W .

Without filtering, the pre-coding vector is chosen as

$$k_{0,s} = \max_{\{k\}} (P_s^k). \quad (6)$$

In the first step of the proposed algorithm, we compare the current optimum channel power with the previous one and introduce a threshold $\zeta > 0$ for signaling a changed pre-coding vector:

$$\tilde{k}_{0,s}(s) = \begin{cases} k_{0,s} & P_s^{k_{0,s}} > (1 + \zeta) P_s^{k_{0,s-1}} \\ k_{0,s-1} & \text{otherwise} \end{cases} \quad (7)$$

In the second step, we then introduce an attenuation filter, which remembers the last F pre-coding indices, $k_{0,s}, s \in \{s_0, s_0 - 1, \dots, s_0 - F + 1\}$ and resets the pre-coding index to the previously used one as long as not at least $F_\zeta, 1 < \zeta < F$ times the new index occurs.

Fig. 6 shows simulation results for different modulation and coding schemes as defined in the single layer channel quality indicator (CQI) table 7G [6] for an additive white Gaussian noise (AWGN) channel with filtering where we have used $\zeta = 0.1, F = 10$ and $F_\zeta = 7$. It can be seen clearly that the error floor occurring in absence of the threshold filter, see Fig. 3, vanishes with the proposed filter design.

For single stream without filtering the changing of the pre-coding vector, a reasonable block error rate can be achieved for coding rates not too high. In case of 64QAM, an error floor starts to occur for coding rates between 0.7 and 0.8 and for 16QAM in the range of 0.7. This implies to apply the suggested filter only in case of used modulation and coding scheme providing coding rates higher than the mentioned thresholds.

B. Limiting 'CQI' to Specific Code Rate if Pre-Coding Is Changed

As demonstrated in section IV-A, the proper data block detection is destroyed by changing the pre-coding vector only for high coding rates. It seems therefore reasonable to reduce the coding rate to a value below the corresponding threshold via corresponding CQI selection if the pre-coding vector is changed. Although this mechanism can be implemented completely on the transmitting side the receiver should know about such an algorithm in order to avoid

too many changes anyway. A corresponding signaling might therefore be useful.

This CQI limitation should be applied only in case of the receiver is not able to correct the switching errors itself (see section III-B). A powerful application of this proposal therefore requires the definition of a new UE class or a parameter indicating the ability of the UE for this correction.

V. CONCLUSION

We have shown that switching the pre-coding vector applied to data transferred from multiple TX antennas leads to a broad distortion of the channel impulse response including the raised cosine filter. In this context, also switching between non-MIMO, MIMO single layer and MIMO dual layer transmission can formally be regarded as switching the pre-coding vector. For high data rates this effect leads to a high probability that data blocks adjacent to the switching border will be lost.

There are several possibilities to avoid or at least to reduce the negative impact, but most of them cause other constraints, e.g. avoiding changes of the pre-coding vector and / or reducing the coding rate. Other measures, e.g., dynamically shortening the disturbance length by reducing the length of root raised cosine filter if the pre-coding vector is changed might be inapplicable at all.

Therefore, the best strategy seems to remove the disturbance induced by the discontinuity in phase by the equalizer. However, if the receiver isn't able to do this, e.g. because of missing pre-coding information, which is signaled encrypted, a reduction of the coding rate to a reasonable value of affected data blocks is an acceptable strategy to avoid systematic block errors.

The calculations were done for HSDPA MIMO, but they are applicable to a wider range and affect many scenarios with MIMO and / or transmit diversity. Although the smooth transition of equalizer weights assumes phase-only pre-coding an extension of this approach might be used if pre-coding with modifications in the amplitude are applied.

Finally we want to mention that the impact of the distortion to LTE systems is low or absent at all because of the cyclic prefix, which is introduced there for other reasons.

ACKNOWLEDGMENT

The Author would like to thank Elke Hallmann and Klaus Knoche for many inspiring discussions during this study.

REFERENCES

- [1] J. G. Proakis and M. Saleh, *Digital Communications*, 5th ed., New York: McGraw-Hill, 2008.
- [2] H. Holma and A. Toskala, *HSDPA/HSUPA for UMTS*, The Atrium: John Wiley & Sons, 2006.
- [3] E. Dahlman, S. Parkvall, J. Sköld, and P. Beming, *3G Evolution, HSPA and LTE for Mobile Broadband*, 1st ed., London: Elsevier, 2007.

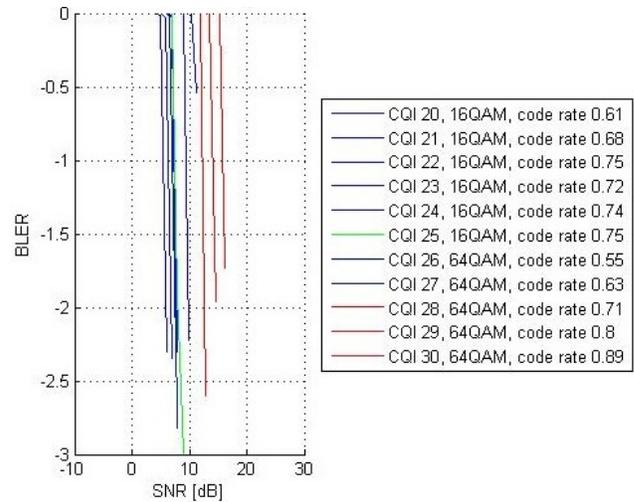


Figure 6. Performance improvement of MIMO Single Stream in AWGN channels by attenuating the changing of the pre-coding vector. The error floor as shown in Fig. 3 can be avoided.

- [4] C. Mehlführer, S. Caban, M. Wrulich, and M. Rupp, "Joint Throughput Optimized CQI and Precoding Weight Calculation for MIMO HSDPA", in Proc. Fourtysecond Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA October 2008, [Online] Available: http://publik.tuwien.ac.at/files/PubDat_167015.pdf. (Accessed: 2013-05-14)
- [5] White paper, "The Evolution of HSPA", 4G Americas, 2011, [Online] Available: http://www.4gamericas.org/documents/4G%20Americas%20White%20Paper_The%20Evolution%20of%20HSPA_October%202011x.pdf. (Accessed: 2013-05-14)
- [6] 3GPP, "Technical specification group radio access network; Physical layer procedures (FDD) (Tech. Spec. 25.214 V7.4.0)", 2007. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/25_series/25.214/. (Accessed: 2013-05-14)
- [7] 3GPP, "Technical specification group radio access network; Physical layer procedures (FDD) (Tech. Spec. 25.214 V11.4.0)", 2012. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/25_series/25.214/. (Accessed: 2013-05-14)
- [8] M. Wrulich, S. Eder, I. Viering and M. Rupp, "Efficient Link-to System Level Model for MIMO HSDPA", Proc. of the 4th IEEE Broadband Wireless Access Workshop, New Orleans, December, 2008.
- [9] W. Aichmann, "Scrambling-Exact Channel Estimation in W-CDMA Systems", in Proc. ICWMC, IARIA Luxembourg June 2011, pp. 12-17. [Online] Available: http://www.thinkmind.org/index.php?view=article&articleid=icwmc_2011_1_30_20028 (Accessed: 2013-05-14)
- [10] W. Aichmann, "Impact of Scrambling-Exact Channel Estimation on HSDPA MIMO", *International Journal On Advances in Telecommunications*, vol. 5, no. 1&2, 2012, pp. 79-89. [Online] Available: http://www.thinkmind.org/index.php?view=article&articleid=tele_v5_n12_2012_8 (Accessed: 2013-05-14)

Cooperative Communications Using Multilevel Bit-Interleaved Coded Modulations

Hadi E. Sawaya

Engineering Faculty - ESIB

Saint Joseph University

B.P. 11-514, Beirut, Lebanon

Email: hadi.sawaya@usj.edu.lb

Abstract—This paper deals with cooperative communications for *non-ergodic* block fading multiple-input multiple-output (MIMO) channels. A 2-level bit-interleaved coded modulation (BICM) with 16-QAM constellation is used in order to obtain a flexible scheme where one or two users can share the resources of the relay node in order to achieve better performance. The block fading channel is assumed to be perfectly known at the receiver but not at the transmitter. The channel between the user and the relay is assumed to be error free. Iterative multistage decoding is done. The performance of the system is derived using Monte Carlo simulation.

Keywords—Cooperative communications; MIMO channel; bit-interleaved coded modulations; multilevel coded modulations.

I. INTRODUCTION

Cooperative communications for wireless networks have recently gained a lot of interest. The idea is to allow the users of the network to cooperate and share resources in order to provide more diversity and more resources like bandwidth and antennas, in order to achieve better capacity, performance, system coverage, etc. Sendonaris *et al.* [1] [2] proposed the concept of cooperative diversity, with decode and forward strategy applied by the cooperative node. T.E. Hunter *et al.* [3] and B. Zhao *et al.* [4] studied the principles of coded cooperation and distributed coding.

Several coded cooperation schemes based on multilevel coded modulations (MLC) [5] or bit-interleaved coded modulations (BICM) [6] were introduced [7][8][9].

In this paper, a multilevel bit-interleaved coded 16-QAM modulation [10] with n_t transmit antennas is applied by the users or nodes in order to send information to destination points with n_r receive antennas.

In the presence of a relay node, a part or the entire data sent by the user is retransmitted (decode and forward) by the relay in order to achieve higher diversity. This depends on whether or not the relay is shared by two users or is transmitting its own data.

In the absence of a relay node, and in order to achieve better performance, the user can reduce the data rate by sending information over one of the two levels of the modulation.

The paper is organized as follows. The channel model is described in Section III. In Section IV, we introduce the 2-level BICM combined to a 16-QAM constellation. This is the modulation used by the users and the relays in order to transmit

data. In Section V, the capacity of the proposed scheme is derived. The iterative multistage decoder (MSD) is studied in Section VI. The different transmission and cooperation strategies are introduced in section VII. In Section VIII, the performance of the different cooperation and non-cooperation schemes are determined using Monte Carlo simulation. Results are compared to the performance of a 1-level BICM and to the outage probability of a *non-ergodic* block fading MIMO channel.

II. RELATED WORKS

In [7], a multilevel cooperation scheme, where each source or relay node uses a level of the modulation, was introduced. However, a frequency selective channel, with a single transmit and a single receive antenna was considered. Moreover, there's no use of bit-interleaved coding and iterative decoding.

In [8], a one level BICM is applied by each user, and iterative decoding is done between users or between the decoder and the constellation. Flexible spectral efficiency is reached by controlling the portion of information data repeated by the relay.

In [9], a cooperative communications scheme, combining BICM and orthogonal frequency division multiplexing (OFDM), is proposed.

In this paper, the main originality is the combination of MLC and BICM into a multilevel coding scheme with a bit-interleaved coded modulation at each level. Moreover, a MIMO channel is considered. Notice that each source or relay uses one level of the modulation in order to transmit its own data, and the other levels for cooperation as in [7]. On the other hand, iterative multistage decoding is applied. This is similar to the second decoding strategy introduced in [8].

III. CHANNEL MODEL

Let us consider a flat fading MIMO channel with n_t transmit and n_r receive antennas. The received and the transmitted signals are related by $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$ where $\mathbf{y} = (y_1, y_2, \dots, y_{n_r})^t$ denotes the vector of complex received signals during any given channel use, $\mathbf{x} = (x_1, x_2, \dots, x_j, \dots, x_{n_t})^t$ denotes the vector of complex transmitted symbols. The superscript t stands for transpose. The superscript h , that will be used later, stands for transpose conjugate. The symbols x_j belong to a Quadrature Amplitude Modulation (QAM) constellation of size $M = 2^m$. The

channel matrix $\mathbf{H} = [h_{i,j}]$ is assumed to be perfectly known at the receiver but not at the transmitter. The fading coefficients $h_{i,j}$ are complex, Gaussian, circular, mutually independent and satisfy $E[|h_{i,j}|^2] = 1$.

We consider the case where the channel matrix \mathbf{H} is fixed in time. This is the so called *non-ergodic* block fading channel. Notice that the *ergodic* fading channel is the one where the channel matrix varies at each channel use. \mathbf{n} denotes the vector of additive white complex Gaussian noise with zero mean and variance $2N_0$.

IV. THE 2-LEVEL BIT-INTERLEAVED CODED MODULATION

The 2-level bit-interleaved coded modulation with n_t transmit antennas is represented in Figure 1. This is a Multilevel Coded (MLC) modulation with a Bit-Interleaved Coded Modulation at each level.

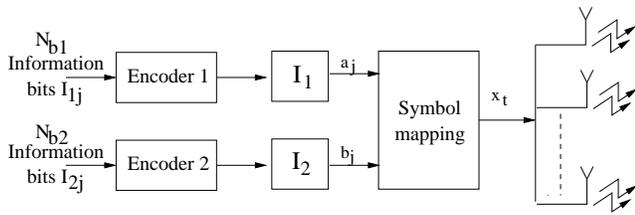


Fig. 1. 2-level BICM transmitter.

At each level, the information bits are encoded into $N_c/2$ coded bits using a non-recursive non-systematic convolutional encoder, then randomly interleaved.

Afterwards, the bits a_j at level 1 and the bits b_j at level 2 are mapped into $N_s = N_c/4$ 16-QAM symbols x_t . Gray labeling of coded bits is applied as in Figure 2.

10	x	x	x	x
11	x	x	x	x
01	x	x	x	x
ab=00	x	x	x	x
ab=00	01	11	10	

Fig. 2. Gray labeled 16-QAM constellation. a \rightarrow level 1. b \rightarrow level 2.

Finally, the block of $N_s = N_c/4$ symbols is divided into signal vectors of length n_t symbols transmitted at each channel use (the signal vector is function of $L = 4.n_t$ coded bits).

The total rate of the system in bits per channel use is equal to $R = 4 \times (R_1 + R_2)$ where R_1 and R_2 are the rates of encoders 1 and 2 respectively. Note that R_1 and R_2 are chosen in a way to achieve the best performance. For large code length, this is done using the capacity design rule [5], which states that the rate R_i at level i is equal to the capacity C_i of the equivalent channel that represents the transmission of the bits of level i (cf. section V).

V. CAPACITY

Let us consider an *ergodic* Rayleigh flat fading MIMO channel. Since \mathbf{H} is known to the receiver and not to the transmitter, it can be considered as part of the output. The capacity is therefore obtained by maximizing the mutual information $I(\mathbf{X}; (\mathbf{Y}, \mathbf{H}))$.

It can be shown after some simple manipulations that this quantity is equal to $E_{\mathbf{H}}\{I(\mathbf{X}; \mathbf{Y}|\mathbf{H} = H)\}$ where H denotes a realization of \mathbf{H} . The mutual information in this last expression is maximized when \mathbf{X} is a zero mean complex circularly symmetric Gaussian vector. This leads to the following equation [11]:

$$C_{MIMO} = E_{\mathbf{H}}\{\log_2(\det(\mathbf{I}_{n_r} + \frac{\rho}{n_t}\mathbf{H}\mathbf{H}^h))\} \quad (1)$$

where $\rho = \frac{E_s}{N_0} = \frac{E\{\mathbf{x}^h\mathbf{x}\}}{2N_0}$, and E_s represents the total transmitted energy per channel use. This is also the total energy at each receiver antenna since all elements $h_{i,j}$ are independent and satisfy $E\{|h_{i,j}|^2\} = 1$.

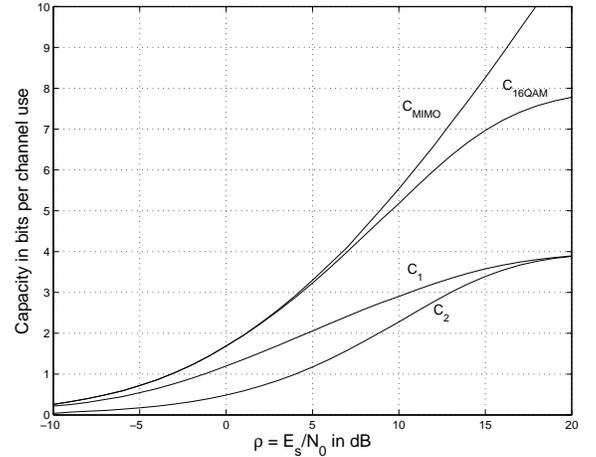


Fig. 3. Capacity of a MIMO channel. 16-QAM constellation. $n_t = n_r = 2$.

Now, in the case of a discrete 16-QAM input, the maximum of the mutual information is achieved by a uniform distribution. The capacity $C_{16QAM} = E_{\mathbf{H}}\{I(\mathbf{X}; \mathbf{Y}|\mathbf{H} = H)\}$ is calculated using the following equation

$$C_{16QAM} = E_{\mathbf{H}}\left(\frac{1}{L} \sum_{i=1}^L \int_y p_H(y|x_i) \log\left(\frac{p_H(y|x_i)}{\frac{1}{L} \sum_j p_H(y|x_j)}\right) dy\right) \quad (2)$$

where, for a fixed realization $\mathbf{H} = H$, the output of the channel \mathbf{Y} is a vector of Gaussian random variables.

$$p_H(\mathbf{y}|\mathbf{x}_i) \propto \exp\left(-\frac{\|\mathbf{y} - H\mathbf{x}_i\|^2}{2N_0}\right) \quad (3)$$

In the case of the 2-level BICM, the random variable \mathbf{X} representing the input of the channel can be written as $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ where \mathbf{X}_1 and \mathbf{X}_2 are the discrete random variables representing the transmission of the coded bits of level 1 and

2 respectively. Applying the chain rule of mutual information [12]:

$$I(\mathbf{X}; \mathbf{Y}) = I(\mathbf{X}_1; \mathbf{Y}) + I(\mathbf{X}_2; \mathbf{Y}|\mathbf{X}_1) \quad (4)$$

we get that the total capacity of the MIMO channel with 16-QAM input is equal to

$$C_{16QAM} = C_1 + C_2 \quad (5)$$

where the quantities $C_1 = E_{\mathbf{H}}\{I(\mathbf{X}_1; \mathbf{Y}|\mathbf{H} = H)\}$ and $C_2 = E_{\mathbf{H}}\{I(\mathbf{X}_2; \mathbf{Y}|\mathbf{X}_1, \mathbf{H} = H)\}$ are the capacities of the equivalent channels representing the transmission of the bits at level 1 and 2 respectively.

These capacities, as well as the capacity of the MIMO channel with continuous input, are represented in Figure 3 as function of the signal-to-noise ratio $\rho = \frac{E_s}{N_0}$.

VI. THE ITERATIVE MULTISTAGE DECODER

The receiver, represented in Figure 4, is divided into two parts. The first part is non iterative and computes the received signal conditional probabilities $p(y_r|c_1, \dots, c_{4.n_t})$ at every antenna r using

$$p(y_r|c_1, \dots, c_{4.n_t}) = \frac{e^{-\frac{\|y_r - \sum_{t=1}^{n_t} h_{t,r} x_t\|^2}{2\sigma^2}}}{2\pi\sigma^2} \quad (6)$$

The second part of the receiver is an iterative multistage decoder (MSD) [5]. MSD consists in decoding the first level of bits a_j then the second level of bits b_j knowing the first level. This is an implementation of the chain rule equation 5.

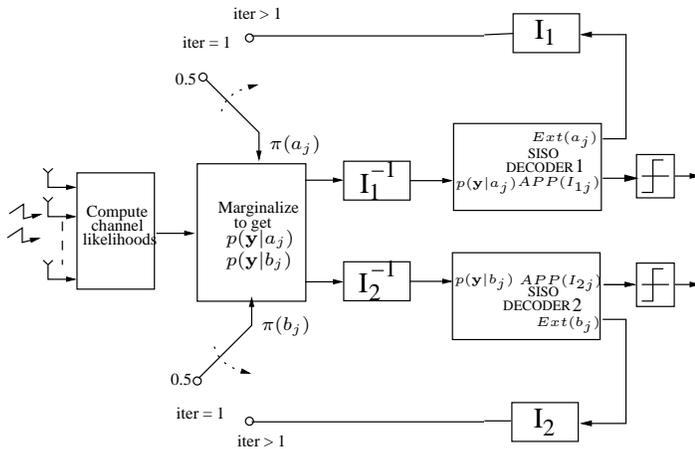


Fig. 4. Two stage iterative MSD decoder.

The decoding at each level is done using a Soft-Input Soft-Output (SISO) decoder that calculates the *a posteriori probability* (APP) associated to a coded bit

$$APP(c_j) \propto Ext(c_j) \times obs(c_j) \quad (7)$$

where $Ext(c_j)$ is the *extrinsic information*, and $obs(c_j)$ is the observation. Notice that iterative decoding is run between the SISO decoder and the "Margenalize" block as in the second method of decoding used in [8]. This is based on the fact that the observation $obs(c_j) = p(y|c_j)$ of a coded bit at level 1 ($c_j = a_j$) or 2 ($c_j = b_j$) depends also on the a priori probabilities $\pi(c_j)$ according to the following equation

$$p(y|c_j) = \sum_{c_i, i \neq j} \left(\prod_{r=1}^{n_r} p(y_r|c_1, \dots, c_{4.n_t}) \prod_{l \neq j} \pi(c_l) \right) \quad (8)$$

VII. TRANSMISSION AND COOPERATION STRATEGIES

Let us consider a user node (U1) that would like to transmit information to a certain destination. The transmission strategy of (U1) depends on the presence or not of a second user (U2) capable to work as a relay node (RN).

In the absence of a relay node (no cooperation), the user can transmit information at a full rate, of R bits per channel use, by sending data over the two levels of the modulation as previously detailed in section IV. However, in the presence of a severe degradation in signal-to-noise ratio, the user can reduce the data rate by sending information over the level 1 only. In this case, the coded bits of the level 2 are assumed to be equal to $b_j = 0$. Analyzing the gray labeling of the 16-QAM, we find that symbol mapping will lead to symbols of a sub QPSK constellation. Therefore, the modulator scheme is now equivalent to a 1-level BICM scheme combined to a QPSK constellation with a reduced rate $R_s = 4 \times R_1$ bits per channel use.

In the presence of a relay node (RN), we only studied the case where the user is transmitting data at a full rate using the two levels of the modulation. Half-duplex relay channel with decode and forward relaying protocol is considered, where the channel between the user and the relay is assumed to be error free. Two cooperation strategies are analyzed.

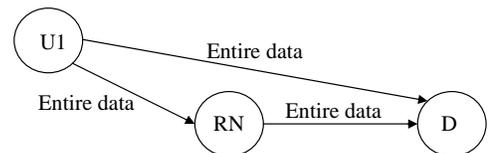


Fig. 5. First cooperation strategy.

In the first strategy, the entire data of the user is retransmitted by the relay using the two levels of the relay (see Figure 5). The cooperation scheme operates as a repetition code where each coded symbols is sent twice over the channel. This is equivalent to a non cooperation scheme with twice the number of received antennas.

Now, in the presence of another user (U3) who would like to cooperate with the same relay (RN) or in case the relay needs to send its own data, half of the coded symbols of user (U1) can be retransmitted by the relay using the level 2 of the relay (RN). The other level is left to the relay (RN) or the user (U3). This is the second cooperation strategy (see Figure 6). Notice that in this case, half of the coded bits a_j of level 1

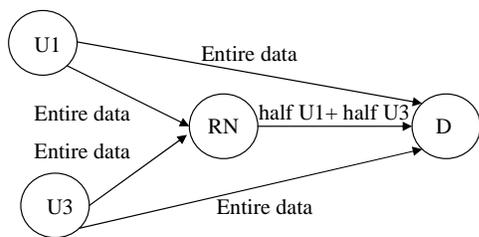


Fig. 6. Second cooperation strategy.

and half the coded bits b_j of level 2 are retransmitted by the relay.

VIII. SIMULATION RESULTS

A *non-ergodic* block fading channel with $n_t = n_r = 2$ transmit and receive antennas is considered. For the 16-QAM constellation, the total rate is equal to $R = 4$ bits per channel use. From Figure 3, we get that the optimal code rates of the encoders at levels 1 and 2, for $R = 4$, are respectively equal to $R_1 = 0.6$ and $R_2 = 0.4$. The length of the coded symbols is $N_s = 100$ symbols ($N_b = 200$ information bits). Encoder 1 is a punctured 16-state non-recursive non-systematic convolutional encoder with octal generators equal to $g = (23, 35)$. Encoder 2 is a rate-2/5 16-state non-recursive non-systematic convolutional encoder with octal generators equal to $g = (27, 71, 52, 65, 57)$. The frame error rate, as function of the signal-to-noise ratio $\frac{E_s}{N_0}$ in dB, is determined using Monte Carlo simulation.

In Figure 7, we validate the good performance of the 2-level BICM by comparing it to the one obtained using a 1-level BICM based on a rate-1/2, 16-state, non-recursive non-systematic convolutional code. We also represented the outage probability $P_{out} = Prob(C_H < R)$ of the non-ergodic block fading channel with continuous and discrete 16-QAM inputs. P_{out} is the probability that the code rate is not supported by the channel and C_H is the capacity for a fixed channel use $\mathbf{H} = H$. This is the optimal performance that can be achieved over the MIMO channel.

The frame error rate of the two cooperation strategies are sketched in Figure 8. We also represented the performance in case of no cooperation for both, full rate ($R = 4$) using a 16-QAM constellation, and reduced rate ($R_s = 2.4$) using the sub QPSK constellation.

In the absence of a relay node, results in Figure 8 show a gain of 2 dB obtained by reducing the code rate to $R_s = 2.4$ bits per channel use. In the presence of a relay node retransmitting the entire data (first cooperation strategy), Figure 8 shows a gain close to 8 dB. However, taking in consideration the energy transmitted by the relay, the total energy transmitted per channel use will be $E_s^t = 2 \times E_s$. Therefore, the real gain is equal to $8 - 10 \times \log_{10}(2) \approx 5$ dB. Notice that the performance is within 2 dB of the optimal performance achieved over the channel with a relay network retransmitting the entire data sent by the user.

Finally, when the relay node retransmit half of the coded symbols (second cooperation strategy), the total energy trans-

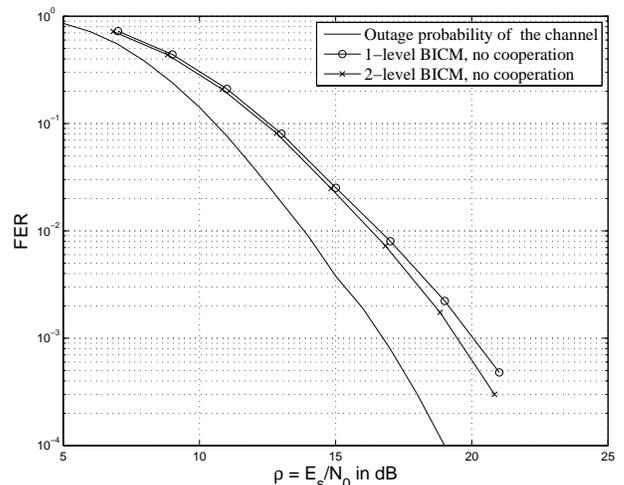


Fig. 7. Frame error rate of the 2-level bit-interleaved coded modulation.

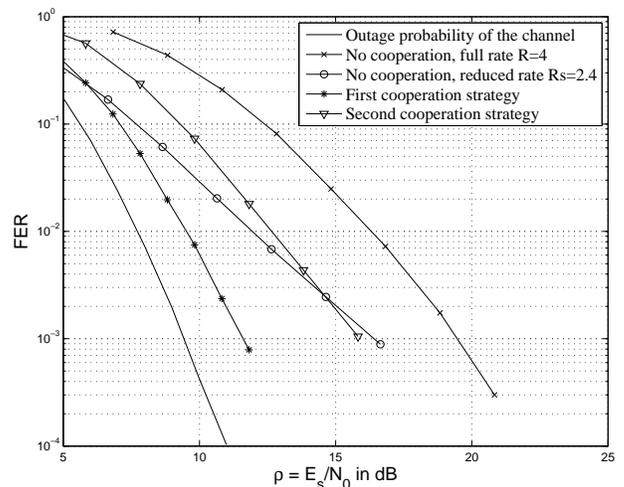


Fig. 8. Frame error rate of the two cooperation strategies compared to non cooperation.

mitted per channel use will be $E_s^t = 1.5 \times E_s$. The real gain is equal to $3.5 - 10 \times \log_{10}(1.5) \approx 1.75$ dB.

IX. CONCLUSION AND FUTURE WORK

In this paper, cooperative communications is done using a 2-level bit-interleaved coded 16-QAM modulation transmitted over a *non-ergodic* block fading channel with $n_t = 2$ transmit and $n_r = 2$ receive antennas. The scheme is flexible allowing cooperation to be done at half and full rate. In both cases, a significant improvement in performance is reached.

Further work could be on studying the cooperation between more than two sources using an L -level BICM ($L > 2$). Analyzing the performance over a frequency selective channel is another future research item.

REFERENCES

- [1] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity, part I: system description," IEEE Transactions on Communications, vol. 51, November 2003, pp. 1927-1938.

- [2] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity, part II: Implementation aspects and performance analysis," *IEEE Transactions on Communications*, vol. 51, November 2003, pp. 1939-1948.
- [3] T. E. Hunter and A. Nosratinia, "Cooperative diversity through coding," *IEEE International Symposium on Information Theory, ISIT, Lausanne, Switzerland, July 2002*, pp. 220.
- [4] B. Zhao and M. Valenti, "Distributed turbo coded diversity for the relay channel," *IEEE Electronics Letters*, vol. 39, May 2003, pp. 786-787.
- [5] U. Wachsmann, R. Fischer, and J. Huber, "Multilevel codes : Theoretical concepts and practical design rules," *IEEE Transactions on Information Theory*, vol. 45, July 1999, pp. 1361-1391.
- [6] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Transactions on Information Theory*, vol. 44, May 1998, pp. 927-946.
- [7] K. Ishii, K. Ishibashi, and H. Ochiai, "Multilevel coded cooperation for multiple sources," *IEEE Transactions on Wireless Communications*, vol. 10, December 2011, pp. 4258-4269.
- [8] H. T. Nguyen, H. H. Nguyen, and T. Le-Ngoc "Bandwidth-efficient Bit-interleaved coded Cooperative Communications," *Wireless Pers Commun* (2008) 46:247-268.
- [9] T. Islam, R. Schober, R. K. Mallik, and V. Bhargava, "Analysis and Design of Cooperative BICM-OFDM Systems," *IEEE Transactions on Communications*, vol. 59, June 2011, pp. 1742-1751.
- [10] H. E. Sawaya, "Two stage bit-interleaved coded modulation for multiple-input multiple-output channels," *IEEE Workshop on Signal Processing Advances in Wireless Communications, SPAWC, Cannes, France, July 2006*.
- [11] E. Telatar, "Capacity of multi-antenna gaussian channels," *AT&T-Bell Labs Internal Technical Memorandum*, June 1995.
- [12] R. Gallager, *Information Theory and Reliable Communication*. J. Wiley, 1968.

An Improved MU-MIMO OFDM Scheme for the 3GPP LTE-Advanced Standard

Wang Yuechao, Wu Zhanji, Wu Guanghao
Beijing University of Post & Telecommunications
Beijing, China

E-mail: wang_yuechao@163.com, wuzhanji@bupt.edu.cn, 604503597@qq.com

Abstract—In order to achieve the required data rates of 1Gbit/s and more in 3GPP LTE-Advanced wireless systems, there is an increased need for high spectral efficiency. Multi-user multiple-input-multiple-output (MU-MIMO) is a promising way to improve frequency usage efficiency. In this paper, a new multi-user MIMO OFDM scheme, named Joint Coding Modulation Diversity (JCMD) is proposed for the 3rd generation partnership project (3GPP) long term evolution advanced (LTE-Advanced). This new scheme can take advantage of the modulation diversity of rotated constellation, the frequency diversity of orthogonal frequency-division multiplexing (OFDM) system and Turbo codes all together. The simulation results show that this proposed scheme can significantly outperform the current Bit Interleaved Coded Modulation (BICM) scheme in the 3GPP LTE-Advanced standard.

Keywords—coded modulation; rotated modulation; modulation diversity.

I. INTRODUCTION

In recent years, demands on high rate wireless communication have increased. Orthogonal Frequency-Division Multiplexing (OFDM) has been proven to be a viable technique, which is proposed for next generation wireless communication system. OFDM is used to transmit data over extremely hostile channel at a comparable low complexity with high data rates [1, 2]. Channel equalization is simplified because OFDM may be viewed as using many slowly modulated narrowband signals rather than one rapidly modulated wideband signal. Compared with the conventional FDM scheme, it can save about half the bandwidth. On the other hand, Multiple-Input Multiple-Output (MIMO) systems have also been one of the key techniques to achieve high rate and high reliability over wireless channels because it can allow modulated symbols to transmit on multiple antennas at the same time to obtain high spectral efficiency. Using sophisticated coding at the transmitter and substantial signal processing at the receiver, the MIMO channel can be provisioned for higher data rates and resistance to multipath fading. A Layered Space-Time (LST) architecture that allows processing of multidimensional signals in the space domain by one dimension processing steps was proposed by Foschini [3]. Bit-Interleaved Coded Modulation (BICM) [4,5,6] is used in most of the existing communication standards (cf. HSPA, IEEE 802.11a/g, IEEE 802.16, DVB-S2, etc.) [7,8,9,10]. In BICM, the

channel encoder and the modulator are separated by a bit-level interleaver, which makes the design simple and flexible. Based on BICM-LST, which is the conventional spectral-efficient spatial multiplexing technology to deal with MIMO fading channel, MIMO-OFDM system can be used to overcome the various challenges that come with the high bit-rate, reliable and power efficient transmission demand in wide-band communication, and has become the foundation of advanced wireless transmission technologies in the current 3GPP LTE-Advanced standard.

J. Boutros et al. [11] studied the uncoded rotated multidimensional modulation schemes over Rayleigh fading channels. To distinguish from the other well-known diversity (time, frequency, code, space), the rotated modulation schemes have an essential modulation diversity and the modulation diversity order is the minimum number of distinct components between any two multidimensional components between any two multidimensional constellation points. The schemes are essentially uncoded and can achieve very high modulation diversity, and can approach Addition White Gaussian Noise (AWGN) error performance over independent flat fading channels without time-dispersion. However, it cannot be directly used for frequency selective fading channels with severe inter-symbol interference (ISI). To make full use of modulation diversity and frequency diversity over the time-dispersion fading channels with ISI, it should cooperate with OFDM. A rotated coding modulation OFDM system was proposed in [12], which extends the modulation diversity in coded OFDM systems. Through the combination of rotating the multiple phase shift keying or multiple quadrature amplitude modulation (MPSK/MQAM) constellation and interleaving the symbol components in space-time-frequency domain, the performance of wireless communications systems can be significantly improved.

A MIMO-OFDM system based on the two-dimensional modulation diversity [12] is proposed for the 3GPP LTE-advanced MU-MIMO schemes. Through jointly optimizing Turbo coded MIMO-OFDM and the rotated coding modulation for each user, the proposed scheme can take full advantage of the coding-gain of Turbo codes, the time and frequency diversity of OFDM system and the spatial diversity of MIMO all together. It is named the joint coding and modulation diversity (JCMD) MU-MIMO OFDM scheme. The rest of this paper is organized as follows.

General multi-user MIMO system model is introduced in Section II. In Section III, the JCMD MU-MIMO

OFDM scheme is proposed. Rotational modulation, the component interleaver and the detection algorithm are introduced in Sections IV, V, and VI, respectively. Section VII describes the procedure of selection of the codebook and the user, which makes the system throughput maximized and Section VIII gives the simulation results. Section IX concludes the paper.

II. MU-MIMO SYSTEM MODEL

In the multi-user MIMO system, which is showed in Figure 1, K users are assumed. We consider a MU-MIMO downlink channel, where M_T transmit antennas are located at the base station and M_{R_i} receive antennas are located at the i^{th} user terminal (UT), $i=1,2,\dots,K$. The total number of receive antennas is as follows

$$M_R = \sum_{i=1}^K M_{R_i}.$$

Let the notation $\{M_{R_1}, \dots, M_{R_K}\} \times M_T$ denote the antenna configuration of the system. Let the MIMO channel of user i be denoted as $\mathbf{H}_i \in \mathbb{C}^{M_{R_i} \times M_T}$. Then, the combined channel matrix is given by

$$\mathbf{H} = [\mathbf{H}_1^T, \mathbf{H}_2^T, \dots, \mathbf{H}_K^T]^T \in \mathbb{C}^{M_R \times M_T} \quad (1)$$

The data vectors $\mathbf{x}_k \in \mathbb{C}^{b_k \times 1}$, $k=1, \dots, K$, for the K users are stacked in the vector $\mathbf{x} = [\mathbf{x}_1^T, \dots, \mathbf{x}_K^T]^T \in \mathbb{C}^{b \times 1}$. The received vector is given by

$$\mathbf{y} = \mathbf{G}(\mathbf{H}\mathbf{w}\mathbf{x} + \mathbf{n}) \quad (2)$$

where $\mathbf{y} = [\mathbf{y}_1^T, \dots, \mathbf{y}_K^T]^T \in \mathbb{C}^{b \times 1}$ is the received data vector, $\mathbf{n} = [\mathbf{n}_1^T, \dots, \mathbf{n}_K^T]^T \in \mathbb{C}^{M_R \times 1}$ is the stacked vector of the zero mean additive white Gaussian noise at the input of the receive antennas. The joint precoding matrix and decoding matrices are denoted by \mathbf{w} and \mathbf{G} , respectively. Define the joint precoder matrix as

$$\mathbf{w} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_K] \in \mathbb{C}^{M_T \times b} \quad (3)$$

where $\mathbf{w}_k \in \mathbb{C}^{M_T \times b_k}$ is the k^{th} user's precoder matrix. Moreover, b_k is the number of data stream sequences transmitted to the k^{th} user, and $b = \sum_{k=1}^K b_k \leq \text{rank}(\mathbf{H}) \leq \min(M_R, M_T)$ is the total number of the transmitter data streams. The decoding matrix \mathbf{G} can be written as

$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathbf{G}_K \end{pmatrix} \in \mathbb{C}^{b \times M_R} \quad (4)$$

where $\mathbf{G}_k \in \mathbb{C}^{b_k \times M_{R_k}}$ is the k^{th} user's decoding matrix.

The received symbol at the k^{th} user can be expressed as $\mathbf{r}_k = \mathbf{H}_k \mathbf{w} \mathbf{x} + \mathbf{n}_k \in \mathbb{C}^{M_{R_k} \times 1}$, and the received data vector is given by $\mathbf{y}_k = \mathbf{G}_k \mathbf{r}_k = \mathbf{G}_k (\mathbf{H}_k \mathbf{w} \mathbf{x} + \mathbf{n}_k) \in \mathbb{C}^{b_k \times 1}$. For the sake of simplicity, we consider a system with

$b_k = N_L, k=1, \dots, K$ and each generic user with J antennas, $M_{R_k} = J, k=1, \dots, K$ in the rest of this paper.

III. JOINT CODING AND MODULATION DIVERSITY MU-MIMO OFDM SCHEME

In [13], the principle of MU-MIMO with per user unitary and rate control (PU2RC) is described. The feedback does not indicate the channel itself, but a preferred precoder from a set of predefined precoders. Based on the conventional MU-MIMO scheme with PU2RC, an improved MU-MIMO OFDM scheme based on modulation diversity is proposed, which is illustrated in Figure 2. The blocks drawn in dotted line are the additional proposed processing based on the conventional MU-MIMO OFDM system in the 3GPP LTE standard.

At the transmitter, the base station firstly selects a group of users in the set of users to be scheduled according to the procedure of PU2RC-MIMO with partial feedback, which will be described in Section VI. Then, the information bits of selected group users are firstly encoded by 3GPP LTE Turbo code [14]. The conventional quadrature amplitude modulation (QAM) is carried out to produce complex-valued symbols by constellation mapper. The conventional modulated symbols are converted into new complex-valued rotated symbols by a rotational modulation mapper. The rotated symbols contain In-phase (I) components and Quadrature (Q) components and subsequently are mapped onto N_L transmission layers. Afterwards, the spatial interleaving and spatial Q-interleaving that will be described in Section V are carried out over each user's N_L transmission layers. The interleaved symbols are then mapped onto different time-frequency resource elements (REs). For the Q-components, an additional time-frequency Q-interleaver is performed. After the codebook-based precoding [15], N_{FFT} -point Inverse Fast Fourier Transform (IFFT) and inserting cyclic prefix (CP) are implemented on each transmit antenna.

At the receiver of each UT, removing CP and N_{FFT} point FFT are carried out first. The minimum mean squared-error (MMSE) equalization in the frequency domain is performed. Afterwards, the received symbols of each spatial stream are obtained after the time-frequency Q-component deinterleaving, time-frequency resource element demapping, spatial Q-deinterleaving, spatial deinterleaving, and layer demapping. For each symbol, the maximum likelihood (ML) rotational demodulation is used to produce the log-likelihood-ratios (LLRs) of encoded bits. Finally, the Turbo decoder utilized the

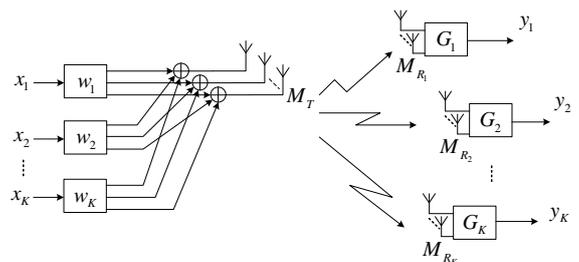


Figure 1. Block diagram of a multi-user MIMO downlink system

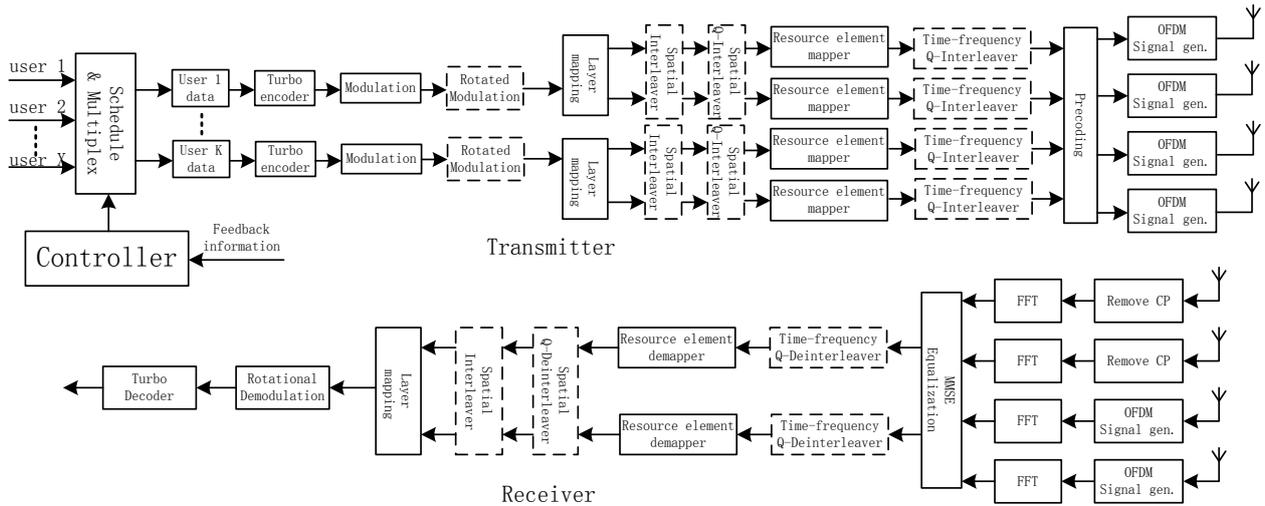


Figure 2. A Transmitter and Receiver Diagram of the JCMD MU-MIMO OFDM scheme

LLRs to decode the information bits.

IV. ROTATIONAL MODULATION

Compared with the usual QAM, rotated constellation can obtain the modulation diversity by rotating some angle [11]. Combining rotating the conventional constellation and interleaving the components, the performance of wireless communications systems can be improved greatly in fading channels.

At the transmitter, coded bits $\mathbf{K} = (k_1, k_2, \dots, k_{N_{bit}})$ are converted into a sequence of complex-valued modulation symbols $\mathbf{D} = (d_1, d_2, \dots, d_{N_{sym}})$ by Gray-mapping [11].

The rotational complex-valued symbol $d_i = X_i + jY_i$ can be obtained by multiplying the conventional modulated complex-valued symbol $d_i = I_i + jQ_i$ with a rotational matrix \mathbf{R} , which is shown as follows.

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} = \mathbf{R} \begin{pmatrix} I_i \\ Q_i \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} I_i \\ Q_i \end{pmatrix} \quad (5)$$

This processing can also be illustrated in Figure 3. In the conventional square QPSK constellation, the I-component and Q-component of one complex-valued modulated symbol just carry one bit, respectively. After the operation of constellation rotation, both I-component and Q-component of each point in the new constellation carry the information of two bits, which means that the information of two bits exists across the I-component and Q-component. After appropriate interleaving of the Q-components, the fading coefficients of the I-components and Q-components are uncorrelated. The modulation diversity order L of a multidimensional signal set is the minimum number of distinct components between any two constellation points. Assuming independent Rayleigh fading channel and the maximum likelihood (ML) detection, multidimensional QAM constellation becomes insensitive to fading channel when the diversity L is large [16].

By adjusting the rotation angle, the optimum

modulation diversity can be obtained to minimize bit error rate (BER). J. Boutros et al. [11] studied the optimum criteria of the uncoded rotation modulation over independent Rayleigh channel, and proposed that the diversity order L and product distance should be maximized [11]. Simulation results show that the best angle of rotation depends on modulation order and coding rate. Also, there is weak correlation between the code rate and the optimum rotation angle, especially from a high coding rate of 3/4 and 8/9. However, it is strongly relevant to the modulation order. For the same code rate, the optimal rotation angle values are different for QPSK, 16QAM and 64QAM, and it decreases as the modulation order increases. Based on the above theoretical analysis and computer simulation, we obtain the optimal rotation angles that are suitable for the improved MU-MIMO OFDM systems based on modulation diversity, which are shown in Table I.

V. Q-COMPONENT INTERLEAVER

A. Spatial interleaver and Spatial Q-component interleaver

The spatial interleaving is the conventional spiral layer interleaving among each user's transmit layers. Let \mathbf{x}_t^i and $\bar{\mathbf{x}}_t^i$ denote the input-symbol and the output-symbol of the i^{th} transmit layer at the t instant. The interleaving is defined as follows.

$$\bar{\mathbf{x}}_t^k = \mathbf{x}_t^i, k = (i+t) \bmod N_L \quad (6)$$

where $k, i \in [0, N_L - 1]$. The spatial Q-component interleaver is carried out after the spatial interleaving to ensure the independence of the I-components and Q-components. Let Q_i^i and \tilde{Q}_i^i denote the input Q-component and the output Q-component of the Q-component interleaver on the i^{th} transmit layer at the t instant. The Q-component interleaving is defined as follows.

TABLE I. THE OPTIMAL ROTATION ANGLES FOR DIFFERENT CODE

RATE			
Code rate	QPSK	16QAM	64QAM
3/4	26.6°	18.4°	14.0°
5/6	26.6°	18.4°	14.0°
8/9	26.6°	18.4°	14.0°

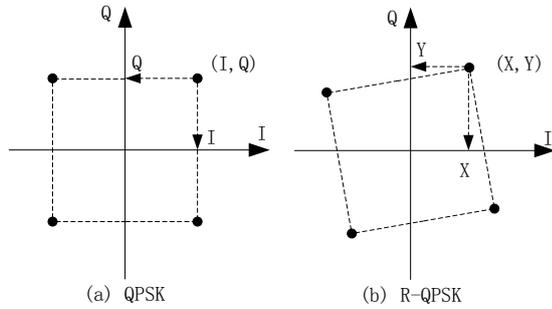


Figure 3. Conventional QPSK constellation (left) and Rotated QPSK constellation

$$\tilde{Q}_i^k = Q_i^k, k = N_L - i - 1 \quad (7)$$

where $k, i \in [0, N_L - 1]$.

B. Time-frequency Q-component interleaver

The time-frequency Q-component interleaving is aimed to make the I-component and the Q-component in one modulated symbol as uncorrelated as possible in the time and frequency domain. We assume the OFDM system has L subcarriers in frequency domain for each user and N_{sym} OFDM symbols in time domain. One Q-component of a complex-valued modulated symbol is mapped onto one resource element $Q(f, t)$ at the f^{th} subcarrier in the t^{th} OFDM symbol. After time-frequency Q-component interleaving, the output is $\tilde{Q}(f, t)$. The Q-component interleaving is defined as follows.

$$\begin{aligned} f' &= (f + L/2) \bmod L \\ t' &= (t + N_{sym}/2) \bmod N_{sym} \end{aligned} \quad (8)$$

where $f \in [0, L-1], t \in [0, N_{sym}-1]$. Thus, the time interval between I-component and Q-component is $N_{sym}/2$ OFDM symbol duration. In frequency domain, the frequency interval is half the bandwidth for each user. So, the time-frequency Q-component interleaver can make full use of the frequency diversity and the time diversity of OFDM system, and it can make the I-components and Q-components as uncorrelated as possible. Based on the use of spatial interleaver, the achievable rate of the proposed scheme can be increased, which is demonstrated as follows.

In (3), \mathbf{w}_k denotes the k^{th} user's precoder matrix, and the equivalent channel matrix between the base station and k^{th} user is expressed as $\mathbf{H}_E = \mathbf{H}_k \mathbf{w}_k \in \mathbb{C}^{J \times N_L}$. Let $q = \min(J, N_L)$. Assuming the transmission power is P and the bandwidth is B , and the q singular values are

$\sqrt{\lambda_1}, \dots, \sqrt{\lambda_q}$ ($\sqrt{\lambda_1} \geq \sqrt{\lambda_2} \geq \dots \geq \sqrt{\lambda_q}$). The total rate each user can achieve is the sum of partial achievable rate on each singular-value-decomposition (SVD) spatial layer of

the MIMO channel [17].

For the system without spatial Q-component interleaver, the fading coefficients of I-component and Q-component in each symbol are identical. The achievable rate can be calculated as follows.

$$C_1 = B \cdot \log_2 \left\{ \left(1 + \frac{\lambda_1 P}{2\sigma^2}\right) \left(1 + \frac{\lambda_2 P}{2\sigma^2}\right) \cdots \left(1 + \frac{\lambda_q P}{2\sigma^2}\right) \right\} \quad (9)$$

For the proposed system, thanks to the spatial Q-component interleaver, the fading coefficients of I-component and Q-component in each symbol are $\sqrt{\lambda_i}$ and $\sqrt{\lambda_{(i+1) \bmod q}}$, $i = 1, \dots, q$, respectively. Thus, the achievable rate of the proposed system can be calculate as follows.

$$C_2 = B \cdot \log_2 \left\{ \left(1 + \frac{\lambda_1 + \lambda_2}{4\sigma^2} P\right) \left(1 + \frac{\lambda_2 + \lambda_3}{4\sigma^2} P\right) \cdots \left(1 + \frac{\lambda_q + \lambda_1}{4\sigma^2} P\right) \right\} \quad (10)$$

When $q=2$, we can easily get the following equations,

$$\begin{aligned} C_1 &= B \cdot \log_2 \left\{ \left(1 + \frac{\lambda_1 P}{2\sigma^2}\right) \left(1 + \frac{\lambda_2 P}{2\sigma^2}\right) \right\} \\ &= B \cdot \log_2 \left\{ 1 + \frac{\lambda_1 + \lambda_2}{2\sigma^2} P + \frac{\lambda_1 \lambda_2}{4\sigma^4} P^2 \right\} \end{aligned} \quad (11)$$

and

$$\begin{aligned} C_2 &= B \cdot \log_2 \left\{ \left(1 + \frac{\lambda_1 + \lambda_2}{4\sigma^2} P\right) \left(1 + \frac{\lambda_2 + \lambda_1}{4\sigma^2} P\right) \right\} \\ &= B \cdot \log_2 \left\{ 1 + \frac{\lambda_1 + \lambda_2}{2\sigma^2} P + \frac{(\lambda_1 + \lambda_2)^2}{16\sigma^4} P^2 \right\} \end{aligned} \quad (12)$$

It can easily obtain $C_1 \leq C_2$, because $(\lambda_1 + \lambda_2)^2 \geq 4\lambda_1 \lambda_2$. Likewise, for $q > 2$, we can get the same result.

VI. DETECTION AND DEMODULATION

The received symbol at the k^{th} user can be expressed as

$$\mathbf{r}_k = \mathbf{H}_k \mathbf{w}_k \mathbf{x}_k + \sum_{i=1, i \neq k}^K \mathbf{H}_k \mathbf{w}_i \mathbf{x}_i + \mathbf{n}_k \quad (13)$$

where the first term on the right hand side is the useful data for k^{th} user. The middle term represents multi-user interference (MUI) and the last term is the noise vector, which contains i.i.d. complex Gaussian random numbers with variance σ^2 . The minimum mean-squared error (MMSE) equalization is used for each antenna of user after the FFT operation. Assuming the perfect channel state information (CSI), the MIMO channel matrix between the base station and user terminal is known to the receiver. Let $\mathbf{H}_k = [\mathbf{h}_1^T, \dots, \mathbf{h}_J^T]^T$ denote a $J \times M_T$ channel for that UT. We use $\tilde{\mathbf{H}}_k = \mathbf{H}_k \mathbf{w}$ to denote the product matrix between the UT's channel and the precoder, where w denotes the precoding matrix. Based on the MMSE criterion, the receiver combining coefficients for the k^{th} user can be defined as follows.

$$\mathbf{G}_k = \tilde{\mathbf{h}}_k^H \left(\tilde{\mathbf{H}}_k \tilde{\mathbf{H}}_k^H + 2K\sigma^2 \mathbf{I}_{M_T} \right)^{-1} \quad (14)$$

$$\tilde{\mathbf{h}}_k = \mathbf{H}_k \mathbf{w}_k = [\tilde{\mathbf{H}}_k]_k = [\mathbf{H}_k \mathbf{w}_k]_k \quad (15)$$

where \mathbf{w}_k denotes the k^{th} user's precoder matrix. $[\cdot]_k$ denotes the column vectors that correspond to k^{th} user. Then, the received data vector is given by

$$\begin{aligned} \mathbf{y}_k &= \mathbf{G}_k \mathbf{r}_k = \mathbf{G}_k \mathbf{H}_k \mathbf{w}_k \mathbf{x}_k + \sum_{i=1, i \neq k}^K \mathbf{G}_k \mathbf{H}_k \mathbf{w}_i \mathbf{x}_i + \mathbf{G}_k \mathbf{n}_k \\ &= \mathbf{A}_k \mathbf{x}_k + \sum_{i=1, i \neq k}^K \tilde{\mathbf{A}}_i \mathbf{x}_i + \mathbf{G}_k \mathbf{n}_k \end{aligned} \quad (16)$$

where $\mathbf{A}_k \in \mathbb{C}^{N_L \times N_L}$, $\tilde{\mathbf{A}}_i \in \mathbb{C}^{N_L \times N_L}$, $i=1, i \neq k$. Let $\mathbf{A}_k^{(m,n)}$ denotes the element on the m^{th} row and the n^{th} column. Likewise, $\tilde{\mathbf{A}}_i^{(m,n)}$ denotes the element $\tilde{\mathbf{A}}_i(m,n)$, and $\mathbf{G}_k^{(m,n)}$ denotes the element $\mathbf{G}_k(m,n)$. For k^{th} user, the symbol on the receive layer m can be expressed as

$$\begin{aligned} y_k^{(m)} &= \{ \mathbf{A}_k^{(m,1)} \mathbf{x}_k^{(1)} + \dots + \mathbf{A}_k^{(m,m)} \mathbf{x}_k^{(m)} \dots + \mathbf{A}_k^{(m,N_L)} \mathbf{x}_k^{(N_L)} \} \\ &+ \{ \sum_{i=1, i \neq k}^K (\tilde{\mathbf{A}}_i^{(m,1)} \mathbf{x}_i^{(1)} + \dots + \tilde{\mathbf{A}}_i^{(m,N_L)} \mathbf{x}_i^{(N_L)}) \} \\ &+ \{ \mathbf{G}_k^{(m,1)} \mathbf{n}_k^{(1)} + \dots + \mathbf{G}_k^{(m,J)} \mathbf{n}_k^{(J)} \} \end{aligned} \quad (17)$$

Assuming the normalized transmit signal power $\sigma_s^2 = 1$, the signal to interference plus noise ratio (SINR) for the symbol on the receive layer m is estimated as follows.

$$SINR_k^m = \frac{|\mathbf{A}_k^{(m,m)}|^2}{\sum_{i=1, i \neq m}^{N_i} |\mathbf{A}_k^{(m,i)}|^2 + \sum_{i=1, i \neq k}^K (|\tilde{\mathbf{A}}_i^{(m,1)}|^2 + \dots + |\tilde{\mathbf{A}}_i^{(m,N_L)}|^2) + 2\sigma^2 \sum_{i=1}^J |\mathbf{G}_k^{(m,i)}|^2} \quad (18)$$

We assume the rotated constellation point in the transmitter is $S = (S_I, S_Q)$, and the corresponding received constellation point after the MMSE detection is $R = (R_I, R_Q)$, the SINR computed in equation (18) is the corresponding SINR for R_I and R_Q , denoted by $SINR_I$ and $SINR_Q$, respectively. Thanks to the above Q-component interleaving, the SINRs of I-component and Q-component in each symbol are different. The rotational demodulation produces the LLRs of encoded bits $\mathbf{K} = [k_1, k_2, \dots, k_l]^T$ as follows.

$$LLR(k_l) = \ln \frac{P(k_l = 0)}{P(k_l = 1)} \quad (19)$$

where k_l denotes the i^{th} bit of one symbol, $l = \log_2 M$. Assuming equal *a priori* probabilities, P is calculated as follows.

$$P = \exp[-(R_I - S_I)^2 \cdot SINR_I] \cdot \exp[-(R_Q - S_Q)^2 \cdot SINR_Q] \quad (20)$$

VII. PRECODING MATRIX AND USER SELECTION ALGORITHM

According to the equation (18), for the k^{th} user, the value of $SINR_k^m$ is not only dependent on its channel

matrix \mathbf{H}_k , but also determined by its corresponding precoding vector \mathbf{w}_k and the other user's precoding vector \mathbf{w}_i . A precoding matrix \mathbf{w}^p will correspond to $K \times N_L$ SINRs. According to the known precoding codebook $\mathbf{W} = \{\mathbf{w}^1, \mathbf{w}^2, \dots, \mathbf{w}^L\}$, then the total number of SINRs for each user is $K \times N_L \times L$. Each user will feedback these SINRs to the base station to perform scheduling.

The scheduling method of maximizing system capacity will be used in this paper. The transmitting side will select K users from the X users to be scheduled, and select the corresponding precoding matrix to maximize system throughput. For the multi-user MIMO system, system capacity can be expressed as

$$C_{MU}(SINR) = \sum_{k=1}^K \sum_{m=1}^{N_L} \log_2(1 + SINR_k^m) \quad (21)$$

The method of choosing optimal precoding matrix that maximizes system capacity can be defined as

$$\mathbf{w} = \arg \max_{\mathbf{w} \in \mathbf{W}} C_{MU}(SINR) \quad (22)$$

where the constraint is that K users correspond to the different precoding vectors in the same precoding matrix, as follows

$$\mathbf{w}_k = [\mathbf{w}]_k \quad (23)$$

The detailed steps are as follows.

- 1) Each user computes $K \times N_L \times L$ SINRs according to their channel matrix and the precoding codebook, and then sends these SINRs to the base station.
- 2) The base station collects all the feedback information. Select K users from the X users to be scheduled and corresponding precoding matrix according to the principle of maximizing system capacity. Each user in the selected users group corresponds to the precoding vectors \mathbf{w}_k in the precoding matrix \mathbf{w} .
- 3) Apply precoding scheme corresponding to the selected group.

VIII. SIMULATION RESULTS

Assuming that each user has two data streams $N_L = 2$, computer simulations are carried out to compare the proposed scheme with the conventional MU-MIMO OFDM scheme in the current 3GPP LTE-Advanced standard. The frame structure, physical resource elements of OFDM system and Turbo code are given in Table II.

Figure 4 depicts the JCMD-vs-BICM frame error rate (FER) performance comparison for 3/4 and 5/6. In the figure, the proposed scheme JCMD MU-MIMO OFDM scheme can obtain obvious SNR gain over the conventional BICM MU-MIMO OFDM scheme in current 3GPP LTE standard.

For QPSK modulation, SNR gains are about 3.52dB and 4.55dB at FER=0.01 for 3/4 and 5/6 code rate, respectively. For 16QAM modulation, SNR gains are about 3.33dB and 3.83dB at FER=0.01 for 3/4 and 5/6 code rate, respectively. Thus, for high code rate, the SNR gain becomes more significant.

TABLE II. SIMULATION PARAMETERS

Parameter	Value
Number of transmit antennas	4
Number of receive antennas	4
Type of transmit precoding	Unitary
Number of UEs	9
Transmission bandwidth	5MHz
Centre frequency	2GHz
DFT Size	512
Sample Period	1/(15000×512)s
Channel Model	ITU-R IMT-Advanced(M.2135)
Mobile Speed	3km/h
Turbo Code Rates	3/4,5/6
Modulation Type	QPSK, 16QAM

REFERENCES

- [1] Van Nee, R., and R. Prasad, "OFDM for wireless multimedia communication," Norwood, MA: Artech House, 2000.
- [2] Prasad, R., "Universal wireless personal communication," Norwood, MA: Artech House, 1998.
- [3] G. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," Bell Labs Technical Journal, Aut. 1996, pp.41-59
- [4] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," IEEE Trans. Commun., vol. 40, no. 3, May 1992, pp. 873-884.
- [5] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," IEEE Trans. Inf. Theory , vol. 44, no. 3, May 1998, pp. 927-946.
- [6] A. Guillén i Fàbregas, A. Martinez, and G. Caire, "Bit-interleaved coded modulation," Foundations and Trends in Communications and Information Theory , vol. 5, no. 1-2, 2008, pp. 1-153.
- [7] 3rd Generation Partnership Project Technical Specifications, available online at <http://www.3gpp.org>.
- [8] IEEE 802.11a Standard. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications.
- [9] IEEE 802.16 Standard. IEEE Standard for Local Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access System.
- [10] ETSI EN. "Second Generation Framing Structure, Channel Coding and Modulation Systems for Broadcasting, Interactive Service, News Gathering and Other Broadband Satellite Applications (DVB-S2)," ETSI EN 302 307 version 1.2.1 Jan. 2012.
- [11] J. Boutros and E. Viterbo, "Signal space diversity: a power and bandwidth efficient diversity technique for the Rayleigh fading channel," IEEE Trans. Inform. Theory, vol.44, July 1998, pp. 1453-1467.
- [12] W. Zhanji and W. Wenbo, "A novel joint coding modulation diversity OFDM system," Global Mobile Congress 2010, Oct. 2010.
- [13] G. Bauch and G. Dietl, "Multi-User MIMO for achieving IMT-Advanced requirements," in Proc. Int. Conf. Telecommun. (ICT), Nov. 2008, pp. 1-7.
- [14] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Multiplexing and channel coding (Release 10), 3GPP TS 36.212, version 10.2.0 (2011-06).
- [15] Samsung, "Downlink MIMO for EUTRA," in 3GPP TSG-RAN WG1, 44 R1-060335, Feb. 2006.
- [16] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 10), 3GPP TS 36.211, version 10.2.0 (2011-06).
- [17] A. Goldsmith, "Wireless communications," Cambridge University Press, Aug. 2005, pp. 299-309.

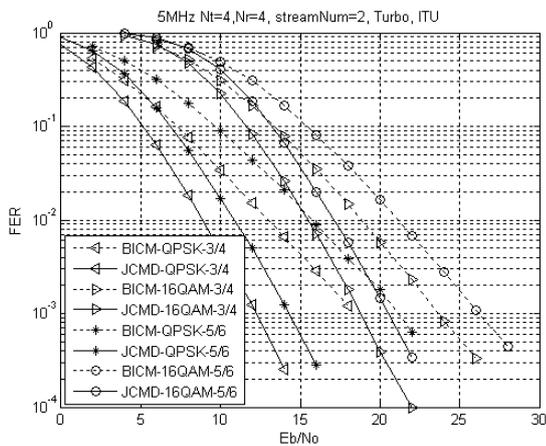


Figure 4. FER performance of JCMD vs BICM MU-MIMO OFDM for 3/4 and 5/6 code rate, QPSK, 16QAM IMT-Advanced

IX. CONCLUSION

In this paper, an improved joint coding and modulation diversity (JCMD) MU-MIMO scheme is proposed. Compared with the current solution BICM, the JCMD scheme only introduces limited complexity, which was brought by the several interleavers. A low-complexity time-frequency component interleaver is proposed to maximize the frequency diversity and the time diversity of OFDM system. The proposed scheme makes full use of Turbo codes coding gain, time and frequency diversity of OFDM system, and the spatial diversity of MIMO.

ACKNOWLEDGMENT

This work is sponsored by National Natural Science Fund (61171101) and National Great Science Specific Projects (2009ZX03003-011-03 & 2010ZX03005-001-02) of P.R. China.

Carrier Frequency Recovery for Oversampled Perfect Reconstruction Filter Bank Transceivers

Siavash Rahimi

Dept. of Electrical and Computer Eng.
McGill University
Montreal, Canada
siavash.rahimi@mail.mcgill.ca

Benoit Champagne

Dept. of Electrical and Computer Eng.
McGill University
Montreal, Canada
benoit.champagne@mcgill.ca

Abstract—In this paper, we consider the problem of data-aided carrier frequency offset (CFO) estimation for filter bank multicarrier systems, with emphasis on oversampled perfect reconstruction filter banks. By exploiting statistical properties of the transmitted pilots in such systems, the maximum likelihood (ML) estimator of the CFO is derived and its performance is investigated numerically for different channel scenarios. The Cramer Rao bound (CRB) on CFO estimator variance for the additive white Gaussian noise (AWGN) channel is also derived as a performance benchmark. Simulation results show that the proposed ML estimator reaches the provided CRB over AWGN channel, while it also exhibits a robust performance in the case of frequency selective channels.

Keywords—Data-aided estimation, carrier frequency offset estimation, maximum likelihood, filter bank multicarrier systems.

I. INTRODUCTION

Due to its desirable characteristics, multicarrier modulation (MCM) is currently the main choice for high speed wireless communications. For instance, one specific form of MCM, orthogonal frequency division multiplexing (OFDM), has been used in many standards, including WiMAX and LTE-Advanced. Recently, to overcome certain limitations of OFDM, alternative forms of MCM have been proposed, which fall into the general category of filter bank multicarrier systems (FBMC). Filtered multitone (FMT) [1] and oversampled perfect reconstruction filter banks (OPRFB) [2], [3] are examples of such systems. However, while FMT and OPRFB have been shown to be less sensitive than OFDM to carrier frequency offset (CFO) [4], some counter measure techniques should be also applied to fully exploit the benefits of MCM in this case.

To mitigate this sensitivity and remove the CFO

effect, limited number of frequency estimation or synchronization algorithms for FBMC systems have been considered. In [5], a *non data-aided* (i.e., blind) CFO estimator for FMT systems is obtained based on the maximum likelihood (ML) principle. Another blind CFO estimator based on the best linear unbiased estimation (BLUE) principle, under the assumption of additive white Gaussian noise (AWGN) channel, is proposed in [6]. Alternatively, a *data-aided* joint symbol timing and frequency synchronization scheme for FMT systems is presented in [7], where the synchronization metric is derived by calculating the time-domain correlation between the received signal and a known pseudo-random training sequence. In [8], a synchronization scheme for data-aided symbol timing and frequency offset recovery is developed by employing the least-squares (LS) approach and exploiting the known structure of a special training sequence whose estimation error does not reach the provided lower bound. Authors in [9], [10] propose CFO estimation schemes based on the ML criterion that are specifically tailored for FBMC systems. While [9] uses a sequence of consecutive pilots, the approach in [10] employs scattered pilots instead, which helps estimate and track channel variations during multicarrier burst transmissions. This method exhibits an improved estimation accuracy when compared to the blind ones, while requiring a moderate pilot overhead and a low complexity.

Although these data-aided methods perform well for FMT systems with root raised cosine prototype filters, they do not demonstrate the same level of accuracy for OPRFB systems. In particular, since OPRFB transceiver employs longer prototype filters, some of the essential assumptions of these methods, e.g., constant CFO effect during the filter support in [9], [10], are not valid in prac-

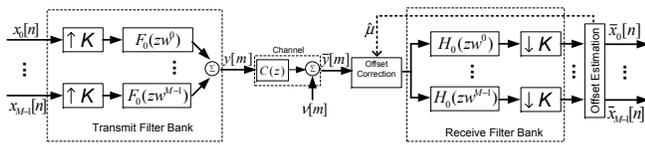


Fig. 1. DFT modulated OPRFB transceiver

tice, which in turn significantly degrades the estimation accuracy as we have been able to observe and therefore, they can not be applied to OPRFB systems that employ longer filters. Consequently, finding a synchronization method well suited for OPRFB systems is of particular interest.

In this paper, we propose a data-aided, ML-based CFO estimation method that is well suited to OPRFB-based MCM systems with longer prototype filters, and investigates its performance numerically for different wireless channel scenarios. By applying judicious simplifications to the log-likelihood function and ignoring the negligible terms, we are able to significantly reduce the implementation complexity of the proposed ML estimator. The Cramer Rao bound (CRB) on CFO estimator variance for the AWGN channel is also derived as a performance benchmark. Results of simulation experiments show that the proposed ML estimator reaches the provided CRB over AWGN channel, while it exhibits a robust performance in the case of frequency selective channels, despite the fact that no channel state information (CSI) is employed.

The paper organization is as follows. The OPRFB system model is outlined in Section II along with a discussion of CFO in such systems. The proposed ML estimator for CFO and the CRB are presented in Section III. The estimator performance is assessed in Section IV. Finally, conclusions are drawn in Section V.

Notations: Bold-faced letters indicate vectors and matrices. Superscript $(\cdot)^*$ denotes the complex conjugation, $\text{Re}[\cdot]$ the real part and $|\cdot|$ the absolute value of a complex number. \mathbf{I}_K denotes the $K \times K$ identity matrix. Moreover, $(\cdot)^T$ represent transpose, $(\cdot)^H$ Hermitian, $(\cdot)^{-1}$ the inverse of a matrix, while $E[\cdot]$ stands for the statistical expectation.

II. PROBLEM FORMULATION

A. OPRFB System Model

We consider a DFT modulated OPRFB transceiver system, as depicted in Fig. 1, where parameters M and

K represent the number of subbands and the upsampling/downsampling factor, respectively, and $K > M$ is assumed. Here, $x_i[n]$ denotes the complex-valued data sequence transmitted on the i th subband at discrete-time nT_s , where $i \in \{0, \dots, M-1\}$, $n \in \mathbb{Z}$, $T_s = F_s^{-1}$ and F_s is the input sampling rate. In DFT modulated FBMC systems, the transmit and receive subband filters can be derived from common prototypes with finite impulse responses (FIR) of length D and respective system functions $F_0(z) = \sum_{n=0}^{D-1} f_0[n]z^{-n}$ and $H_0(z) = \sum_{n=0}^{D-1} h_0[n]z^n$. For convenience in analysis, $H_i(z)$ is assumed non-causal although in practice, causality can be restored simply by introducing an appropriate delay in the receiver. Defining $w = e^{-j2\pi/M}$, the transmit and receive filters for the i th subband are respectively obtained as

$$F_i(z) = F_0(zw^i), \quad H_i(z) = H_0(zw^i). \quad (1)$$

In this work, the filter length D is restricted to be a multiple of M and K , i.e., $D = d_P P$, where P denotes the least common multiple of M and K and d_P is a positive integer. As proposed in [2], [3], to enforce the perfect reconstruction (PR) property, the paraconjugates of the transmit filters are employed as receive filters, i.e., $h_i[n] = f_i^*[n]$. Therefore, PR can be expressed as

$$\sum_{q=-\infty}^{\infty} f_j[q - pK] f_i^*[q - nK] = \delta_{ij} \delta_{np}, \quad (2)$$

where δ_{ij} denotes the Kronecker delta function. As shown in Fig. 1, the transmitter output signal at discrete-time mT_s/K is given by

$$y[m] = \sum_{i=0}^{M-1} \sum_q x_i[q] f_i[m - qK], \quad (3)$$

where the range of the summation over q is delimited by the finite support of the subband FIR filter, $f_i[m]$.

We assume that during a time interval equal to the processing delay of the transceiver system (i.e., $2DT_s/K$), the transmission channel can be modelled as a linear time-invariant system with FIR $c[l]$ of length Q and corresponding system function $C(z) = \sum_{l=0}^{Q-1} c[l]z^{-l}$. The channel output is corrupted by an AWGN sequence $\nu[m]$, with zero-mean and variance $E[|\nu[m]|^2] = \sigma_\nu^2$, assumed to be statistically independent from the input data. The input-output relationship of the noisy channel can therefore be expressed as

$$\bar{y}[m] = \sum_{l=0}^{Q-1} c[l]y[m-l] + \nu[m], \quad (4)$$

where $\bar{y}[m]$ denotes the received baseband discrete-time signal. On the receiver side, $\bar{y}[m]$ is passed through a bank of analysis filters and downsampled by K . Accordingly, for each subband, the reconstructed signal $\bar{x}_i[n]$ can be written as

$$\bar{x}_i[n] = \sum_q \bar{y}[q] f_i^*[q - nK]. \quad (5)$$

B. Effect of Carrier Frequency Offset

In practice, there often exists a mismatch between the carrier frequency in the receiver and the transmitter, denoted as CFO. In this case, the received signal $\bar{y}[m]$ can be modelled as [6]–[8]

$$\bar{y}[m] = e^{j2\pi\mu m} \sum_{l=0}^{Q-1} c[l] y[m-l] + \nu[m], \quad (6)$$

where μ is a normalized CFO with respect to the subband spacing $F_s K/M$. Upon substitution of (3) and (6) into (5), the reconstructed signal for the i th subband, $\bar{x}_i[n]$, can be written in terms of the input signals $x_j[n]$, for $j \in \{0, \dots, M-1\}$, as

$$\bar{x}_i[n] = \sum_p \sum_{j=0}^{M-1} \Gamma_{i,j}^{n,p}(\mu) x_j[p] + \nu_i[n], \quad (7)$$

where $\Gamma_{i,j}^{n,p}(\mu)$ and $\nu_i[n]$ are defined as

$$\Gamma_{i,j}^{n,p}(\mu) = \sum_{l=0}^{Q-1} \sum_q e^{j2\pi\mu q} c[l] f_j[q-l-pK] f_i^*[q-nK], \quad (8)$$

$$\nu_i[n] = \sum_q \nu[q] f_i^*[q-nK]. \quad (9)$$

Here, the complex factor $\Gamma_{i,j}^{n,p}(\mu)$ (8) characterizes the interference level of the p th input sample from the j th subband on the n th output sample of the i th subband, in the presence of CFO with magnitude μ . We note that for $|n-p| > (D+Q)/K$, due to the finite support of $f_i[n]$, $\Gamma_{i,j}^{n,p}(\mu) = 0$; accordingly, the range of the sum over p in (7) is finite. The term $\nu_i[n]$ (9) represents the additive noise passed through the i th subband of the receive filter bank. This term has zero-mean and, due to the PR property imposed on $f_i[n]$, its covariance is given by

$$E[\nu_i[q] \nu_j^*[p]] = \delta_{ij} \delta_{qp} \sigma_\nu^2. \quad (10)$$

Considering the reconstructed signal $\bar{x}_i[n]$ in (7), it appears that even if the channel could be perfectly equalized, which is equivalent to $c[0] = 1$ and $c[l] = 0$ for

$l \neq 0$, the presence of the CFO term $e^{j2\pi\mu q}$ in the interference factors $\Gamma_{i,j}^{n,p}(\mu)$ (8) would render the transceiver system non-PR. That is, the terms $\Gamma_{i,j}^{n,p}(\mu) x_j[p]$ would be non-zero for $j \neq i$ or $p \neq n$, and this in turn would result in a loss of performance in the data transmission process. It is worth to mention that in previous work [9], [10], it is assumed that the CFO factor $e^{j2\pi\mu q}$ can be taken out of the summation in (8) and consequently the interference terms $\Gamma_{i,j}^{n,p}(\mu) x_j[p]$ when $j \neq i$ or $p \neq n$ are negligible, which does not hold for the OPRFB systems. Our interest therefore lies in the development of a suitable, data-aided ML-based approach for the estimation of the CFO parameter μ .

As seen from Fig. 1, once a suitable estimate of μ is available, say $\hat{\mu}$, it can be used to compensate the CFO at the receiver front-end and thereby avoid its deleterious effects. In this paper, we focus on a simplified model of the noisy channel, i.e., AWGN for which the above condition on the channel coefficients $c[l]$ is satisfied, but extensions of our proposed approach to more complex time dispersive channels with joint equalization and CFO recovery is possible.

III. FREQUENCY OFFSET ESTIMATION

In this section, we first derive a novel CFO estimator based on the ML principle, which employs known transmitted pilots. We then propose a number of practical simplifications in the calculation of the associated log-likelihood function (LLF) that result in a lower implementation complexity for this estimator. Finally, the CRB on the variance of unbiased CFO estimators is derived as a performance benchmark.

A. ML Estimator

As indicated above, we consider a simplified AWGN channel model (i.e., $C(z) = 1$) in the derivation of the proposed ML-based CFO estimator; consequently, the resulting approach will not require the use of a priori CSI. In this special case, (8) reduces to

$$\Gamma_{i,j}^{n,p}(\mu) = \sum_q e^{j2\pi\mu q} f_j[q-pK] f_i^*[q-nK]. \quad (11)$$

In this work, we define a data frame as the set of M subband inputs $x_i[n]$ ($i \in \{0, 1, \dots, M-1\}$) entering the transmit filter bank at time n . We assume that within a burst of N consecutive frames, say from $n = 0$ to $N-1$, a total of L frames, with time indices $0 \leq t_0 < t_1 < \dots < t_{L-1} \leq N-1$, are selected for

the transmission of pilot tones. At any given time t_n , a subset of S subbands, with indices $0 \leq s_0 < s_1 < \dots < s_{M-1} \leq M-1$, are dedicated to the transmission of a unit-energy pilot symbol $p_{s_i}[t_n]$. We therefore consider a rectangular lattice of $N_P = LS$ pilot tones distributed over the time-frequency plane. However, our approach can be applied to other distributions of pilot symbols. Without loss in generality (since the pilot symbols are known to the receiver), we set $p_{s_i}[t_n] = 1$ for all pair (s_i, t_n) . Let $z_{s_i}[t_n]$ denote the reconstructed signal corresponding to the transmitted pilot $p_{s_i}[t_n]$. From (7), it follows that

$$z_{s_i}[t_n] = \gamma_{s_i}^{t_n}(\mu) + \nu_{s_i}[t_n], \quad (12)$$

where we defined

$$\gamma_{s_i}^{t_n}(\mu) = \sum_p \sum_{j=0}^{M-1} \Gamma_{s_i, j}^{t_n, p}(\mu) \quad (13)$$

In order to express (12) in compact vector form, we introduce:

$$\mathbf{z}_{s_i} = [z_{s_i}[t_0], z_{s_i}[t_1], \dots, z_{s_i}[t_{L-1}]]^T \quad (14)$$

$$\boldsymbol{\lambda}_{s_i}(\mu) = [\gamma_{s_i}^0(\mu), \gamma_{s_i}^1(\mu), \dots, \gamma_{s_i}^{L-1}(\mu)]^T \quad (15)$$

$$\boldsymbol{\nu}_{s_i} = [\nu_{s_i}[t_0], \nu_{s_i}[t_1], \dots, \nu_{s_i}[t_{L-1}]]^T \quad (16)$$

Therefore, we can write (12) as

$$\mathbf{z}_{s_i} = \boldsymbol{\lambda}_{s_i}(\mu) + \boldsymbol{\nu}_{s_i}. \quad (17)$$

Moreover, by arranging, we can write

$$\mathbf{Z} = \boldsymbol{\Lambda}(\mu) + \mathbf{W} \quad (18)$$

where

$$\mathbf{Z} = [\mathbf{z}_{s_0}^T, \mathbf{z}_{s_1}^T, \dots, \mathbf{z}_{s_{S-1}}^T]^T, \quad (19)$$

$$\boldsymbol{\Lambda}(\mu) = [\boldsymbol{\lambda}_{s_0}(\mu)^T, \boldsymbol{\lambda}_{s_1}(\mu)^T, \dots, \boldsymbol{\lambda}_{s_{S-1}}(\mu)^T]^T, \quad (20)$$

$$\mathbf{W} = [\boldsymbol{\nu}_{s_0}^T, \boldsymbol{\nu}_{s_1}^T, \dots, \boldsymbol{\nu}_{s_{S-1}}^T]^T. \quad (21)$$

As a consequence of the AWGN model assumption, it follows that \mathbf{W} is a zero-mean Gaussian random vector with diagonal covariance matrix $\mathbf{C}_W = E[\mathbf{W}\mathbf{W}^*] = \sigma_\nu^2 \mathbf{I}$, where \mathbf{I} is the identity matrix. Accordingly, for a given value of the unknown CFO parameter μ , the observation vector \mathbf{Z} in (18) is also Gaussian with mean $\boldsymbol{\Lambda}(\mu)$ and covariance $\sigma_\nu^2 \mathbf{I}$. The probability density function (PDF) of \mathbf{Z} , say $f(\mathbf{Z}; \mu)$ can therefore be formulated and subsequently maximized to produce the desired estimate

of μ . Take the natural logarithm of this PDF, the LLF [11] can be expressed (up to a constant term) in the form

$$\begin{aligned} \log(f(\mathbf{Z}; \mu)) &= -\frac{1}{\sigma^2} [\mathbf{Z} - \boldsymbol{\Lambda}(\mu)]^H [\mathbf{Z} - \boldsymbol{\Lambda}(\mu)] \\ &= -\frac{1}{\sigma^2} \sum_{i=0}^{S-1} \sum_{n=0}^{L-1} |z_{s_i}[t_n] - \gamma_{s_i}^{t_n}(\mu)|^2 \end{aligned} \quad (22)$$

Finally the ML estimator of the CFO can be written as:

$$\hat{\mu} = \arg \max_{\mu \in \mathcal{M}} \{\log(f(\mathbf{Z}; \mu))\}, \quad (23)$$

where \mathcal{M} is the search range for μ . According to (22), maximization of the LLF attempts to find the CFO μ , such that the skewed pilots by this hypothetical μ best match (in the LS sense) the reconstructed pilot data at the output of the receive filter bank.

B. Simplifications of $\Gamma_{i,j}^{n,p}(\mu)$

Here, we propose two simplifications of $\Gamma_{i,j}^{n,p}(\mu)$ to speed up the calculation of the (22). First consider (11), which includes a summation over the length D (often large) of the prototype filter $f_0[q]$. Recall that $f_i[q] = f_0[q]w^{-iq}$, therefore, (11) can be first simplified as

$$\Gamma_{i,j}^{n,p}(\mu) = w^{K(pj-in)} \varphi_{i-j}^{n,p}(\mu), \quad (24)$$

where

$$\varphi_{\Delta}^{n,p}(\mu) = \sum_q e^{j2\pi\mu q} f_0[q-pK] f_0^*[q-nK] w^{q\Delta}. \quad (25)$$

By this implementation, instead of calculating $\Gamma_{i,j}^{n,p}(\mu)$ for all the M^2 possible pairs (i, j) , it is sufficient to compute $\varphi_{\Delta}^{n,p}(\mu)$ for $2M-1$ possible different values of $i-j = \Delta \in \{-M+1, \dots, M-1\}$ and find the corresponding $\Gamma_{i,j}^{n,p}(\mu)$ by a multiplication as in (24). Therefore, we can roughly reduce the number of operations needed to compute the terms $\Gamma_{i,j}^{n,p}(\mu)$ by a factor of $M/2$.

Due to the excellent spectral containment of the prototype filters, we can assume that the main source of the CFO-induced interference on each target subband is due to the first few neighbouring subbands, and that the interference from more distant subbands is negligible. Therefore, as the second proposed simplification, to derive the total interference from the other subbands on the subband with pilot index s_i , it is sufficient to only factor in the contribution from the two neighbouring subbands on each side of the s_i th one. As a result, (13) is approximated as

$$\gamma_{s_i}^{t_n}(\mu) \approx \sum_p \sum_{j=s_i-2}^{s_i+2} \Gamma_{s_i, j}^{t_n, p}(\mu) \quad (26)$$

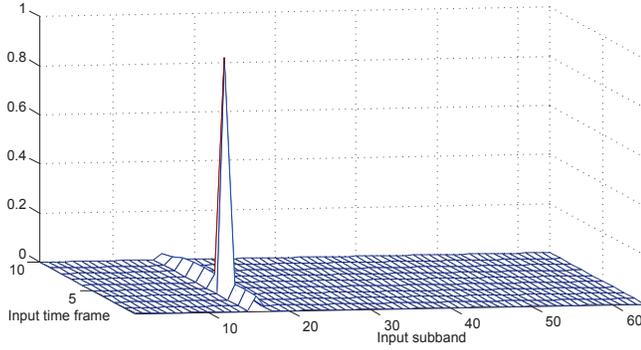


Fig. 2. Interference level $\Gamma_{i,j}^{n,p}(\mu)$ of p th input sample from j th subband on the n th output sample of the i th subband ($p \in \{0, 1, \dots, 10\}$, $j \in \{0, 1, \dots, 63\}$, $n = 4$, $i = 16$ and $\mu = 0.08$)

C. Cramer Rao Bound

Next, we derive a compact expression for the CRB on the variance of an unbiased data-aided CFO estimator obtained over the AWGN channel. Considering $\partial \mathbf{C}_W / \partial \mu = \mathbf{0}$, the Fisher $\mathcal{I}(\mu)$ [11] is

$$\begin{aligned} \mathcal{I}(\mu) &= 2\text{Re} \left[\frac{\partial \mathbf{\Lambda}(\mu)^H}{\partial \mu} \mathbf{C}_Z^{-1} \frac{\partial \mathbf{\Lambda}(\mu)}{\partial \mu} \right] \\ &= \frac{2}{\sigma^2} \sum_{i=0}^{S-1} \sum_{n=0}^{L-1} \left| \frac{\partial \gamma_{s_i}^{t_n}(\mu)}{\partial \mu} \right|^2, \end{aligned} \quad (27)$$

where $\text{Re}[\cdot]$ represents the real part of its argument and

$$\frac{\partial \gamma_{s_i}^{t_n}(\mu)}{\partial \mu} = j2\pi \sum_p \sum_{j=0}^{M-1} \sum_q q e^{j2\pi \mu q} f_j[q-l-pK] f_{s_i}^*[q-t_nK]. \quad (28)$$

Therefore, we can obtain the CRB on the variance of an unbiased CFO estimator $\hat{\mu}$ as

$$\text{Var}(\hat{\mu}) \geq \frac{1}{\mathcal{I}(\mu)} = \left(\frac{2}{\sigma^2} \sum_{i=0}^{S-1} \sum_{n=0}^{L-1} \left| \frac{\partial \gamma_{s_i}^{t_n}(\mu)}{\partial \mu} \right|^2 \right)^{-1}. \quad (29)$$

It can be seen that this CRB is inversely proportional to the signal-to-noise ratio (SNR), or proportional to the noise variance σ^2 . Moreover, it generally depends on the number of the observed pilots $N_p = LS$. Specifically, it is a decreasing function of both L and S .

IV. RESULTS AND DISCUSSION

In this section, the performance of the proposed ML estimator of the CFO is assessed and compared with the CRB (29). In addition to the AWGN channel (where $Q = 1$ and $c[0] = 1$), we consider a frequency selective channel consisting of $Q = 5$ independent Rayleigh-fading taps with an exponentially decaying power delay profile, i.e., $E[|c[l]|^2] = \beta e^{-\alpha l}$ for $l \in \{0, \dots, Q-1\}$, $\alpha = 0.5$, and β is a constant such that $\sum_{l=0}^{Q-1} E[|c[l]|^2] =$

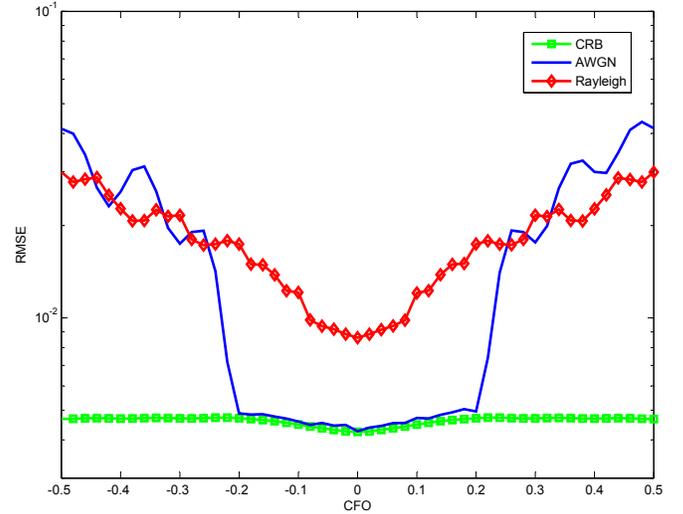


Fig. 3. RMSE of CFO estimator $\hat{\mu}$ versus μ (SNR=10dB, $L = 10$)

1. The results are reported for an OPRFB system with $M = 64$ subbands, up/downsampling factor $K = 72$ and real prototype filter of length $D = 1728$, designed based on the method [2], [3]. Results are presented for different values of the SNR, defined as E_s/N_0 , where $E_s = E[|x_i[n]|^2]$ is the input symbol energy and $N_0 = \sigma^2$ is the variance of the channel induced Gaussian noise. Moreover, L known pilots are inserted at the start of the transmitted burst on all the available subbands ($S = M$).

Figure 2 justifies the assumptions made in Section III-B about the cross-channel interference. In this figure, the level of interference $\Gamma_{i,j}^{n,p}(\mu)$ from the p th input sample of the j th subband on the n th output sample of the i th subband is plotted for $p \in \{0, 1, \dots, 10\}$, $j \in \{0, 1, \dots, 63\}$, $n = 4$, $i = 16$ and $\mu = 0.08$. It is evident that only a few subbands surrounding the target subband are contributing as interference sources.

The root mean square error (RMSE) of the proposed estimator, i.e., $\sqrt{E[|\hat{\mu} - \mu|^2]}$, is shown in Figure 3 versus the true CFO μ for SNR=10dB and $L = 10$ pilot frames. Here, the AWGN acquisition range (i.e., the CFO values where the algorithm's RMSE coincides with the CRB) is observed to be $|\mu| < 0.2$ (i.e., 20% of subband spacing). Clearly, in this interval, the RMSE of the proposed estimator over the AWGN channel is almost independent of the CFO. It is worth mentioning that the results reported in [10] show an acquisition range $|\mu| < 0.1$ when implemented for FMT transceivers, whereas in the OPRFB context, it leads to unsatisfactory results (i.e., RMSE level of the order of 10%) due to its underlying assumptions. Moreover, the proposed method also

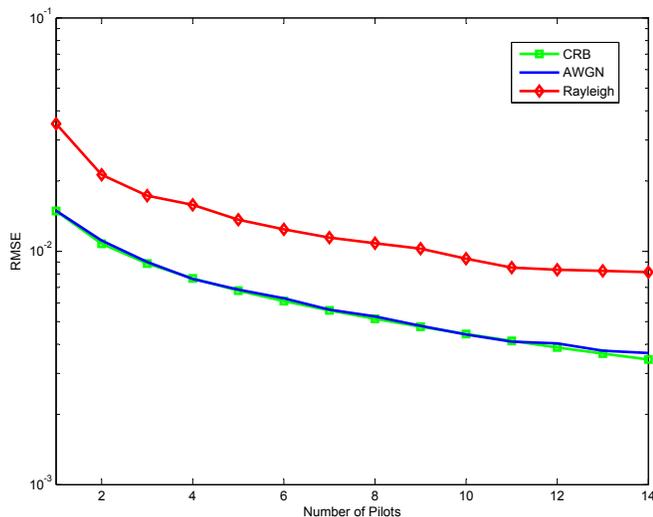


Fig. 4. RMSE of CFO estimator $\hat{\mu}$ versus L (SNR=10dB, $\mu = 0.08$)

exhibits a robust performance over frequency selective Rayleigh channel, with an RMSE of about 1% within the range $|\mu| < 0.1$.

Sensitivity of the proposed estimator to L , the number of the pilots being used, is depicted in Figure 4 with SNR= 10dB and CFO $\mu = 0.08$. As expected, by increasing L , a better estimation accuracy can be achieved. In addition, over AWGN channel, the CRB is always attainable. Figure 5 exhibits the RMSE performance of the proposed estimator as a function of SNR with true CFO $\mu = 0.08$ and $L = 10$. Interestingly, in contrast to the result for the CFO synchronization method reported in [8], the proposed ML estimator can achieve a performance very close to the CRB over a wide range of SNR values..

V. CONCLUSION

In this paper, we considered the problem of data-aided CFO estimation and recovery for FBMC systems with emphasis on OPRFBs. By exploiting statistical properties of inserted pilots transmitted by such systems over an AWGN channel, the ML estimator for the CFO was derived. The complexity of the proposed estimator is considerably reduced by identification of the insignificant LLF terms and consequently neglecting them in the estimation. This method was tested over AWGN and frequency selective channels with various CFOs, SNRs and number of pilots. The results show that over the AWGN channel, the proposed estimator exhibits a performance close to the CRB with wider acquisition range compared to other methods, whereas its performance remains satisfactory over frequency selective channels.

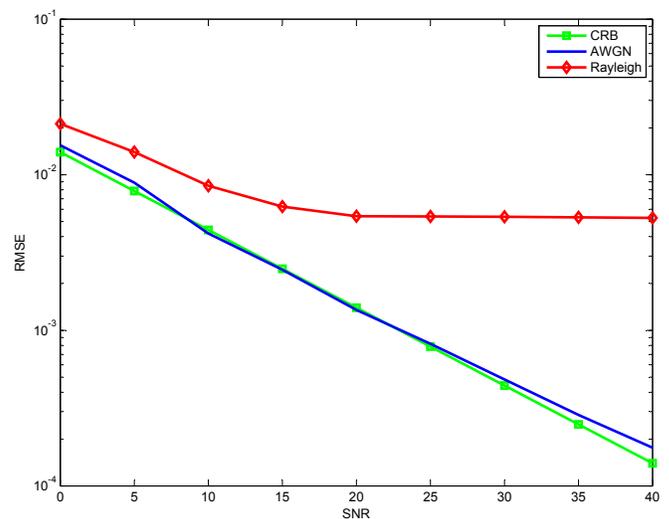


Fig. 5. RMSE of CFO estimator $\hat{\mu}$ versus SNR (CFO=0.08, $L = 10$)

REFERENCES

- [1] G. Cherubini, E. Eleftheriou, and S. Olcer, "Filtered multitone modulation for very high-speed digital subscriber lines," *IEEE J. Sel. Areas Commun.*, vol. 20, pp. 1016–1028, Jun. 2002.
- [2] S. Rahimi and B. Champagne, "Perfect reconstruction DFT modulated oversampled filter bank transceivers," in *Proc. EU-SIPCO*, Barcelona, Spain, Aug. 2011, pp. 1588–1592.
- [3] —, "Oversampled perfect reconstruction DFT modulated filter banks for multi-carrier transceiver systems," *Signal Processing*, in press.
- [4] —, "On the robustness of oversampled filter bank multi carrier systems against frequency offset," in *Proc. ISWCS*, Paris, France, Aug. 2012, pp. 944–948.
- [5] V. Lottici, M. Luise, C. Saccomando, and F. Spalla, "Blind carrier frequency tracking for filterbank multicarrier wireless communications," *IEEE Trans. Commun.*, vol. 53, pp. 1762 – 1772, Oct. 2005.
- [6] T. Fusco, A. Petrella, and M. Tanda, "Blind CFO estimation for noncritically sampled FMT systems," *IEEE Trans. Signal Process.*, vol. 56, pp. 2603 –2608, June 2008.
- [7] A. Tonello and F. Rossi, "Synchronization and channel estimation for filtered multitone modulation," in *Proc. WPMC*, Abano Terme, Italy, Sep. 2004, pp. 590–594.
- [8] T. Fusco, A. Petrella, and M. Tanda, "Data-aided symbol timing and CFO synchronization for filter bank multicarrier systems," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 2705–2715, May 2009.
- [9] M. Carta, V. Lottici, and R. Reggiannini, "Frequency recovery for filter-bank multicarrier transmission on doubly-selective fading channels," in *Proc. ICC*, Glasgow, Scotland, June 2007, pp. 5212–5217.
- [10] V. Lottici, R. Reggiannini, and M. Carta, "Pilot-aided carrier frequency estimation for filter-bank multicarrier wireless communications on doubly-selective channels," *IEEE Trans. Signal Process.*, vol. 58, pp. 2783 –2794, May 2010.
- [11] S. Kay, *Fundamentals of Statistical Signal Processing, Vol. I: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice Hall, 2001.

A Semi-Analytical Performance Prediction of Turbo Coded SC-FDMA

Fatima Ezzahra Naamane¹, Mostafa Belkasmi²

Departement of Communication Networks

ENSIAS

Rabat, Morocco

Email: ¹naamane20@gmail.com, ²m.belkasmi@um5s.net.ma

Mohamed Et-tolba

Departement of Communications Systems

INPT

Rabat, Morocco

Email: mohamed.ettolba@gmail.com

Abstract—Single carrier frequency division multiple access (SC-FDMA) is the access technique which is used in the long term evolution (LTE) uplink in order to reduce the peak-to-average power ratio (PAPR). The LTE uplink uses this technique in joint application with turbo coding and high order modulations to achieve peak data rates up to 86 Mbps within 20 MHz bandwidth. This paper introduces a semi analytical method for predicting the turbo coded SC-FDMA performance in terms of bit error rate (BER). Simulation results have shown that the BER obtained with the proposed method is close to that measured by Monte-Carlo simulation method.

Keywords—SC-FDMA; LTE turbo code; Monte Carlo Method; BER Simulation; PDFs

I. INTRODUCTION

LTE [1] is designed to support peak data rates up to 300 Mbps and 86 Mbps for downlink and uplink, respectively, within 20 MHz bandwidth. In terms of system performance, it targets to improve user throughput, coverage and spectrum efficiency. In order to reach this performance, the LTE radio interface uses orthogonal frequency division multiple access (OFDMA) [2] as access technique for its robustness against both inter-symbol interference and channel impairments.

The LTE uplink transmission is based on SC-FDMA [3] to its reduced PAPR compared to OFDM. These features are used in joint application with multiple input multiple output (MIMO), adaptive modulation and coding (AMC) [4] and hybrid automatic repeat request (Hybrid ARQ) [5]. For reliable data transmission, the LTE uplink transmitter consists of a combination of error detection code, channel coding, interleaving, and SC-FDMA modulator. To meet the LTE requirement, the channel coding is performed, using a 1/3 rate turbo encoder which is often decoded with an iterative soft input soft output (SISO) decoder. The turbo encoder internal interleaver is based on quadratic permutation polynomials (QPP) to support high data rates. The performance of LTE uplink physical layer can be measured in terms of BER, block error rate (BLER) or throughput. In this work, our focus is BER evaluation which has been made by simulating the 3GPP LTE uplink transmitter, the transmission channel and the receiver. This simulation is done using Monte-Carlo (MC) method. However, it has been proven that it is prohibitive and time consuming. To makeup for this limitation of MC simulation method, semi-analytical performance prediction has been proposed in several studies.

Bohdanowicz [6] proposed the importance sampling (IS) method for BER prediction. It has been found that for simple memoryless systems (e.g. a BPSK modem [7]), the efficiency of the IS technique is high and its implementation is relatively

easier. However, its accuracy can be severely degraded, especially that of the coded systems.

For such systems, several solutions of performance prediction problem for complex system are carried out in several studies. In [8], a low complexity prediction technique for Turbo-Like code has been proposed. It is based on estimating the probability density function (pdf) of the log likelihood ratio (LLR) at the output of the decoder. The BER prediction is made assuming that the probability density functions of the decoder output LLRs are normal densities.

A semi-analytical approach of the BER prediction has been presented by Saoudi et al. [9]. The authors have proposed to predict the BER by using the known kernel estimator of the pdf of the soft decision made at the output of the detector. This method assumes that any prior information on the pdf of received samples is available at the receiver side. Moreover, its accuracy depends on the estimation of the smoothing parameter which is very important in the prediction process. Saoudi et al. [10] proposed an unsupervised soft BER prediction method for any digital communications systems. This technique considers that the pdf at the detector output is estimated with a Gaussian Kernel. The accuracy of this prediction method is very sensitive to smoothing parameter especially for high signal to noise ratios (SNRs).

In this paper, we propose a semi-analytical BER prediction method which is based on the pdf estimation using Gaussian kernel. We assume that no knowledge on the distribution of the received soft samples is currently available. In the proposed method, we have derived a new expression of the smoothing parameter which takes into account the histogram of the soft samples at the detector output. It has been shown that our method exhibits a significant accuracy in term of BER compared to Monte Carlo simulation method.

The remainder of the paper is organized as follows. In Section II, we present the coded SC-FDMA system. We introduce the problem of error probability derivation in Section III. Section IV details the probability density function estimation using kernel method and how one should select the bandwidth h to optimize the properties of the probability density function. Simulations and numerical results are given in Section V. Finally, the paper is concluded in Section VI.

II. CODED SC-FDMA SYSTEM

The coded SC-FDMA system is presented in Figure 1; it consists of a source, a turbo code LTE, modulation mapper, modulation SC-FDMA, transmission channel and a receiver.

The coded SC-FDMA system will be the case of study the performance of the proposed method for BER prediction.

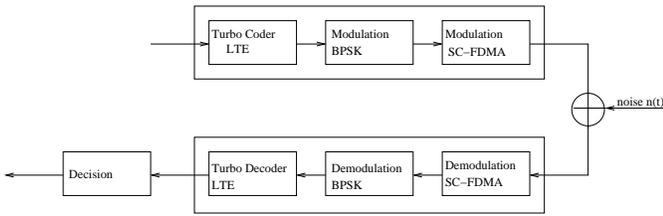


Fig. 1: Coded SC-FDMA System

A. Analytical form of SC-FDMA signal

This technique consists of distributing a large number of carriers, not directly as the source symbols in OFDM, but their frequency representation after having spread over the band of the system [11].

The signal of user k to the output of the system will be given by the expression:

$$S^k(t) = \sum_{n \in \Omega_n^k} U_n^k p(t - nT_s) e^{j2\pi f_n t} \quad (1)$$

- Let $\{f_n\}_{0 < n < N-1}$ all orthogonal carrier frequencies modulated system, and f_c the center frequency (RF frequency) of the transmitted signal in the channel. It has the following relationship:

$$f_n = f_c + n\Delta f \quad (2)$$

Or $\Delta f = \frac{1}{T_s}$ is the spacing between sub-channels, with T_s the duration of a symbol.

- $\{U_n^k\}_{n \in \Omega_n^k}$, frequency representation of the symbols of the modulation block of user $k \in [0, 1, \dots, L-1]$ with DFT obtained after the modulation.

We recall that the spectral spreading factor of the system is denoted L and the maximum number of users that can communicate simultaneously in the system.

- Ω_n^k represents the set of Q sub-carriers modulated by user k .
- $p(t)$ shows the shaping filter.

B. SC-FDMA demodulation

The principle of the demodulation of the SC-FDMA system is to demodulate the signal on each sub-carrier system. As a result, the received signal is first reduced to baseband, before being sampled for the digital signal processing. After removing the guard interval, a demodulator DFT provides the symbols modulating each carrier. An equalizer is then implemented, as in the technical SC / FDE, in order to eliminate the contribution of the channel on each subcarrier signal, and there by recover the symbol frequency. A demodulator IDFT can then retrieve the source symbols of the system. The signal received at user k receiver on symbol

duration is written as follows:

$$y^k(t) = \sum_{n \in \Omega_n^k} U_n^k \int [h_n^k(t - \tau) p(t - nT_s) e^{j2\pi f_n t} d\tau \quad (3)$$

C. Turbo Coder LTE

The LTE system [12] has adopted a new structure for the turbo encoder. This is an improvement to turbo encoder interleaver by a new permutation polynomial based deterministic interleaver, called QPP interleaver and anti-interleaver with advantages beyond other interleavers. The encoder is characterized by a new structure simple, flexible operation and the most important is that parallel turbo decoding and register contention problem is solved successfully, which effectively increases the efficiency of the high-speed block parallel Turbo decoding [13].

In addition, the receiver contains an LTE turbo decoder based on the theory of iterative decoding. This is an important feature in the turbo- decoding, so decoding complexity lineally increases, with the size of the sequence information. In order to achieve a better decoding performance, component decoding must adopt soft input soft output (SISO) algorithm. MaxMAP decoding algorithm and SOVA decoding algorithm are two kinds of common soft-input soft-output Turbo decoding methods [14][15].

III. BIT ERROR PROBABILITY DERIVATION

To derive the bit error probability, we consider the general digital communication system presented in Figure 2. It consists of a source, a transmitter, transmission channel and a receiver. The source is considered to be digital and delivers the information bits $b \in \{0, 1\}$. These bits are processed by a transmitter which can include channel coding, interleaving, and modulation. After that, the information bits at the output of the transmitter are transmitted over a channel.



Fig. 2: General system model

For simplicity, the channel is assumed to be Gaussian [7]. The channel output is delivered to the receiver, which tries to detect the information bits from a noisy signal by using a detector, a sampling process and a decision. Due to the channel effect, the receiver can make a wrong decision on information bits at its output \tilde{b} . So, it is important to measure the communication system efficiency. The most popular mean to do this is the BER evaluation. According to figure above, the bit error probability is written as :

$$p_e = P_1.Pr(\tilde{b} = 0 \setminus b = 1) + P_0.Pr(\tilde{b} = 1 \setminus b = 0) \quad (4)$$

where $P_k, k = 0, 1$, is the probability that $b = k$. $Pr(\cdot)$ is the conditional probability. In terms of the decision threshold,

the probability in (4) can be rewritten as follows:

$$p_e = P_1 \cdot Pr(y(t_0) < \mu \setminus b = 1) + P_0 \cdot Pr(y(t_0) \geq \mu \setminus b = 0) \quad (5)$$

where t_0 is the sampling time, μ is the decision threshold and:

$$\begin{cases} y(t_0) \geq 0; & \text{if } \tilde{b}_i = 1 \\ y(t_0) < 0; & \text{if } \tilde{b}_i = 0 \end{cases}$$

The conditional probabilities in (5) can be evaluated by integrating the probability density functions of the random variable y . If $f(y)$ denotes the pdf of y , the error probability is expressed as:

$$p_e = P_1 \cdot \int_{-\infty}^{\mu} f_1(y) dy + P_0 \cdot \int_{-\mu}^{+\infty} f_0(y) dy \quad (6)$$

So, to predict the error probability P_e , one has to estimate the probability density $f(y)$.

IV. PROBABILITY DENSITY FUNCTION AND ERROR PROBABILITY ESTIMATION :

A. Kernel estimator

Several types of non parametric estimation approaches are suggested to estimate a probability density function. This is due to the recent development in statistics theory. The most known of these methods is the kernel estimator [16] which we have adopted in this work. For a given set S of N received samples y_1, y_2, \dots, y_N , the kernel estimator of the probability density function $\tilde{f}(y)$ is [17][18]:

$$\tilde{f}(y) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{y - y_i}{h}\right) \quad (7)$$

where h is the smoothing parameter and $K(\cdot)$ is the kernel function. To guaranty that $\tilde{f}(y)$ is a density; the kernel is a function that satisfies $\int_{-\infty}^{+\infty} K(u) du = 1$. Moreover, it is assumed to be symmetric about 0. In this work, we used the Gaussian kernel, which satisfies the above properties. The estimated density $\tilde{f}(y)$ can be rewritten as:

$$\tilde{f}(y) = \frac{(Nh)^{-1}}{\sqrt{2\pi}} \sum_{i=1}^N e^{-(\frac{y-y_i}{h})^2/2} \quad (8)$$

To get the expression of the estimated bit error probability p_e , we divide the set of observed samples into two subsets S_0 and S_1 . The first subset contains N_0 observed samples, which correspond to the transmission of $b = 0$. The second subset consist of N_1 observed samples when $b = 1$ is transmitted. In this manner, and by substituting the probability density $f(y)$ by its estimate $\tilde{f}(y)$, the estimated bit error probability is expressed as:

$$\tilde{p}_e = \frac{P_0}{N_0} \sum_{i=1}^{N_0} Q\left(\frac{-(y_i)_0}{h_0}\right) + \frac{P_1}{N_1} \sum_{i=1}^{N_1} Q\left(\frac{(y_i)_1}{h_1}\right) \quad (9)$$

where $h_k, k = 0, 1$ is the smoothing parameter and $Q(\cdot)$ denotes the complementary unit cumulative Gaussian distribution, that is,

$$Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt. \quad (10)$$

From (9), it is very clear that the accuracy of bit error probability estimation depends on the choice of the optimal smoothing parameter.

B. Choice of optimum smoothing parameter h

The mean integrated squared error (MISE) criterion is one of several methods used for selecting the optimum bandwidth. In MISE-based method, this optimum bandwidth is obtained by minimizing the MISE, which is expressed as:

$$MISE = E\left[\int_{-\infty}^{+\infty} [\tilde{f}(y) - f(y)]^2 dy\right] \quad (11)$$

The bandwidth that mimimizes MISE is given by:

$$h_{opt} = arg_h \min (MISE(h)) \quad (12)$$

Under additional assumptions:

$$\lim_{N \rightarrow \infty} h = 0; \quad \lim_{N \rightarrow \infty} Nh = 0; \quad (13)$$

Several types of MISE criterion have been suggested in litterature. Hereafter, we details the most populer ones.

1) Kernel based MISE Criterion:

For Kernel based MISE, we can use the normal reference method in selecting the bandwidth h for kernel estimator. As in [10], the optimum smoothing parameter is expressed as:

$$h_{opt,Kernel} = C1(f)C2(K)N^{-1/5} \quad (14)$$

$$C1(f) = \left[\int_{-\infty}^{+\infty} [f''(x)]^2 dx\right]^{-1/5} \quad (15)$$

$$C2(K) = \frac{\left[\int_{-\infty}^{+\infty} [K(x)]^2 dx\right]^{1/5}}{\left[\int_{-\infty}^{+\infty} [x^2 K(x)]^2 dx\right]^{2/5}} \quad (16)$$

From (14), it is clear that the optimum parameter depends on the unknown pdf and also on the kernel $K(\cdot)$. For a Gaussian kernel, we get $C2(K) = (2\sqrt{\pi})^{-1/5}$ and for the normal reference method $C1(f) = (8\sqrt{\pi}/3)^{1/5}\sigma$, yielding:

$$h_{opt,Kernel} = (4/3)^{1/5}\sigma N^{-1/5} = 1.06\sigma N^{-1/5} \quad (17)$$

Other way are used to select the bin size h . We focus on the normal reference method that uses the histogram of the PDF. Our approach is based on the latter method of calculating h . This method will be detailed in the following paragraph.

2) Histogram based MISE Criterion:

The histogram is used to measure the probability of observing a particular interval length [19]. It is a way to estimate the pdf by taking origin x_0 and a bin width h and define the bins of the histogram as the intervals $[x_0 + mh, x_0 + (m + 1)h]$ for positive and negative integers m . The histogram estimate of the pdf is then defined by:

$$H_N(x) = \frac{1}{Nh} (\text{number of } x_i \text{ in the same bin as } x) \quad (18)$$

Based on the same principle in the previous section, the MISE between $H_N(x)$ and the true pdf is written as:

$$MISE = \varepsilon \left[\int_{-\infty}^{+\infty} [H_N(x) - f(x)]^2 dx \right] \quad (19)$$

For a very large N ($N \rightarrow \infty$), the value of h needed in (18) can be shown [20] as:

$$h_{opt,Hist} = C1(f)N^{-1/3} \quad (20)$$

$$C1(f) = 6^{1/3} \left[\int_{-\infty}^{+\infty} [f''(x)]^2 dx \right]^{-1/3} \quad (21)$$

As can be seen, the optimum smoothing parameter depends only on the unknown pdf f . We need to find $C1(f)$ and under the assumption that $f \sim N(\mu, \sigma^2)$. This gives a simple data-based strategy for choosing the bin width h .

$$C1(f) = (24\sqrt{\pi})^{-1/3} \sigma \quad (22)$$

The optimum smoothing parameter is finally written as:

$$h_{opt,Hist} = (24\sqrt{\pi})^{-1/3} \sigma N^{-1/3} \quad (23)$$

In practice, the choice of an efficient method for the calculation of h ; for an observed data sample is a more complex problem, because of the effect of the bandwidth on the shape of the corresponding estimator. If the bandwidth is small, we will obtain an under-smoothed estimator, with high variability. On the contrary, if the value of h is important, the resulting estimator will be very smooth and farther from the function that we are attempting to estimate.

An example is drawn in Figure 3, where we show kernel estimators using the kernel function (the standard Gaussian density) and two different values for the bandwidth. The data sample consists of 1000 random numbers of an exponential distribution.

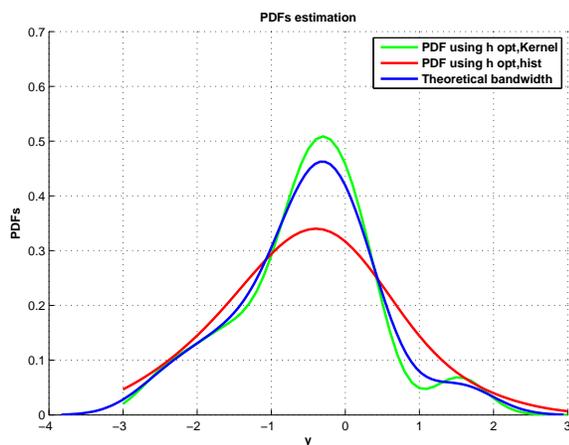


Fig. 3: Estimated densities for bandwidths chosen using different methods

For the data from a SC-FDMA system, the kernel estimators using the kernel function (the standard Gaussian density) and two different values for the bandwidth provides the result in Figure 4.

Figure 4 shows Gaussian kernel density estimates based on two different bandwidths for a sample of 500 data points from the SC-FDMA system. The second method MISE of bandwidth calculation h is of a good performance and can

obtain an under smoothed estimator as the Kernel based MISE criterion.

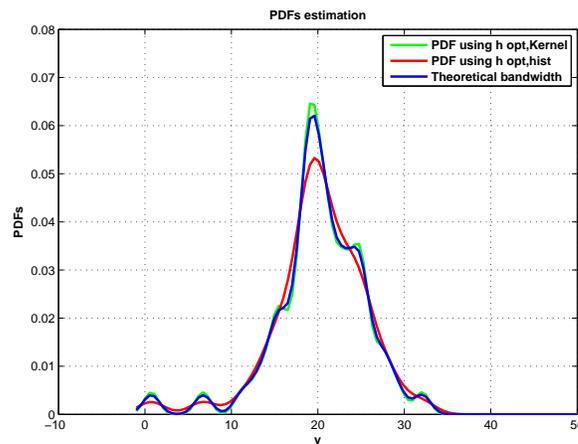


Fig. 4: Estimated densities for bandwidths chosen using different methods

Our practice shows that the histogram based MISE Criterion appears to be a suitable method for the choice of the bandwidth. As this method is based on estimating the histogram of samples, we have compared it to another method that is also based on the estimation of MISE using Kernel estimator. It is seen that the histogram based MISE Criterion outperforms the Kernel based MISE Criterion in terms of squared errors.

Thus, we choose the histogram based MISE Criterion to study the performance of the new BER estimation method proposed in this paper.

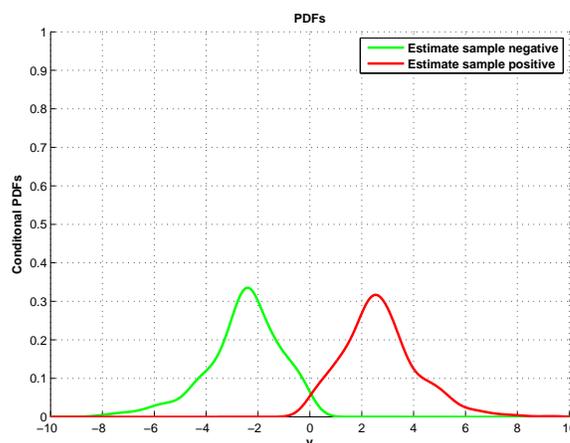


Fig. 5: Estimated Conditional pdfs for $SNR = 4dB$

Once the optimal smoothing parameter is calculated, we can present both the estimated conditional pdfs for $(y_i)_0$ and $(y_i)_1$ by using (9). We consider that the number of soft outputs

whose simulator result is shown in Figure 5 is $N = 1024$ observations whose simulation result is shown in Figure 5.

V. SIMULATIONS AND NUMERICAL RESULTS

To evaluate the performance of the proposed new method of estimating the BER, we consider two systems to simulate, namely BPSK system and a coded SC-FDMA system.

A. Validation of BER Prediction

In this section, we have shown that the new proposed BER estimator (Figure 6) provides the same performance as the Monte Carlo method whose data sample consists of 1000 random numbers of an exponential distribution.

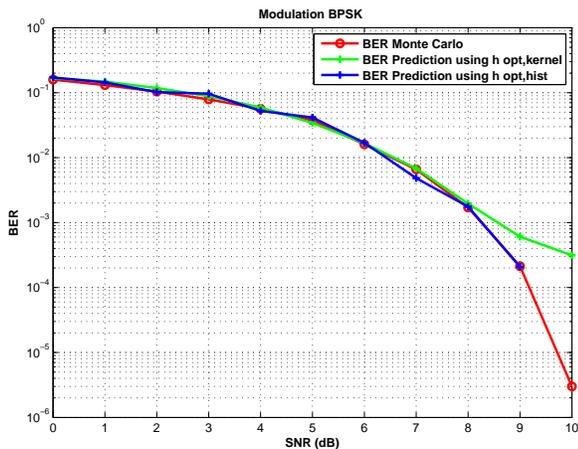


Fig. 6: BER performance for BPSK

This validation encourages us to study this performance for a coded SC-FDMA system.

B. BER prediction results for SC-FDMA

The objective of this section is to study and evaluate the implementation of the SC-FDMA technology encoded by a turbo encoder LTE on a Gaussian channel.

Our focus is on the behavior of the estimator proposed on the transmission bits by a coded SC-FDMA system. In such a system, it is difficult to have a reliable estimate of BER using MC-aided techniques with a limited number of soft observations and in the regions where the SNR is very high. In other words, the BER estimation performances of coded SC-FDMA system are studied in LTE uplink simulation system under Gaussian Channel.

In Figure 7, we can see that the new proposed BER estimator (9) provides the same performance as the Monte Carlo method, based on the perfect knowledge of the transmission bits. The proposed method is characterized, however, by the lack of knowledge of the received samples distribution of a SC-FDMA demodulation. We also present in Figure 7, the BER estimation performance using turbo encoded with SC-FDMA technique. The proposed technique provides reliable estimates, comparable to the Monte Carlo technique.

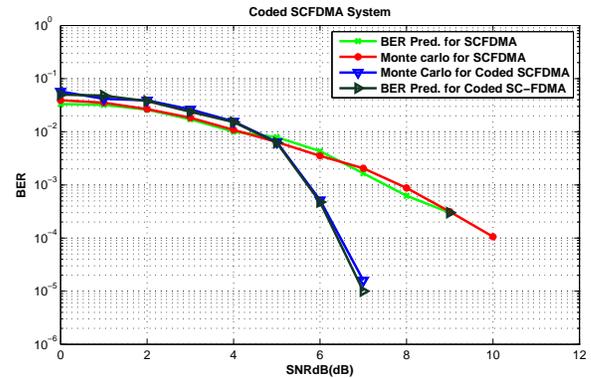


Fig. 7: BER performance comparison for Turbo Coded SC-FDMA

Note that the new technique allows a reliable estimate of BER for SNR up to $8dB$ values, while the MC-technique is unable to do so. It stops at $SNR = 8dB$ due to the very limited number of bits of information transmitted. This last conclusion presents a major advantage of the new estimator proposed for digital communications.

VI. CONCLUSION

The purpose of this article is to address the problem of estimation of BER for digital communication systems. In this context, we have proposed a semi-analytical method for predicting the turbo coded SC-FDMA performance in terms of bit error rate (BER), which is based on the pdf estimation using Gaussian kernel. We have assumed that no knowledge on the distribution of the received soft samples is available.

In the proposed method, we have derived a new expression of the smoothing parameter, which takes into account the histogram of the soft samples at the detector output. After, application to turbo coded BPSK modulation and coded SC-FDMA; we have concluded that have the same performance with either Monte Carlo technique (MC) or the new proposed BER estimator.

REFERENCES

- [1] H. Holma and A. Toskala, "LTE for UMTS OFDMA and SC-FDMA Based Radio Access", John Wiley & Sons Ltd, 2009
- [2] J. Xinzhu, "Channel Estimation Techniques of SC-FDMA", Department of Physics and Electrical Engineering, 2007
- [3] H. G. Myung and D. J. Goodman, "Single Carrier FDMA A New Air Interface For Long Term Evolution", John Wiley & Sons Ltd, 2008
- [4] S. Zarei, "LTE: Channel Coding and Link Adaptation", 2009
- [5] J. C. Ikuno, S. Schwarz and M. Simko, "LTE Rate Matching Performance with Code Block Balancing", European Wireless Conference (EW0 11), 2011
- [6] A. Bohdanowicz, "On Efficient BER Evaluation of Digital Communication Systems via Importance Sampling", Proc. Symposium on Communications and Vehicular Technology (SCVT 01), Delft, The Netherlands, August 2001, pp. 61-67.
- [7] J. G. Proakis, "Digital Communications", 3rd Edition, McGraw-Hill, 1994
- [8] A. Abedi and A.K. Khandani, "A new method for performance evaluation of bit decoding algorithms using statistics of the log likelihood ratio", Journal of the Franklin Institute, vol. 345, 2008, 60-74, doi:10.1016/j.jfranklin.2007.06.002.

- [9] S. Saoudi, M. Troudi, and F. Ghorbel, "An Iterative Soft Bit Error Rate Estimation of Any Digital Communication Systems Using a Non-parametric Probability Density Function", *EURASIP Journal on Wireless Communications and Networking*, Hindawi Publishing Corporation, vol.2009, January 2009, pp. 1-17, doi:10.1155/2009/512192.
- [10] S. Saoudi, T. Ait-Idir, and Y. Mochida, "A Novel Non-Parametric Iterative Soft Bit Error Rate Estimation Technique for Digital Communications Systems ", *Proc. IEEE International Conference on Communications (ICC 11)*, June 2011, pp. 1-6, doi:10.1109/icc.2011.5962691
- [11] Y. S. Yameogo, "Etudes de Nouvelles Techniques d'Estimation et d'Egalisation du Canal Adapt au Systeme SC-FDMA", pp 19-21, 2011,tel-00657895, version 1.
- [12] 3GPP TS 36.212 V11.0.0 (2012-09), Multiplexing and Channel Coding.
- [13] Y. Li, X. Gui, and F. Ye, "Analysis of BLER Performance for LTE Uplink Baseband Simulation System", *Journal of Computational Information Systems (CIS 12)* 8:7, April 2012 , pp 2691-2699.
- [14] D. Liu, "Research on Key Techniques of Turbo Code and the Application of Turbo Principle", *Changsha National University of Defense Technology*, pp. 35-45, June 2004.
- [15] J. Yuan and W. Wang, "Turbo Decoding Algorithm and Its Fixed-point Simulation", *Computer Simulation*, pp 91-94, 2005.
- [16] Z. Zambom and R. Dias, "A Review of Kernel Density Estimation with Applications to Econometrics", *International Econometric Review (IER 2013)*, vol.5, pp. 20-42, April 2013.
- [17] M. Rozenblatt, "Remarks on Some Non-Parametric Estimates of a Density Function", *Annals of Mathematical Statistics*, vol.27, pp 832-837, September 1956.
- [18] E. Parzen, "On Estimation of a Probability Density Function and Mode", *Annals of Mathematical Statistics*, vol. 33, pp 1065-1076, September 1962.
- [19] W. Zucchini, "Applied Smoothing Technique, part1 : Kernel Density Estimation", October 2003.
- [20] D. Agnew and C. Constable, "chapter 9: Non Parametric Density Function Estimation", *Preliminary Course Outline for SIO 223A*, Fall 2008.

The Effect of Human Bodies on Path Loss Model in an Indoor LOS Environment

Young-Hoon Kim and Seong-Cheol Kim

Seoul Nation University
INMC and SOEE
Seoul, Korea

e-mail: kebi0907@gmail.com, sckim@maxwell.snu.ac.kr

Abstract— This paper deals with the effect of the presence of people sitting in an indoor Line-of-Sight (LOS) environment on a Ultra-Wideband (UWB) channel. To assess this, we selected four scenarios with no people, 5, 10, and 20 people in the room. We created two regions in the room, one near two side walls and the other in the center aisle. In each scenario, we measured the channel at 24 receiving points from a fixed transmitting point in the room. At each receiving point, the receiver was moved around to 9 local positions to obtain a local average. In this paper, the considered UWB channel parameters are a frequency-independent pathloss model and a frequency-dependent pathloss model. We find that the pathloss exponent for the region of the center aisle decreases as more people are in the room, while the pathloss exponent for the region near the side walls increases when there are more people in the room. We also study the effect of the frequency on the pathloss characteristic. The results suggest that the effect of the presence of people on UWB channels should be considered when assessing the performance of UWB systems.

Keywords—Path loss model; Human Bodies; UWB; Indoor; LOS.

I. INTRODUCTION

UWB systems are commonly defined as systems that have either more than 20% relative bandwidth or more than 500MHz absolute bandwidth. It is well known for UWB systems to have many advantages, such as low complexity, a low cost, resistance to severe multipath fading, and the capability of a fine time resolution. There are numerous applications of UWB systems, such as personal area networks (PANs), sensor networks, geo-location sensors, and emergency communications. UWB systems using a relatively large bandwidth have to use low power so as not to cause interference in the neighboring communication systems. This feature makes UWB systems deployable in LOS and weak non-line-of-sight (NLOS) environments in which the signal undergoes less attenuation. A classroom is an example of a LOS environment, and furniture (e.g., desks and chairs) as well as the people in the classroom are factors that change the UWB channel. In previous work, there were three types of channel variation by people: 1) the depth and duration of shadow fading due to pedestrians moving in the vicinity of such links [1-3], 2) the effect of the presence of humans on wireless personal area networks (WPANs) in which one end of the link is located either close to or on a person [4-7], and 3) the effect of the presence of a human on

wireless body area networks (WBANs) in which both ends of the link are located either close to or on a person [8-10]. The above-mentioned papers deal with channel variation by only a person. Another recent paper [11] assesses the channel variation depending on how many seats are occupied by passengers in airplane, but this environment, which commonly involves metal material, is different from an indoor environment in a building.

This paper addresses UWB channel variation depending on the presence of people sitting in an indoor LOS environment. From the measurement data, we obtain a frequency-independent pathloss model and a frequency-dependent pathloss model. The paper introduces a frequency-dependent pathloss model, which includes the effect of the frequency on the pathloss characteristic. This can be useful for multiband orthogonal frequency-division multiplexing (MB-OFDM), as it divides the entire frequency band into several sub-bands with a bandwidth of 528 MHz.

This paper is organized as follows: Section II presents the channel measurement system and the measurement scenario. In Section III, we describe the channel parameters, i.e., the frequency-independent pathloss and frequency-dependent pathloss. Finally, the paper is ended with a summary and conclusion in Section IV.

II. MEASUREMENT METHODOLOGY

A. Measurement System

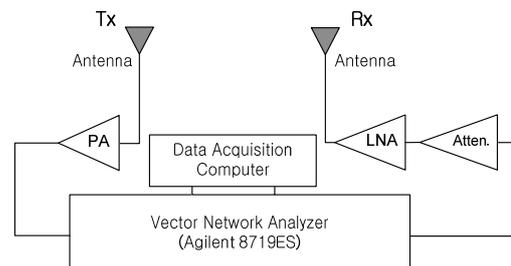


Figure 1. Block diagram of the measurement system.

In this paper, we measured the UWB channel using a frequency-domain channel sounding method for channel characterization with a vector network analyzer (VNA). The VNA (Agilent 8719ES [12]) transmits 801 discrete tones that are uniformly spaced from 3.1 to 4.7 GHz with a frequency interval of 2 MHz, requiring 400 ms for one sweep. This

frequency interval allows us to measure a multipath with a maximum excess delay of 500 ns, and the bandwidth of 1.6 GHz gives a time resolution of less than 0.01 ns. The measurement system is described in Fig. 1. The same dipole antennas with a gain of 2 dBi are used on both the transmitting and receiving sides and are located on 1.5-m-high tripods. A power amplifier (PA) with a gain of 25 dB and a low-noise amplifier (LNA) with a gain of 27 dB are used on the transmitting and receiving sides, respectively. To eliminate the effect of the antennas, the PA, the LNA and the cables, all measured data are calibrated in an anechoic chamber.

B. Measurement Scenario

To analyze the effect of the presence of people on a UWB channel, we use four scenarios: 1) a room with desks and chairs, 2) a room with desks, chairs and five people, 3) a room with desks, chairs and 10 people, 4) a room with desks, chairs and 20 people. We divide a room into two regions, the first near both side walls, Region A, and the second in the center aisle, Region B, as described in Fig. 2. In Fig. 2, 'Occupied' denotes people sit on the chair; otherwise this is 'Unoccupied' and the signs of the desk and chair is given. All scenarios are carried out in the same room of which the wall material is commonly concrete and where there are two large glass windows on one side wall. The locations of the transmitting and receiving antennas are identical in all scenarios, as illustrated in Fig. 2. At each of 24 locations in the room for each scenario, the channel responses are measured at nine local points arranged in a 3x3 square grid, as shown in Fig. 2 (a). 100 frequency responses were collected at each local position.

III. PATH LOSS MODEL

A. Frequency-independent Pathloss Model

In the conventional narrowband system, the pathloss model needs to calculate the link budget of the system and to minimize the interference in the neighboring systems. However, the UWB system is expected to require a pathloss model that accounts for the frequency component due to its much wider bandwidth than a conventional system, as discussed in Section III.B. First, in this section, we utilize the frequency-independent pathloss model using equation (1) to show the effect of the presence of people on the pathloss calculated with only the distance,

$$PL_{dB}(d) = PL_{dB}(d_0) + 10n \log_{10}(d/d_0) + S \quad (1)$$

where $PL_{dB}(d_0)$ is the pathloss at the reference distance d_0 (which is 1m in this paper), d is the separation between the transmitter and the receiver, n is the pathloss exponent, and S is related to the degree of large-scale fading with a zero-mean Gaussian distributed random variable (in dB) with a standard deviation of σ_S (also in decibels) [13]. $PL_{dB}(d_0)$, and σ_S in (1) are averaged over a 1.6 GHz bandwidth, and

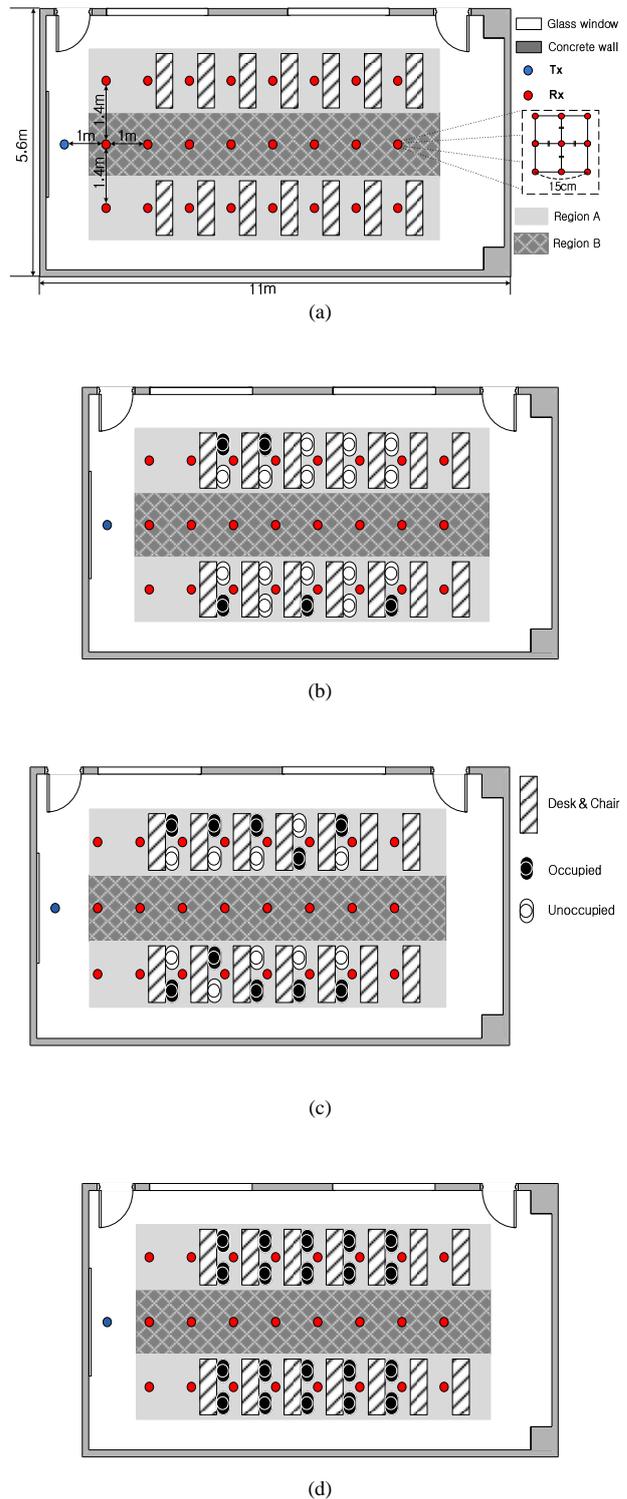


Figure 2. Floor plans and receiver locations in (a) a room with desks and chairs, (b) a room with desks, chairs and five people (c) a room with desks, chairs and 10 people, and (d) a room with desks, chairs and 20 people.

n is computed using the minimum mean square error algorithm. Fig. 3 shows the path loss and their linear regression model in Region A and Region B. The parameters of the regression model are summarized in Table I.

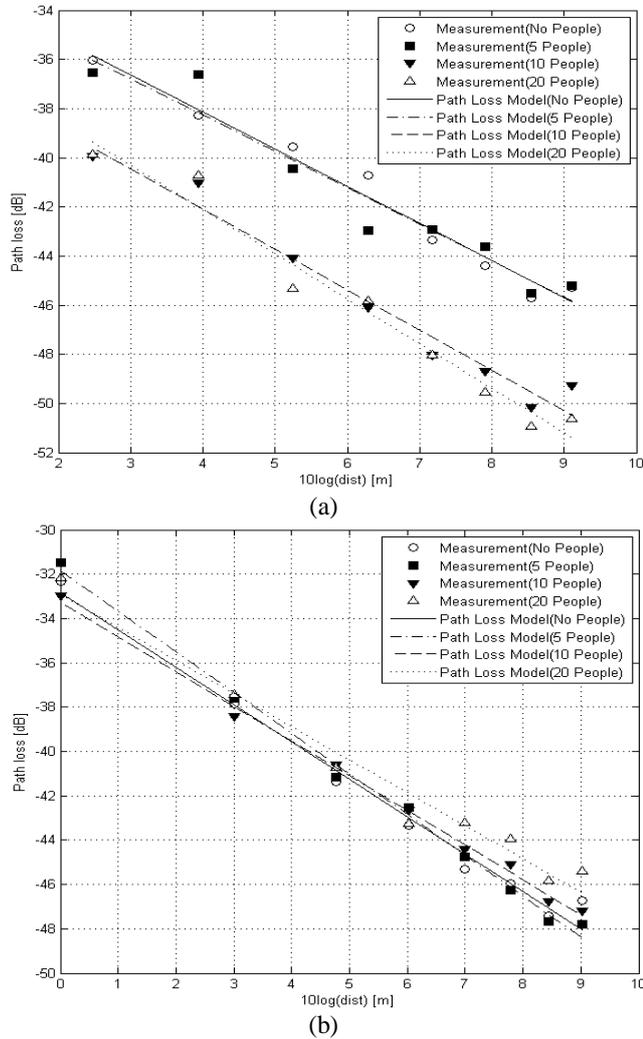


Figure 3. Path loss model of (a) Region A (Near wall) and (b) Region B (Center aisle)

Table I shows the variations of the pathloss exponent with the different scenarios in Region A and Region B. Before analyzing the effect of people on the pathloss exponent, for two regions with no people, Region A in which the desks and chairs are located, has a smaller pathloss exponent than Region B, which corresponds to the center aisle. This difference is due to the desks and chairs. For Region A, corresponding to the area near both of the side walls, the pathloss exponent increases when there are more people in the room. However, it decreases for Region B in the middle of the room. This difference is caused by the difference in the main propagation mechanism of the received signal: In Region A, as the number of people increases, the number of positions in which the direct path is

TABLE I. EMPIRICAL PATHLOSS PARAMETERS

Scenario	n		PL(d0) [dB]		σ_s [dB]	
	Region A (Near wall)	Region B (Center aisle)	Region A (Near wall)	Region B (Center aisle)	Region A (Near wall)	Region B (Center aisle)
no people	1.51	1.68	32.06	32.83	0.48	0.47
5 people	1.48	1.83	32.34	31.84	0.66	0.34
10 people	1.64	1.57	35.51	33.25	0.54	0.24
20 people	1.82	1.49	34.86	32.91	0.63	0.61

blocked increases. In addition, some reflected paths through both side walls and the ceiling are affected by the existing people, and some diffraction paths exist around them [14]. In contrast, in Region B, the direct path is not affected by people because there are LOS paths in all positions, but some reflected paths are blocked or attenuated by people. This different propagation mechanism results in different tendencies of the pathloss exponent in Region A and Region B.

B. Frequency-dependent Pathloss Model

In the conventional narrowband system, it is sufficient to represent the pathloss model in formula (1) in section A, but the UWB system requires the pathloss model considering the frequency dependency due to its wide bandwidth. For this reason, the frequency-dependent pathloss property has been discussed in many studies [15-19]. In this paper, we analyze the frequency-dependent pathloss property through the frequency-dependent pathloss model as introduced by Jinwon Choi et al. [15]. The frequency-dependent pathloss is expressed using the following modified expression of (2)

$$PL_{dB}(d, f) = PL_{dB}(d_0) + 10n(f) \log_{10}(d / d_0) + S \quad (2)$$

where $n(f)$ is the frequency-dependent pathloss exponent. The other components are identical to those in (1).

For UWB systems in particular, the MB-OFDM scheme proposes that the assigned frequency bands should be divided into sub-bands having bandwidths of 528 MHz [20].

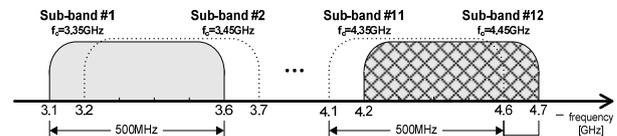


Figure 4. Twelve sub-bands with bandwidths of 500 MHz

To design an efficient MB-OFDM system, the effect of the frequency on UWB signals with different frequency bands should be characterized. For this, we obtained a practical pathloss exponent formula to express the loss as a function of the frequency. The pathloss exponent variation with the frequency is characterized by taking the average of the pathloss exponent over a 500-MHz overlapped window

bandwidth whose center frequency is incremented from 3.35 to 4.45 GHz in steps of 100 MHz, as shown in Fig. 4. In this model, the pathloss exponent averaged over each sub-band is denoted as ns_k at the k th sub-band for $k = 1, 2, \dots, 12$.

The variation of ns_k is expressed as a function of the center frequency of each sub-band.

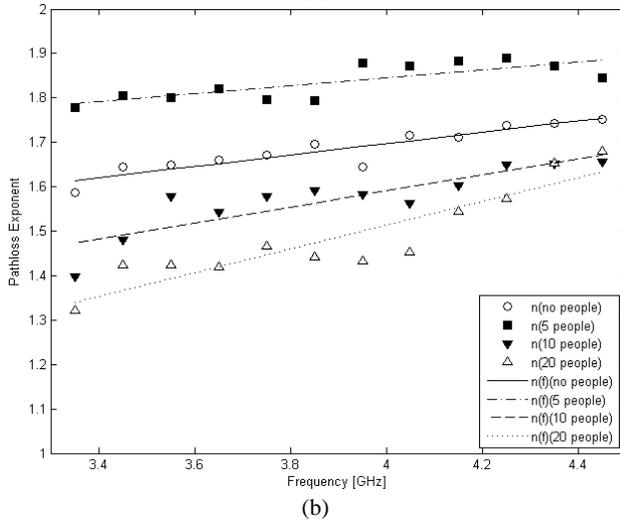
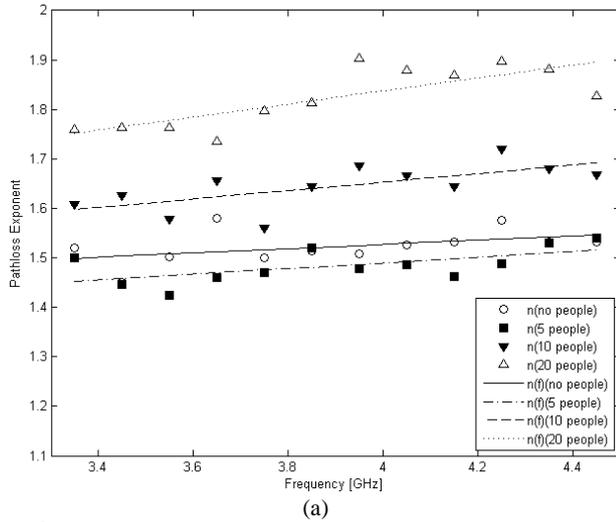


Figure 5. ns_k and their linear regression models in (a) Region A (Near wall) and (b) Region B (Center aisle)

The measured ns_k ' values are illustrated in Fig. 5. In Region A near the side walls, ns_k ' values of the scenario with 5 people are similar to those of the scenario with no people, but ns_k increases in the scenarios with 10 and 20 people. This is because the blocked multipath increases as more people are in the region. In Region B which is the center aisle of the room, ns_k ' values of the scenario with 5

TABLE II. LINEAR REGRESSION COEFFICIENTS OF (3)

Scenarios	Region A (Near wall)		Region B (Center aisle)	
	a	B	a	b
No people	0.04	1.36	0.13	1.19
5 people	0.06	1.26	0.09	1.48
10 people	0.09	1.31	0.18	0.87
20 people	0.13	1.31	0.27	0.45

people are larger than those of the scenario with no people. This situation arises because the locations which are near to the transmitter have less blocked multipath by people, but the locations far from the transmitter have more blocked multipath. But ns_k ' values of the scenarios with 10 and 20 people are smaller than the other scenarios because the difference of the path loss with the distance. As shown in Fig. 5, ns_k increases with the center frequency of the sub-band and it can be regressed as a linear function. A linear regression model of the pathloss exponent with the frequency is obtained as follows:

$$ns(f_c) = a \times f_c + b \quad (3)$$

where f_c is the center frequency of the sub-band (in gigahertz).

The statistical representatives of ns_k ' values and the linear regression coefficients a and b of (3) are shown in Table II. In both regions, the pathloss exponents increase as the frequency increases, but Region B shows faster growth of the pathloss exponents than Region A for the same scenario. These differences between Region A and Region B are the result of different main propagation mechanisms, of which the main difference is that the direct path is blocked by the people in Region A, whereas this is not the case in Region B. In addition, the slope of the pathloss exponents, a , increases as more people are in the room in the two regions. This means that the frequency response is affected by the presence of people in the room.

IV. CONCLUSION

To show how presence of people affects on a UWB channel in an indoor LOS environment, we selected four scenarios in the same room. The frequency-domain channel sounding method was used for channel characterization from 3.1 to 4.7 GHz. From the results, the presence of people causes channel variation. For the frequency-independent pathloss model, the pathloss exponent for the region of the center aisle decreases as more people are in the room, while the pathloss exponent for the region near both side walls increases as the number of people in the room increases. For the frequency-dependent pathloss model that is useful for the

MB-OFDM scheme in which the entire frequency band is divided into several sub-bands with a bandwidth of 528 MHz, the pathloss exponent increases as the frequency increases, but the region of the center aisle in the room experiences faster growth than the region near both side walls. In summary, this paper shows that the presence of people substantially affects radio-wave propagation in an indoor LOS environment and should be considered when characterizing the performance of UWB systems. This finding will be helpful to those who want to validate the results of software simulations of radio-wave propagation in an indoor LOS environment.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2012-0000162)

REFERENCES

- [1] R. Ganesh and K. Pahlavan, "Effects of traffic and local movements on multipath characteristics of an indoor radio channel," *Electron. Lett.*, vol. 26, no. 12, 7 Jun. 1990, pp. 810-812.
- [2] K. I. Ziri-Castro, W. G. Scanlon, and N. E. Evans, "Prediction of variation in MIMO channel capacity for the populated indoor environment using a radar cross-section-based pedestrian model," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, May 2005, pp. 1186-1194.
- [3] K. I. Ziri-Castro, N. E. Evans, and W. G. Scanlon, "Propagation modeling and measurements in a populated indoor environment at 5.2 GHz," in *Proc. AusWireless 2006*, 13-16 Mar. 2006, pp. 1-8.
- [4] S. L. Cotton and W. G. Scanlon, "Characterization and modeling of the indoor radio channel at 868 MHz for a mobile bodyworn wireless personal area network," *IEEE Antennas Wireless Propag. Lett.*, vol. 6, Dec. 2007, pp. 51-55.
- [5] T. B. Welch et al., "The effects of the human body on UWB signal propagation in an indoor environment," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 9, Dec. 2002, pp. 1778-1782.
- [6] A. Kara, "Human body shadowing variability in short range indoor radio links at 3-11 GHz," *Int. J. Elec.*, vol. 96, no.2, Feb. 2009, pp. 205-211.
- [7] J. Karedal, A. J. Johansson, F. Tufvesson, and A. F. Molisch, "Shadowing effects in MIMO channels for personal area networks," *IEEE VTC 2006 Fall*, 25-28 Sep. 2006, pp. 1-5.
- [8] A. Fort, J. Ryckaert, C. Desset, P. De Donecker, P. Wambacq, and L. Van Biesen, "Ultra-wideband channel model for communication around the human body," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 4, Apr. 2006, pp. 927-933.
- [9] S. L. Cotton and W. G. Scanlon, "A statistical analysis of indoor multipath fading for a narrowband wireless body area network," in *Proc. IEEE PIMRC'06*, Sep. 2006, pp. 1-5.
- [10] A. Alomainy, Y. Hao, A. Owaldally, C. G. Parini, Y. I. Nechayev, C. C. Constantinou, and P. S. Hall, "Statistical analysis and performance evaluation for on-body radio propagation with microstrip patch antennas," *IEEE Trans. Antennas Propag.*, vol. 55, no. 1, Jan. 2007, pp. 245-248.
- [11] Chiu S. and Michelson D.G., "Effect of Human Presence on UWB Radiowave Propagation within the Passenger Cabin of a Midsize Airliner," *IEEE Trans. Antennas Propag.*, vol. 58, no. 3, 2010, pp. 917-926.
- [12] 8719ES S-parameter Vector Network Analyzer, [Online]. Available: <http://www.home.agilent.com/en/pd-1000002244%3Aeapg%3Apro-pn-8719ES/s-parameter-vector-network-analyzer-135-ghz?cc=CA&lc=eng>, [retrieved: Jun., 2013]
- [13] T.S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 2002.
- [14] Zasowski T., Meyer G., Althaus F., and Wittneben, A., "UWB signal propagation at the human head," *IEEE Trans. Micro. Theory and Technol.*, vol.54, no. 4, 2006, pp. 1836-1845.
- [15] Jinwon Choi, Noh-Gyoung Kang, Yu-Suk Sung, Jun-Sung Kang, and Seong-Cheol Kim, "Frequency-Dependent UWB channel characteristics in office environments," *IEEE Trans. Veh. Tech.*, Vol. 58, Sept. 2009, pp. 3102-3111.
- [16] A. F. Molisch, K. Balakrishnan, D. Cassioli, C.-C. Chong, S. Emami, A. Fort, J. Karedal, J. Kunisch, H. Schantz, U. Schuster, and K. Siwiak, "IEEE 802.15.4a channel model—Final report," *Tech. Rep. Doc. IEEE 802.15-04-0662-02-004a*, 2005.
- [17] C. C. Chong, Y. E. Kim, S. K. Yong, and S. S. Lee, "Statistical characterization of the UWB propagation channel in indoor residential environments," *Wireless Commun. Mobile Comput.*, vol. 5, no. 5, Aug. 2005, pp. 503-512.
- [18] R. C. Qiu and I. T. Lu, "Wideband wireless multipath channel modeling with path frequency dependence," in *Proc. IEEE ICC*, Dallas, TX, Jun. 1996, pp. 277-281.
- [19] R. C. Qiu and I. T. Lu, "Multipath resolving with frequency dependence for broadband wireless channel modeling," *IEEE Trans. Veh. Technol.*, vol. 48, no. 1, Jan. 1999, pp. 273-285.
- [20] MBOA, MultiBand OFDM Physical Layer Proposal for IEEE Task Group 3a, Sep. 2004. [Online]. Available: <http://www.wimedia.org>, [retrieved: Sept., 2010]

Equal Gain Combining SC-FDMA Performance over Correlated Shadowed Rice Land Mobile Satellite Channel

Jyoti Gangane¹, Mari-Carmen Aguayo-Torres², Juan Sánchez-Sánchez³, Francisco Blázquez-Casado⁴

^{1,2,4}Communications Engineering Dept, ³Ericsson SLU

^{1,2,4}University of Málaga

^{1,2,3,4}Málaga, Spain

¹ganganejyoti@gmail.com, ^{2,4}{aguayo, fbc}@ic.uma.es, ³juan.jesus.sanchez.sanchez@ericsson.com

Abstract— Single Carrier Frequency Division Multiple Access (SC-FDMA) has shown robustness to multipath fading and similar characteristics to those of Orthogonal Frequency Division Multiple Access (OFDMA). The main advantage of SC-FDMA over OFDMA is its lower Peak to Power Average Ratio (PAPR). Because of this, Single Carrier Frequency Division Multiple Access has been proposed for transmission in Land Mobile Satellite (LMS) Systems. The link level performance of LMS system is diminished by rapid amplitude and phase variations of the received signal. These fluctuations are caused by multipath propagation and attenuation due to shadowing and can be alleviated by the use of diversity at receiver. In this paper, we have investigated SC-FDMA performance for Equal Gain Combining (EGC) diversity over a correlated Shadowed Rice LMS channel, a generalized Rice channel in which the Line of Sight is following Nakagami-m distribution and received echoes following an exponentially decreasing Power Delay Profile. We show that Interleaved SC-FDMA performance is better than Localized SC-FDMA.

Keywords- OFDMA; SC-FDMA; LOS; LMS; Shadowed Rice.

I. INTRODUCTION

Satellite communication is growing fast in order to follow unpredictable increase in demand for improved quality of service, higher capacity and ubiquitous connectivity. Satellite networks with the terrestrial networks are showing fast and strong development. In this regard, the use of multicarrier modulation and Multiple Input Multiple Output (MIMO) [1] technology are very promising candidates to reduce bandwidth requirement and enhance capacity and data rates. MIMO technology offers many advantages, such as i) Multiuser diversity gain, (ii) Spatial multiplexing gain (iii) Coding gain, and Interference reduction [1].

Orthogonal Frequency Division Multiplexing (OFDM) has demonstrated to be competent technology for wireless communication. The major advantage of OFDM is its relatively simple method of handling frequency selective fading channels which are normally observed in wireless channels [2]. However, OFDM transmission has high Peak to Average Power Ratio (PAPR) which increases the power back up for operation of Linear Power Amplifier. There are many techniques to reduce the OFDM PAPR [3][4] and the outcome of that is DFT Spread OFDM [5].

Single Carrier Frequency Division Multiple Access (SC-FDMA) is a closely related transmission scheme with the same multipath fading mitigation characteristics as that of OFDM [6][7]. Though SC-FDMA has shown worse performance with Rayleigh fading [8], its performance has improved with increase in Line of Sight (LOS) in Rician fading [9].

SC-FDMA performs DFT spreading in the transmitter, before going to Inverse Discrete Fourier Transform (IDFT) [5]. The DFT operation in transmitter helps SC-FDMA to spread the energy of individual modulation symbols over a number of subcarriers in the DFT block. The transmitters in SC-FDMA system use different subcarrier to transmit information symbols. However, they transmit the subcarriers sequentially, instead of in parallel. In relation to OFDM, this arrangement reduces fluctuation in the transmitted waveform. Therefore, SC-FDMA signals have naturally lower PAPR. However, in a cellular system with severe multipath fading, the SC-FDMA signals arrive at the base station with inter symbol interference (ISI). At the receiver, adaptive frequency domain equalization is used to cancel ISI.

Depending on how the DFT spread symbols are mapped onto the subcarriers in the IDFT block, mapping is classified into two types: localized and interleaved. In Localized-SC-FDMA (L-SC-FDMA), each terminal uses a set of adjacent subcarriers to transmit its symbols [10]. In Interleaved-SC-FDMA (I-SC-FDMA), the subcarriers used by the terminals are spread over the entire signal band, and occupied subcarriers are equidistant from each other.

Diversity combining techniques such as Maximal Ratio Combining (MRC), Equal Gain Combining (EGC) and Selection Combining (SC) are used to mitigate fading in wireless channel. It is known that equal gain combining receiver yields similar performance to MRC, with lower implementation complexity. In an EGC combiner, the output of different diversity branches are first co-phased and then weighted equally before combining to give resultant output [11]. Co-phasing is required to avoid signal cancellation.

The EGC performance analysis is found in [12]. A complete summary of most of the linked work is found in [13]. Frequency-Domain (FD) Amplify and Forward (AF) single relay SC-FDMA performance with EGC is studied in [14]. In [15], we studied the Equal Gain combining SC-FDMA performance over unshadowed Rice Land Mobile Satellite (LMS) channel.

The rest of the paper is organized as follows. Section II presents the system model. Section III describes the channel model used to carry out simulation. In Section IV, simulation results are presented. Finally, Section V provides the conclusion.

II. CHANNEL MODEL

The link level performance of LMS systems strongly depends on the transmission channel between the satellite and the mobile user. The statistical models for narrowband LMS channels can be classified into two types of fading: shadowed and unshadowed Rice fading. When the mobile receiver has a clear path to the satellite it is called unshadowed fading. In case of shadowed fading, the line of sight path is blocked by terrain, vegetation, or human-made structures.

The first shadowed Rice model was proposed by Loo [16], where Line Of Sight (LOS) component follows lognormal distribution. In this paper, we have used the model discussed in [17], where they assumed that the power of LOS component is a gamma random variable. G. L. Stuber et al. [18] discussed that the square root of a gamma variable has Nakagami-m distribution.

In the used channel model, the scattered component of the received signal follows Rayleigh distribution and LOS follows a Nakagami-m distribution [17] as follows:

$$P_X(x) = \frac{x}{b_o} \exp\left(\frac{-x^2}{2b_o}\right) \quad (1)$$

$$P_Y(y) = \frac{2m^m}{\Gamma(m)\Omega^m} y^{2m-1} \exp\left(\frac{-my^2}{\Omega}\right) \quad (2)$$

where $2b_o = E[X^2]$ is the average power of scatter component, $\Gamma(\cdot)$ is the gamma function, $m = E[Y^2]/Var[Y^2] \geq 0$ with $Var[\cdot]$ as the variance, and $\Omega = E[Y^2]$ is the average power of the LOS component. The shadowed Rice PDF [17] in terms of a Nakagami parameters is as follows:

$$P_R(r) = E_Y \left[\frac{r}{b_o} \exp\left(\frac{r^2 + Z^2}{2b_o}\right) I_0\left(\frac{Zr}{b_o}\right) \right] \quad (3)$$

where E_Y is the expectation with respect to Y , and $I_n(\cdot)$ is the n^{th} order modified Bessel's function of the first kind.

In the proposed satellite channel model, a fixed and sufficiently large number of rays are assumed, thus eliminating the need to use a Poisson distribution. Power Delay Profile (PDP) describes the numbers and position of echoes, as well as their average power. In Land Mobile system it is usual to consider an exponential PDP [19] with parameter τ_{avg} such that the scattered received power for l^{th} echo is as follows:

$$P_l = \frac{1}{\tau_{avg}} e^{-\left(\frac{\tau_l}{\tau_{avg}}\right)} \quad (4)$$

For a rural environment, a typical value of τ_{avg} is less than $1 \mu\text{s}$ and less than $2 \mu\text{s}$ for a suburban area [20]. The value of τ_{avg} used in this paper is $1.17 \mu\text{s}$.

III. SYSTEM MODEL

Figure 1 shows the block diagram of an EGC SC-FDMA transmitter-receiver. The input bits to be transmitted are converted in complex symbols (e.g. BPSK [8] or QAM [8]) using digital modulation techniques. The N_c -point DFT operation converts these complex symbols into precoded complex symbols \mathbf{X}_{N_c} . The precoded symbols are then mapped on a subset of different allocated subcarriers per user, i.e., N_c out of M sub-carriers in which the total system bandwidth is divided. The mapping can be Interleaved (I-SC-FDMA) or Localized (L-SCFDMA). After subcarrier mapping, a cyclic prefix (CP) is added and complex symbols are transmitted. We assume that CP is long enough so that the non zero echoes are fitted into it.

At the receiver, after the removing the CP, the M-DFT operation converts the received symbol from the time domain to the frequency domain. The received signal is given by

$$\mathbf{Y}_n = \mathbf{H}_n \mathbf{X} + \boldsymbol{\eta}_n \quad (5)$$

where $n \in \{1, 2, \dots, N_R\}$ represents the antenna index, N_R being the number of antennas at the receiver, $\boldsymbol{\eta}_n$ is a noise vector whose entries are i.i.d. complex Gaussian $\mathcal{CN}(0, N_0)$ and \mathbf{H}_n represents the $M \times M$ diagonal matrix whose entries are the channel frequency response as seen by the N_R antennas for each allocated subcarrier [21]. These N_R signals are frequency combined using EGC [21].

$$\mathbf{Y} = \frac{\sum_{n=1}^{N_R} \mathbf{Y}_n * \exp(-j * \text{angle}(\text{diag}(\mathbf{H}_n)))}{\sum_{n=1}^{N_R} |\mathbf{H}_n|} \quad (6)$$

after taking N_c -IDFT of \mathbf{Y} , it is given to the detector. The output of detector is estimate of the input bits.

IV. PERFORMANCE RESULTS

In this section, simulation results are carried to evaluate BER performance for the SIMO (1×2) EGC I-SC-FDMA and L-SC-FDMA over shadowed Rice LMS channel. The parameters used for simulation are shown in Tables I and II.

TABLE I. SHADOWED RICE CHANNEL PARAMETERS FOR DIFFERENT FADING I

	b_o	m	Ω
Light shadowing	0.158	19.4	1.29
Heavy shadowing	0.063	0.739	8.97e-04
Average shadowing	0.126	10.1	0.835

Different shadowed conditions, Heavy, Light and Average are considered while carrying out the simulation. The SIMO (1×2) EGC SC-FDMA performance has been evaluated for different antenna correlation factor, number of antennas at receiver, and allocated subcarriers.

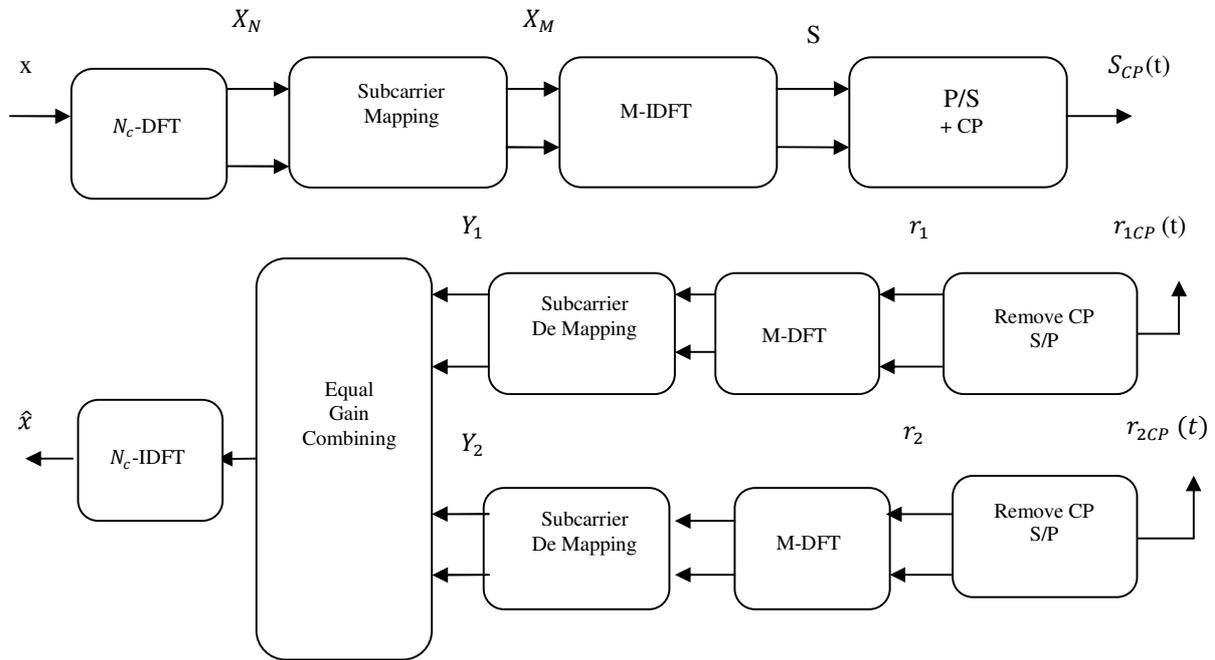
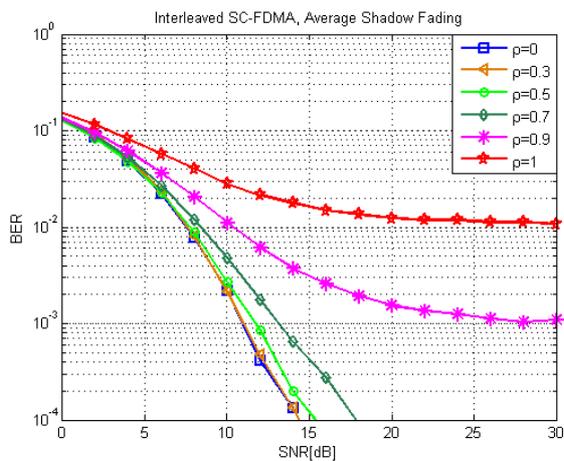


Figure 1. EGC SC-FDMA Transmitter Receiver scheme

TABLE II SIMULATION PARAMETERS

FFT Size	1024
Modulation Techniques	QPSK
Carrier Frequency	2.00GHz
System Bandwidth	20MHz
Channel model	Satellite LMS with exponential PDP
Number of Antennas Used	2, 4
Equalizers used	EGC
τ_{avg}	1.17 μ s



Figures 2 and 3 show the BER performance of Interleaved SC-FDMA and Localized SC-FDMA for Average and Light fading. Interleaved SC-FDMA performance over Average and Light fading is better than that of Localized SC-FDMA. In light fading, the LOS component has more power than the multipath component. The light and average shadowed Rician distribution can be represented by the normal Rician distribution with a Rice factor of 5.3 and 4.1 dB, respectively [22].

Figure 4 shows the BER performance of Interleaved SC-FDMA for Heavy fading. In case of Heavy fading the Line of Sight component has very low power, it behaves as good as Rayleigh fading.

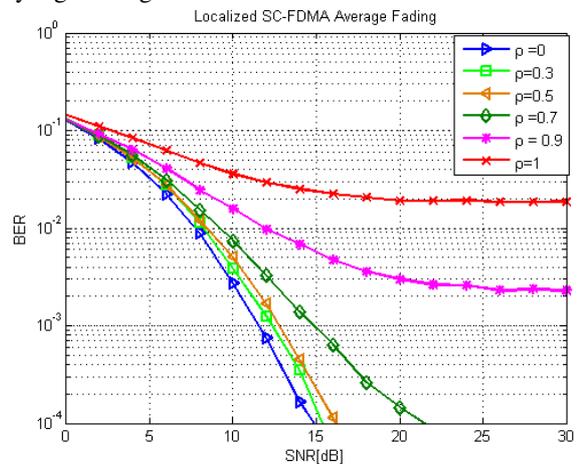


Figure 2: BER performance of Interleaved SC-FDMA and Localized SC-FDMA for different antenna correlation factor: Average Shadow fading

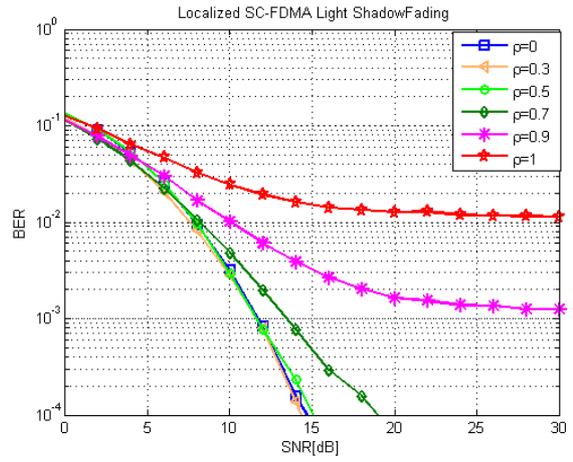
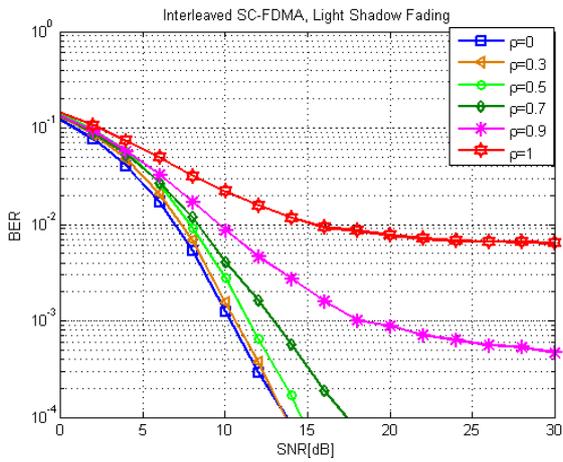


Figure 3: BER performance of Interleaved SC-FDMA and Localized SC-FDMA for different antenna correlation factor: Light Shadow fading

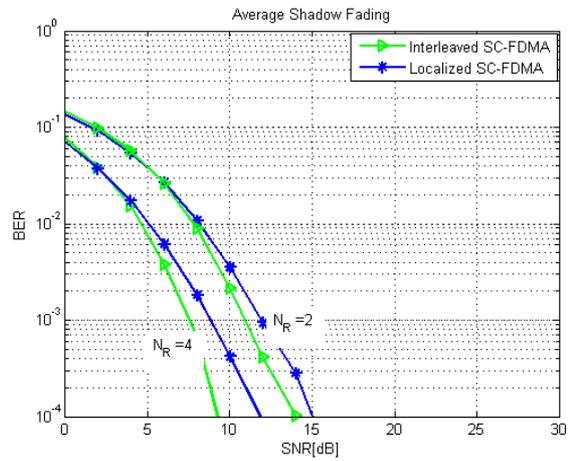
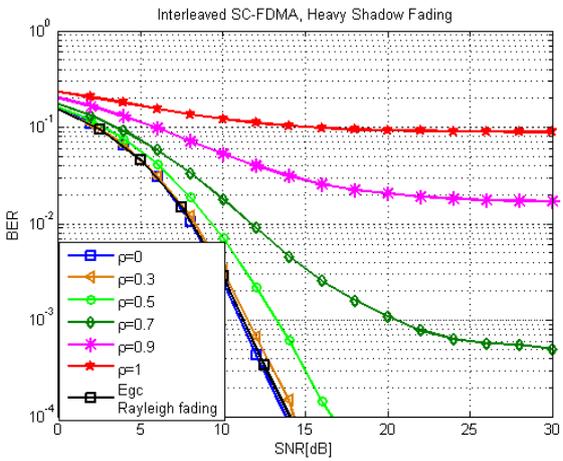


Figure 4: BER performance of Interleaved SC-FDMA for different antenna correlation factor and different antennas at receiver: Heavy & Average Shadow fading

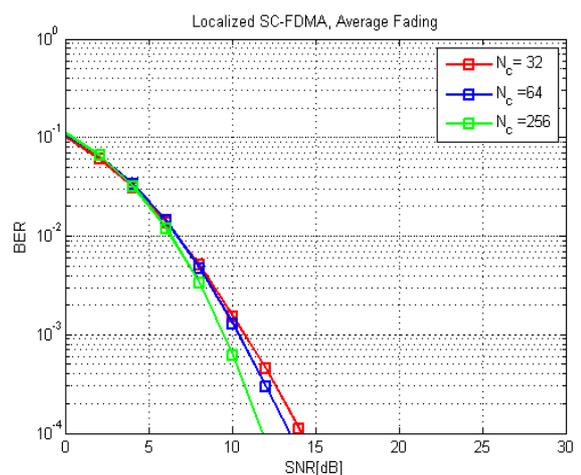
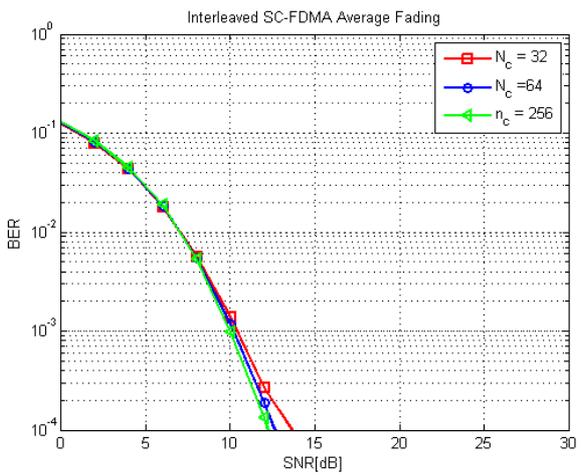


Figure 5: BER performance of Interleaved SC-FDMA and Localized SC-FDMA for different allocated subcarriers: Average Shadow fading

Results have shown that performance is similar below certain antenna correlation threshold, for Light fading threshold is 0.5. Note that the existence of LOS (quite powerful for Light fading) always provokes certain correlation between signals received at different antennas, thus the lack of spatial correlation does not further improve the BER performance. The results in Figure 4 shows that a higher number of receiving antennas improve BER performance over Average fading, as expected.

In Figure 5, the performance for different number of allocated subcarriers over Average fading shows that BER performance is not greatly modified by N_c .

V. CONCLUSION

In this paper, we have investigated BER performance of SIMO (1×2) EGC for Interleaved and Localized SC-FDMA over shadowed Rice LMS channel where LOS component follows Nakagami-m distribution. The performance is studied under light, average and heavy fading. It is observed that, as expected, in all three cases EGC diversity shows better performance as compared to SISO SC-FDMA. In case of Heavy fading, the performance is worse because LOS component has low power. The antenna correlation effect in all three types of fading shows that the more antenna correlation, the more errors at the detection. However, existence of LOS always provokes certain correlation between signals received at different antennas thus improvement for lower spatial correlation or using more antennae is not high. In general, Interleaved SC-FDMA BER performance is better than Localized SC-FDMA in shadowed Rice fading. The Interleaved SC-FDMA can be best transmission scheme for uplink satellite communication with EGC receiver diversity technique. Further, it is recommended to study the performance of MIMO SC-FDMA over Land Mobile Satellite channel.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support of the European Commission through Erasmus Mundus External Cooperation Window "Mobility for Life" and express their gratitude to the Spanish Government (Plan Nacional de I+D+I, TEC2010-18451) and Sinhgad Technical Education Society (STES), Pune and University of Pune (India), and to Universidad de Málaga – Campus de Excelencia Internacional Andalucía Tech.

REFERENCES

- [1] P. Arapoglou, K. Liolis, M. Bertinelli, A. Panagopoulos, P. Cottis, and R. Gaudenzi, "MIMO Over Satellite: A Review," IEEE Communications Surveys and Tutorials, Vol. 13, No. 1, May 2010, pp. 27-51
- [2] J. Bingham. "Multicarrier modulation for data transmission: An idea whose time has come," IEEE Communication Magazine, May 1990, pp. 5-14
- [3] P. Eetvelt, G. Wade, and M.Tomlinson, "Peak to average power reduction for OFDM schemes by selective scrambling," in Electronic Letters, Vol. 32, Oct. 1996,1963-1964
- [4] Y. Rahmatallah, N. Bouaynaya, and S. Mohan "Bit Error Rate Performance of Linear Companding Transforms for PAPR Reduction in OFDM Systems," in IEEE Global Communications Conference (GLOBECOM 2011), Houston, Texas, December 2011, pp. 1-5
- [5] M. Nisar, H. Nottensteiner, and T. Hindelang, "On performance limits of DFT spread OFDM systems," Mobile and Wireless Communications Summit, 2007, 16th IST, July 2007, pp. 1-4
- [6] A. Czyliwik, "Comparision between adaptive OFDM and Single Carrier modulation with frequency domain equalization," in Proc. Vehicular Technology Conference 1997 Phoenix, AZ, vol. 2, May 1997, pp.865-869
- [7] H. Sari, G Karam, and I. Jeanclaude, "Transmission techniques for digital terrestrial TV broadcasting," IEEE Commun. Mag., vol. 33, no. 2, pp. Feb. 1995, pp. 100-109
- [8] J. Sánchez-Sánchez, "Analysis of SC-FDMA and OFDMA Performance over Fading Channels," Ph.D. dissertation, Universidad de Málaga (Spain), May 2011. [Online]. Available: <http://hdl.handle.net/10630/4689>
- [9] J. Gangane, M. Aguayo-Torres, J. Sánchez-Sánchez, and U. Fernández-Plazaola, "SC-FDMA Performance over Rician Channel," Proc. 6th IEEE International Conference on Broadband and Biomedical Communications (IB2Com), November 2011, Melbourne, Australia, pp. 91-96
- [10] U. Sorger, I. Broeckan, and M. Schnell, "IFDMA- A New Spread Spectrum Multiple Access Scheme, Multicarrier Spread Spectrum," Kulwer, 1997, pp. 111-118
- [11] A. Annamalai, C. Tellambura, and V. Bhargava, "Equal -gain diversity receiver performance in wireless channels," IEEE Trans. Communications, Vol. 48, Oct. 2000, pp. 1732-1745
- [12] N. Beaulieu and A. Abu-Dayya, "Analysis of equal gain diversity on Nakagami fading channels," IEEE Trans. Commun. Vol.39, No. 2, Feb. 1991, pp.225-234
- [13] M. Simon and M. Alouni, Digital Communication over Fading Channels, 2nd ed., New York:,Wiley, 2004
- [14] J. Zang, L. Yang, and I. Hanzo, "Multi -user performance of the amplify and forward single relay assisted SC-FDMA uplink," in Proc. IEEE Vehicular Technology Conference 2009 Fall, Sept. 2009, pp.1-5
- [15] J. Gangane, M. Aguayo-Torres, J. Sanchez, and S. Wagh, "Equal Gain Combining (EGC) SC-FDMA Performance over Land Mobile Satellite (LMS) Rice Fading Channel," Proc. IEEE International Conference on Computational Intelligence and Computing Research (IEEE ICCIC), Dec. 2012, Coimbatore, India, pp.1-5
- [16] C. Loo, "A statistical model for a land mobile satellite link," IEEE Trans. Veh. Technol, Vol. 34, August 1985, pp. 122-127
- [17] A. Abdi, W. Lau, M. Alouni, and M. Kaveh, "A new simple model for land mobile satellite channels: first-and second order statistics," IEEE Trans. Wireless Comm., Vol. 2, No. 3, May 2003, pp. 519-528
- [18] G. L. Stuber, Principles of Mobile Communication, Boston, MA: Kluwer, 1996
- [19] F. Fontan et al, M. Castro, J. Kunisch, J. Pamp, E. Zollinger, S. Buonomo, P. Baptista, and B. Arbesser, "A versatile framework for a narrow and wide band statistical propagation model for the LMS channel," IEEE Trans. Broadcasting, Vol. 43, Dec.1997, pp.431-458
- [20] J. Rees, "Measurements of the wide band radio channel characteristics for rural, residential, and suburban areas," IEEE Trans. Veh., Technol., Vol. 36, Feb. 1987, pp. 2-6
- [21] A. Goldsmith, Wireless Communications, Cambridge University Press, 2005
- [22] D. Richard, V. Nee, S. Howard, and R. Prasad, "Direct -Sequence Spread Spectrum in a Shadowed Rician Fading Land Mobile Satellite Channel," IEEE Journal on Selected Areas in Communications, Vol.10, No. 2. February 1992, pp. 350-357

Linear Node Movement Patterns in MANETS

Mohammed Alzaylaee, John DeDourek, Przemyslaw Pohec

Faculty of Computer Science

University of New Brunswick

Fredericton, Canada

e-mails: {m.alzaylaee, dedourek, pohec}@unb.ca

Abstract— MANETs are mobile, self-configuring networks of wireless mobile devices that have no fixed infrastructure configuration. Movement of the nodes affects the operation of a MANET. Because of the node movement MANETs need to rely on robust routing protocols. We focus our investigation on the impact of movement patterns on operation of a MANET. Of particular interest is linear movement patterns similar to those observed in some man-made objects. We investigate the operation of a MANET under four different linear movement patterns by simulating a MANET running the AODV routing protocol. The best results were obtained with nodes moving perpendicularly to the geographical direction of packet forwarding.

Keywords- mobility models, MANET; M2ANET; ns-2; AODV

I. INTRODUCTION

A network which links several devices and relies on radio signal frequency is known as a wireless network. A Mobile Ad Hoc Network (MANET) is connected by radio links and consists of self-reliant mobile nodes [1]. The nodes move and operate without any central control as individual autonomous systems. They move freely in any direction and links among these devices change frequently. The dynamic movement pattern that these nodes follow influences the overall network performance. Different movement patterns can be categorized with respect to mobility models [2][3][4]. Examples of realistic models that have been suggested include obstacle mobility and pathway mobility [5]. In these models, there are pre-determined pathways and obstacles, which determine the movement of nodes and propagation of signals in a wireless network.

Linear node movement occurs in practice in man-made system. Examples of these include: movement of machines on a factory floor, bus routes following a city grid, and most interestingly some plane routes. For example, typical flight corridors for transatlantic flights show close to parallel paths of nodes (planes) as illustrated in Figure 1. In this paper, we present the results of investigation into different mobility models with practical applications for MANETs. Instead of using the most common random mobility (exemplified by setdest in ns-2) we propose moving nodes along mostly straight paths, possibly adjusting direction periodically. Such a movement pattern should be easier to realize in practice, as the nodes would be moving essentially following a predefined track, like a train moving on rails

between two stations back and forth.

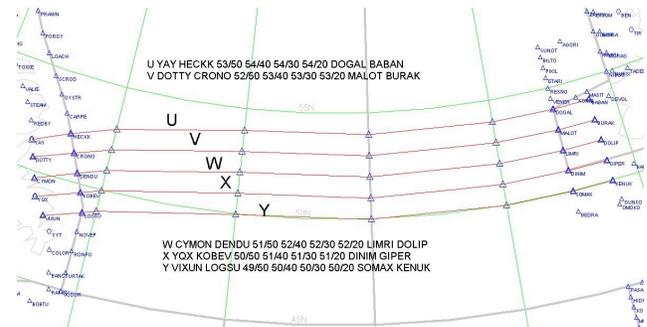


Figure 1. North Atlantic Tracks for the eastbound crossing on the evening of May 4, 2006 [6].

In Section II, we present background on MANETs and principal routing protocols. Different linear movement patterns for MANETs and simulation of these movements in ns-2 are discussed in Section III. Experiments with different linear movement patterns are in Section IV. Finally, we present the experimental results in Section V, followed by conclusion and future work.

II. STATE OF THE ART

A MANET is comprised of interconnected nodes, which make use of communication paths that are allowing multi-hop activity. They offer distinct advantages and are versatile for some particular applications and environments. There are no fixed or prerequisite base stations or infrastructures; therefore, their creation and usage is not time consuming and can occur at any given point in time and at any place. MANETs have a fault-resilient nature, given that they are not operating a single point of failure and are very flexible. The deletion and addition of new nodes, forming new links are a normal part of operation of a MANET [1][7][8]. A group of nodes can facilitate communication between distant stations forming a mobile medium, as introduced in [9]. These benefits of MANETs have resulted in many applications in hostile environments including military, police, rescue and other hostile or disorganized environments.

MANET research and deployment are aided by the fact that they use small and relatively inexpensive wireless communication devices. Simulators like ns-2 include models for wireless nodes, links and protocols and can be used for experimenting with MANETs [10]. The simulations rely on numerous parameters, which include not only the patterns of communication but also the model for mobility.

Future expectations are that MANETs will be deployed in various scenarios that tend to have complex connectivity dynamics and node mobility. A good example would be a MANET on a battlefield wherein the soldiers' movement is dictated by their commander. For a citywide MANET, maps or obstacles limit the movement of the nodes. Note that mobility of the node and its properties are specific to the application in question; therefore, mobility properties that are widely varying possess the ability to greatly impact the performance of the various routing protocols. To cope with these specific requirements of MANETS, highly adaptive protocols based on flooding [11][12][13] and on dynamic routing [14][15][16] were developed. In our paper, we propose and analyze a number of variants of motion with spatial dependencies: our vertical and horizontal motions are similar to the Column Mobility Model (CMM), and square and specified area motions are similar to the Reference Point Group Mobility (RPGM) introduced in [17]. In addition to different motion types, we also investigate the relation between the direction of the motion itself and the geographical direction of the data transfer. For experiments in our research we considered using one of the two popular MANET routing protocols supported in ns-2 simulator: Dynamic Source Routing (DSR) and Ad hoc On-Demand Distance Routing (AODV). Both DSR and AODV share a similar on-demand behavior route discovery for ad hoc networks, but with different mechanics for routing. However, AODV outperforms the DSR routing protocol when simulating a large number of nodes as in our experiment [7]. This is why AODV is used in modeling MANETs in this report.

III. INVESTIGATION OF SELECTED LINEAR MOVEMENT PATTERNS

Setdest [10] is a tool built into ns-2 [10] and uses a random mobility model to generate random movements for the nodes using a pause and move strategy. In this strategy, a next movement and speed are independent of the previous move. In this paper, we investigate the MANET performance under a scenario with constrained node movement. More specifically, we investigate the node movements that are still random but confined to a single direction, a line, and are similar to some real movements. For example a rail car can move only forward and back but not sideways. We assume that the MANET is composed of many such nodes, but it is used to carry the data between two defined (and fixed) nodes. This type of a model is called Mobile Medium Ad Hoc Network (M2ANET) and is

based on the concept of mobile medium introduced in [9]. The references to “vertical” and “horizontal” are based on the movement direction as observed in the ns-2 simulator on a computer screen. Horizontal corresponds to the movement of the nodes in the general geographic direction of packet forwarding, and vertical corresponds to the path perpendicular to the one defined above.

A. Vertical motion

In this motion, each node is initially assigned an X coordinate randomly, which stays the same during the simulation, while the Y coordinates changes for each move. Each node is assigned a random destination located on the vertical line defined by the node X coordinate and moves to this destination at a random speed. The process is repeated until a node reaches the destination.

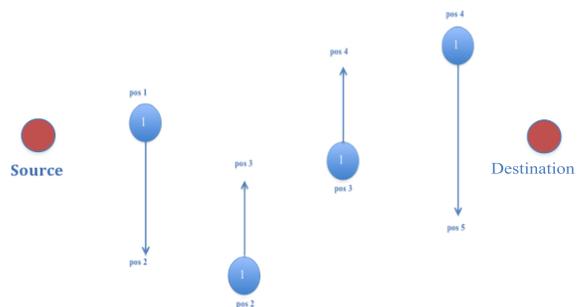


Figure 2. Vertical motion.

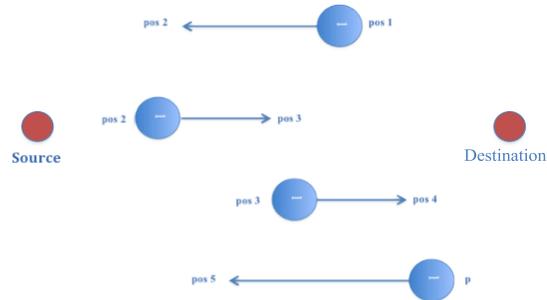


Figure 3. Horizontal motion.

B. Horizontal motion

This movement motion does exactly the opposite of the vertical movement. The Y's coordinate remains the same, while the X's coordinate changes for each move during the simulation time. This type of motion is similar to the movement of planes on the path shown in Figure 1, with the exception that some of the planes (if they were controlled by random number generators) would be turning back midway.

C. Square motion

This movement motion allows the nodes to move in a square path with different speeds for each node during the simulation time. This square path is defined to be a 100*100 square in the experimental area and is the same for each node. Each node takes a random position at first. Then it

moves to the right by adding 100 to its initial position. Then it moves up by 100 from its previous position. Finally, it moves to the left by adding 100 to the previous position. A full square path is followed by each node and repeated until the end of the simulation time.

D. Specified area motion

This movement motion keeps the nodes moving randomly based on values generated the Java Math library function `Math.random()` in a specified area (200*200 in our experiment). Each node has an initial random position at the start of the simulation. Then each node moves to a new randomly assigned position but does not go beyond 200*200 from the node's first initial position during the simulation time. (The displacement along each axis is calculated mod 200).

E. Simulation environment

The simulation time for our experiment is set to be 1000 sec in a topology of size 1000m * 1000m. We tested different node density starting from 5 nodes up to 100 nodes, which includes the two stationary nodes. We run the simulation 10 times for each node density and then we calculate the average of these runs. Node transmissions are simulated using the 802.11 ns-2 model, and the packets are forwarded using the AODV protocol.

For comparison, we also simulated random motion using the standard ns-2 `setdest` utility. For performance evaluation, since we always have the same number of packets sent, we only need to quantify packets that are received at the destination node. This is the same as using the packet delivery as a main performance metric in the experiment.

IV. EXPERIMENTAL RESULTS

A. Base case scenario with `setdest` results

The base case scenario in our experiment used the `setdest` utility to generate random movements for the nodes. The experiment for this movement was run 10 times for each node density and the average of packet delivery was calculated. In Figure 2, the three curves represent the number of packets received at three different speeds. The graph shows that at high densities the number of packets received at the destination node decreases as the maximum speed increases from 25, 50 to 500 m/s. This result is based on the random movement generated from `setdest` in ns-2.

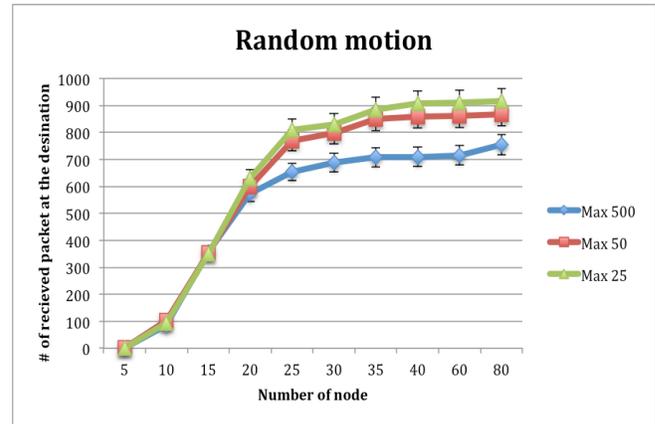


Figure 4. Random motion results.

B. Vertical motion result

In this scenario, where the nodes are moving only up and down (i.e., perpendicular to the geographical direction between the source and destination) during the simulation with random speeds, we observe the average packets received at the destination node density during 10 experiments. Figures 5 and 6 illustrate the comparison of different motions in low and high average speed. The graphs show that the number of received packets increases with the increase in number of mobile nodes (which increase node density), up to a point when certain number of nodes is reached (35 in our experiments). At this point the node density is no longer the primary factor in delivery of packets. After this node density is reached, the node speed becomes the primary deciding factor in how many packets are delivered.

C. Horizontal motion result

The horizontal motion, as explained before, allows the nodes to move only in the direction of (i.e., defined by a line between) the two stationary nodes, which is in the horizontal motion experiment would be right and left. In this experiment, we also run the simulation 10 times for each node density, and calculated the average of packets received. As illustrated in Figures 5 and 6, the overall trends are similar to the vertical motion, except that the vertical motion reaches a higher packet delivery, which is particularly pronounced at high node speeds. A closer inspection of the simulation runs revealed that in case of horizontal motion some mobile nodes had their initial random position assigned too far from the two stationary nodes that these nodes were out of range to deliver the cbr packets.

For both motions that we create, Figure 5 shows no difference for the delivery rate at lower speed, whereas with the higher speed, we can see from Figure 6 that vertical motion has a better performance based on the packet delivery.

D. Square path motion result

In the square path, we combined the vertical and horizontal motions into one set of moves and we limited the extent of movement to a square path. Therefore, each node moves only 100*100 from its initially random position and this would have an impact on the network connectivity at low node densities. Indeed, the observed performance (Figures 5 and 6) was lower than horizontal and vertical motion at lower node densities, and similar to these two at higher node densities.

E. Specified area motion result

Specified area motion keeps the nodes move randomly but in limited square shaped area, which is in our experiment 200m * 200m from the topology size 1000m *1000m. With the specified area motion, the nodes are also moving in a confined area. Despite the randomness of the movement, the observed performance is the worse of the four movement types investigated in this paper.

F. Comparing four different motions

At both high and low speeds the specified area motion type delivers the least number of packets and vertical motion delivers the most. Square motion is more reliable at high speeds than horizontal motion, but less reliable at low speeds. This is because square motion performs both vertical and horizontal motion and at lower speeds suffers loss of connectivity during the horizontal motion and vertical motions that occur while out of range. For all motions at high node density, movement speed becomes the primary factor in reliability and with a fewer packet delivered at high speed.

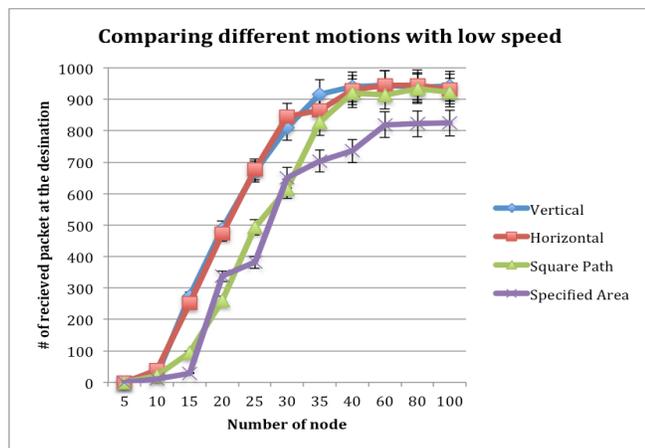


Figure 5. Comparing different motions with low speed.

From our experiment, we notice that when there is no motion among the nodes then, depending on the initial positions of the nodes, either zero percent (no path from source to destination exist) or 100 percent (path exists) of

the packets will be delivered. In this case, the possibility of receiving all the packets rises with increased node density. As soon as the nodes start to move, the packet delivery starts to decrease. When we have the nodes distributed randomly in our experiment topology, there are no packets received when we have very low node density. Once we have 20 or more nodes distributed randomly, there will be either zero or 100 percent of the packets received. The higher the node density, the better are the chances of having the packets delivered.

V. CONCLUSION AND FUTUR WORK

In this paper, we presented four different node movement patterns for use in MANETs. They include: the horizontal movement in which the nodes move in the geographical direction of forwarding packets, the vertical movement (perpendicular to the above), the movement in a square paths, and the movement in a specified area. We observed that the movement of the nodes affects the performance of packet delivery in the simulated MANET.

As expected, we observed that the speed of the movement affects the performance of the MANET in all movement types. More packets are dropped with larger maximum speeds at higher node densities. For example, for the horizontal motion the average packet delivery in the simulation with 40 nodes was 93% at low speed and only 80% at high speed. Our experiments show that the vertical motion (i.e., motion perpendicular to the direction of forwarding packets) is better than the horizontal motion (i.e., motion in the direction of forwarding packets) at a given node density. For example, in experiments at high speed, 89% of packets are delivered in a network with vertical node movements vs. 80% in case of the horizontal movement (both in the simulation of a network with n=40 nodes). The horizontal motion causes the most displacement for the node relative to the source and destination nodes. Once the nodes are out of range by horizontal motion, they continue to be out of range and do not forward data.

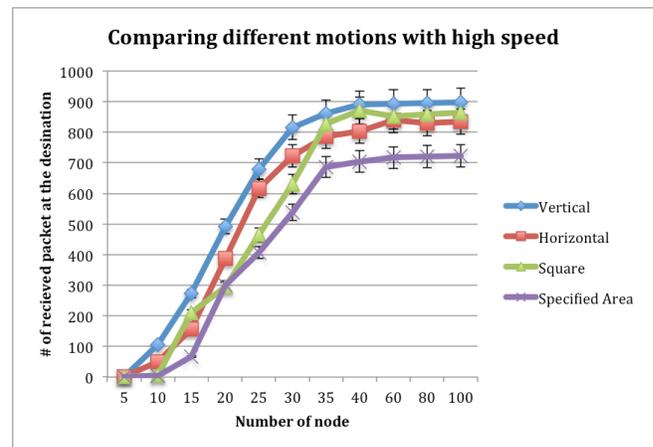


Figure 6. Comparing different motions with high speed.

Based on our results, we suggest further testing of all the movements with changing multiple parameters, such as increasing or lowering the experimental area, changing the size of the specified area size in one of the movement models, changing the routing protocols, varying speed, and node density, etc. Furthermore, other movement patterns can be investigated in simulation and in an actual physical MANET experiments. It would also be of interest to model the movement of all transatlantic flights over the North Atlantic and see if a MANET network could be established if all the planes were equipped with suitable transceivers.

ACKNOWLEDGMENT

This work is sponsored and funded by the Ministry of Higher Education of Saudi Arabia through the Saudi Arabian Cultural Bureau in Canada.

REFERENCES

- [1] S. Basagni, M. Conti, S. Giordano and I. Stojmenovic (Ed.). *Mobile Ad Hoc Networking*. IEEE/Wiley Press, 2004.
- [2] F. Bei, and A. Helmy, "*A survey of mobility models in wireless Ad hoc Networks*". California: University of California", USA, 2004.
- [3] M. Dorigo, V. Maniezzo, and A. Coloni, "The Ant System: Optimization by a colony of cooperating agents", *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 26 (1), 1996, pp. 29-41.
- [4] A. B. McDonald and T. F. Znati, "A mobility-based framework for adaptive clustering in wireless ad hoc networks", *IEEE Journal on Selected Areas in Communications*, Vol 17, No. 8, Aug 1999, pp. 1466-1487.
- [5] D. P. Agrawal and Q. A. Zeng, *Introduction to wireless and mobile systems*, third edition. Thomson Engineering, 2011.
- [6] North Atlantic Tracks. (2012, October 18). In *Wikipedia, The Free Encyclopedia*. Retrieved 06:36, June 5, 2013, from http://en.wikipedia.org/w/index.php?title=North_Atlantic_Tracks&oldid=518517067.
- [7] M. R. Pearlman and Z. J. Hass, "Determining the optimal configuration for the zone routing protocol," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 8, pp. 1395-1414, Aug. 1999.
- [8] S.Sarkar, T.G.Basavaraju et al,"Ad hoc Mobile Wireless Networks, Principles, Protocols and Applications", Auerbach Publications - 2008.
- [9] J. DeDoutre and P. Pochee, (2011), *M₂ANET: a Mobile Medium Ad Hoc Network, Wireless Sensor Networks: Theory and Practice*, WSN 2011, Paris, France, pp. 1-4.
- [10] T. Issariyakul and E. Hossain, "Introduction to Network Simulator NS2 ", Springer, ISBN: 978-0-387-71759-3, 2009
- [11] S. Pleisch, M. Balakrishnan, K. Birman, R. van Renesse, "MISTRAL: Efficient Flooding in Mobile Networks", *Proc. ACM MobiHoc*, pp. 1-12, 2006.
- [12] P. Chan, and Y. Dong,, "Performance Evaluation on URL Distribution using Flooding in Mobile Ad Hoc Networks (MANETs)", *Southeast Conference, Kennesaw, Kennesaw, GA, USA, 2005*, pp. 70-80
- [13] S. Murthy, and J.J Garcia-Luna-Aceves, "An efficient routing protocol for wireless networks", *ACM Mobile Networks and Applications Journal, Special issue on Routing in Mobile Communication Networks*, 1996, pp. 183-197.
- [14] G. Di Caro and M. Dorigo "Mobile agents for adaptive routing," *System Sciences, Proceedings of the Thirty-First Hawaii International Conference on*, Volume: 7 , 6-9 Jan 1998 Page(s): 74 -83.
- [15] V. Park and M.S. Corson, "A Highly Adaptive Distributed Routing Algorithm for Mobile Wireless Networks," *Proc. IEEE INFOCOM '97 (April 1997)*,.pp. 1405-1413.
- [16] C. E. Perkins and P. Bhagwat "Highly dynamic destination sequenced distance vector routing (DSDV) for mobile computers" *SIGCOMM '94—Computer Communications Review*, 24(4):234–244, 1994.
- [17] N. Aschenbruck, E. G. Padilla, and P. Martini, "A survey on mobility models for performance analysis in tactical mobile networks", *Journal of Telecommunications and Information Technology*, vol.2, 2008, pp. 54-61.

A Proposed Real-Time Scheduling Algorithm for WiMAX Networks

Nada M. El-Shennawy

*Computers and Automatic Control Dept
Faculty of Engineering, Tanta University
Tanta, Egypt
Nada_elshennawy@f-eng.tanta.edu.eg*

Mostafa A. Youssef

*Computer Science and Engineering
E-JUST
Alexandria, Egypt
moustafa@cs.umd.edu*

Mohamed N. El-Derini

*Vice Rector for Education and Student Affairs
Pharos University
Alexandria, Egypt
nazih.elderini@pua.edu.eg*

Mahmoud M. Fahmy

*Computers and Automatic Control Dept.
Faculty of Engineering, Tanta University
Tanta, Egypt
mfn_288@hotmail.com*

Abstract- Over the past few years, there has been a rapid growth of new services offered to end users on the Internet, such as online video games, video conferences, and multimedia services. WiMAX networks are one prominent viable solution for wireless broadband access that provides last-mile access to the Internet. To satisfy the Quality of Service (QoS) requirements of the applications in an acceptable way, an efficient scheduling algorithm is needed. In the literature, attention was focused on throughput and delay only. Jitter, though of great significance, was not taken into account. In this paper, we consider jitter, in addition to throughput and delay, in order to formulate a three-term dynamic weight function. The jitter and delay terms, in particular, are weighted by specific weighting factors, whose values in real-time applications are different from those in non-real-time applications. Simulation results are obtained by OPNET, and it is shown that the proposed algorithm outperforms two famous previously published algorithms.

Keywords- *WiMAX; QoS; Jitter; Scheduling algorithms; IEEE 802.16*

I. INTRODUCTION

Worldwide interoperability for microwave access (WiMAX) networks are a broadband wireless access network technology, designed according to IEEE 802.16 standard [1, 2, 3]. These promising networks possess a multitude of advantageous features such as high data rate, large spanning area, and provision for achieving the required Quality of Service (QoS) of real-time applications. They act as a convenient medium for delivering vital services to end users on the Internet, such as video conferences, online video games, and multimedia services to end users.

The QoS plays a major role in determining network performance. It has three main parameters, namely,

throughput, delay, and jitter [4, 5]. A scheduling algorithm is needed to allocate the bandwidth to applications in such a way as to maximize throughput and minimize delay and jitter. The scheduling algorithm should be simple, fair, and efficient.

A good survey about scheduling algorithms in WiMAX networks is presented by So-In et al. [6]. Dhrona et al. [7] have made a comprehensive performance study of uplink scheduling algorithms in point-to-multipoint WiMAX networks, where simulation analysis was carried out using average delay, average throughput, fairness and frame utilization. Recently, Kumar and Gupta [8] perform another comparative descriptive analysis for various scheduling algorithms in WiMAX networks.

Among notable scheduling algorithms for WiMAX networks are [9, 10, 11]: Weighted Fair Queuing (WFQ), Random Early Detection (RED), Fair Queuing (FQ), Deficit Round Robin (DRR), Round Robin (RR), Weighted Round Robin (WRR), and First-In First-Out (FIFO). We have to choose the algorithm which guarantees the best performance. In this respect, an algorithm with dynamic bandwidth allocation is usually recommended. It is also mentioned in [12] that weighted scheduling algorithms are preferred for satisfaction of QoS requirements. The reason is that the weight corresponds to the number of time slots to be allocated to the service class. This number of slots is fixed for each WiMAX frame; hence the weight representing the number of slots is preferably to be an integer. This means that we do not actually need algorithms such as DRR [11, 12] in which floating point numbers are used. Further, the resulting algorithm will be much less sophisticated.

Other recent attempts have been made by Ali and Dimiyati [13] and El-Shinnawy et al. [14]. In [13], a scheduling algorithm has been developed on the basis of the

number of bandwidth requests from Non-Real-Time Polling Service (nrtPS). In [14], a priority scheduling algorithm based on jitter, minimum rate, delay, and class type has been introduced. It was the first time for jitter to appear in a scheduling algorithm, but the treatment was confined to priority algorithms.

Jitter [15] is defined as a measure of the variability over time of the packet latency across a network. It is a very important QoS factor in the assessment of network performance. The cause of jitter occurrence is that a packet can get queued or delayed somewhere in the network. Increasing the jitter value beyond a certain threshold leads to missing packets and serious audio problems in real-time applications.

In the present paper, we include jitter explicitly in a dynamic weight function. To our knowledge, the weight function so formulated implies a new concept. The performance attained is better than other algorithms, in terms of average throughput, average delay, and average jitter.

The rest of this paper is organized as follows. In Section II, an overview of WiMAX networks is given. Scheduling algorithms are reviewed in Section III. Section IV presents the details of the proposed approach. Simulation results are introduced in Section V. Finally, conclusions and trends for future work are reported in Section VI.

II. ON WiMAX NETWORKS

WiMAX networks have two basic operation modes [6, 7, 8]: point-to-multipoint (PMP) connection and mesh connection. In the PMP mode, the communications between all subscriber stations (SSs) are organized and passed through the base station (BS), while in the mesh mode, the communication can be achieved directly between subscriber stations. WiMAX provides five service classes to support the variation of QoS requirements for different applications [1, 6, 7]. The first class is an unsolicited grant service (UGS), which is used to support real-time applications with constant data rate such as VoIP (Voice over Internet Protocol) without silence suppression. The second class is an extended real-time polling service (ertPS) class, which is designed to support real-time applications with variable data rate such as VoIP with silence suppression. The real-time polling service (rtPS) is the third service class type. rtPS focuses on real-time applications with variable data rate such as a Moving Picture Experts Group (MPEG) compressed video. The fourth service class is a non-real-time polling service (nrtPS) class, which is designed for non-real-time variable bit rate traffic such as file transfer protocol (FTP). The fifth class is the best effort (BE) at, which applications do not make use of any specific QoS requirements.

WiMAX is based on the principles of orthogonal frequency division multiplexing (OFDM) [1], which is a suitable modulation access technique for non-line-of-sight (NLOS) conditions with high data rates. However, in WiMAX the various parameters pertain to the physical layer, such as the number of subcarriers, pilots, and guard band. The WiMAX physical (PHY) layers defined in IEEE 802.16

standard are: wireless MAN-SC (single carrier), wireless MAN SCa, wireless MAN-OFDM (orthogonal frequency division multiplexing) and wireless MAN-OFDMA (orthogonal frequency division multiple access). Details about these layers can be found in [1, 2]. OFDMA WiMAX frame is divided into two subframes: uplink subframe and downlink subframe separated by a Transmit-receive Transition Gap (TTG) and a Receive-transmit Transition Gap (RTG). The structure of the WiMAX frame is shown in Figure 1 [1, 2, 6].

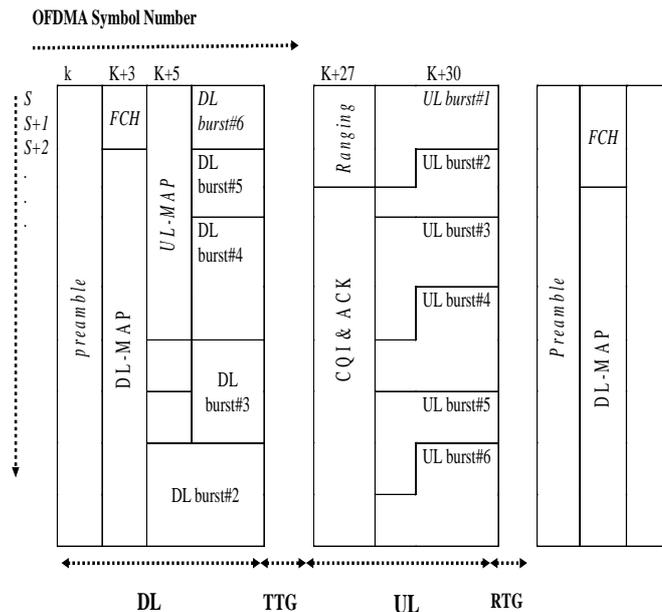


Figure 1. WiMAX Frame Structure

WiMAX provides a Media Access Control (MAC) layer that uses a grant request mechanism to authorize the exchange of data. Thus, a better exploitation of the radio resources, in particular with smart antennas, and independent management of the traffic of every user is allowed [16]. MAC layer in WiMAX has actions [1, 2, 6], i.e., provides QoS, responsible of security and key management, and provides power saving mode and idle mode operations. MAC layer is divided into three sub-layers. First, a convergence sub-layer is designed as a link between the higher layers and WiMAX MAC layer. This is done by mapping data from the upper layers to the appropriate MAC layer. Second, a common part sub-layer, which is responsible for bandwidth allocation, connection establishment and maintenance for all QoS requirements. Third, a security sub-layer, which is developed for authentication, security key exchange, and encryption. To ensure good performance of WiMAX networks for the different requirements of QoS in real-time applications, a suitable bandwidth allocation algorithm is needed. In the starting of each WiMAX frame, the scheduling algorithm computes the bandwidth allocation

for each subscriber station to send this information in UL-MAP.

III. EARLIER SCHEDULING ALGORITHMS

To meet the QoS requirements of multimedia applications, a scheduling algorithm is needed to allocate the bandwidth to users to satisfy upper bounds on delay and jitter and to maximize throughput. Scheduling algorithms can be classified into two categories [6]: channel-aware algorithms, and channel-unaware algorithms. In channel-aware algorithms, channel information, such as signal strength, signal-to-noise ratio, and received signal power, affects the bandwidth allocation decision. In channel-unaware algorithms, however, no channel information is used. Many aware schedulers are proposed in the literature: examples are modified largest weighted delay first (M-LWDF) [17] and Link Adaptive largest weighted throughput (LWT) [18].

Also, the channel-unaware schedulers are the subject of many research papers. In [6, 10], the weighted family RR algorithm is proposed. This algorithm assigns one allocation for each connection in each serving cycle. In [6, 7, 10] WRR assigns a weight value to each connection then serves connections according to their allocated bandwidth based on weight. The main problem of WRR is that when the traffic has a variable packet size, it provides incorrect percentage of bandwidth allocation. DRR [11, 16] solves the problem of WRR by using two variables for each queue, deficit counter (DC) and quantum (Q). Deficit weighted round robin (DWRR) [18] is the same as DRR but with a new weight variable for each queue and the Q value depends on the weight value. Another modification on DRR, named modified deficit round robin (MDRR) [20], is operated in the same way as DRR but with adding a new parameter called a queue priority.

The above-mentioned scheduling algorithms have the following drawbacks. First, in weighted scheduling algorithms, the bandwidths are assigned statically and do not vary with the burst changes. Second, no enough attention is given to jitter causing problems in real-time applications. Finally, priority scheduling algorithms caused starvation in low priority classes. According to [6] and to the best of our knowledge, no scheduling algorithms take jitter delay into account in weighting function when taking the bandwidth allocation decision.

IV. THE PROPOSED ALGORITHM

In WiMAX networks, the BS is responsible for the scheduling of service classes for uplink and downlink directions. The scheduling algorithm works on the bases of the bandwidth requests of SSs in the uplink direction. The proposed approach is used as an uplink scheduling algorithm in the MAC layer of BS. This approach is a type of weighted scheduling algorithms with a dynamic weight equation defined in terms of the parameters: throughput and delay as well as jitter. These parameters characterize the QoS of the application at hand.

For each type of applications, the importance of these parameters is varying. In real-time applications which belong to rtPS service class in WiMAX, the QoS parameters are all

important and none of them can be dispensed with. But, in non-real-time applications, which belong to nrtPS class in WiMAX, throughput is the only important parameter, since non-real-time applications are insensitive to delay and jitter.

The problem under consideration is concerned with the development of a real-time scheduling algorithm for WiMAX networks. The bandwidth is to be allocated among n queues; that is, n subscriber stations. The proposed method depends on the formulation of a dynamic weight function in terms of the three QoS parameters: throughput, delay, and jitter. To this end, a weight W_i is assigned to queue i as a positive factor of the form:

$$W_i = \frac{N_i}{\sum_{j=1}^n N_j}, \quad i=1,2,\dots,n \quad (1)$$

In (1), N_i is expressed as the sum of three terms corresponding to contributions of throughput, delay, and jitter, respectively. Specifically, we propose the following formula for a weight function N_i :

$$N_i = T_i + D_i + J_i, \quad i=1,2,\dots,n \quad (2)$$

The first term T_i , in (2), is the fractional throughput contribution to N_i , defined as:

$$T_i = \frac{X_i}{\sum_{j=1}^n X_j} \quad (3)$$

where X_i is the minimum reserved traffic rate for queue i . The second term D_i is the fractional delay contribution

$$D_i = \frac{(\alpha_i Y_i / L_i)}{\sum_{j=1}^n (\alpha_j Y_j / L_j)} \quad (4)$$

where Y_i is a time-varying average delay, L_i is the given maximum latency, and α_i is a positive delay weighting factor. In (4), the ratio Y_i/L_i (less than unity) expresses the proportion of the delay of a particular queue relative to the maximum acceptable delay of the network. Further, the ratio Y_i/L_i is weighted by a factor α_i , whose value varies according to the subscriber station (value of i). This is justifiable since each subscriber station is devoted to a particular application. The third term J_i is the fractional jitter contribution,

$$J_i = \frac{(\beta_i Z_i / K_i)}{\sum_{j=1}^n (\beta_j Z_j / K_j)} \quad (5)$$

where Z_i is a time-varying average jitter, K_i is the given maximum jitter and β_i is a positive jitter weighting factor. The terms in (5) can be interpreted in the same way as in (4).

Equation 2 is valid for both real- and non-real time applications; this implies that the weighting factors α_i and β_i should take on different values of the two types of applications. The values of α_i and β_i for real-time applications should be greater than those for non-real-time

applications. The reason is the fact that real-time applications are more highly sensitive to delay and jitter. There is no apparent way for systematically determining the value of α_i and β_i . Therefore, we resort to a trial-and-error method. The criterion for the choice of the values of α_i and β_i depends on the performance of the algorithm in WiMAX networks. We begin with arbitrary initial values of α_i and β_i , and estimate the network performance in terms of throughput, delay, and jitter. When the performance is not satisfactory, the values of α_i and β_i are changed in a prescribed random manner, until a satisfactory network response is arrived at. These final values of α_i and β_i are then fixed and made use of in the bandwidth allocation operation of the algorithm. The simulation results to follow demonstrate the idea.

The computational scheme of the proposed algorithm is summarized in the following consecutive steps:

- 1) Values for the delay weighting factor α_i and the jitter weighting factor β_i are selected.
- 2) For each queue, get the values of Y_i and Z_i .
- 3) Calculate the values of T_i (in (3)), D_i (in (4)), and J_i (in (5)).
- 4) Calculate the three-term weight function N_i according to (2).
- 5) Calculate the weight W_i by virtue of (1).
- 6) Divide the bandwidth of the uplink subframe among the n queues based on the relationship:

$$(BW)_i = W_i * (UL)_{BW}$$
 (6)
 where BW_i is the bandwidth reserved to queue i and UL_{BW} is the total bandwidth of the uplink subframe.
- 7) The value of the bandwidth of each queue is sent to SS.
- 8) The service for the queue is continued until the bandwidth is ended.
- 9) The service is moved between the queues using round robin mechanism.

V. SIMULATION SETTING AND EXPERIMENTS

Simulation in this paper is performed by the OPNET simulator [21]. The network used consists of four WiMAX service classes: ertPS, rtPS, nrtPS and BE with applications: VoIP, video conference, FTP and HTTP, respectively. The traffic parameters for each service class are listed in table I.

TABLE I. TRAFFIC PARAMETERS

Service class	Minimum reserved traffic rate in bps	Maximum sustained traffic rate in bps	Maximum latency in msec	Maximum jitter in msec
ertPS	25000	64000	20	150
rtPS	64000	500000	30	160
nrtPS	45000	500000	100	300
BE	1000	64000	N/A	N/A

The simulation results are obtained using several scenarios by varying the number of SSs. Each scenario consists of one BS serving a number of SSs in PMP mode of operation. The frame duration is 5 msec, with 50% for each uplink and downlink subframe. A random topology in 1000 x 1000 m square space is used. The number of SSs varies from 10 to 60 with ratio 2:3:3:2 SSs for service classes ERTPS:RTPS:NRTPS:BE, respectively. The proposed weighted scheduling algorithm is compared with both MDRR [20] and WRR [6, 7, 10]. The throughput, delay, and jitter are considered as performance metrics. The simulation time is 10 minutes

Figures 2, 3, and 4 show the simulation results for the average throughput, average delay, and average jitter, respectively, as functions of the number of SSs, for the proposed algorithm together with the other two algorithms WRR and MDRR, it is clear from these figures that the proposed approach exhibits a better performance than WRR and MDRR since it has:

- A higher throughput
- A lower delay
- A lower jitter

In Figure 2, it is to be noted that the differences between the average throughput values in the three scheduling algorithms are not appreciable, because the throughput in the three algorithms is defined using the same concept.

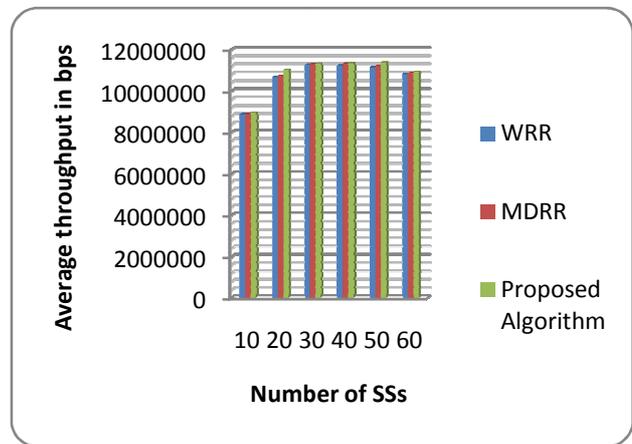


Figure 2. WiMAX average throughput vs. number of SSs

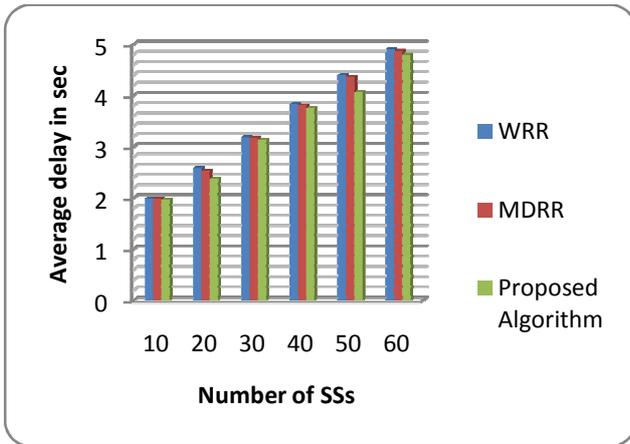


Figure 3. WiMAX average delay vs. number of SSs

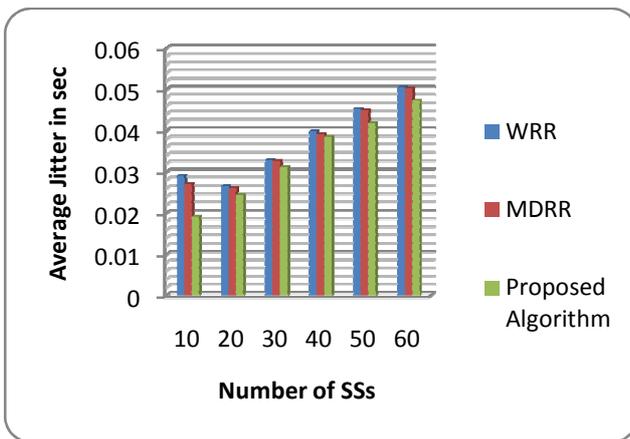


Figure 4. WiMAX average jitter vs. number of SSs

The relative superiority of the proposed algorithm can be attributed to the inclusion of jitter, in addition to throughput and delay, in the weight function N_i (equation 2). Besides, the values of the weighting factors α_i and β_i are chosen in such away that both delay and jitter are given greater attention in real-time applications than in non-real-time applications. In this specific application, it is found that the best possible values of α_i and β_i are in the ratio 1:5 and 1:3, respectively, in non-real-time and real-time applications.

VI. CONCLUSION AND FUTURE WORK

An uplink dynamic channel-unaware weighted scheduling algorithm for WiMAX networks has been proposed. It has advantages of conceptual soundness and computational simplicity. A weight function, for each queue,

is formulated as the sum of the respective contributions of the QoS parameters: throughput, delay, and jitter. The inclusion of jitter is the essential modification that results in a more comprehensive weight function. Weighting factors α_i and β_i are introduced in the expressions of delay and jitter contributions, respectively. The values of these factors in the real-time applications should be different from those in non-real-time applications. There is no obvious strategy to choose the values of α_i and β_i ; therefore, a trial-and-error technique is resorted to. The proposed algorithm is simulated using OPNET. Comparison is made with two powerful algorithms: WRR and MDRR. The results demonstrate that our algorithm outperforms the other two, with respect to throughput, delay, and jitter, as functions of the number of subscriber stations.

In a future research work, a systematic method for the choice of α_i and β_i will be devised. In addition, the algorithm will be extended to the channel-aware case. An interesting challenging task will be focused on the application of the algorithm to the newly established Long-Term Evolution (LTE) networks [22].

REFERENCES

- [1] J. G. Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of WiMAX: understanding broadband wireless networking*, Pearson Education Inc, Feb 2007.
- [2] IEEE 802.16e, *IEEE Standard for local and metropolitan area networks, Air Interface for Fixed Broadband Wireless Access Systems, Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and Corrigendum 1*, Feb 2006.
- [3] C. Cicconetti, A. Erta, L. Lenzini, and E. Mingozzi, "Performance evaluation of the IEEE 802.16 MAC for QoS support," *IEEE Trans. Mob. Comput.*, 6, (1), 2007, pp. 26–38.
- [4] R. A. Talwalkar and M. Ilyas, "Analysis of Quality of Service (QoS) in WiMAX Networks," *16th IEEE International Conference on Networks, ICON 2008*, pp. 1-8.
- [5] A. Iera, A. Molinaro, and S. Pizzi, "Channel-Aware Scheduling for QoS and Fairness Provisioning in IEEE 802.16/WiMAX Broadband Wireless Access Systems," *IEEE Network* 21, vol. 5, 2007, pp. 34-41.
- [6] C. So-In, R. Jain, and A.K. Tamimi, "Scheduling in IEEE 802.16e Mobile WiMAX Networks: Key Issues and Survey," *IEEE Journal on Special Areas in Communications (JSAC)*, vol. 27, no. 2, 2009.
- [7] P. Dhrona, N. Abu Ali, and H. Hassanein, "A performance study of scheduling algorithms in Point-to-Multipoint WiMAX networks," *IEEE Conference on Local Computer Networks LCN*, 2008, pp. 843-850.
- [8] B. Kumar, and P. Gupta, "Scheduling Algorithms in WiMAX Network," *Second International Conference on Advanced Computing and Communication Technologies*, 2012, pp. 457-462.
- [9] L. Jin-Cherng, C. Chun-Lun, and L. Cheng-Hsiung, "Performance Evaluation for Scheduling Algorithms in WiMAX Network," *International Conference on Computer Technology and Development, ICCTD '09*, 2009, pp. 68-74.

- [10] A. Al-Howaide, A. Doulat, and Y. Khamayseh, "Performance Evaluation of Different Scheduling Algorithms in WiMAX," *International Journal of Computer Science, Engineering and Applications (IJCEA)*, vol. 1, no. 5, 2011, pp. 81-94
- [11] H. Guesmi, S. Maaloul, and R. Tourki, "Design of Scheduling Algorithm for QoS Management on WiMAX Networks," *Computer Science and Engineering*, vol. 2, 2011, pp. 43-50.
- [12] A. Sayenko, O. Alenan, J. Karhula, and T. Hamalainen " Ensuring the QoS requirements in 802.16 scheduling," *Proceedings of the 9th ACM international symposium on Modeling analysis and simulation of wireless and mobile systems*, 2006, pp. 108 - 117.
- [13] D. Ali and K. Dimiyati, "Threshold based Cyclic Polling (TbCP): An Uplink Scheduling Algorithm for Mobile WiMAX Systems," *International Journal of Information and Electronics Engineering*, vol. 1, no. 1, 2011, pp. 1-8.
- [14] A. El-Shinnawy, M. Elgazzar, and A. Nassar, "A Multi-Factor Scheduling Algorithm for WiMAX Systems," *International Journal of Information and Electronics Engineering*, vol. 2, no. 3, 2012, pp. 347-351.
- [15] C. Demichelis and P. Chimento, "IP Packet Delay Variation Metric for IP Performance Metrics (IPPM)," RFC 3393, November 2002.
- [16] S. Omerovic, "WiMAX Overview," Technical report, University of Ljubljana, Slovenia, 2006.
- [17] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *IEEE Commun. Mag.* Vol. 39, 2001, pp. 150-154.
- [18] Y. Zhang and S. Liew, "Link-adaptive largest-Weighted-throughput packet scheduling for real-time traffics in wireless OFDM networks," in *Proceedings of the IEEE Global Telecommunications Conf., St. Louis, MO*, vol. 5, 2005, pp. 5-9.
- [19] J. Lakkakorpi, A. Sayenko, and J. Moilanen, "Comparison of Different Scheduling Algorithms for WiMAX Base Station: Deficit Round-Robin vs. Proportional Fair vs. Weighted Deficit Round-Robin," *IEEE Wireless Communications and Networking Conference, WCNC*, 2008, pp. 1991-1996.
- [20] H. Kaur and G. Singh, "Implementation and Evaluation of Scheduling Algorithms in Point-to-Multipoint Mode in WiMAX Networks," *International Journal of Computer Science and Technology (IJCST)*, vol. 2, 2011, pp. 540-546.
- [21] OPNET Technologies, Opnet modeler Software Package available: www.opnet.com.
- [22] S. Sesia, I. Toufik, and M. Baker, *LTE - The UMTS Long Term Evolution: From Theory to Practice*, John Wiley & Sons Ltd., 2011.

On the Performance of Selective Decode and Forward Relaying over Imperfectly Known Fading Relay Channels

Ali Zarei Ghanavati and Daniel C. Lee
 School of Engineering Science
 Simon Fraser University
 Burnaby, Canada
 Email: azareigh@sfu.ca, dchlee@sfu.ca

Abstract—In this paper, we investigate the effects of imperfect channel estimation on the performance of a Selective-Decode-and-Forward (SDF) relay-assisted communication system. In the system studied, a pilot symbol-assisted modulation (PSAM) scheme is used along with a channel estimation scheme based on minimum mean square error (MMSE). In particular, we derive an approximate expression for the bit error probability (BEP) in the presence of channel estimation error for the SDF cooperative protocol. Numerical simulations are presented to show that the derived approximate BEP expression is very close to the actual BEP. We also provide power-allocation that optimally assigns power constraint to training and data transmission phases to minimize the BEP.

Keywords—cooperative communication; fading; channel estimation

I. INTRODUCTION

Recent advances in information and wireless technologies have led to growing interests in the development of multiple-input and multiple-output (MIMO) systems, which improves spectral and power efficiency of wireless networks. However, installing multiple antennas at transmitter and/or receiver results in size expansion of the devices, which is not practical in many wireless applications. A cooperative communication system can be considered as an alternative to MIMO systems. Cooperative techniques can exploit cooperative diversity by means of providing several copies of a signal that have experienced channel gains with low correlations [1][2][3]. Furthermore, early research in cooperative communication has shown that energy efficient transmitters operating in relay networks help extend the battery life [4].

According to the way in which the information is transmitted from the source terminal to the relay terminals and the way it is processed at the relay terminals, the existing cooperative protocols can usually be divided into three types – such as decode-and-forward (DF) protocols, selective decode-and-forward (SDF) protocols and amplify-and-forward (AF) protocols [1][2][3][5].

In the DF protocol, the relay terminals decode the received signal, then the sequence will be re-encoded and it will be forwarded to the destination terminal. In the SDF protocol, the relay terminals detect the signal and check if the detected

signal contains any error. If the information sequence is considered error free, then the sequence will be re-encoded by either the same or a different code before it will be finally transmitted to the destination terminal. In AF protocols, the relay terminals simply re-transmit a scaled version of the signal that they receive from the source terminal to the destination terminal.

Most of the existing work on cooperative communications has assumed perfect knowledge of the channel fading coefficients at the receiver side, which is an overly optimistic assumption, and does not match with reality. For instance, W. Su et al. [6] and A. S. Ibrahim et al. [7] assume that the receiver has access to perfect information about the channel and the transmitter knows the instantaneous channel gain without the phase component. Various expressions for the symbol error probability of a cooperative communication system have been derived in [8][9][10][11] under the assumption of perfect channel state information (CSI) at both the relay and destination. Although research results based on these assumptions provide valuable insights, in practical systems these coefficients must be estimated and then used in the detection process. Especially, in mobile applications, the assumption of perfect channel knowledge is unwarranted as randomly varying channel conditions are learned by the receivers imperfectly.

In recent work (e.g., [12][13]), the effects of the channel estimation error on the bit error rate (BER) performance of cooperative communication systems have been studied by using a simple model for the channel estimation error, where the variance of the channel estimation error is assumed to be fixed for all values of the signal-to-noise ratio (SNR). Y. Wu et al. [14] investigate effects of channel estimation errors on the symbol-error-rate (SER) performance of a cooperative communication system operating in AF mode. J. Zhang et al. [15] assume that the receiver estimates the channel imperfectly, based on the pilot signal sent by the transmitter. Moreover, both the source and the relay perform an optimization on the power to be allocated to the pilot and data, assuming that achievable rate is the only factor to be optimized.

To the best of our knowledge, no previous work exists

on the BEP performance of selective decode-and-forward relay communication system that uses pilot symbol-assisted modulation for channel estimation.

In the present paper, we derive a lower bound for BEP performance of binary phase shift keying (BPSK) modulation for the selective decode-and-forward relay communication system. We then apply the lower bound expression of BEP as the performance metric of the system, and solve a power allocation problem to optimally allocate power to training and data sequences.

The rest of the paper is organized as follows. We first introduce our system model and channel estimation in Section II. In Section III, the BEP expression is derived for the system under consideration. Simulation results are presented in Section IV, followed by conclusion and future work in Section V.

Notation: $E\{\cdot\}$, $\{\cdot\}^*$, \hat{x} denote the expected value, the complex conjugate and the estimate of the variable x , respectively.

II. SYSTEM MODEL AND CHANNEL ESTIMATION

We consider a three-node relay network, which consists of a source, relay, and destination node. This relay network model is depicted in Figure 1. We assume BPSK transmission over flat fading channels. Note that this assumption is imposed only for the convenience of notation. The analysis in this paper can also be applied, after some simple modifications, to the case wherein the source and the relay transmit using M-ary phase shift keying (MPSK) modulation. Let h_{sd} , h_{sr} and h_{rd} represent the source-destination, source-relay, and relay-destination channel gains, respectively. Each node-to-node channel gain is modelled by a zero-mean complex Gaussian random variable with variance σ_h^2 . Furthermore, each channel is assumed to be constant during the frame transmission where each frame consists of a fixed number of symbols. A practical relay node with low cost usually cannot transmit and receive signal at the same time in the same frequency band. We assume that nodes transmit under half-duplex constraint in the same frequency band.

As in [1], the transmission protocol can be described as follows: First, the source transmits data to both relay and destination with power P_s . The received signals at the relay and the destination can be written as:

$$y_{sr} = \sqrt{P_s}h_{sr}x + n_{sr} \quad (1)$$

$$y_{sd} = \sqrt{P_s}h_{sd}x + n_{sd} \quad (2)$$

where n_{sr} and n_{sd} represent the additive noise terms, x represents the transmitted symbol with unit average energy, i.e., $E\{|x|^2\} = 1$ and P_s is the source power (energy per symbol time) for data transmission. In the second time slot, if the relay correctly detects the received message in the first time slot, then it will forward it to the destination; otherwise, the relay will remain silent to avoid the propagation of errors [7]. Thus, the received signal at the destination in the second

time slot is in the following form

$$y_{rd} = \theta\sqrt{P_r}h_{rd}x + n_{rd} \quad (3)$$

where θ can be either 0 or 1 indicating whether the relay was silent or not, n_{rd} is additive noise term, and P_r is the relay power (energy per symbol time) for data transmission. At the end of the second time slot, the destination will combine the desired signals from the source and the relay, if any, and attempt to detect the symbol.

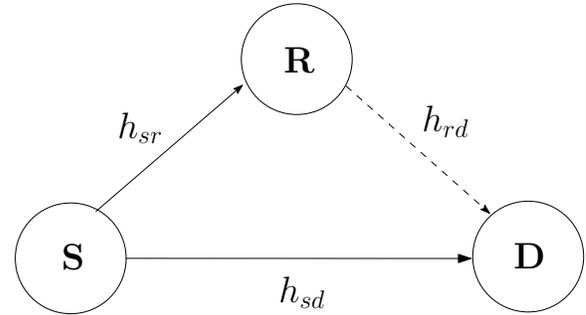


Fig. 1. Three-node relay network model

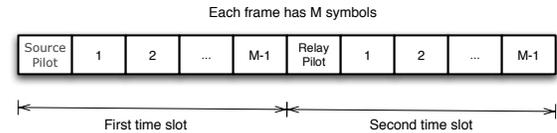


Fig. 2. Transmission structure in a block of M symbols

The transmission block consists of two phases—training phase and data transmission phase. We assume that the communication between nodes is through single-input-single-output (SISO) channel. For channel gain estimation of a SISO system, only one pilot symbol is required [15][16]. To reduce the variance of channel estimation errors, one can either increase the number of pilot symbols or use a higher pilot transmit power. As in [15], we assume that only one pilot symbol is used to estimate the channel coefficient. We point out that since block fading channels are considered, the accuracy of MMSE channel estimation depends only on the total training signal power regardless of the training duration, and increasing the number of training symbols in each block results in decreasing the throughput. We assume that, in the considered scheme, the source and relay can allocate power to pilot phase and data phase in different proportions. At the beginning of the frame, the source sends a pilot symbol, which we denote as s_p , to the relay and destination. The received signals, r_{sr}^p and r_{sd}^p can be expressed as follows

$$r_{sr}^p = h_{sr}s_p + n_{sr}$$

$$r_{sd}^p = h_{sd}s_p + n_{sd} \quad (4)$$

where r_{sr}^p and r_{sd}^p are the received pilot signals at the relay and at the destination, respectively. Then, the relay sends pilot symbol s'_p to the destination. The received pilot signal at the destination is as follows

$$r_{rd}^p = h_{rd}s'_p + n_{rd} \quad (5)$$

The noise terms n_{sd} , n_{sr} , and n_{rd} are modelled as zero mean complex Gaussian random variables with equal variance N_0 ($N_0/2$ per real dimension).

The system conveys source information as a sequence of frames, and each frame in the first time slot accommodates M symbols sent from the source—one symbol is dedicated for pilot and the rest are for data transmission (Figure 2). The second time slot of the frame accommodates M symbols sent from the relay—one symbol is dedicated for pilot and the rest are for data transmission (Figure 2). The energy allocated to a frame for the source's transmission is denoted by P_1 , and the energy allocated to a frame for the relay's transmission is denoted by P_2 (that is, if the relay ends up transmitting in that frame). Parameter α denotes the fraction of source's energy allocated for transmission during the training phase, and β denotes the fraction of relay's energy allocated for transmission during the training phase. Thus, $|s_p|^2$ and $|s'_p|^2$, the transmit energies for the training phase are equal to αP_1 (if transmitted by the source) and βP_2 (if transmitted by the relay), respectively, where $0 < \alpha, \beta < 1$. Source power (energy per symbol time) for data transmission (P_s) and the relay power for data transmission (P_r) can be obtained as:

$$P_s = \frac{(1-\alpha)P_1}{M-1}, \quad P_r = \frac{(1-\beta)P_2}{M-1} \quad (6)$$

The relayed communication system considered in the present paper assumes that the relay and the destination both estimate the wireless channel gain from the received pilot signal using MMSE channel estimation method.

MMSE estimate of the channel is obtained at the relay and destination by using $\hat{h}_{sr} = \mathbb{E}\{h_{sr}r_{sr}^{p*}\} \cdot (\mathbb{E}\{r_{sr}^p r_{sr}^{p*}\})^{-1} \cdot r_{sr}^p$, $\hat{h}_{sd} = \mathbb{E}\{h_{sd}r_{sd}^{p*}\} \cdot (\mathbb{E}\{r_{sd}^p r_{sd}^{p*}\})^{-1} \cdot r_{sd}^p$ and $\hat{h}_{rd} = \mathbb{E}\{h_{rd}r_{rd}^{p*}\} \cdot (\mathbb{E}\{r_{rd}^p r_{rd}^{p*}\})^{-1} \cdot r_{rd}^p$ [15]. Based on [17], we can write

$$h_{ij} = \hat{h}_{ij} + e_{ij} \quad (7)$$

where e_{ij} is the channel estimation error modelled as a zero mean complex Gaussian random variable with variance $\sigma_{e_{ij}}^2$ and we have [17]:

$$\begin{aligned} \hat{h}_{ij} &\sim \mathcal{CN}\left(0, \frac{\sigma_h^4 |s_p|^2}{\sigma_h^2 |s_p|^2 + N_0}\right) \\ e_{ij} &\sim \mathcal{CN}\left(0, \frac{\sigma_h^2 N_0}{\sigma_h^2 |s_p|^2 + N_0}\right) \end{aligned} \quad (8)$$

where $|s_p|^2$ is the power allocated to pilot symbol, which is equal to αP_1 or βP_2 , depending on whether the source or relay is transmitting and $\mathcal{CN}(\cdot, \cdot)$ denotes complex Gaussian distribution. It can be easily shown that e_{ij} and \hat{h}_{ij} are

uncorrelated random variables and since they have jointly Gaussian distribution, they are also independent.

In the next section, we will use the analysis presented so far to derive the BEP expression.

III. BEP ANALYSIS AND OPTIMAL POWER ALLOCATION

In SDF protocols, the relay forwards a symbol only if the relay has high confidence that the symbol has been decoded correctly, and remains silent otherwise. It is reasonable to add a module for such decision making. From the destination node's vantage point, that decision made by the relay can only be guessed. Traditionally, three methods can be used to detect the presence of a signal [18]: energy detector, matched filter, and cyclostationary feature detection. Energy detection, which is the most popular method, is suboptimal and non-coherent and can be simply implemented [19]. Matched filter is a coherent detection that maximizes the signal to noise ratio. Cyclostationary feature detection exploits the inherent periodicity of the received signal. Besides all the methods mentioned above for detecting the presence of the signal, flag-based methods are also of interest [20]. However, sending a flag signal has the drawback of consuming additional energy and bandwidth. In this section, we assume that the destination knows θ in (3) perfectly. In other words, at each time slot, the destination knows whether the relay is in transmission mode or not. Then, we derive the optimal signal detection rule and a closed-form BEP expression for the system that uses MMSE channel estimator. Using (7), the received signals at the destination during two time slots can be written as

$$\begin{aligned} y_{sd}^i &= \sqrt{P_s} \hat{h}_{sd} x_i + \sqrt{P_s} e_{sd} x_i + n_{sd}^i \\ y_{rd}^i &= \theta_i \sqrt{P_r} \hat{h}_{rd} x_i + \theta_i \sqrt{P_r} e_{rd} x_i + n_{rd}^i \end{aligned} \quad (9)$$

where $i = 1, 2, \dots, M-1$ and $\theta_i \in \{0, 1\}$. We consider a method in which symbols are detected individually (i.e., in a symbol-by-symbol fashion), which can be employed by a system that does not have sufficient memory space to store and consider the preceding symbols. With θ known by the destination, optimal detection rule to be used by the destination can be written as

$$\hat{x} = \arg \max_{x \in \{-1, +1\}} p(y_{sd}, y_{rd} | x, \hat{h}_{sd}, \hat{h}_{rd}, \theta) \quad (10)$$

By using (9), conditioned on the estimated channel gains (\hat{h}_{sd} and \hat{h}_{rd}), transmitted symbol (x) and θ , it can be easily seen that y_{sd} and y_{rd} are two independent complex Gaussian random variables with means $\sqrt{P_s} \hat{h}_{sd} x$, $\theta \sqrt{P_r} \hat{h}_{rd} x$ and variances $P_s \sigma_{e_{sd}}^2 + N_0$, $\theta P_r \sigma_{e_{rd}}^2 + N_0$, respectively. The decision rule becomes

$$\begin{aligned} \hat{x} = \arg \max_{x \in \{-1, +1\}} &\left(\frac{1}{\pi(P_s \sigma_{e_{sd}}^2 + N_0)} e^{-\frac{|y_{sd} - \sqrt{P_s} \hat{h}_{sd} x|^2}{P_s \sigma_{e_{sd}}^2 + N_0}} \right. \\ &\left. \times \frac{1}{\pi(\theta P_r \sigma_{e_{rd}}^2 + N_0)} e^{-\frac{|y_{rd} - \theta \sqrt{P_r} \hat{h}_{rd} x|^2}{\theta P_r \sigma_{e_{rd}}^2 + N_0}} \right) \end{aligned} \quad (11)$$

The decision rule is simply the minimum-distance decision rule, where

$$\begin{aligned} \hat{x} &= \arg \min_{x \in \{-1, +1\}} \left(\frac{|y_{sd} - \sqrt{P_s} \hat{h}_{sd} x|^2}{P_s \sigma_{e_{sd}}^2 + N_0} \right. \\ &\quad \left. + \frac{|y_{rd} - \theta \sqrt{P_r} \hat{h}_{rd} x|^2}{\theta P_r \sigma_{e_{rd}}^2 + N_0} \right) \\ &= \arg \max_{x \in \{-1, +1\}} \operatorname{Re} \left\{ \left(\frac{\sqrt{P_s}}{P_s \sigma_{e_{sd}}^2 + N_0} \hat{h}_{sd}^* y_{sd} \right. \right. \\ &\quad \left. \left. + \frac{\theta \sqrt{P_r}}{P_r \sigma_{e_{rd}}^2 + N_0} \hat{h}_{rd}^* y_{rd} \right) x \right\} \end{aligned} \quad (12)$$

In order to implement this decision rule, we can use maximal ratio combining (MRC) that treats the estimated channels as true channels [21]

$$y^{\text{MRC}} = \frac{\sqrt{P_s}}{P_s \sigma_{e_{sd}}^2 + N_0} \hat{h}_{sd}^* y_{sd} + \frac{\theta \sqrt{P_r}}{P_r \sigma_{e_{rd}}^2 + N_0} \hat{h}_{rd}^* y_{rd} \quad (13)$$

Conditioned on the channel gains, by substituting (9) in (13) and after some simple manipulations, the SNR at the receiver can be written as follows [24]

$$\gamma^{\text{MRC}} = \frac{P_s}{P_s \sigma_{e_{sd}}^2 + N_0} |\hat{h}_{sd}|^2 + \frac{\theta P_r}{P_r \sigma_{e_{rd}}^2 + N_0} |\hat{h}_{rd}|^2 \quad (14)$$

We now examine a closed form expression of BEP. Conditioned on \hat{h}_{sd} , \hat{h}_{rd} , and θ , the BEP can be written as [24]

$$P(e|\hat{h}_{sd}, \hat{h}_{rd}, \theta) = Q \left(\sqrt{2 \left(A |\hat{h}_{sd}|^2 + \theta B |\hat{h}_{rd}|^2 \right)} \right) \quad (15)$$

where

$$A = \frac{P_s}{P_s \sigma_{e_{sd}}^2 + N_0}, \quad B = \frac{P_r}{P_r \sigma_{e_{rd}}^2 + N_0} \quad (16)$$

We denote $|\hat{h}_{sd}|^2$ and $|\hat{h}_{rd}|^2$ by X and Y respectively. In accordance with the complex Gaussian channel model, X and Y are exponentially distributed; we can therefore write

$$\begin{aligned} f_X(x) &= \lambda_{sd} e^{-\lambda_{sd} x}, \quad x \geq 0, \quad \lambda_{sd} = \frac{1}{\sigma_{\hat{h}_{sd}}^2} \\ f_Y(y) &= \lambda_{rd} e^{-\lambda_{rd} y}, \quad y \geq 0, \quad \lambda_{rd} = \frac{1}{\sigma_{\hat{h}_{rd}}^2} \end{aligned} \quad (17)$$

where

$$\lambda_{sd} = \frac{\alpha P_1 \sigma_h^2 + N_0}{\alpha P_1 \sigma_h^4}, \quad \lambda_{rd} = \frac{\beta P_2 \sigma_h^2 + N_0}{\beta P_2 \sigma_h^4} \quad (18)$$

Conditioned on θ and using (15), (17), and $Q(x) = \frac{1}{\pi} \int_0^{\frac{\pi}{2}} e^{-\frac{x^2}{2 \sin^2 \varphi}} d\varphi$ [24], BEP can be written as

$$\begin{aligned} P(e|\theta) &= \int_0^\infty P(e|x, y, \theta) f_X(x) f_Y(y) dx dy \\ &= \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \int_0^\infty \int_0^\infty \exp \left(-\frac{Ax + \theta By}{\sin^2 \varphi} \right) f_X(x) f_Y(y) dx dy d\varphi \end{aligned} \quad (19)$$

Conditional probability of error can be rewritten as

$$\begin{aligned} P(e|\theta) &= \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \int_0^\infty \lambda_{sd} e^{-\left(\frac{A}{\sin^2 \varphi} + \lambda_{sd}\right)x} dx \\ &\quad \times \int_0^\infty \lambda_{rd} e^{-\left(\frac{B\theta}{\sin^2 \varphi} + \lambda_{rd}\right)y} dy d\varphi \\ &= \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \frac{\lambda_{sd} \sin^2 \varphi}{\lambda_{sd} \sin^2 \varphi + A} \cdot \frac{\lambda_{rd} \sin^2 \varphi}{\lambda_{rd} \sin^2 \varphi + B\theta} d\varphi \end{aligned} \quad (20)$$

After some simple manipulation, we can write

$$\begin{aligned} P(e|\theta) &= \frac{1}{2} + \frac{1}{\pi} \frac{1}{\frac{B\theta}{\lambda_{rd}} - \frac{A}{\lambda_{sd}}} \left[\int_0^{\frac{\pi}{2}} \frac{\left(\frac{A}{\lambda_{sd}}\right)^2}{\sin^2 \varphi + \frac{A}{\lambda_{sd}}} d\varphi \right. \\ &\quad \left. - \int_0^{\frac{\pi}{2}} \frac{\left(\frac{B\theta}{\lambda_{rd}}\right)^2}{\sin^2 \varphi + \frac{B\theta}{\lambda_{rd}}} d\varphi \right] \end{aligned} \quad (21)$$

By using the result from [25, p. 177, 2.562.1], which is

$$\int \frac{dx}{a + b \sin^2 x} = \frac{\operatorname{sign} a}{\sqrt{a(a+b)}} \tan^{-1} \left(\sqrt{\frac{a+b}{a}} \tan x \right),$$

and

$$P(e) = P(e|\theta = 0)P(\theta = 0) + P(e|\theta = 1)P(\theta = 1) \quad (22)$$

and after some manipulations, the probability of error can be expressed as in (23) at the top of the next page. We point out that the probability that the relay detects the symbol with error (i.e., $P(\theta = 0)$) can be written as $P(\theta = 0) = \frac{1}{2} \left(1 - \sqrt{\frac{A}{A + \lambda_{sd}}} \right)$.

Remark: In the derivation of (23), it is assumed that the relay is perfectly capable of diagnosing whether there has been an error in the detection of the symbol transmitted by the source. Based on this diagnosis, if the detection was erroneous the relay would not forward that symbol to the destination in the second time slot. In a practical system, error detection scheme (e.g., cyclic redundancy check) could be implemented. In the following, we prove (23) is a lower bound of the BEP for a more practical system in which the relay uses a block error detection scheme:

I) The case in which the relay mistakenly concludes that there is a symbol that is detected wrong in the frame (while all the symbols were detected correctly at the relay): In this case, at the second time slot, the relay remains silent and the whole frame is kept from being forwarded. In the practical system, the final symbol detection at the destination is made only based on the received signal at the first time slot, which results in a higher BEP than the ideal system of Section II.

II) The case in which the relay's block error detection mistakenly concludes that all the detected symbols in the frame are error-free (while there is a symbol that is detected with error at the relay): In the practical system, the relay forwards the frame which includes some symbol different from the source's and the frame may contain some corrected decoded symbols. In the ideal system of Section II, which leads to (23), only correctly decoded symbols will be forwarded. For the

$$P(e) = \frac{1}{4} \left[1 + \sqrt{\frac{A}{A + \lambda_{sd}}} \right] \left[1 + \frac{\lambda_{sd}\lambda_{rd}}{B\lambda_{sd} - A\lambda_{rd}} \left(\frac{A}{\lambda_{sd}} \sqrt{\frac{A}{A + \lambda_{sd}}} - \frac{B}{\lambda_{rd}} \sqrt{\frac{B}{B + \lambda_{rd}}} \right) \right] + \frac{1}{4} \left[1 - \sqrt{\frac{A}{A + \lambda_{sd}}} \right]^2 \quad (23)$$

correctly decoded symbols, the destination's symbol detection, which combines the signal from the source and the signal from the relay, will result in the same probability of error in both the practical system and the ideal system. For a symbol that is detected by the relay with error, in the practical system the destination will end up combining the signal from the source and a signal carrying a different symbol from the relay, while in the ideal system of section II the destination will detect it based only on the signal from the source. Therefore, the ideal system will have less likelihood of making a symbol (bit) error.

III) In the case in which the relay correctly concludes that the frame contains a symbol that is decoded with error at the relay: In this case, in the system the whole frame is kept from being forwarded. In the ideal system, only the corrected decoded symbols will be forwarded, and for these symbols the ideal system will combine both source's and relay's signals. Therefore, the ideal system has a lower chance of making error.

IV) The case in which the relay correctly concludes that the frame contains no symbol error: In both the practical system and the ideal system of Section II, the whole frame is forwarded to the destination. In this case, both systems will have the same detection mechanism of individual symbols in the destination. Finally, from I, II, III, and IV, it can be easily concluded that (23) is a lower bound for a system that uses block error detection at the relay.

In order to optimally allocate power to training and data transmission phases so that (23) is minimized, we formulate an optimization problem as

$$(\alpha_{opt}, \beta_{opt}, r_{opt}) = \arg \min_{\alpha, \beta, r} P_e(\alpha, \beta, r)$$

subject to : $0 \leq r \leq 1$, $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$ (24)

where $r = P_1/P$ is the ratio of source power constraint to total power constraint (i.e., $P_1 = r \cdot P$ and $P_2 = (1 - r) \cdot P$).

IV. SIMULATION RESULTS

In this section, Monte-Carlo simulation results are presented and compared with (23). Matlab was used for Monte-Carlo simulation, and 10^8 transmitted symbols were drawn from the BPSK constellation in order to estimate the BEP. The node-to-node channels are modelled by zero mean independent Gaussian random variables with unit variance.

Figure 3 shows the BEP analysis proposed in this paper in comparison with the simulation results for $M=4, 8, 16$ and 64 where M is the frame length. Parameters α, β and $r = P_1/P$ are set to $0.30, 0.30$ and 0.61 respectively. As the figure shows, all the simulation curves and analytical expression are very close. Note that the energy allocated to each data symbol can be obtained from (6). In order

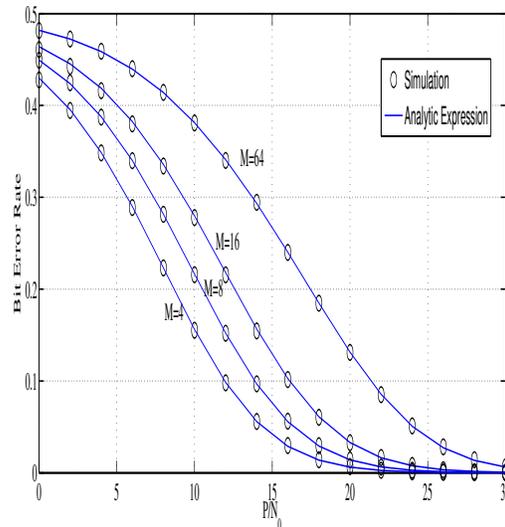


Fig. 3. BEP per P/N_0 for $\alpha=\beta=0.3$, $r = P_1/P = 0.61$ and different frame lengths

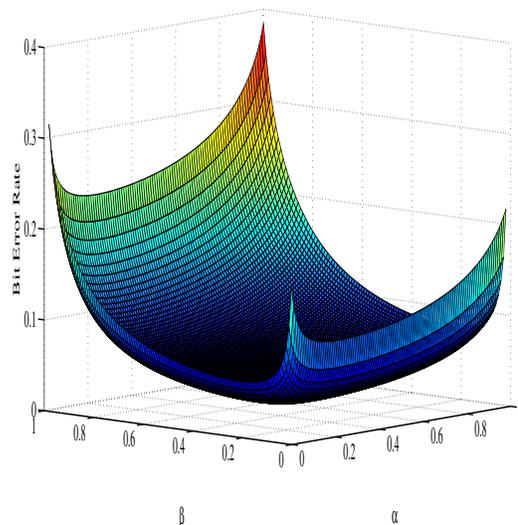


Fig. 4. BEP per α and β , $P/N_0=15$ dB, $P_1/P = 0.61$ and $M=4$

to numerically compare the BEP curve obtained by simulation and the BEP curve obtained by analytical expression, for each frame length (M) we define the ratio d_M as $d_M = (\text{BEP}_{\text{simulation}} - \text{BEP}_{\text{analytical expression}}) / \text{BEP}_{\text{simulation}}$. For $P/N_0=15$ dB, $\alpha=\beta=0.30$ and $r = P_1/P=0.61$, we have $d_{64}=0.0005$, $d_{16}=0.0011$, $d_8=0.0011$, and $d_4=0.0010$. Also

TABLE I
RESULTS OF OPTIMIZATION FOR M= 4, 8, 16 AND 64 UNDER $P/N_0 = 15$ dB and 25 dB

		M= 4	M=8	M=16	M= 64
$P/N_0 = 15$ dB	r	0.61	0.61	0.71	1
	α	0.35	0.29	0.24	0.18
	β	0.35	0.29	0.26	-
$P/N_0 = 25$ dB	r	0.61	0.61	0.61	0.61
	α	0.34	0.24	0.19	0.13
	β	0.34	0.20	0.16	0.13

for $P/N_0=30$ dB we have: $d_{64}=0.0010$, $d_{16}=0.0024$, $d_8=0.0075$, and $d_4=0.0053$. These results indicate that there is little difference between the values of BEP obtained from simulation and analytical expression.

In Figure 4, the BEP is plotted against both α and β for $P/N_0=15$ dB, $r = 0.61$, and $M=4$. From the derived BEP expression (23), it is apparent that for a fixed total power, the power allocation among pilot and data affects the BEP performance. If we allocate too much power for pilots, channel estimation error will be reduced but because of low data SNR, detection of the data in noise is more difficult. On the other hand, lower power for pilots results in poor channel estimation and thus in poor detection [22], [23]. The minimum value of BEP under $P/N_0=15$ dB, $r = 0.61$ and $M=4$, turns out to be at $\alpha = \beta = 0.35$.

In Table I, the optimum values for α , β and r for $M= 4, 8, 16$ and 64 under $P/N_0 = 15$ dB and 25 dB are tabulated. To obtain the optimal values of α , β and r , an optimization problem was formulated as in (24) and solved.

V. CONCLUSION AND FUTURE WORK

We have investigated the impact of channel estimation error on the performance of a selective decode-and-forward relay communication network that uses a practical model for channel estimation. We have derived a lower bound expression for the BEP performance of considered scheme. We presented numerical simulation to show the proposed approximate analytical formulation is very close to the actual BEP. It is observed that channel estimation error causes loss of BEP performance. For a given power budget, this loss is decreased, by optimal power allocation between pilot symbol and data symbols based on the lower bound expression for BEP.

This paper assumed that the destination node perfectly knows whether the relay at each time slot is in silent mode or not. We are currently designing the destination's schemes for detecting whether the relay has forwarded the symbols in the frame or not. The study of these schemes and analyzing the BEP performance are left for future research.

REFERENCES

- [1] J. Laneman and G. Wornell, "Cooperative diversity in networks: algorithms and architecture," Massachusetts Inst. of Tech., Cambridge, MA.: PhD dissertation, 2002.
- [2] R. U. Nabar, H. Bolcskei, and F. W. Kneubuhler, "Fading relay channels: performance limits and space-time signal design," IEEE Journ. on Sel. Areas in Commun., no. 6, Aug. 2004, pp. 1099-1109.
- [3] 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, Dec. 2004, pp. 3062-3080.
- [4] J. N. Laneman and G. W. Wornell, "Energy-efficient antenna sharing and relaying for wireless networks," IEEE WCNC, Sep. 2000, pp. 7-12.
- [5] T. Wang, A. Cano, G. B. Giannakis, and J. N. Laneman, "High-performance cooperative demodulation with decode-and-forward relays," IEEE Trans. Commun., vol. 55, no. 6, Apr. 2007, pp. 830-841.
- [6] W. Su, A. K. Sadek, and K. J. R. Liu, "Cooperative communication protocols in wireless networks: performance analysis and optimum power allocation," Wireless Personal Commun. (Springer), vol. 44, no. 2, Jan. 2008, pp. 181-217.
- [7] A. S. Ibrahim, A. K. Sadek, W. Su, and K. J. R. Liu, "Cooperative communications with relay-selection: when to cooperate and whom to cooperate with?," IEEE Trans. on Wireless Commun., vol. 7, July 2008, pp. 2814-2827.
- [8] M. O. Hasna and M. S. Alouini, "A performance study of dual-hop transmissions with fixed gain relays," IEEE Trans. on Wireless Commun., vol. 3, no. 6, Nov. 2004, pp. 1963-1968.
- [9] P. A. Anghel and M. Kaveh, "Exact symbol error probability of a cooperative network in a Rayleigh-fading environment," IEEE Trans. on Wireless Commun., vol. 3, no. 5, Sep. 2004, pp. 1416-1421.
- [10] A. Ribeiro, X. Cai, and G. B. Giannakis, "Symbol error probabilities for general cooperative links," IEEE Trans. on Wireless Commun., vol. 4, no. 3, Mar. 2005, pp. 1264-1273.
- [11] Y. Li and S. Kishore, "Asymptotic analysis of amplify-and-forward relaying in Nakagami-fading environments," IEEE Trans. Wireless Commun., vol. 6, no. 12, Dec. 2007, pp. 4256-4262.
- [12] Y. Wu and M. Patzold, "Performance analysis of amplify-and-forward cooperative systems with channel estimation errors," The Eleventh IEEE International Conference on Commun. Systems (ICCS), Nov. 2008, pp. 1620-1624.
- [13] A. S. Ibrahim and K. J. R. Liu, "Mitigating channel estimation error with timing synchronization tradeoff in cooperative communications," IEEE Trans. on Signal Processing, vol. 58, Jan. 2010, pp. 337-348.
- [14] Y. Wu and M. Patzold, "Performance analysis of cooperative communication systems with imperfect channel estimation," IEEE International Conf. on Communications (ICC), June 2009, pp. 1-6.
- [15] J. Zhang and M. C. Gursoy, "Achievable rates and resource allocation strategies for imperfectly known fading relay channels," EURASIP Journal on Wireless Communications and Networking, vol. 2009, Article ID 458236, 2009.
- [16] S. Bhashyam, A. Sabharwal, and B. Aazhang, "Feedback gain in multiple antenna systems," IEEE Trans. on Commun., vol. 50, no. 5, May 2002, pp. 785-798.
- [17] M. C. Gursoy, "An energy efficiency perspective on training for fading channels," IEEE International Symposium on Information Theory (ISIT '07), June 2007, pp. 1206-1210, Nice, France.
- [18] D. Cabric, S. M. Mishra, and R. W. Brodersen, "Implementation issues in spectrum sensing for cognitive radios," Asilomar Conference on Signals, Systems, and Computers, 2004.
- [19] S. Atapattu, C. Tellambura, and H. Jiang, "Relay based cooperative spectrum sensing in cognitive radio networks," IEEE Global Telecommun. Conf. (GLOBECOM), Nov.-Dec. 2009.
- [20] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," IEEE Journal on Selected Areas in Commun., vol. 24, Mar. 2006, pp. 659-672.
- [21] R. You, H. Li, and Y. Bar-Ness, "Diversity combining with imperfect channel estimation," IEEE Trans. on Commun., Vol. 53, No. 10, Oct. 2005, pp. 1655-1662.
- [22] K. Ahmed, C. Tepedelenhoglu, and A. Spanias, "Effect of channel estimation on pair-wise error probability in OFDM," IEEE Conference on Acoustics, Speech and Signal Processing, vol. 4, May 2004, pp. 745-748.
- [23] A. Z. Ghanavati, U. Pareek, S. Muhaidat, and D. Lee, "On the performance of imperfect channel estimation for vehicular Ad-Hoc network," IEEE Vehicular Technology Conference Fall (VTC), September 2010.
- [24] M. K. Simon and M. S. Alouini, Digital communication over fading channels, New York, John Wiley & Sons, 2000.
- [25] I. S. Gradshteyn and I. M. Ryzhik, Table of integrals, series and products, Academic Press, London, 1980, 5th ed.

Fast Association Scheme over IEEE 802.15.4 based Mobile Sensor Network

Pranesh Sthapit, Yeon-Sang Choi,
Goo-Rak Kwon, and Jae-Young Pyun
Department of Information and Communication Engineering
Chosun University, Korea
pranesh@chosun.kr, cys@chosun.ac.kr, grkwon@chosun.ac.kr,
and jypyun@chosun.ac.kr (corresponding author)

Suk-seung Hwang
Department of Mechatronics Engineering
Chosun University, Korea
hwangss@chosun.ac.kr

Abstract—IEEE 802.15.4 is designed mainly for static low duty cycle wireless personal area network and its capability to support mobility has not been extensively explored yet. We observed that the amount of time required for the association process is the key reason for making IEEE 802.15.4 unable to handle mobility. In a beacon-enabled network, mobile node can only learn about the existence of any neighbor by listening to incoming beacons. Nevertheless, as neither the channel on which the neighbor operates nor its frequency of beaconing are known, nodes have to scan every available channel, which can take from few to several seconds depending upon the beacon interval. In this paper, we propose a new fast association technique, which prevents nodes from scanning multiple channels. In our proposed scheme, by scanning just a single channel, node can learn about all the coordinators working in different channels. Our single channel scan scheme is able to decrease the association time of IEEE 802.15.4 operation in 2.4 GHz by 32 times. Experiment results have verified that our scheme works well in the mobile environment.

Keywords—node association; multi-channel; IEEE 802.15.4

I. INTRODUCTION

Incorporating mobility in wireless personal area network (WPAN or PAN) is one of the most desirable features today, but it raises many new challenges. WPANs widely use IEEE 802.15.4 due to its low energy consumption, low cost, and small size. In WPAN, node mobility is expected to facilitate numerous applications, from home health-care and medical monitoring to target detection [1]. Wireless body area network (WBAN) is one such area where node mobility is frequent and IEEE 802.15.4 has been proposed as its possible MAC in various literatures [2][3]. There are, however, also multiple other cases in which an IEEE 802.15.4 node needs to discover coordinator in their proximity and establish communication with them. Such a scenario requires very efficient discovery schemata. Although the standard MAC protocol IEEE 802.15.4 has proven to satisfy low data-rate and low power-consumption requirements, the support for mobility is desirable in various WPAN applications [1]. In IEEE 802.15.4, there are mainly two types of devices: full function device (FFD) and reduced function device (RFD). FFD can support all the network functions and can operate as a PAN coordinator or an end device. RFD can only be used as an end device. The IEEE 802.15.4 standard supports three kinds of topology: star, peer-to-peer, and cluster tree topologies, which can operate on beacon and non-beacon-enabled modes.

Node association defines the procedure with which nodes can become a member of WPAN. In beacon-enabled networks, end devices first discover coordinator(s) by listening to the

beacons and then join them by using association messages. However, once a node moves out of the radio range of its parent, it loses synchronization with its parent and finds a new parent.

This paper focuses on the problem of how a node having no initial information about the surrounding can organize fast and energy efficient discovery of other WPANs in a beacon-enabled network. In such case, node can only learn about the existence of any neighbor by listening to their beacons. There are 16 channels in the 2.4 GHz band, 10 in the 915 MHz band, and 1 channel in the 868 MHz band allocated for IEEE 802.15.4 [4]. Nevertheless, as the channel on which the neighbor operates are unknown, nodes have to scan every single available channel in their operational band before getting associated. In this paper, we introduce a novel association scheme called dedicated beacon channel (DBC) in which beacon frame is broadcast in a dedicated channel. Our multichannel solution allows a node to acquire network information about all the coordinators in the vicinity by scanning just a single channel. DBC can decrease the association time of IEEE 802.15.4 by 32 times. However, application of DBC is not limited to IEEE 802.15.4 and can be used in any beacon-enabled network. Furthermore, with the application of DBC scheme, now IEEE 802.15.4 does not require separate channel scans such as ED scan, active scan, passive scan, and orphan scan for learning different status of channels. Our single DBC scheme does functions of all channel scans with minimal time. However, we propose only the modification of association scheme to provide support for mobility where as keeping intact the original features such as flexibility, scalability, adaptability, and low power consumption of original protocol as it is. The proposed scheme is a complete and sophisticated protocol that will be beneficial in the next generation mobile WPAN applications.

The remainder of this paper is organized as follows. Section II presents the brief description on association procedure of beacon-enabled IEEE 802.15.4 followed by related works in Section III. Section IV shows the operation and the main features of the proposed association scheme. The numerical and simulation analysis of the proposed scheme has been described in Sections V and VI. Finally, we conclude the paper in Section VII.

II. ASSOCIATION IN BEACON-ENABLED IEEE 802.15.4

In the beacon-enabled mode, communication is synchronized and controlled by a network coordinator, which transmits periodic beacons. The beacon contains information related to PAN identification, synchronization, and superframe structure.

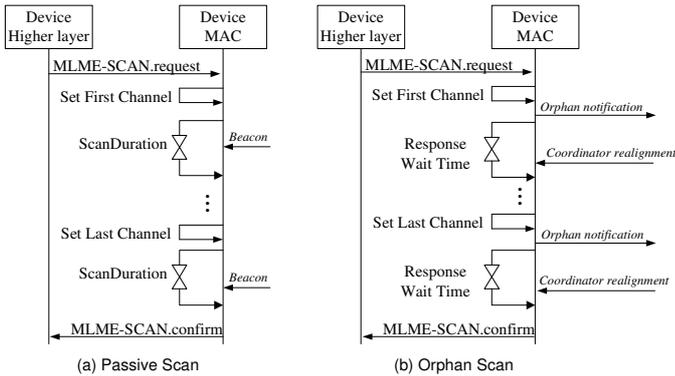


Fig. 1. Channel scanning mechanism in beacon-enabled IEEE 802.15.4.

The superframe may consist of active and inactive periods. The active part of the superframe consists of two groups: the contention access period (CAP) and an optional contention free period (CFP). All communications must take place during the active part, and devices can sleep to conserve energy in the inactive part. The structure of superframe is determined by coordinators using two parameters: superframe order (SO) and beacon order (BO). SO is used to determine the length of superframe duration, while BO defines the beacon interval. The duty cycle of nodes within PAN depends on BO and SO. Increasing both BO and SO, increases transmission latency and decreases system throughput due to intensive channel contention, whereas simultaneously decreasing BO and SO increases energy consumption and reduces available bandwidth [5].

Any FFD could be a PAN coordinator. FFD should perform the energy detection (ED) to detect the peak energy of a channel and choose an appropriate channel for data transmission. In each channel, ED scan is performed for the duration of t_{scan} symbols. Then, it carries out an active scan to locate any coordinator transmitting beacon frames within its personal operating space (POS). During the active scanning, it first sends out a beacon request command and waits for the duration of t_{scan} symbols. If a beacon could not be detected in t_{scan} , the FFD believes that there is no coordinator in its POS and can construct its own PAN by broadcasting its periodic beacons. After a PAN has been initialized, the other devices in the POS of the PAN can communicate with the coordinator and associate with this PAN. In order to start association, an end device needs to know PAN's physical channel, coordinator ID, addressing mode, and PAN ID. In beacon-enabled IEEE 802.15.4, two types of channel scanning operations are performed by end devices. During scan, nodes are deprived from data communication and must discard all data frames received. (i) *Passive scan* is performed by the node, which is just turned on or who failed to find its coordinator through the orphan scan. As shown in Fig. 1(a), during passive scan, a device search for beacon frame in each channel for the duration of t_{scan} symbols and records the beacon frames received in each channel. If no beacon is detected, the device starts another passive scan after a period of time. (ii) *Orphan scan* allows a device to attempt to relocate its coordinator following a loss of synchronization. The device shall first send the orphan notification command frame, and waits for coordinator realignment command frame for at most $macResponseWaitTime$ symbols as

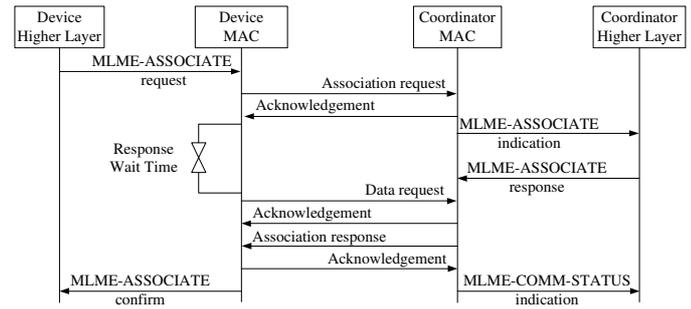


Fig. 2. Association message exchange in IEEE 802.15.4.

shown in Fig. 1(b). This procedure is repeated until it receives the coordinator realignment frame or all the available channels are scanned. If the device is unable to find its parent though the orphan scan, it looks for a new parent by performing passive scan. Once the node finds the suitable coordinator, it can start association procedure by sending a request for association. The message sequence exchanged during the association is shown in Fig. 2. If the association request is permitted, the device will successfully associate with the PAN and get its own PAN address to communicate with others. Thus, in beacon-enabled IEEE 802.15.4, a node association requires channel scanning followed by the association message exchange, which is really a time consuming procedure. For a mobile node, association procedure is even worse because the mobile node has to go through the orphan scan first followed by the passive scan and then only can start the association message exchange. We use the term re-association to explicitly denote the association procedure used by the mobile node. We have observed from our study that if somehow this whole association duration be shortened to some tolerable level, IEEE 802.15.4 can be used in the mobile sensor network applications.

III. RELATED WORKS

There are some efforts done to minimize the association duration in IEEE 802.15.4. In IEEE 802.15.4e [6], optional fast association (FastA) is defined, which allows a device to associate in a reduced duration. However, most of the efforts are limited to decreasing the duration of association message exchange or coordinator discovery, whereas the channel scanning part has been left untouched. The association duration can be drastically shortened if somehow nodes are prevented from scanning every single available channel. Similarly, there are several works on multichannel solutions, but are limited to throughput improvement or beacon collision avoidance [7][8][9]. To our best knowledge, we are the first to propose the multichannel solution for the association procedure.

In [10], Zhang et al. proposed an improved association scheme called Simple Association Process (SAP) that eliminates the redundant primitives, thus, decrease the packet collisions and the association delay. In order to account for real-time traffic, Meng et al. [11] proposed a new association scheme in which channel scanning is stopped as soon as one of the discovered PAN coordinators is estimated worthy to initiate the association process. Although, this scheme prevents nodes from scanning all available channels, but still nodes need to scan multiple channels before finding the suitable coordinator.

Similarly, there are other works, which focus on the neighbor discovery for quick association. In [12], algorithms

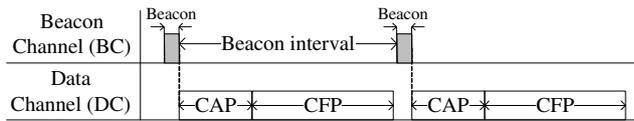


Fig. 3. Channel switching mechanism of the proposed DBC scheme.

are proposed for the optimized discovery of IEEE 802.15.4 static and mobile networks operating in multiple frequency bands and with different beacon intervals. In [13], a scheme to increase coordinator connectivity time with mobile nodes is presented for IEEE 802.15.4 beacon-enabled networks. Nodes use time-stamp of received beacons during the scan, along with link quality to determine the appropriate coordinator for association. Other mobility management schemes for cluster-tree based WPAN have been proposed by Chaabane et al. [14] and Bashir et al. [15]. These approaches use the speculative algorithm for node association based on LQI. Based on LQI value, mobile node anticipates cell change before the loss of connection and tries to associate with the next coordinator. However, in all the cases, nodes have to scan multiple channels to find coordinators.

IV. PROPOSED SCHEME

In our proposed DBC scheme, we exploit the channel switching capability of IEEE 802.15.4 radio hardware. In DBC, we proposed the use of two channels: beacon channel (BC) and data channel (DC). BC is used for transmission and reception of the beacon frames only where as the rest of the communication is done in the DC. A fixed dedicated channel is assigned for beacon whereas PAN coordinators are flexible to choose their own DC. Node switches its radio channel to the beacon channel during its beacon period and then returns to its original data channel at the end of the beacon period as shown in Fig. 3. Thus, transmission and reception of the beacon frame is done in the separate beacon channel. With this small and simple modification in the original IEEE 802.15.4, now the nodes do not need to scan each and every channel for the association. The entire network information can be learned by just scanning the BC only as shown in Fig. 4. IEEE 802.15.4 has 4 types of scanning, i.e., ED scan, active scan, orphan scan, and passive scan. Our scheme allows IEEE 802.15.4 to work solely with the single passive scan only. Our scheme has the potential to decrease the association time not just by percentage, but by many folds that we will be demonstrating in coming sections. DBC scheme can be used in any network topology. However, in case of peer-to-peer and cluster tree topologies, we assume that IEEE 802.15.4 successfully avoids beacon collisions.

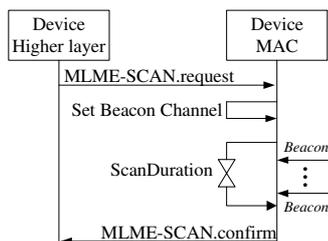


Fig. 4. The passive scan in the proposed DBC scheme.

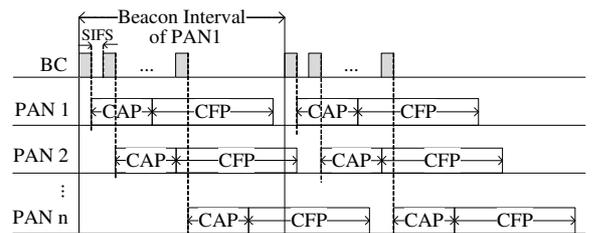


Fig. 5. Co-existence of multiple PANs.

A. Modification in Beacon Frame

In order to realize our proposed scheme, the data channel of PAN should be conveyed in the beacon frame. For this, either a new field (data channel of 1 byte) can be added in the beacon frame as shown in Fig. 6 or the data channel information may be piggyback in beacon payload or some other fields.

Frame Control	Sequence Number	Addressing Fields	Auxiliary Security Header	Superframe Specification	GTS Fields	Pending Address Fields	Data Channel	Beacon Payload	FRC
---------------	-----------------	-------------------	---------------------------	--------------------------	------------	------------------------	--------------	----------------	-----

Fig. 6. Addition of data channel field in the beacon frame of IEEE 802.15.4.

B. Initialization of a PAN coordinator

Unlike in original IEEE 802.15.4, nodes do not perform ED scan and active scan for initialization of a PAN coordinator. In our proposed scheme, when a FFD is initialized, it performs the passive scan in the beacon channel only for the durations of t_{scan} (Fig. 4). At the end of the passive scan, the FFD will have the clear picture of all the working PANs in the surrounding and can join the suitable PAN. If a beacon could not be detected in the scan duration, the FFD believes that there is no coordinator in its POS and can construct its own PAN by broadcasting its periodic beacons. The FFD is flexible to select its own non-overlapped DC, but it must broadcast its beacon in the BC. In the case where a number of PANs coexist in an adjacent area, (i.e., apartments or buildings having independent PAN), once the FFD finds out the available PANs through the passive scan, it schedules its beacon period with the minimum gap of the short interframe space (SIFS) than the last beacon frame it received, and chooses the unused DC from the available pool as shown in Fig. 5.

C. Association of network devices

A device that wants to associate with a PAN can be a new (just turn on) or an orphan device. In IEEE 802.15.4, there are separate procedures for the association of new and orphan devices. However, in our proposed scheme, there is only the passive scan and both devices perform the same procedure. An unassociated node performs passive scan in the BC for the duration of t_{scan} symbols. At the end of the passive scan, the end device will have clear information about all the coordinator in the surrounding and may join the suitable coordinator.

V. NUMERICAL ANALYSIS

Let $aBaseSuperFrameDuration$ be the number of symbols forming a superframe when $SO = 0$ and t_{scan} symbols be the time spent to scan a channel. In IEEE 802.15.4, it takes equal duration to perform ED, active, and passive scan on a channel and is given by,

$$t_{scan} = aBaseSuperFrameDuration \times (2^{BO} + 1). \quad (1)$$

A. PAN Initialization

In IEEE 802.15.4, a PAN coordinator performs ED scan and active scan in all available n channels. But, in case of DBC, only passive scan is performed in the BC for the duration of t_{scan} . Thus, the total initialization time of a PAN coordinator for IEEE 802.15.4 and DBC is,

$$PAN_{802.15.4_init} = 2n \times t_{scan}, \quad (2)$$

$$PAN_{DBC_init} = t_{scan}. \quad (3)$$

Thus, comparing (2) and (3), DBC decreases the initialization of PAN coordinator by the factor of $2n$.

B. Association

The total time spent for association is the sum of time spent in the channel scan and the time spent in the association message exchange ($Asso_{msg}$). A device that wants to associate with a PAN can be a new or orphan. IEEE 802.15.4 has separate procedures for the association of new and orphan devices. The total time spend for association by a newly joining node for both protocols are given by,

$$t_{802.15.4_asso} = n \times t_{scan} + Asso_{msg}. \quad (4)$$

$$t_{DBC_asso} = t_{scan} + Asso_{msg}. \quad (5)$$

From (4) and (5), DBC is able to decrease the association time of newly joining node by almost the factor of n .

In beacon-enabled IEEE 802.15.4, an orphan node performs the orphan scan on each channel for duration of $macResponseWaitTime$ ($32 \times aBaseSuperFrameDuration$) symbols until its parent is found or all n channels are scanned. Upon the failure of orphan scan, a new parent is searched through passive scan as mentioned above. Thus, the total time spent for re-association is given by,

$$t_{802.15.4_reasso} = n \times macResponseWaitTime + t_{802.15.4_asso}. \quad (6)$$

Thus, comparing (5) and (6), DBC decreases the re-association time of a node by many folds.

C. Numerical example

Assuming network parameters from Table II, we get $aBaseSuperFrameDuration$ of 15.36 ms and $macResponseWaitTime$ and $Asso_{msg}$ of 0.49 sec [4]. Thus, using these values and above equations and assuming $BO = 3$, the association time for both IEEE 802.15.4 and DBC is shown in Table I. From the values obtained in the table, we can conclude that DBC reduces the association time of IEEE 802.15.4 by significant amount and makes association duration independent of the number of available channels. We calculated, theoretically how many times DBC can decrease the PAN initialization and re/association time of the original IEEE 802.15.4 for all values of BO ($BO = SO$), and the achieved graph is shown in Fig. 7.

TABLE I
TOTAL TIME SPENT ON ASSOCIATION.

Channels	PAN Initialization		Association		Re-Association	
	802.15.4	DBC	802.15.4	DBC	802.15.4	DBC
3	0.82 s	0.14 s	0.90 s	0.63 s	2.37 s	0.63 s
10	2.76 s	0.14 s	1.87 s	0.63 s	6.77 s	0.63 s
16	4.42 s	0.14 s	2.70 s	0.63 s	10.54 s	0.63 s

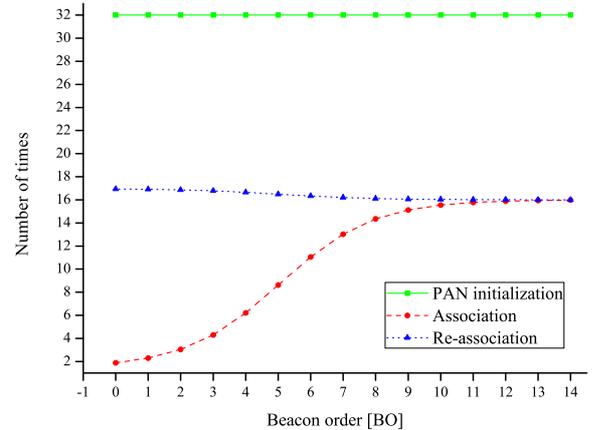


Fig. 7. No. of times DBC improves the association duration of IEEE 802.15.4.

VI. PERFORMANCE EVALUATION

The simulation analysis is performed using network simulator NS-2. 10 nodes are deployed in a 50x50m field with the PAN coordinator in the center as shown in Fig. 8, where arrow heads indicate the direction of movement of the mobile end device (node 9). All coordinators broadcast beacon, and BO is same for all nodes. In the simulation time of 100 secs, node 9 starts data transmission and at the simulation time of 110 secs, it starts to move. The mobile node continuously moves while transmitting data to PAN coordinator. The simulation ends when the mobile node comes to its original position. There are total 8 cell changes before the mobile node comes to complete rest. In all the simulations, SO is the same as BO. Network parameters are listed in Table II.

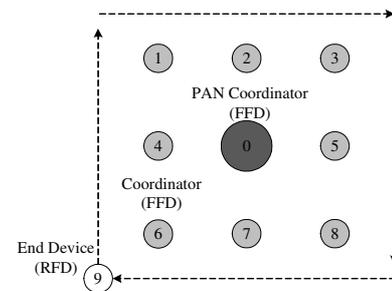


Fig. 8. Network topology used in the simulation.

TABLE II
NETWORK PARAMETERS AND VALUES.

Parameter	Value
Reception power	56.5 mW
Transmission power	48 mW
Idle power	2.79 mW
Sleep power	30 μ W
Radio range	10 m
Routing	AODV
Frequency band	2.4 GHz
Radio data rate	250 kbps
Number of Channels	16
Data rate	2 kbps
Traffic	CBR
Buffer size	10 packets
Packet size	50 B

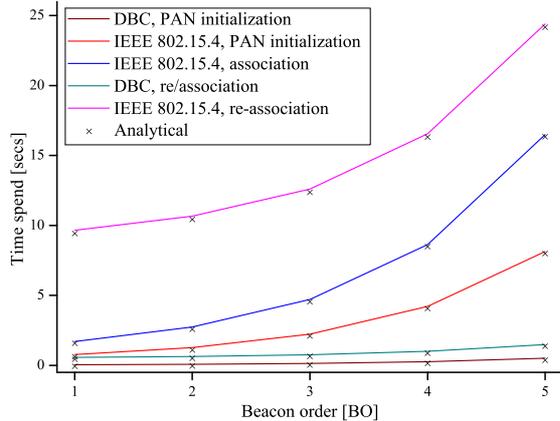


Fig. 9. Total time spent for association at different beacon intervals.

A. Association Time

Fig. 9 shows the PAN initialization time and the node re/association times for various values of BO obtained from numerical analysis and NS-2 simulations. In the NS-2 implementation, the PAN coordinator performs the passive scan for PAN initialization in the beacon-enabled mode. Similarly, for the assigned BO, except the PAN coordinator, other nodes use $BO + 1$ internally as the parameter for the scan duration. The numerical equations from Section V were adjusted accordingly. In all cases, the analytical results match well with the simulation results. As shown in the figure, the time required by DBC for re/association is much lower because only the BC is scanned for the association procedure. However, in the case of IEEE 802.15.4, it scans all available 16 channels spending significant amount of time and energy. Also, from the figure, we can see that the time required for both re/association is same for DBC because there is no orphan scan, and the same passive scan procedure is used for both re/association. At $BO = 3$, DBC is able to decrease the PAN initialization time by 15.92 times, the node association time by 6.19 times and the re-association time by 16.59 times, which is a great achievement in itself.

The energy spent by the PAN coordinator to start a PAN and for the node re/association under various BO is shown in Fig. 10. Since the time spent by the DBC is much lesser, it has a direct proportional impact in the energy consumption. At $BO = 3$, DBC is able to decrease the PAN initialization energy by 7 times, the node association energy by 7.5 times and the re-association energy by 17.75 times. Thus, from above observations, we can conclude that DBC is able to save both energy and time by the significant amount.

B. Association Success Rate

Success rate of association is calculated as the ratio of successful associations to the total number of possible associations. In our simulation model, there were total 8 associations possible before the mobile node comes to rest. DBC increases the successful association rate of a mobile node by providing quick passive discovery of a coordinator. Percentage of successful associations at different BO(s) and node speeds are shown in Fig. 11. We observed that in case of IEEE 802.15.4, even a slight mobility has a significant negative impact on association. At the human walking speed

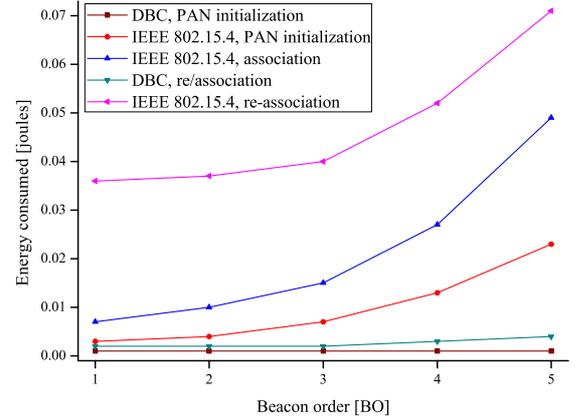


Fig. 10. Energy required for association at different beacon intervals.

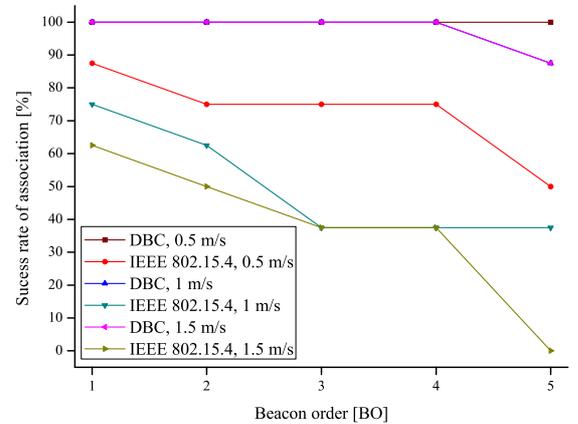


Fig. 11. Association success rate at 0.5, 1, and 1.5 m/s.

of 1.5 m/s, IEEE 802.15.4 has poor success rate of association even at the lower values of BO, and at $BO = 5$ nodes were completely unable to associate. However, in the case of DBC, association rate was 100% until $BO = 4$ and even at $BO = 5$, DBC could successfully performed 7 associations out of 8. However, at node speed of 1.5 m/s, DBC was completely unable to associate at $BO \geq 9$.

C. Packet delivery ratio and throughput

Fig. 12 shows the throughput observed at the PAN coordinator in packet delivery ratio (PDR). Because of the prompt re/association capability of DBC, mobile node can transmit most of the generated data to the coordinator increasing the overall throughput of the network. We can see in Fig. 12 that PDR of DBC is much better than that of IEEE 802.15.4 for various node speeds, which corresponds to the better throughput achieved. At node speed of 1.5 m/s and $BO = 5$, the PDR of IEEE 802.15.4 was just 20% that is also due to the fact that the mobile node 9 gets some opportunity to transmit data through node 6 before it starts to move and cannot associate then after. However, the PDR of DBC was of 82% for the same scenario.

The effect of cell change on throughput observed at the PAN coordinator at the node speed of 1 m/s and at the data rate of 2 Kbps is shown in Fig. 13. Coordinators in the network are using beacon interval corresponding to $BO = 3$. The red

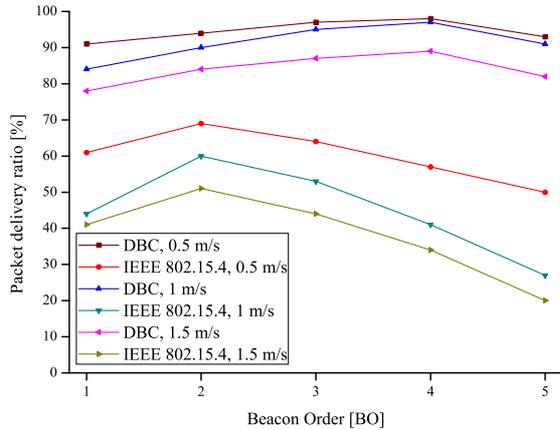


Fig. 12. PDR at different beacon intervals using DBC and IEEE 802.15.4 at 0.5, 1, and 1.5 m/s.

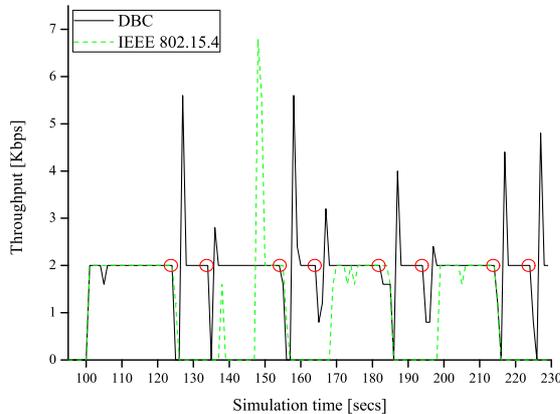


Fig. 13. Throughput observed at BO = 3, mobility of 1 m/s, and data rate of 2 Kbps. Red circle indicates the loss of synchronization by node 9.

circles in the figure indicate the time when the node 9 lost synchronization with the parent coordinator. Throughput of DBC drops while performing cell change because packets can't be transmitted in the periods of passive discovery and association. However, nodes quickly get associated with a new coordinator and restart data transmission. Nodes can buffer packets. Therefore, buffered data are also transmitted after new association is completed, resulting into increased throughput immediately after association. In some cell changes, throughput does not even drop to zero and regains its default value verifying prompt association property of DBC. However, in case of IEEE 802.15.4, nodes were unable to associate in every cell change and out of 8, only 3 successful associations were achieved resulting in poor performance. The PDR observed were of 53% and 96% respectively for IEEE 802.15.4 and DBC.

VII. CONCLUSION

In this paper, we proposed a new association scheme for IEEE 802.15.4 called DBC, which can decrease both time and energy required for association. To achieve the above mentioned advantages, the proposed DBC uses a dedicated channel for beacon transmission, depriving nodes from scanning all the available channels and looking for beacon. Our analytical and simulation results demonstrated that our scheme

is highly efficient in terms of both energy and time. With the implementation of our scheme, we give new direction for IEEE 802.15.4 to be able to handle mobility. However, in this paper we assumed, there is no beacon collision. As future directions, we envision to study and provide solution for beacon-collision avoidance in a dense network.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2011-0013062).

REFERENCES

- [1] K. Zen, D. Habibi, A. Rassau, and I. Ahmad, "Performance evaluation of IEEE 802.15.4 for mobile sensor networks," Proc. of the 5th IEEE WOCN '08, India, May 2008, pp. 1-5.
- [2] N. F. Timmons and W. G. Scanlon, "Analysis of the performance of IEEE 802.15.4 for medical sensor body area networking," in Proc. IEEE SECON, Oct. 2004 pp. 16-24.
- [3] N. F. Timmons and W.G. Scanlon, "An adaptive energy efficient MAC protocol for the medical body area networks," in Proc. of IEEE Wireless VITAE, May 2009, pp. 587-593.
- [4] IEEE 802.15.4, "Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (LR-WPANs)," IEEE, Sept. 2006.
- [5] Y. Huang, A. Pang, and H. Hung, "A Comprehensive Analysis of Low-Power Operation for Beacon-Enabled IEEE 802.15.4 Wireless Networks," IEEE Trans. on Wireless Communications, vol. 8, no. 11, 2009, pp. 5601-5611.
- [6] IEEE P802.15.4e-2012, "Part 15.4: Low-Rate Wireless Personal Area Networks (WPANs), Amendment 1: MAC sub-layer," Feb. 2012.
- [7] F. Osterlind and A. Dunkels, "Approaching the maximum 802.15.4 multi-hop throughput," in Proc. 5th ACM Workshop Embedded Netw. Sens., Charlottesville, VA, Jun. 2008, pp. 1-6.
- [8] E. Toscano and L.L. Bello, "Multiplechannel Superframe Scheduling for IEEE 802.15.4 Industrial Wireless Sensor Networks," IEEE Trans. on Industrial Informatics., vol. 8, no. 2, May 2012, pp. 337-350.
- [9] B.Y. Shih, C.J. Chang, A.W. Chen, and C.Y. Chen, "Enhanced MAC channel selection to improve performance of IEEE 802.15.4," International Journal of Innovative Computing, Information and Control, vol.6, no.12, 2010, pp.5511-5526.
- [10] F. Zhang, F. Wang, B. Dai, and Y. Li, "Performance Evaluation of IEEE 802.15.4 Beacon-enabled Association Process," in Proc. 22nd Int. Conf. Adv. Inf. Netw. Appl., Mar. 2008, pp. 541-546.
- [11] F. Meng and Y. Han, "A New Association Scheme of IEEE 802.15.4 for Real-time Applications," in Proc. Wireless Communications, Networking and Mobile Computing, 2009, pp. 1-5.
- [12] N. Karowski, A.C. Viana, and A. Wolisz, "Optimized Asynchronous Multi-channel Neighbor Discovery," in Proc. Infocom, 2011, pp. 536-540.
- [13] K. Zen, D. Habibi, and I. Ahmad, "Improving Mobile Sensor Connectivity Time in the IEEE 802.15.4 Networks," Proc of Telecommunication Networks and Applications Conference (ATNAC), 2008, pp. 317-320.
- [14] C. Chaabane, C. Pegatoquet, M. Auguin, and M. D. Jemaa, "Energy optimization for mobile nodes in a cluster tree IEEE 802.15.4/ZigBee network," Proc of IEEE Computing, Communications and Applications Conference (ComComAp), 2012, pp. 328-333.
- [15] F. Bashir, W.S. Baek, P. Sthapit, D. Pandey, and J.Y. Pyun, "Coordinator Assisted Passive Discovery for Mobile End Devices in IEEE 802.15.4," in Proc. of IEEE CCNC, Jan. 2013, pp. 601-604.

An Application of Second-Order Reed-Muller Codes for Multiple Target Localization in Wireless Sensor Networks

Thu Lam Ngoc Nguyen*, Jaejin Lee*, and Yoan Shin*
 *School of Electronic Engineering, Soongsil University, Seoul, Korea
 Email: {ngocthu, zlee, yashin}@ssu.ac.kr

Abstract—Compressive sensing is a new signal processing technique for efficient reconstruction of an n -dimensional signal from m ($m \ll n$) measurements. Most of compressive sensing researches are based on randomization, while research on deterministic sampling is essential for practical implementation. In this paper, we study $m \times n$ deterministic binary sensing matrices using second-order Reed-Muller codes, which satisfy a statistically restricted isometry property with reduced complexity for an application of multiple target localization in wireless sensor networks. We formulate multiple target locations as a sparse matrix, then exploit received signal strength information to recover noisy information using the deterministic sensing matrices and greedy algorithms to locate each target. The simulation results show that our scheme also achieves high accuracy in terms of localization errors when compared to traditional approaches with the random sensing matrices.

Keywords—wireless sensor network; multiple target localization; compressive sensing; deterministic sensing matrix; Reed-Muller codes.

I. INTRODUCTION

Wireless sensor networks can play a key role in target tracking, monitoring, environmental sensing and some other applications. Many approaches for network localization based on local sensor information have been developed with low-cost, low-power and small size constraints [1]–[8]. In this paper, we consider a scenario that the nodes are randomly deployed in a large area, and we determine multiple target locations based on the Received Signal Strength (RSS) information from their neighbors. Some of existing localization algorithms for this scenario are inefficient, since they require a large number of data between transmitter and receiver [9]–[14], [18]. Fortunately, the compressive sensing can help us to overcome these problems. The goal of compressive sensing is to recover an unknown signal vector $\mathbf{x} \in \mathbb{R}^n$ from linear measurement \mathbf{y} obtained by

$$\mathbf{y} = \Phi \mathbf{x}, \quad (1)$$

where $\Phi = \{\Phi_i\}_{i=1}^m \in \mathbb{R}^{m \times n}$ is the sensing matrix. The most concern is when the number of measurement m is much smaller than n , i.e., $m \ll n$. In this case, finding an exact solution \mathbf{x} based on the measurement \mathbf{y} is an ill-posed problem since the system of equations is under-determined. However, we can deal with it by finding an approximation of \mathbf{x} by solving this problem as

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \Phi \mathbf{x} = \mathbf{y}, \quad (2)$$

where $\|\mathbf{x}\|_0 = |\text{supp}(\mathbf{x})|$, and a vector is called k -sparse if it has at most k nonzeros elements. The compressive sensing technique guarantees exact recovery of the original signal \mathbf{x} with high probability if the sensing matrices satisfy the Restricted Isometry Property (RIP). That is, for a fixed k , there exists a small number $\delta_k \in (0, 1)$ such that

$$(1 - \delta_k) \|\mathbf{x}_k\|_2^2 \leq \|\Phi \mathbf{x}_k\|_2^2 \leq (1 + \delta_k) \|\mathbf{x}_k\|_2^2 \quad (3)$$

for any k -sparse \mathbf{x}_k . Hence, the problem (2) can be solved either by using greedy algorithms such as Basis Pursuit (BP) [15], Orthogonal Matching Pursuit (OMP) [16], [17], or replaced by solving for sparse signal via l_1 minimization as

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \Phi \mathbf{x} = \mathbf{y}. \quad (4)$$

In traditional compressive sensing approach, researchers have used random projection for the sensing matrices Φ to obtain the measurement \mathbf{y} , since the RIP can be satisfied with some random matrices with their entries following Gaussian process, Bernoulli process, etc. [19], [20]. Thus, a k -sparse signal $\mathbf{x} \in \mathbb{R}^n$ can be exactly reconstructed from m measurements. However, random matrices have many drawbacks: significant space requirement for storage, no efficient algorithm to verify the RIP, hard to deployment in many applications, to name a few. To this end, designing deterministic sensing matrices is essential for practical implementation. Recently, many advantages of deterministic sensing matrices have been shown. The most of these advantages is their fast and efficient reconstruction nature. In [21], Calderbank constructed some statistical RIP conditions such as Statistical Restricted Isometry Property (StRIP) and Uniqueness-guaranteed Statistical Restricted Isometry Property (UStRIP). These are weaker versions of the RIP that allow to construct deterministic sensing matrices. In [22], DeVore gave a generalization of construction via algebraic curves over finite fields. The author constructed binary sensing matrices of size $p^2 \times 2^{p+1}$ by using polynomials over finite field \mathbb{F}_p . This idea has been developed in many researches [23]. By choosing appropriate algebraic curves, these deterministic sensing matrices were better than DeVore's one. In [24], an application of coding theory in compressive sensing was presented, where a fast reconstruction algorithm for deterministic compressive sensing using second-order Reed-Muller codes was proposed. The matrix Φ is said to satisfy the StRIP (k, δ) if

$$\Pr \left\{ \left| \|\Phi \mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2 \right| \leq \delta \|\mathbf{x}\|_2^2 \right\} \geq 1 - \epsilon, \quad (5)$$

holds with probability exceeding $1 - \delta$, and we assume that \mathbf{x} distributes uniformly among k -sparse vectors. They showed that if Φ satisfies the StRIP respect to the parameters ϵ and δ , high probability reconstruction is also guaranteed. The deterministic sensing matrices formed by Reed-Muller codes, Bose-Chaudhuri-Hocquenghem (BCH) codes and some others can achieve this StRIP condition.

In this paper, we study the construction of deterministic sensing matrices formed by second-order Reed-Muller codes and how to apply this theory to multiple target localization in wireless sensor networks. We formulate each target location as a sparse vector in the discrete spatial domain. Then, we measure the RSS information from the targets and apply the construction of deterministic sensing matrix formed by second-order Reed-Muller codes as the measurement matrix. These matrices satisfy the StRIP, so that the approximated solution of (2) can be obtained by using a recovery algorithm in the last step.

The organization of the paper is as follows. In Section II, we explain a motivation of developing real-valued second-order Reed-Muller codes for deterministic sensing matrices in compressive sensing. In Section III, we formulate the multiple target localization problem as an application of compressive sensing by using the sensing matrices dealt in Section II. Numerical results are considered in Section IV, followed by concluding remarks in Section V.

II. REAL-VALUED SECOND-ORDER REED-MULLER CODES IN DETERMINISTIC SENSING MATRIX CONSTRUCTION

A. Main construction

Recall that for any two binary vector $\mathbf{a} = (a_0, \dots, a_{p-1})$ and $\mathbf{b} = (b_0, \dots, b_{p-1})$ in \mathbb{Z}_2^p , the inner product is defined as

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{a}^T \mathbf{b} = \sum_{i=1}^{p-1} a_i b_i \pmod{2}, \quad (6)$$

where $(\cdot)^T$ denote the transpose operation. The second-order Reed-Muller code is given as follows.

$$\phi_{\mathbf{P}, \mathbf{b}}(\mathbf{a}) = \frac{(-1)^{w(\mathbf{b})}}{\sqrt{2^p}} i^{(2\mathbf{b} + \mathbf{P}\mathbf{a})^T \mathbf{a}} \quad (7)$$

where \mathbf{P} is a $p \times p$ binary symmetric matrix, \mathbf{b} is a $p \times 1$ binary vector in \mathbb{Z}_2^p and $w(\mathbf{b})$ is the weight of \mathbf{b} , i.e., number of bit-1 entries. For given matrix \mathbf{P} and vector \mathbf{b} , the second-order Reed-Muller code is a $2^p \times 1$ vector. For implementation purposes, the matrices \mathbf{P} are set as all-zero matrices or the matrices with zero-diagonals. Thus, there is only $2^{p(p-1)/2}$ matrices \mathbf{P} satisfying this condition, which are $\{\mathbf{P}_1, \dots, \mathbf{P}_{2^{p(p-1)/2}}\}$ and the functions $\{\phi_{\mathbf{P}, \mathbf{b}}(\mathbf{a})\}$ are real-valued. The set

$$\mathcal{F}_{\mathbf{P}} = \{\phi_{\mathbf{P}, \mathbf{b}} | \mathbf{b} \in \mathbb{Z}_2^p\} \quad (8)$$

forms a basis of \mathbb{Z}_2^p . The inner product on $\mathcal{F}_{\mathbf{P}}$ is defined as follows. For any two vectors $\phi_{\mathbf{P}, \mathbf{b}}$ and $\phi_{\mathbf{P}', \mathbf{b}'}$ in $\mathcal{F}_{\mathbf{P}}$

$$\langle \phi_{\mathbf{P}, \mathbf{b}}, \phi_{\mathbf{P}', \mathbf{b}'} \rangle = \begin{cases} \frac{1}{\sqrt{2^q}} & 2^q \text{ times,} \\ 0 & 2^p - 2^q \text{ times,} \end{cases} \quad (9)$$

where $q = \text{rank}(\mathbf{P} - \mathbf{P}')$. The deterministic sensing matrix in [25] has the form

$$\Phi_{RM} = \begin{bmatrix} \mathbf{U}_{\mathbf{P}_1} & \mathbf{U}_{\mathbf{P}_2} & \cdots & \mathbf{U}_{\mathbf{P}_{2^{p(p-1)/2}}} \end{bmatrix}_{2^p \times 2^{p(p+1)/2}} \quad (10)$$

where $\mathbf{U}_{\mathbf{P}_i}$ is unitary matrix corresponding to $\mathcal{F}_{\mathbf{P}_i}$, $i = 1, \dots, 2^{p(p-1)/2}$. Note that if we set $m = 2^p$ and $n = 2^{p(p+1)/2}$, we get an $m \times n$ sensing matrix Φ_{RM} . The reconstruction problem using this matrix is to reconstruct the k -sparse vector \mathbf{x} from the data \mathbf{y} given by

$$\mathbf{y} = \Phi_{RM} \mathbf{x}. \quad (11)$$

In [26], the Delsarte-Goethals sets $DG(p, r)$ provide some restricted conditions for set of matrices \mathbf{P} . The set $DG(p, r)$ is a collection of $p \times p$ binary symmetric matrices with property that for any distinct matrices $\mathbf{P}, \mathbf{Q} \in DG(p, r)$, the rank of $\mathbf{P} + \mathbf{Q}$ is greater or equal to $p - 2r$. This implies that these sets are nested

$$DG(p, 0) \subset DG(p, 1) \subset \cdots \subset DG\left(p, \frac{p-1}{2}\right). \quad (12)$$

The set $DG(p, 0)$ is called Kerdock set [27]. Setting \mathbf{P} to range over $DG(p, (p-1)/2)$, the sensing matrices made from the matrices \mathbf{P} are the matrices of size $2^p \times 2^{p(r+2)}$.

B. Matrices with construction guarantee

Since the deterministic designs are based on the implemented and practical aspects, we focus on the sensing matrices whose entries are ± 1 by removing the normalization factor of $1/\sqrt{2^p}$ in Φ_{RM}

$$\hat{\Phi} = \sqrt{m} \Phi_{RM}. \quad (13)$$

Let us denote $\mu(\mathbf{A})$ as the largest magnitude of entries \mathbf{A} .

$$\mu(\mathbf{A}) = \max_{k,j} |\mathbf{A}_{k,j}|. \quad (14)$$

Thus $\mu(\hat{\Phi}) = 1$. For a fixed signal $\mathbf{x} \in \mathbb{R}^n$ where $\|x\|_0 = k$ the recovery in (2) is exact with high probability for the number of observations $m \geq C \cdot k \cdot \log n$ where C is a known small constant. We have

$$\hat{\Phi}^* \hat{\Phi} = 2^{p(p+1)/2} \mathbf{I} = n \mathbf{I},$$

where $(\cdot)^*$ denotes conjugate operation. For a small value of $\delta \in (0, 1)$, the eigenvalues of $\hat{\Phi}^* \hat{\Phi}$ are close to n with high probability. Thus, $\|\frac{1}{n} \hat{\Phi}^* \hat{\Phi} - \mathbf{I}\|_2 \leq 1/n$, and we have

$$\Pr\left\{\left|\|\hat{\Phi} \mathbf{x}\|^2 - \|\mathbf{x}\|^2\right| \leq \delta \|\mathbf{x}\|^2\right\} \geq 1 - \frac{1}{n}, \quad (15)$$

Hence, the matrix $\hat{\Phi}$ satisfies the StRIP with sparsity k and $\epsilon = \frac{1}{n}$. We can find some further information on binary symmetric matrices formed by second-order Reed-Muller codes in [25], [28], [29].

C. Examples

Let $p = 2$, then $\mathbb{Z}_2^2 = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$. There is only $2^{2(2-1)/2} = 2$ binary symmetric matrices \mathbf{P} of size 2×2 satisfying the condition. These are

$$\mathbf{P}_1 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \mathbf{P}_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad (16)$$

Thus, the corresponding unitary matrices $\mathbf{U}_{\mathbf{P}_1}$ and $\mathbf{U}_{\mathbf{P}_2}$ are

$$\mathbf{U}_{\mathbf{P}_1} = \frac{1}{2} \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \mathbf{U}_{\mathbf{P}_2} = \frac{1}{2} \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ -1 & -1 & -1 & -1 \end{bmatrix}$$

Hence, we get the deterministic sensing matrix $\hat{\Phi}$ as

$$\hat{\Phi} = \begin{bmatrix} 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \end{bmatrix}_{2^2 \times 2^3}$$

III. AN APPLICATION ON MULTIPLE TARGET LOCALIZATION IN WIRELESS SENSOR NETWORKS

A. Problem formulation

Consider an area which is divided into a discrete grid with n points. Denote k as the number of targets which are located in this area. Each target is $n \times 1$ vector whose elements are zeros, except 1 at the index of grid point where target is located. With k targets, we get a matrix of target locations over the grid as

$$\boldsymbol{\theta} = [\boldsymbol{\theta}_1 \quad \boldsymbol{\theta}_2 \quad \cdots \quad \boldsymbol{\theta}_k]_{n \times k}. \quad (17)$$

We take m measurements respect to each target under matrix Ψ , which will be explained in the next subsection. Then the RSS signals are given by

$$\underbrace{\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_k \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} \Psi_1 & 0 & \cdots & 0 \\ 0 & \Psi_2 & \cdots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \Psi_k \end{bmatrix}}_{\Psi} \underbrace{\begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \\ \vdots \\ \boldsymbol{\theta}_k \end{bmatrix}}_{\boldsymbol{\theta}} \quad (18)$$

We can describe the compression procedure as follows.

$$\underbrace{\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_k \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} \hat{\Phi}_1 & 0 & \cdots & 0 \\ 0 & \hat{\Phi}_2 & \cdots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \hat{\Phi}_k \end{bmatrix}}_{\hat{\Phi}} \underbrace{\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_k \end{bmatrix}}_{\mathbf{x}} + \underbrace{\begin{bmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \\ \vdots \\ \mathbf{n}_k \end{bmatrix}}_{\mathbf{n}}. \quad (19)$$

Our goal is to find all the locations of these targets with an accurate, fast and efficient algorithm with a small value of m .

B. Localization process

The matrix form of (19) is

$$\mathbf{y} = \hat{\Phi} \boldsymbol{\Psi} \boldsymbol{\theta} + \mathbf{n}. \quad (20)$$

These matrices $\boldsymbol{\Psi}$, $\hat{\Phi}$ are generated as follows.

- RSS matrix $\boldsymbol{\Psi} = \text{diag}\{\Psi_1, \dots, \Psi_k\}$ is made from using the radio propagation channel model

$$\begin{aligned} \{\Psi_i\}_{uv} &= PL(d_{uv}) \\ &= PL(d_0) - 10n_p \log_{10} \left(\frac{d_{uv}}{d_0} \right), u, v = 1, \dots, n, \end{aligned} \quad (21)$$

where $d_0 = 1\text{m}$ is the reference distance, d_{uv} is the real distance between transmitter and receiver in meters, $PL(d_0)$ is computed using the free space path loss equation, and n_p is the path loss component.

- In the sensing matrix $\hat{\Phi} = \text{diag}\{\Phi_1, \dots, \Phi_k\}$, each $\Phi_i (i = 1, \dots, k)$ is generated by the matrices satisfying the StRIP, as we have discussed in the previous section.

According to the compressive sensing theory, the localization problem is stated as the recovery of a sparse signal \mathbf{x}_i ($i = 1, \dots, k$) from measurement \mathbf{y}_i , which is equivalent to reconstruct target location $\boldsymbol{\theta}_i$ from \mathbf{y}_i . Assume that $\|\mathbf{n}\|_2 \leq \epsilon$ where ϵ is a small positive constant. Since each signal \mathbf{x}_i is represented by a sparsity basis, each sparse vector $\boldsymbol{\theta}_i$ can be found either by solving the following l_1 -minimization problem

$$\hat{\boldsymbol{\theta}} = \arg \min \|\boldsymbol{\theta}\|_1 \quad \text{subject to} \quad \mathbf{y} = \boldsymbol{\Theta} \boldsymbol{\theta}, \quad (22)$$

where $\boldsymbol{\Theta} = \hat{\Phi} \boldsymbol{\Psi}$, or by solving the convex optimization program by calling the OMP algorithm as

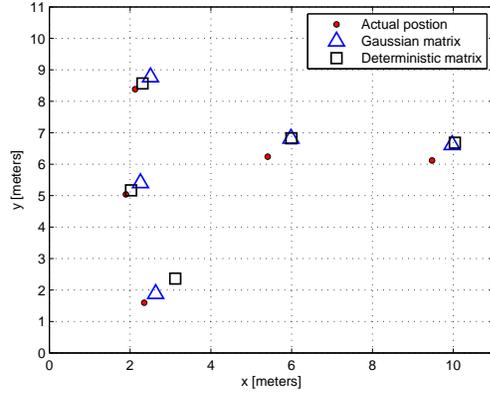
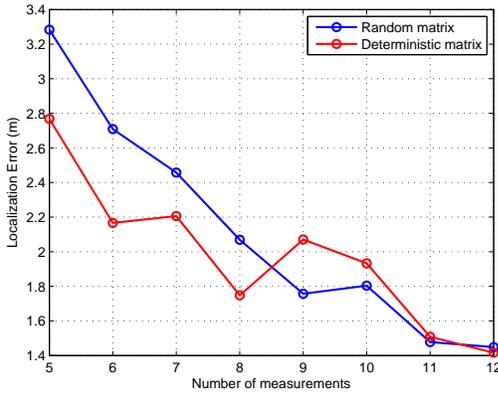
$$\hat{\boldsymbol{\theta}} = \text{OMP}(\mathbf{y}, \boldsymbol{\Theta}, \epsilon). \quad (23)$$

The main idea of the OMP algorithm is to find the columns of the matrix $\boldsymbol{\Theta}$ whose linear combination is close to \mathbf{y} . The OMP is more simpler and faster than other alternatives. In this paper, we have applied this algorithm to simulate and to generate sensing matrices $\hat{\Phi}$. To improve the performance and to find the exact target locations in the grid points, a threshold has been defined to select the largest component in the location vector of n components with minimum overall distance error.

IV. NUMERICAL RESULTS

In this section, we examine the performance of multiple target localization using compressive sensing under random matrices and deterministic matrices formed by second-order Reed-Muller codes in an indoors environment. We randomly deployed M sensors in an area with the size of $10\text{m} \times 10\text{m}$ with N grid points, and placed the targets by randomly selecting k grid points in uniform manners. We added Gaussian noises with zero mean and standard deviation of 0.05 to the observation \mathbf{y} . We used the Average Localization Error (ALE) to quantify the localization accuracy, which is defined as

$$\text{ALE}(\mathbf{p}) = \frac{1}{k} \sqrt{\|\mathbf{p} - \mathbf{p}^*\|_2^2}, \quad (24)$$


 (a) Position recovery performance with $M = 8$.


(b) Average localization error respect to 5 targets.

Fig. 1. Localization of 5 targets under random sensing matrix and deterministic sensing matrix.

where \mathbf{p} is the actual point and \mathbf{p}^* is the estimated point. Each run is collected 100 times. In the simulations, each RSS information $\Psi_i (i = 1, \dots, k)$ was obtained by

$$\{\Psi_i\}_{uv}(d) = -46.2 - 10n_p \log_{10}(d), u, v = 1, \dots, n, \quad (25)$$

Each location was observed over 100 simulations.

Figure 1 shows the position recovery performance. The area is divided by 64×64 grid points. According to the compressive sensing approach, with 5 targets, the number of RSS measurements required was at least $M > 2k = 10$ in the random i.i.d. Gaussian matrix case and 8 in the proposed deterministic matrix for exact solution recovery. Figure 1(a) shows the position recovery for 5 targets with $M = 8$. It shows that the proposed scheme achieves good performance as the traditional scheme does. Note that only deterministic sensing matrices are practically feasible when considering actual implementation. Figure 1(b) shows the ALE versus the number of measurements. The ALE of the traditional random approach was more dramatically decreased than the proposed one when the number of measurement increased.

Figure 2 shows the ALE with M randomly selected measurements in a given set of sensors. M was to set at 8 and 32 in this simulation. In the case of $M = 8$, the ALE of the proposed scheme is smaller than the traditional case as

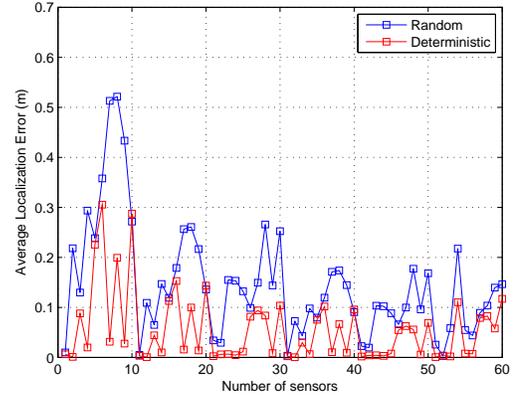
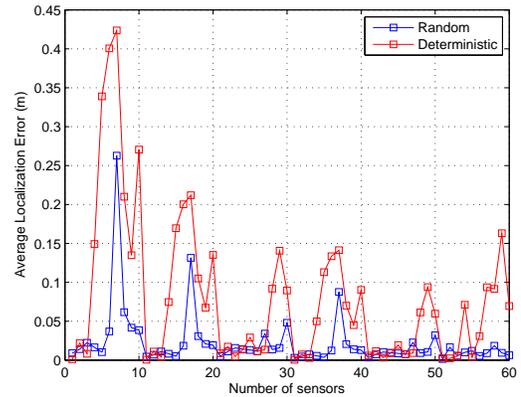

 (a) 8×64 sensing matrix.

 (b) 32×64 sensing matrix.

Fig. 2. Average localization error.

seen in Figure 2(a). However, when $M = 32$ we observe the opposite results as in Figure 2(b). Since M is large, the independence among columns of the deterministic sensing matrices is not guaranteed, while the RIP holds with random matrices in this case. Thus, the perfect reconstruction by the proposed method may be not guaranteed, and performance by the random matrices is better than the proposed one when the number of measurements becomes large.

For classical approach, each sensor must be recorded n measurements, which brings large communication cost, especially in large-scale networks. Based on the idea of reducing cost in compressive sensing theory and the advantages of deterministic sensing matrices formed by the second-order Reed-Muller codes on recovery, our method reduces the overall communication bandwidth requirement per sensor, and achieve high localization accuracy. However, the trade-off between high level of accuracy and low computational cost should be considered as well.

V. CONCLUDING REMARKS

In this paper, we presented an approach for multiple target localization in wireless sensor networks using deterministic sensing matrices. We begin with problem formulation and present a localization method from sparse measurement based

on compressive sensing theory. Constructing a sparse measurement matrix is one of the most difficult part during this process. We investigated second-order Reed-Muller codes and applied them to form the measurement matrices in our problem. A key advantage of compressive sensing with these matrices is that it admits a fast reconstruction algorithm, especially for basis pursuit, and depends only on number of measurements m and sparsity k , not depends on the signal length n , in addition to their deterministic structure. Numerical results show that these matrices also guarantee to recover approximated solutions as the traditional schemes do, especially when the signal vectors are very sparse. We expect that this type of matrices will be useful for various localization applications in wireless sensor networks.

ACKNOWLEDGMENT

This research was partly supported by Mid-career Researcher Program through NRF grant funded by MEST, Korea (No. 2012-0005330), and by the MSIP, Korea in the ICT R&D Program 2013 (KCA-2012-12-911-01-107).

REFERENCES

- [1] Y. Shang, W. Ruml, Y. Zhang, and M. Fromherz, "Localization from connectivity in sensor networks," *IEEE Trans. Parallel Distributed Syst.*, vol. 15, no. 11, pp. 961-974, Nov. 2004.
- [2] N. Patwari, J. N. Ash, and S. Kyperountas, "Locating the nodes: Cooperative localization in wireless sensor networks," *IEEE Sig. Proc. Mag.*, vol. 22, no. 4, pp. 54-69, July 2005.
- [3] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Computer Networks*, vol. 38, no. 4, pp. 393-422, 2002.
- [4] J. Liu, Y. Zhang, and F. Zhao, "Robust distributed node localization with error management," *Proc. ACM MobiHoc 2006*, pp. 250-261, Florence, Italy, May 2006.
- [5] X. Ji and H. Zha, "Sensor positioning in wireless ad-hoc sensor networks with multidimensional scaling," *Proc. IEEE INFOCOM 2004*, pp. 2652-2661, Hong Kong, China, Mar. 2004.
- [6] N. Patwari and A. O. Hero III, "Manifold learning algorithms for localization in wireless sensor networks," *Proc. IEEE ICASSP 2004*, vol. 3, pp. 857-860, Montreal, Canada, May 2004.
- [7] N. Patwari, A. O. Hero III, M. Perkins, N. S. Correal, and R. J. Odea, "Relative location estimation in wireless sensor networks," *IEEE Trans. Sig. Proc.*, vol. 51, no. 8, pp. 2137-2148, Aug. 2003.
- [8] G. Mao, B. Fidan, and B. D. O. Anderson, "Wireless sensor network localization techniques," *Computer Networks*, vol. 51, no. 10, pp. 2529-2553, July 2007.
- [9] L. Doherty, L. E. Ghaoui, and S. J. Pister, "Convex position estimation in wireless sensor networks," *Proc. IEEE INFOCOM 2002*, vol. 3, pp. 1655-1663, San Francisco, USA, Apr. 2002.
- [10] P. Biswas, T. C. Liang, K. C. Toh, Y. Ye, and T. C. Wang, "Semidefinite programming approaches for sensor network localization with noisy distance measurements," *IEEE Trans. Auto. Sci. Eng.*, vol. 3, no. 4, pp. 360-371, Oct. 2006.
- [11] Q. Shi, C. He, L. Jiang, and J. Luo, "Sensor network localization via nondifferentiable optimization," *Proc. IEEE GLOBECOM 2008*, New Orleans, USA, Dec. 2008.
- [12] R. Zetik, S. Jovanoska, and R. Thomä, "Simple method for localisation of multiple tag-free targets using UWB sensor network," *Proc. IEEE ICUWB 2011*, Bologna, Italy, Sept. 2011.
- [13] A. Kushki, K. N. Plataniotis, and A. N. Venetsanopoulos, "Kernel-based positioning in wireless local area networks," *IEEE Trans. Mobile Comp.*, vol. 6, no. 6, pp. 689-705, June 2007.
- [14] V. Cevher, P. Boufounos, R. G. Baraniuk, A. C. Gilbert, and M. J. Strauss, "Near-optimal Bayesian localization via incoherence and sparsity," *Proc. IPSN 2009*, San Francisco, USA, Apr. 2009.
- [15] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Info. Theory*, vol. 55, no. 5, pp. 2230-2249, May 2009.
- [16] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301-321, 2009.
- [17] J. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Info. Theory*, vol. 53, no. 12, pp. 4655-4666, Dec. 2007.
- [18] V. Cevher, M. F. Duarte, and R. G. Baraniuk, "Distributed target localization via spatial sparsity," *Proc. EUSIPCO 2008*, Lausanne, Switzerland, Aug. 2008.
- [19] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Info. Theory*, vol. 52, no. 2, pp. 489-509, Feb. 2006.
- [20] D. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp. 1289-1306, Apr. 2006.
- [21] R. Calderbank, S. Howard, and S. Jafarpour, "Construction of a large class of deterministic sensing matrices that satisfy a statistical isometry property," *IEEE Trans. Info. Theory*, vol. 4, no. 2, pp. 358-374, Apr. 2010.
- [22] R. DeVore, "Deterministic constructions of compressed sensing matrices," *Jour. Complexity*, vol. 23, no. 4-6, pp. 918-925, 2007.
- [23] S. Li, F. Gao, G. Ge, and S. Zhang, "Deterministic construction of compressive sensing matrices via algebraic curves," *IEEE Trans. Info. Theory*, vol. 58, no. 8, pp. 5035-5041, Aug. 2012.
- [24] M. A. Tsfasman and S. G. Vladu, *Algebraic-Geometric Codes*, in *Math. Appl. (Soviet Series)*, vol. 58, Kluwer Academic Publishers, 1991.
- [25] S. Howard, A. Calderbank, and J. Searle, "A fast reconstruction algorithm for deterministic compressive sensing using second order Reed-Muller codes," *Proc. IEEE CISS 2008*, pp. 11-15, Princeton, USA, Mar. 2008.
- [26] P. Delsarte and J. M. Goethals, "Alternating bilinear forms over GF(q)," *Jour. Combinatorial Theory*, vol. 19, pp. 26-50, 1975.
- [27] A. Kerdock, "A class of low-rate nonlinear binary codes," *Info. & Control*, vol. 20, pp. 182-187, 1972.
- [28] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, North-Holland Publishing, 1977.
- [29] A. R. Hammons, P. V. Kumar, A. R. Calderbank, N. J. A. Sloane, and P. Sole, "The \mathbb{Z}_4 -linearity of Kerdock codes, Preparata, Goethals, and related codes," *IEEE Trans. Info. Theory*, vol. 40, no. 2, pp. 301-319, Mar. 1994.

Bit Error Rate Analysis of MIMO Schemes in LTE Systems

Ali Jemmali, Jean Conan, Mohammad Torabi
 Department of Electrical Engineering,
 École Polytechnique de Montréal, Montréal, QC, Canada.
 {ali.jemmali, jean.conan, mohammad.torabi}@polymtl.ca

Abstract— In this paper, a Bit Error Rate (BER) analysis is presented for Multiple-Input Multiple-Output (MIMO) schemes in the 3GPP Long Term Evolution (LTE) system. Analytical expressions for the average BER of the system are derived over flat Rayleigh fading channels for two different MIMO schemes as defined in LTE, assuming M-ary quadrature amplitude modulation (M-QAM) schemes and are evaluated numerically. Monte-Carlo simulation results of the LTE system are also provided to verify the accuracy of the mathematical analysis. It is shown that the results obtained from Monte-Carlo simulations match closely with those obtained from the derived mathematical formulas.

Keywords- Performance Analysis, MIMO, LTE, M-QAM Modulation.

I. INTRODUCTION

To increase the capacity and speed of wireless communication systems, a new wireless data networks has been emerged and has been standardized by the 3rd Generation Partnership Project (3GPP). This new standard is a natural evolution to the existing second (2G) and third (3G) generation wireless networks in order to respond to the growing demand in terms of data rates and speed and marketed as 4G Long Term Evolution (LTE). In LTE, data throughput and the speed of wireless data are increased by using a combination of new methods and technologies like Orthogonal Frequency Division Multiplexing (OFDM) and Multiple-Input Multiple-Output (MIMO) techniques.

In the downlink, the LTE transmission is based on Orthogonal Frequency Division Multiple Access (OFDMA), known as a technique of encoding digital data on multiple carrier frequencies. It was shown that OFDMA is an efficient technique to improve the spectral efficiency of wireless systems. By converting the wide-band frequency selective channel into a set of several flat fading subchannels, OFDM technique becomes more resistant to frequency selective fading than single carrier systems. As OFDM signals are in time and frequency domain, they allow adding frequency domain scheduling to time domain scheduling. In LTE, for a given transmission power, the system data throughput and the coverage area can be optimized by employing Adaptive Modulation and Coding (AMC) techniques. The role of the user scheduler at the transmitter side is to assign the data rate for each user

according to the channel conditions from the serving cell, the interference level from other cells, and the noise level at the receiver side.

In LTE standard, the use of MIMO has been considered as an essential technique in order to achieve the target in terms of data throughput and reliability. MIMO is known to be a very powerful technique to improve the system performance of wireless communication systems. The diversity and multiplexing modes are the two main modes of operation of multiple antennas systems. The principle of diversity mode is based on transmitting the same signal over multiple antennas and hence to improve the reliability of the system by the diversity gain. In this mode, the mapping function of transmit symbols used at the transmit antennas is called Space Time Block Code (STBC). On the other hand, multiplexing mode uses two or more different spatial streams and send them through two different antennas, consequently, the data rate can be improved.

To study the performance of LTE systems a MATLAB based downlink physical layer simulator for Link Level Simulation (LLS) has been developed in [1] [2]. A System Level Simulation of the Simulator is also available [3]. The goal of developing the LTE simulator was to facilitate comparison with the work of different research groups and it is publicly available for free under academic non-commercial use license [2]. The main features of the simulator are adaptive coding and modulation, MIMO transmission and scheduling. As the simulator includes many physical layer features, it can be used for different applications in research [3]. In [4], the simulator was used to study the channel estimation of OFDM systems and the performance evaluation of a fast fading channel estimator was presented. In [5] and [6], a method for calculating the Precoding Matrix Indicator (PMI), the Rank Indicator (RI), and the Channel Quality Indicator (CQI) were studied and analyzed with the simulator.

In this paper, the Bit Error Rate (BER) analysis of two transmit diversity schemes known as Space Frequency Block Codes (SFBC) and Frequency Switched Transmit Diversity (FSTD) MIMO schemes in LTE system for M-QAM modulation scheme are presented in terms of SNR using the moment generating function of the SNR. The results obtained from

the analysis are then compared to the results of Monte-Carlo simulation using the Link Level LTE simulator [1] [2].

The remainder of this paper is organized as follows. In Section II, we present the system and channel model used in the simulation. In Section III, we present a performance analysis for the average BER of SFBC and FSTD MIMO schemes. The numerical and simulation results and discussions are presented in Section IV. Finally, Section V concludes the paper.

II. SYSTEM MODEL

In this section, the structure of the OFDM LTE signal is described. The OFDM signal has a time and a frequency domains. In the time domain, the LTE signal is composed of successive frames. Each frame has a duration of $T_{\text{frame}} = 10$ msec. Each frame is divided into 10 subframes with equal length of 1 msec. Each subframe consists of two equal length time-slots with a time duration of $T_{\text{slot}} = 0.5$ msec. For a normal cyclic prefix length, each time-slot consists of $N_s = 7$ OFDM symbols. In the frequency domain, the OFDM technique converts the LTE wideband signal into several narrowband signals. Each narrowband signal is transmitted on one subcarrier frequency. In LTE, the spacing between subcarriers is fixed to 15 KHz. Twelves adjacent subcarriers, occupying a total of 180 KHz, of one slot forms the so-called Resource Block (RB). The number of Resource Blocks in an LTE slot depends on the allowed system bandwidth. The minimum number of RB is equal to 6 corresponding to 1.4 MHz system bandwidth. For 20 MHz system bandwidth (Maximum Allowed bandwidth in LTE) the number of RB is equal to 100. In a MIMO system with M_R receive antennas and M_T transmit antennas, the relation between the received and the transmitted signals on subcarrier frequency k ($k \in 1, \dots, K$), at sampling instant time n ($n \in 1, \dots, N$) is given by

$$\mathbf{y}_{k,n} = \mathbf{H}_{k,n} \mathbf{x}_{k,n} + \mathbf{n}_{k,n} \quad (1)$$

where $\mathbf{y}_{k,n} \in C_{M_R \times 1}$ is the received vector, $\mathbf{H}_{k,n} \in C_{M_R \times M_T}$ represents the channel matrix on subcarrier k at instant time n , $\mathbf{x}_{k,n} \in C_{M_T \times 1}$ is the transmit symbol vector and $\mathbf{n}_{k,n} \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I})$ is a white, complex valued Gaussian noise vector with variance σ_n^2 . Assuming perfect channel estimation, the channel matrix and noise variance are considered to be known at the receiver. A linear equalizer filter given by a matrix $\mathbf{F}_{k,n} \in C_{M_R \times M_R}$ is applied on the received symbol vector $\mathbf{y}_{k,n}$ to determine the post-equalization symbol vector $\mathbf{r}_{k,n}$ as follows [6]

$$\mathbf{r}_{k,n} = \mathbf{F}_{k,n} \mathbf{y}_{k,n} = \mathbf{F}_{k,n} \mathbf{H}_{k,n} \mathbf{x}_{k,n} + \mathbf{F}_{k,n} \mathbf{n}_{k,n}. \quad (2)$$

The Zero Forcing (ZF) or Minimum Mean Square Error (MMSE) design criterion [7] are typically used for the linear receiver and the input signal vector is normalized to unit power. In MIMO-OFDM systems, the key factor of link error prediction and performances is the signal to noise ratio (SNR)

which represents the measurement for the channel quality information. In this study, the SNR is defined as follows [1]:

$$\gamma_{k,n} = \frac{\|\mathbf{H}_{k,n} \mathbf{x}_{k,n}\|_F^2}{N_T \sigma_n^2} \quad (3)$$

where $\mathbf{x}_{k,n}$ is the transmitted symbol vector, $\|\cdot\|_F^2$ is the squared Frobenius norm of a matrix.

III. AVERAGE BER PERFORMANCE ANALYSIS

In the following, we present a BER performance analysis for the 2×1 MIMO SFBC and 2×2 MIMO SFBC systems, over slow fading channels. Using the Moment Generating Function (MGF)-based approach we obtain closed-form expressions for the average BER performance of the system. We then present numerical evaluation results obtained from the closed-form expressions. Finally, we present Monte-carlo simulations to verify the accuracy of our analysis.

A. BER Analysis of SFBC

In LTE, the transmit diversity techniques are defined only for 2 and 4 transmit antennas and one data stream. When two eNodeB antennas are available for transmit diversity operation, the Space Frequency Block Code (SFBC) is used [8]. SFBC is based on the well known Space Time Block Codes (STBC), also known as Alamouti codes for two transmit antennas [9]. STBC is employed with the UMTS and it operates on pairs of adjacent symbols in the time domain. Since the signal in LTE is two dimensional (time and frequency domains) and the number of available OFDM symbols in a subframe is not always an even number, the direct application of STBC is not straightforward. In LTE, for SFBC transmission, the symbols are transmitted from two eNodeB antenna ports on each pair of adjacent subcarriers as follows [8]:

$$\begin{bmatrix} y^{(0)}(1) & y^{(0)}(2) \\ y^{(1)}(1) & y^{(1)}(2) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \\ -x_2^* & x_1^* \end{bmatrix} \quad (4)$$

where $y^{(p)}(k)$ denotes the symbols transmitted on the k^{th} subcarrier from antenna port p . An important characteristic of such codes is that the transmitted signal streams are orthogonal and a simple linear receiver is required for optimal performances.

Since OFDM converts the multipath channel into N frequency flat fading channel, we first derive the BER expressions over flat Rayleigh fading channels, given by $P_b(E)$. Then, the overall average BER over N subcarriers, in each case can be calculated from

$$BER = \frac{1}{N} \sum_{k=1}^N P_{b,k}(E) \quad (5)$$

where the index k (subcarrier index) is ignored for the sake of brevity. In addition, the impact of cyclic prefix in OFDM is assumed to be negligible.

For the 2×1 SFBC MIMO scheme, the probability density function of the SNR for each subcarrier is given by a chi-square distribution function as follows [10]:

$$f(\gamma) = \frac{2}{\bar{\gamma}^2} \gamma e^{-\frac{2}{\bar{\gamma}} \gamma} \quad (6)$$

where $\bar{\gamma}$ is the average SNR per symbol given by $\bar{\gamma} = E_s/N_0$.

The moment generating function (MGF) can be determined using the following equation:

$$M_{\bar{\gamma}}(s) = \int_0^{\infty} e^{-s\gamma} f(\gamma) d\gamma. \quad (7)$$

Inserting (6) into (7) and solving the integral yields

$$M_{\bar{\gamma}}(s) = \frac{4}{\bar{\gamma}^2 (s + \frac{2}{\bar{\gamma}})^2}. \quad (8)$$

The average BER expression for M-QAM modulation scheme can be obtained from [11] (equation (8.111; Page 255))

$$P_b(E) \cong B \sum_{i=1}^{\sqrt{M}/2} \frac{1}{\pi} \int_0^{\pi/2} M_{\bar{\gamma}}(A_{i,\theta}) d\theta \quad (9)$$

where $A_{i,\theta} = \frac{(2i-1)^2}{2 \sin^2 \theta} \frac{3}{(M-1)}$ and B is defined by

$$B = 4 \left(\frac{\sqrt{M}-1}{\sqrt{M}} \right) \left(\frac{1}{\log_2 M} \right). \quad (10)$$

Then, using the MGF expression in (8), we obtain

$$M_{\bar{\gamma}}(A_{i,\theta}) = \frac{4}{\bar{\gamma}^2 \left(\left[\frac{(2i-1)^2}{2 \sin^2 \theta} \frac{3}{(M-1)} \right] + \frac{2}{\bar{\gamma}} \right)^2}. \quad (11)$$

Substituting (11) into (9) and after some manipulations, we obtain

$$P_b(E) \cong B \sum_{i=1}^{\sqrt{M}/2} \frac{1}{\pi} \int_0^{\pi/2} \left(\frac{\sin^2 \theta}{\sin^2 \theta + c_i} \right)^2 d\theta \quad (12)$$

where $c_i = \frac{3(2i-1)^2}{2(M-1)} \frac{\bar{\gamma}}{2}$.

The average BER performance as a function of $\bar{\gamma} = E_s/N_0$ can be evaluated by numerical evaluation of the integral in (12) for M-QAM modulation schemes. Alternatively, by solving the integral, we obtain a closed-form expression for the average BER of M-QAM modulation as follows

$$P_b(E) \cong B \sum_{i=1}^{\sqrt{M}/2} \mathcal{I}_2(\pi/2, c_i) \quad (13)$$

where the closed-form expression for $\mathcal{I}_2(\cdot, \cdot)$ can be obtained from [11](eq.5A.24) as follows

$$\begin{aligned} \mathcal{I}_n(\phi, D) &= \frac{1}{\pi} \int_0^{\phi} \left(\frac{\sin^2 \theta}{\sin^2 \theta + D} \right)^n d\theta, \quad -\pi \leq \phi \leq \pi \\ &= \frac{\phi}{\pi} - \frac{\beta}{\pi} \left\{ \left(\frac{\pi}{2} + \tan^{-1} \alpha \right) \sum_{q=0}^{n-1} \binom{2q}{q} \frac{1}{(4(1+D))^q} \right. \\ &\quad \left. + \sin(\tan^{-1} \alpha) \sum_{q=1}^{n-1} \sum_{p=1}^q \frac{T_{pq}}{(1+D)^q} [\cos(\tan^{-1} \alpha)]^{2(q-p)+1} \right\} \end{aligned} \quad (14)$$

where $T_{pq} = \binom{2q}{q} \left[\binom{2(q-p)}{q-p} 4^p [2(q-p)+1] \right]^{-1}$, $\beta = \sqrt{\frac{D}{1+D}} \operatorname{sgn} \phi$, and $\alpha = -\beta \cot \phi$.

B. BER Analysis of FSTD

In LTE, the frequency space code, designed for 4 transmit antennas is defined as follows:

$$\begin{bmatrix} y^{(0)}(1) & y^{(0)}(2) & y^{(0)}(3) & y^{(0)}(4) \\ y^{(1)}(1) & y^{(1)}(2) & y^{(1)}(3) & y^{(1)}(4) \\ y^{(2)}(1) & y^{(2)}(2) & y^{(2)}(3) & y^{(2)}(4) \\ y^{(3)}(1) & y^{(3)}(2) & y^{(3)}(3) & y^{(3)}(4) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & 0 & 0 \\ 0 & 0 & x_3 & x_4 \\ -x_2^* & x_1^* & 0 & 0 \\ 0 & 0 & -x_4^* & x_3^* \end{bmatrix}. \quad (15)$$

For the 4×2 FSTD MIMO scheme, we can show that the instantaneous SNR of the system, for k -th subcarrier, is equivalent to that for a 2×2 STBC MIMO system. Therefore, the probability density function of the SNR is given by a chi-square distribution function as follows [10]:

$$f(\gamma) = \frac{8}{3\bar{\gamma}^4} \gamma^3 e^{-\frac{2}{\bar{\gamma}} \gamma}. \quad (16)$$

In this case, the MGF expression can be obtained by substituting (16) into (7), which yields

$$M_{\bar{\gamma}}(s) = \frac{16}{\bar{\gamma}^4 (s + \frac{2}{\bar{\gamma}})^4}. \quad (17)$$

Similarly to the SFBC case discussed in previous Section, inserting (17) into (9), the average BER expression with M-QAM modulation for FTSD can be written as

$$P_b(E) \cong B \sum_{i=1}^{\sqrt{M}/2} \frac{1}{\pi} \int_0^{\pi/2} \left(\frac{\sin^2 \theta}{\sin^2 \theta + c_i} \right)^4 d\theta \quad (18)$$

where $c_i = \frac{3(2i-1)^2}{2(M-1)} \frac{\bar{\gamma}}{2}$, and the integral can be calculated numerically.

Alternatively, by solving the integral, we obtain a closed-form expression for the average BER of M-QAM modulation as follows

$$P_b(E) \cong B \sum_{i=1}^{\sqrt{M}/2} \mathcal{I}_4(\pi/2, c_i) \quad (19)$$

where the closed-form expression for $\mathcal{I}_4(\cdot, \cdot)$ can be obtained from (14).

Finally, for the sake of comparisons, we express the average BER of the SISO case, that has been derived for Rayleigh fading channels for M-QAM signals [11] (eq. 8.112; Page 256), as follows:

$$P_b(E) \cong B/2 \sum_{i=1}^{\sqrt{M}/2} \left(1 - \sqrt{\frac{1.5(2i-1)^2 \bar{\gamma} \log_2 M}{M-1 + 1.5(2i-1)^2 \bar{\gamma} \log_2 M}} \right) \quad (20)$$

where B is defined earlier.

IV. SIMULATION AND ANALYTICAL RESULTS

In this section, we provide the results obtained from the mathematical expressions derived in this paper. Monte-Carlo simulation results are also provided to show the accuracy of the analysis. The common simulation settings for Monte-Carlo simulations are summarized in Table I.

The average BER performance as a function of $\bar{\gamma} = E_s/N_0$ for SISO and MIMO schemes are shown in Fig. 1 and Fig.2. Fig. 1 shows the results for 16-QAM modulation and Fig.2 presents the results for 64-QAM modulation. It can be seen that the average BER performances of 16-QAM and 64-QAM schemes at high SNRs decrease by factors $\bar{\gamma}^1$, $\bar{\gamma}^2$, and $\bar{\gamma}^4$, for SISO, 2×1 , and 4×2 cases, respectively. Thus, the diversity order (slope of the curves) are equal to 1, 2 and 4, respectively, for the considered cases. As stated earlier, since in 4×2 FSTD, at each time-slot/frequency-slot 2 out of 4 transmit antennas are in use, therefore the diversity order will be $2 \times 2 = 4$. In fact, the corresponding average BER curve for 4×2 FSTD is somehow like the classical 2×2 STBC system, when the channel is not a time-varying channel.

From the figures it is clear that the BER performance improves as the number of transmit or receive antennas increases, as expected. It can be observed that the negative slope of the BER curve for the SISO case is equal to 1, meaning that the diversity order for the SISO case is equal to 1, as expected. The second curves in Fig.1 and Fig.2 represent the BER results of the 2×1 diversity scheme. Asymptotically, the slope of these curves can be observed to be equal to 2, which corresponds to the diversity order of 2×1 system. An SNR (E_s/N_0) gain improvement can also be observed compared to the SISO scheme. From Fig.1, it can be observed that to achieve the BER value of 10^{-3} , the 2×1 diversity scheme needs about 10 dB less in E_s/N_0 , compared to the SISO case.

TABLE I
SIMULATION SETTINGS

Parameter	Setting
Transmission Schemes	SISO; 2×1 SFBC; 4×2 FSTD
Bandwidth	5 MHz
Simulation length	5000 subframes
Channel Type	Flat Rayleigh
Channel knowledge	Perfect
CQI	9 (16-QAM) and 16 (64-QAM)

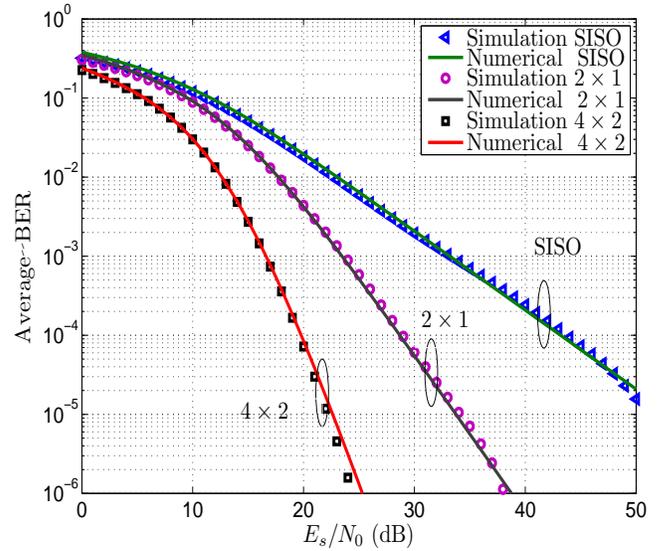


Fig. 1. Numerical Evaluation and Monte-Carlo Simulations of the average BER for 16-QAM modulation.

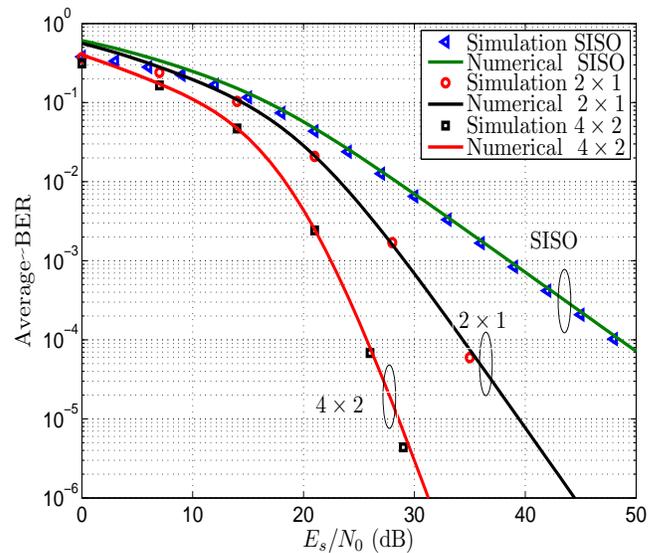


Fig. 2. Numerical Evaluation and Monte-Carlo Simulations of the average BER for 64-QAM modulation.

In Fig.2 for 64-QAM modulation, the BER of 10^{-3} is achieved at $E_s/N_0 = 39$ dB in SISO configuration, however the same value of BER is achieved with only at $E_s/N_0 = 29$ dB in the 2×1 diversity scheme. Thus a SNR gain of 10 dB is clearly observed for the 2×1 diversity scheme. The BER results of the 4×2 diversity scheme for both modulation schemes 16-QAM and 64-QAM are also shown. As described earlier, in high SNR region the slope of that curve tends to be equal to 4. This value corresponds to the diversity order of a 2×2 system.

Finally, it can be observed from Fig.1 and Fig.2 that numerical evaluation results obtained from BER formulas match closely to the BER results obtained from Monte-Carlo simulations. This verifies the accuracy of the mathematical analysis.

V. CONCLUSION

In this paper, we have presented an average BER performance analysis for MIMO schemes in the 3GPP Long Term Evolution (LTE) system over a flat Rayleigh fading channel. The theoretical analysis for two different MIMO schemes in a 5 MHz bandwidth LTE system were presented. To verify the accuracy of the analysis the results of Monte-Carlo simulation for the studied schemes were provided and compared with the theoretical analysis. To show the BER performance improvement in the MIMO schemes, the performance of a SISO configuration was also presented. The results show a good agreement between numerical results and Monte-Carlo simulation results.

REFERENCES

- [1] C. Mehlführer, M. Wrulich, J. C. Ikuno, D. Bosanska, and M. Rupp, "Simulating the long term evolution physical layer," in *Proc. of the 17th European Signal Processing Conference (EUSIPCO 2009)*, Glasgow, Scotland, Aug. 2009.
- [2] Online, available <http://www.nt.tuwien.ac.at/ltesimulator>.
- [3] J. Ikuno, M. Wrulich, and M. Rupp, "System level simulation of lte networks," in *Vehicular Technology Conference (VTC 2010-Spring), 2010 IEEE 71st*, may 2010, pp. 1–5.
- [4] M. Simko, C. Mehlführer, M. Wrulich, and M. Rupp, "Doubly dispersive channel estimation with scalable complexity," in *Smart Antennas (WSA), 2010 International ITG Workshop on*, feb. 2010, pp. 251–256.
- [5] S. Schwarz, M. Wrulich, and M. Rupp, "Mutual information based calculation of the precoding matrix indicator for 3gpp umts/lte," in *Smart Antennas (WSA), 2010 International ITG Workshop on*, feb. 2010, pp. 52–58.
- [6] S. Schwarz, C. Mehlführer, and M. Rupp, "Calculation of the spatial preprocessing and link adaption feedback for 3gpp umts/lte," in *Wireless Advanced (WiAD), 2010 6th Conference on*, june 2010, pp. 1–6.
- [7] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2008.
- [8] S. Sesia, T. Issam, and M. Backer, *LTE The UMTS Long Term Evolution From Theory To Practice*. John Wiley, 2011.
- [9] S. Alamouti, "A simple transmit diversity technique for wireless communications," *Selected Areas in Communications, IEEE Journal on*, vol. 16, no. 8, pp. 1451–1458, oct 1998.
- [10] M. Torabi and D. Haccoun, "Performance Analysis of Joint User Scheduling and Antenna Selection Over MIMO Fading Channels," *IEEE Signal Process. Lett.*, vol. 18, no. 4, pp. 235–238, April 2011.
- [11] M. S. Alouini and M. K. Simon, *Digital Communications over Fading Channels: A Unified Approach to Performance Analysis*. Wiley, 2000.

The Performance Analysis of Complex SSC/MRC Combiner in Rice Fading Channel

Dragana Krstić

Department of Telecommunications,
Faculty of Electronic Engineering,
University of Niš
Niš, Serbia
dragana.krstic@elfak.ni.ac.rs

Petar Nikolić

Tigartyres,
Pilot, Serbia
nikpetar@gmail.com

Aleksandar Stevanović

Mechanical Technical School 15th of May
Niš, Serbia
aleksandar.stevanovic_mts@yahoo.com

Goran Stamenović

Tigar, Pirot, Serbia
goran.stamenovic@tigar.com

Abstract— In this paper, the complex Switch and Stay Combining/Maximal Ratio Combining (SSC/MRC) combiner will be considered. The output signal is observed at two time moments in the presence of Rice fading. Both combiners, SSC such as MRC, are dual branches. The probability density function (PDF) at the output of the complex combiner will be obtained and the bit error rate (BER) for binary phase shift keying (BPSK) modulation will be determined. The results will be shown graphically in some figures. The improvement of using this kind of complex SSC/MRC combiner, with regard to classical MRC and SSC combiners at one time moment, is emphasized.

Keywords— Bit error rate, Probability density function; Complex SSC/MRC combiner; Rice fading; two time moments

I. INTRODUCTION

The fading is one of the most important reason for system performance degradation in mobile communication. Many different communication systems are subjected to fading caused by multipath propagation due to reflection, refraction and scattering by buildings, trees and other large structures [1]. Because of that received signal is a sum of different signals that arrive via different propagation paths.

Some statistical models are used in the literature to describe the fading envelope of the received signal such as Rayleigh and Rician distributions [2]. They are used to characterize the envelope of faded signals over small geographical areas (short term fading). The log-normal and gamma distribution are used when much wider geographical areas is involved.

There are several ways to reduce fading influence on system performances without increasing the signal power and channel capacity. The diversity reception techniques are used extensively in fading radio channels to reduce the effects of fading on system performances [3].

The use of diversity is significant if there is statistical independence in the fading of the received signal in each of the diversity branches. The assumption of statistical

independence between the diversity channels is valid only if they are sufficiently separated [2]. In mobile radio systems the signals at the mobile station become decorrelated as the antenna separation increases. In space diversity systems an antenna separation of 30 to 50 wavelengths is required to have correlation coefficients strictly between zero and one-third, in which case, for a two-channel maximal-ratio system in a Rayleigh-fading environment, the effect of correlation may be ignored [4].

However, there are other cases of practical interest where the assumption of statistical independence is not valid [5]. A long time ago Al-Hussani and Al-Bassion studied the effect of correlation on the performance of a dual-branch MRC combiner for correlated Nakagami fading channel. They found that for the Nakagami-fading environment and for a worst case fading condition and identical signal-to-noise ratio (SNR) in each of two branches, the performance difference between a single channel and dual channels system increases from 3 to 24 dB as the correlation coefficient decreases from unity to zero.

There are several kinds of diversity combining schemes such as Maximal-Ratio Combining (MRC), which is the optimal combining scheme [3]. In this combiner signals from all inputs are summed. Because MRC requires cognition of the channel fading parameters, it is the most complicated combining model [3].

The next by performances is Equal Gain Combining (EGC) [6], and then Selection Combining (SC) and Switch and Stay Combining (SSC), with lower performances. These combining models are simpler and cheaper and they are very often implemented in practice whereas SC and SSC combining models do not require signal combining and fading envelope evaluation [7].

SSC is simplification of the system complexity, but with losing in quality. In this model, the receiver selects an antenna until its value drops below predetermined threshold. Then, the receiver switches to other antenna and stay with it for the next time slot, regardless the channel quality of that

antenna is above or below the threshold. In the literature, mainly dual SSC schemes have been analyzed [8].

By the authors' knowledge in the new open literature these problems are not treated by sampling in two time moments, except by this group of authors. We derived the expressions for joint probability density function of the SSC combiner output signal in the presence of different fading distributions in two time moments (for example in [9] for log-normal) and based on them determined the performance analysis of SSC/SC combiner at two time moments in the presence of log-normal and Rayleigh fading in [10], [11]. The bit error rates for SSC/MRC combiner at two time moments in the presence of log-normal, Rayleigh and Hoyt fading we determined in [12]-[14], respectively.

In this paper, the probability density function and the bit error rate, based on it, for SSC/MRC combiner output signal in the channel with Rice fading and sampling at two time moments during one time interval, will be considered. The system is more complicated compared to classical MRC and SSC systems at one time moment, but with better performances. That imply that bit error rate can be increased and transmit power can be reduced comparing to classical systems.

This paper is organized as follows: Section II presents related works; Section III gives the complex SSC/MRC system model and the probability density functions derivations, such as the bit error rate of the SSC/MRC combiner output signal at two time moments. Sections IV shows the numerical results obtained for performances introduced in previous sections. Finally, the main gains obtained in the paper are pointed out in Section V, Conclusion.

II. SYSTEM MODEL

The model of the SSC/MRC combiner with two inputs, considering in this paper, is presented in Fig. 1.

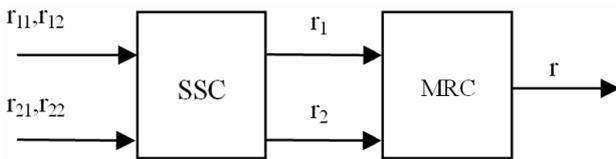


Figure 1. Complex dual SSC/MRC combiner.

The complex SSC/MRC combiner with two branches at two time moments is considered.

At the inputs of the first part of complex combiner the signals are r_{11} and r_{21} at first time moment and they are r_{12} and r_{22} at second time moment. The output signals from SSC part of complex combiner are r_1 and r_2 . The first index represents the branch ordinal number and the other one signs the time moment observed. The indices at the output signal correspond to the time moments considered. The SSC combiner output signals r_1 and r_2 , are the inputs for the MRC combiner. Finally, the overall output signal is r .

The joint probability density function of correlated signals r_1 and r_2 at the SSC combiner output, at two time moments, Rice distributed and with same parameters σ_i and A_i [15], is obtained in closed form from expressions in [9, eq. (3) - (6)] as

$$C_1(r, \sigma, A) = \frac{1}{\sigma^2} e^{-\frac{A^2}{\sigma^2(1+\rho)}} \sum_{i, l_1, l_2, l_3=0}^{\infty} \varepsilon_i \cdot \frac{1}{\rho^{i+2l_1}} \cdot \frac{1}{l_1! l_2! l_3! (i+l_1)! (i+l_2)! (i+l_3)! (2\sigma^2(1-\rho^2))^{2i+l_1+l_2+2l_3}} \cdot A^{2i+2l_2+2l_3} (1-\rho)^{2i+2l_2+2l_3} r^{2i+2l_1+2l_3+1} \cdot e^{-\frac{r^2}{2\sigma^2(1-\rho^2)}} \gamma\left(i+l_1+l_2+1, \frac{r^2}{2\sigma^2(1-\rho^2)}\right) \quad (1)$$

$$C_2(r_1, r_2, \sigma, A) = \frac{r_1 r_2}{\sigma^4(1-\rho^2)} e^{-\frac{r_1^2+r_2^2+2A^2(1-\rho)}{2\sigma^2(1-\rho^2)}} \cdot \sum_{i=0}^{\infty} \varepsilon_i I_i\left(\frac{\rho r_1 r_2}{\sigma^2(1-\rho^2)}\right) I_i\left(\frac{A r_1}{\sigma^2(1+\rho)}\right) I_i\left(\frac{A r_2}{\sigma^2(1+\rho)}\right) \quad (2)$$

where $\varepsilon_i = \begin{cases} 1, & i=0 \\ 2, & i>0 \end{cases}$, parameters σ_i and A_i are noise variance and amplitude, respectively.

For $r_1 < r_T, r_2 < r_T$ it is:

$$p_{r_1 r_2}(r_1, r_2) = P_1 C_1(r_1, \sigma_2, A_2) C_1(r_2, \sigma_1, A_1) + P_2 C_1(r_1, \sigma_1, A_1) C_1(r_2, \sigma_2, A_2) \quad (3)$$

For $r_1 \geq r_T, r_2 < r_T$:

$$p_{r_1 r_2}(r_1, r_2) = P_1 C_1(r_1, \sigma_1, A_1) \frac{r_2}{\sigma_2} e^{-\frac{r_2^2+A_2^2}{2\sigma_2^2}} I_0\left(\frac{r_2 A_2}{\sigma_2}\right) + P_1 C_1(r_1, \sigma_2, A_2) C_1(r_2, \sigma_1, A_1) + P_2 C_1(r_1, \sigma_2, A_2) \frac{r_2}{\sigma_1} e^{-\frac{r_2^2+A_1^2}{2\sigma_1^2}} I_0\left(\frac{r_2 A_1}{\sigma_1}\right) + P_2 C_1(r_1, \sigma_1, A_1) C_1(r_2, \sigma_2, A_2) \quad (4)$$

For $r_1 < r_T, r_2 \geq r_T$:

$$p_{r_1 r_2}(r_1, r_2) = P_1 (1 - Q_1(A/\sigma_1, r_1/\sigma_1)) C_2(r_1, r_2, \sigma_2, A_2) + P_1 C_1(r_1, \sigma_2, A_2) C_1(r_2, \sigma_1, A_1) + P_2 (1 - Q_1(A/\sigma_2, r_1/\sigma_2)) C_2(r_1, r_2, \sigma_1, A_1) +$$

$$+ P_2 C_1(r_1, \sigma_1, A_1) C_1(r_2, \sigma_2, A_2) \quad (5)$$

For $r_1 \geq r_T, r_2 \geq r_T$:

$$\begin{aligned} p_{r_1 r_2}(r_1, r_2) &= P_1 C_2(r_1, r_2, \sigma_1, A_1) + \\ &+ P_1 C_1(r_1, \sigma_1, A_1) \frac{r_2}{\sigma_2^2} e^{-\frac{r_2^2 + A_2^2}{2\sigma_2^2}} I_0\left(\frac{r_2 A_2}{\sigma_2^2}\right) + \\ &+ P_1 C_1(r_1, \sigma_2, A_2) C_1(r_2, \sigma_1, A_1) + \\ &+ P_1 (1 - Q_1(A/\sigma_1, r_1/\sigma_1)) C_2(r_1, r_2, \sigma_2, A_2) + \\ &+ P_2 C_2(r_1, r_2, \sigma_2, A_2) + \\ &+ P_2 C_1(r_1, \sigma_2, A_2) \frac{r_2}{\sigma_1^2} e^{-\frac{r_2^2 + A_1^2}{2\sigma_1^2}} I_0\left(\frac{r_2 A_1}{\sigma_1^2}\right) + \\ &+ P_2 C_1(r_1, \sigma_1, A_1) C_1(r_2, \sigma_2, A_2) + \\ &+ P_2 (1 - Q_1(A/\sigma_2, r_1/\sigma_2)) C_2(r_1, r_2, \sigma_1, A_1) \quad (6) \end{aligned}$$

The outputs of SSC combiner are used as inputs for MRC combiner.

Total conditional signal value at output of the MRC combiner, for equally transmitted symbols of L branch MRC receiver, is given by [1]

$$r = \sum_{l=1}^L r_l \quad (7)$$

For coherent binary signals the conditional BER $P_b(e|\{r_l\}_{l=1}^L)$ is given by [3]

$$P_b(e|\{r_l\}_{l=1}^L) = Q(\sqrt{2gr}) \quad (8)$$

where Q is one-dimensional Gaussian Q-function [3]

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt \quad (9)$$

Gaussian Q-function can be defined using alternative form as [3, 16]

$$Q(x) = \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\frac{x^2}{2\sin^2\phi}\right) d\phi \quad (10)$$

Using the alternative representation of Gaussian-Q function (10), the conditional BER can be expressed as

$$P_b(e|\{r_l\}_{l=1}^L) = \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\frac{gr}{\sin^2\phi}\right) d\phi = \frac{1}{\pi} \int_0^{\pi/2} \prod_{l=1}^L \left(-\frac{gr_l}{\sin^2\phi}\right) d\phi \quad (11)$$

The unconditional BER is obtained by

$$P_b(e) = \underbrace{\int_0^\infty \int_0^\infty \dots \int_0^\infty}_{L} P_b(\{r_l\}_{l=1}^L) \prod_{l=1}^L p_{r_1, r_2, \dots, r_L}(r_1, r_2, \dots, r_L) dr_1 dr_2 \dots dr_L \quad (12)$$

Substituting (11) in (12), $P_b(e)$ is obtained as

$$P_b(e) = \underbrace{\int_0^\infty \int_0^\infty \dots \int_0^\infty}_{L} \frac{1}{\pi} \int_0^{\pi/2} \prod_{l=1}^L \left(-\frac{gr_l}{\sin^2\phi}\right) d\phi p_{r_1, r_2, \dots, r_L}(r_1, r_2, \dots, r_L) dr_1 dr_2 \dots dr_L \quad (13)$$

For dual branch MRC combiner, $P_b(e)$ is

$$P_b(e) = \underbrace{\int_0^\infty \int_0^\infty}_{L} \frac{1}{\pi} \int_0^{\pi/2} \prod_{l=1}^L \left(-\frac{gr_l}{\sin^2\phi}\right) d\phi p_{r_1, r_2, \dots, r_L}(r_1, r_2, \dots, r_L) dr_1 dr_2 \dots dr_L \quad (14)$$

Substituting (3 - 6) in (14), $P_b(e)$ of SSC/MRC combiner can be obtained as:

$$\begin{aligned} P_b(e) &= \frac{1}{\pi} \int_0^{\pi/2} \int_0^\infty \int_0^\infty dr_1 dr_2 d\phi \left(-\frac{gr_1}{\sin^2\phi}\right) \left(-\frac{gr_2}{\sin^2\phi}\right) \cdot \\ &\cdot [P_1 C_1(r_1, \sigma_2, A_2) C_1(r_2, \sigma_1, A_1) + \\ &+ P_2 C_1(r_1, \sigma_1, A_1) C_1(r_2, \sigma_2, A_2)] + \\ &+ \frac{1}{\pi} \int_0^{\pi/2} \int_0^\infty \int_0^\infty dr_1 dr_2 d\phi \left(-\frac{gr_1}{\sin^2\phi}\right) \left(-\frac{gr_2}{\sin^2\phi}\right) \cdot \\ &\cdot \left[P_1 C_1(r_1, \sigma_1, A_1) \frac{r_2}{\sigma_2^2} e^{-\frac{r_2^2 + A_2^2}{2\sigma_2^2}} I_0\left(\frac{r_2 A_2}{\sigma_2^2}\right) + \right. \\ &+ P_1 C_1(r_1, \sigma_2, A_2) C_1(r_2, \sigma_1, A_1) + \\ &+ P_2 C_1(r_1, \sigma_2, A_2) \frac{r_2}{\sigma_1^2} e^{-\frac{r_2^2 + A_1^2}{2\sigma_1^2}} I_0\left(\frac{r_2 A_1}{\sigma_1^2}\right) + \\ &+ P_2 C_1(r_1, \sigma_1, A_1) C_1(r_2, \sigma_2, A_2) \left. \right] + \\ &+ \frac{1}{\pi} \int_0^{\pi/2} \int_0^\infty \int_0^\infty dr_1 dr_2 d\phi \left(-\frac{gr_1}{\sin^2\phi}\right) \left(-\frac{gr_2}{\sin^2\phi}\right) \cdot \\ &\cdot [P_1 (1 - Q_1(A/\sigma_1, r_1/\sigma_1)) C_2(r_1, r_2, \sigma_2, A_2) + \\ &+ P_1 C_1(r_1, \sigma_2, A_2) C_1(r_2, \sigma_1, A_1) + \\ &+ P_2 (1 - Q_1(A/\sigma_2, r_1/\sigma_2)) C_2(r_1, r_2, \sigma_1, A_1) + \\ &+ P_2 C_1(r_1, \sigma_1, A_1) C_1(r_2, \sigma_2, A_2) \left. \right] + \end{aligned}$$

$$\begin{aligned}
 & + \frac{1}{\pi} \int_{r_1}^{\infty} \int_{r_1}^{\infty} \int_0^{\pi/2} dr_1 dr_2 d\phi \left(-\frac{gr_1}{\sin^2 \phi} \right) \left(-\frac{gr_2}{\sin^2 \phi} \right) \cdot \\
 & \quad \cdot [P_1 C_2(r_1, r_2, \sigma_1, A_1) + \\
 & \quad + P_1 C_1(r_1, \sigma_1, A_1) \frac{r_2}{\sigma_2^2} e^{-\frac{r_2^2 + A_2^2}{2\sigma_2^2}} I_0\left(\frac{r_2 A_2}{\sigma_2^2}\right) + \\
 & \quad + P_1 C_1(r_1, \sigma_2, A_2) C_1(r_2, \sigma_1, A_1) + \\
 & \quad + P_1 (1 - Q_1(A/\sigma_1, r_1/\sigma_1)) C_2(r_1, r_2, \sigma_2, A_2) + \\
 & \quad + P_2 C_2(r_1, r_2, \sigma_2, A_2) + \\
 & \quad + P_2 C_1(r_1, \sigma_2, A_2) \frac{r_2}{\sigma_1^2} e^{-\frac{r_2^2 + A_1^2}{2\sigma_1^2}} I_0\left(\frac{r_2 A_1}{\sigma_1^2}\right) + \\
 & \quad + P_2 C_1(r_1, \sigma_1, A_1) C_1(r_2, \sigma_2, A_2) + \\
 & \quad + P_2 (1 - Q_1(A/\sigma_2, r_1/\sigma_2)) C_2(r_1, r_2, \sigma_1, A_1)] \quad (15)
 \end{aligned}$$

III. NUMERICAL RESULTS

Some values of the bit error rate for different types of combiners and correlation parameters, are presented in Fig. 2 and 3, where it is assumed that both inputs have the same channel parameters. It is adopted that r_t is the optimal threshold for the SSC decision [3]:

$$r_t = \sigma \sqrt{\frac{\pi}{2}} e^{-\frac{A^2}{4\sigma^2}} \left[\left(1 + \frac{A^2}{2\sigma^2}\right) I_0\left(\frac{A^2}{4\sigma^2}\right) + \frac{A^2}{2\sigma^2} I_1\left(\frac{A^2}{4\sigma^2}\right) \right] \quad (16)$$

The family of curves for the BER is shown in Fig. 2 versus different distribution parameter A .

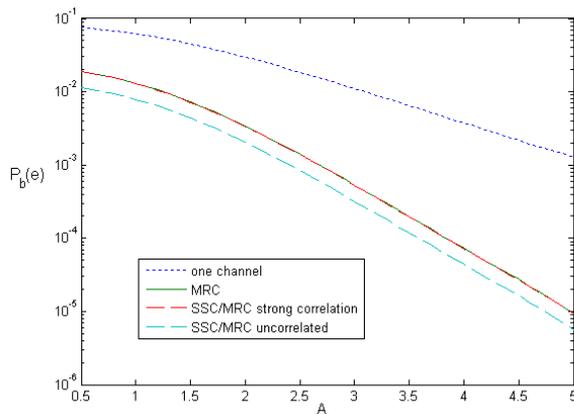


Figure 2. Bit error rate for different types of combiners versus parameter A for $\sigma=1$

It consists of four curves: one for one channel receiver, the other for MRC combiner at one time moment, third for SSC/MRC combiner at two time moments for uncorrelated case, and fourth for very strong correlation.

We can see that SSC/MRC combiner has remarkably better performances for uncorrelated case than MRC combiner considered at one time moment. For $\rho=1$ the BER curves for complex SSC/MRC combiner and for MRC combiner coincide with each other. Thus, it is visible that utilization of this complex SSC/MRC combiner provides better performance of the system for uncorrelated signals, but for strong correlation it is not economic to use complex combiner.

The influence of correlation to the bit error rate of complex SSC/MRC combiner is presented in Fig. 3. The benefits of using this type of combiner increases with decreasing of correlation between input signals.

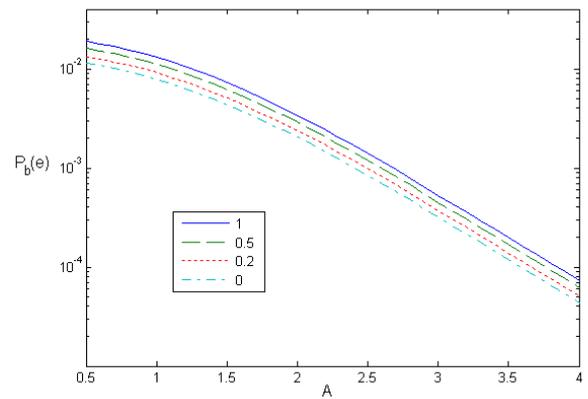


Figure 3. Bit error rate for SSC/MRC combiner parameter A for $\sigma=1$ and for different values of ρ

It is obvious that using of complex SSC/MRC combiner results in better performance of the system, because the BER for uncorrelated SSC/MRC combiner decreases regarding MRC combiner. Also, it is apparent that there is no economic reason for the use of complex SSC/MRC combiner in the case of strong correlations between input signals since then characteristics are not better than for simple MRC combiner.

IV. CONCLUSIONS

The probability density function of output signal at two time moments, for complex dual SSC/MRC combiner, is determined in this paper in closed form expressions. The bit error probability is calculated on that basis.

This is important for the system performance determination for system deciding by two samples. First, the joint probability density function of the SSC combiner output signal at two time moments is derived and used as input of MRC combiner. The obtained results for bit error rate are calculated and shown graphically. The performance improvement of SSC/MRC combiner at two time moments, comparing with classical SSC and MRC combiners, is

described and pointed out for different values of correlation coefficient.

ACKNOWLEDGMENT

This work has been funded by the Serbian Ministry for Science under the projects III-44006 and TR-33035.

REFERENCES

- [1] G. L. Stuber, Principles of Mobile Communications, Norwell, MA: Kluwer Academic Publishers, 1996.
- [2] H. Suzuki, "A statistical model for urban radio propagation," IEEE Trans. Commun., vol. COM-25, July 1977, pp. 673-680.
- [3] M. K. Simon and M. S. Alouni, Digital Communication over Fading Channels, Second Edition, Wiley-Interscience, A John Wiley&Sons, Inc., Publications, New Jersey, 2005.
- [4] D. G. Brennan, "Linear diversity combining techniques," in Proc. IRE, vol. 47, June 1959, pp. 1075-1102.
- [5] E. Al-Hussaini and A. Al-Bassiouni, "Performance of MRC diversity systems for the detection of signals with Nakagami fading," IEEE Trans. Commun., vol. COM-33, Dec. 1985, pp. 1315-1319.
- [6] D. V. Bandur, M. V. Bandur, and M. Stefanović, "Output Signal Characteristics of a Post-detection EGC Combiner with Two ASK Input Signals in the Presence of Fading and Gaussian Noise", Electronics and Electrical Engineering (Elektronika ir elektrotehnika), no. 2(82), T180 Telecommunications engineering, 2008, pp. 41-44.
- [7] N. Sekulović, M. Stefanović, D. Drača, A. Panajotović, and M. Zdravković, "Switch and stay combining diversity receiver in microcellular mobile radio system", Electrical Review (Przeglad Elektrotechniczny), vol. 86, no. 2, Dec. 2010, pp. 346-350.
- [8] A. Panajotović, M. Stefanović, and D. Drača, "Cochannel Interference Effect on BEP Performance of SSC Receiver in Correlated Rician Fading", Journal of the Franklin Institute, vol. 347, no. 7, 2010, pp. 1242-1252.
- [9] D. Krstić, P. Nikolić, F. Destović, and M. Stefanović, "The Joint Probability Density Function of the SSC Combiner Output Signal in the Presence of Log-Normal Fading", Electronics and Electrical Engineering, (Elektronika ir Elektrotehnika), ISSN 1392-1215, No. 3(109), pp. 11-16, 2011.
- [10] M. Stefanović, P. Nikolić, D. Krstić, and V. Doljak, "Outage probability of the SSC/SC combiner at two time instants in the presence of lognormal fading", Przeglad Elektrotechniczny (Electrical Review), ISSN 0033-2097, R. 88 NR 3a/2012, March 2012, pp.237-240.
- [11] P. Nikolić, D. Krstić, M. Milić, and M. Stefanović, "Performance Analysis of SSC/SC Combiner at Two Time Instants in The Presence of Rayleigh Fading", Frequenz. Volume 65, Issue 11-12, Nov. 2011, pp. 319-325.
- [12] D. Krstić, P. Nikolić, G. Stamenović, and M. Stefanović, "Bit error rate for SSC/MRC Combiner at Two Time Instants in The Presence of log-normal Fading", Facta Universitatis. Series Automatic Control and Robotics, ISSN 1820-6417, Vol.10, No 1, 2011, UDC 621.396.94 621.395.38 519.724, pp. 83-95.
- [13] D. Krstic, P. Nikolic, and D. Radenkovic, "The Performances of Complex SSC/MRC Combiner in the Presence of Rayleigh Fading", Network Protocols and Algorithms, vol.4, No.3, 2012. DOI: 10.5296/npa.v4i3.2055
- [14] D. Krstić, P. Nikolić, G. Stamenović, and M. Stefanović, "The Bit Error Rate for Complex SSC/MRC Combiner at Two Time Instants in the Presence of Hoyt Fading", International Journal on Advances in Telecommunications, vol. 5, n 1&2, 2012, http://www.iariajournals.org/telecommunications/tele_v5_n12_2012_paged.pdf, pp. 69-78.
- [15] P.S. Bithas, G.K. Karagiannidis, and N.C. Sagias, "Performance Analysis of a Class of GSC Receivers Over Nonidentical Weibull Fading Channels", IEEE Trans. Veh. Technol., vol. 54, Nov. 2005, pp. 1963-1970.
- [16] J.W. Craig, "A new, simple and exact result for calculating the probability of error for two-dimensional signal constellations," IEEE MILCOM'91 Conf. Rec., Boston, MA, pp. 25.5.1-25.5.5.

Intercell Interference Coordination for the ePDCCH in LTE-Advanced Macrocellular Deployments

David González G, Mario Garcia-Lozano, Silvia Ruiz Boqué
 Escola d'Enginyeria de Telecomunicació i Aeroespacial (EETAC)
 Universitat Politècnica de Catalunya
 Barcelona, Spain
 Emails: david.gonzalez.g@ieee.org, {mariogarcia, silvia}@tsc.upc.edu.

Abstract—This paper investigates several schemes to improve the performance of the enhanced Physical Downlink Control Channel (ePDCCH) in Long Term Evolution Advanced (LTE-A) networks by means of Intercell Interference Coordination (ICIC). Given the flexible design of the ePDCCH, based on frequency division multiplexing, static ICIC techniques such as Soft Frequency Reuse (SFR) can be applied and hence, performance degradations at cell edges can be avoided in contrast to its antecesor, the Physical Downlink Control Channel (PDCCH) in LTE. The study is focused in realistic/irregular deployments, where the amount of intercell interference received at different cells varies considerably making very difficult the task of homogenizing the performance of the ePDCCH over the coverage area. In order to address this problem, the proposed multiobjective scheme adjusts the configuration of SFR at cell level. The problem formulation includes several performance metrics including spectral efficiency, cell edge performance, consumption/amount of control resources and energy requirements. The results reveal that the proposed scheme is able to (1) reduce the average consumption of control resources and, (2) minimize energy needs without penalizing the capacity of data channels.

Index Terms—Long Term Evolution Advanced; LTE-A; Soft Frequency Reuse; SFR; Enhanced Physical Downlink Control Channel; ePDCCH; Multiobjective Optimization

I. INTRODUCTION

According to the conclusions in [1], the mobile Internet mass market becomes a reality. The findings of this survey indicate that a tremendous number of Internet users do it through mobile devices, a 69%, from which 61% use smartphones. Mobile operators have answered to this challenge by investing on promising technologies such as Long Term Evolution (LTE) and its evolution, LTE-Advanced (LTE-A) [2]. Indeed, reliable studies forecast 234 commercial LTE networks in 83 countries by the end of 2013 [3].

In this context, it is expected that significant efforts are being placed on LTE-A, the system called to fulfill the expectations of users and industry in the medium term. LTE-A features an interesting set of novelties with respect to LTE such as wider bandwidths, enhanced downlink and uplink transmission, relaying, support of heterogeneous networks and Machine-Type communications among others [2]. However, all these innovations require reliable means to convey an increasing amount of control information. Thus, in 3GPP Release 11, the need for enhanced capabilities for the Physical Downlink Control Channel (PDCCH) was identified [4]. To be

precise, the design of the PDCCH in LTE is much less flexible than the one in data channels. The structure and operation of the PDCCH is described in [5], but basically, there do not exist mechanisms to perform neither frequency domain scheduling nor Intercell Interference Coordination (ICIC) over the PDCCH and hence, low Signal to Interference plus Noise Ratio (SINR) levels at cell edges, a well know issue in Orthogonal Frequency Division Multiple Access (OFDMA), degrade the performance of the PDCCH. Since the control information carried by the PDCCH is highly sensitive, LTE defines some mechanisms to guarantee the required reliability. The most important one is based on Aggregation Levels (ALs), which consists in grouping several Control Channel Elements (CCEs), the basic control information unit, in order to transmit the PDCCH using more robust transmission formats, i.e., lower coding rates. However, higher ALs increase the consumption of CCEs, thus reducing the capacity of the PDCCH. This situation is critical in scenarios with a large number of users using low-rate services such as VoIP as they tend to heavily load the PDCCH. This issue has been analyzed in [6] and [7].

LTE-A provides alternative protection mechanisms for the PDCCH: carrier aggregation plus cross carrier scheduling [8] in the frequency domain and Almost Blank Subframes (ABSs) [9] in the time domain. However, while cross carrier scheduling is not an option for legacy users, ABSs severely penalizes the capacity and hence, its usage has been reserved for Heterogeneous Networks (HetNets). Thus, in the light of these observations, a new enhanced PDCCH (ePDCCH) was introduced in the Release 11 [10]. The ePDCCH employs Frequency Division Multiplexing (FDM) and hence, it allows frequency domain ICIC. In addition, it is compatible with legacy carriers providing more signaling capacity and it can operate in Multicast-Broadcast Single Frequency Network (MBSFN) subframes [11].

However, given its recent appearance (Release 11, 2012), few studies about the ePDCCH have been reported. Indeed, most of the work done about the ePDCCH has been focused on comparing the performance of its baseline design against the conventional PDCCH. The study presented by Einhaus et al. [12] demonstrates that the ePDCCH outperforms the PDCCH in terms of achievable SINR levels mainly due to its inherent capability to perform frequency domain resource allocation. The work presented by Yi et al. [13] is concentrated on the

design of the search space, i.e., how to allocate the enhanced CCEs (eCCEs) [10], the basic control information unit defined for the ePDCCH, in the physical resources devoted for such purpose. Other related works, such as [8] and [9], as indicated before, are just focused on the mechanisms introduced in the Release 10 such as cross carrier scheduling and ABS. To the best of the authors' knowledge, no work has investigated static ICIC mechanisms applied to the ePDCCH.

Thus, several ICIC strategies based on Soft Frequency Reuse (SFR) [14] are investigated as alternatives to protect the ePDCCH in the context of realistic/irregular macrocellular deployments. Two different multiobjective optimization frameworks are introduced and analyzed. The proposed schemes adjust the operational parameters of SFR and the amount of resources allocated to the ePDCCH in order to optimize several performance metrics such as spectral efficiency, cell edge performance, average consumption of eCCEs, amount of control resources and energy requirements. Therefore, the study presented herein is unique in the sense that it:

- Introduces several effective SFR-based optimization frameworks for the ePDCCH. In fact, not only the performance of this channel is studied, the work also analyses the impact on the capacity of data channels. As a consequence, interesting tradeoffs and design insights are identified.
- Provides means, due to its multiobjective nature, to obtain several network configurations instead of one single solution. This feature is important, because it allows mobile operators to select different configurations according to time-varying network conditions such as load and/or traffic patterns.

The rest of the paper is organized as follows: the next section introduces the system model and provides a brief introduction to the structure of the ePDCCH and the operation of SFR. Section III describes the multiobjective optimization framework and proposed schemes. Finally, the paper is closed with the analysis of numerical results and conclusions in Sections IV and V, respectively.

II. BACKGROUND

A. System Model

This study considers the downlink of a LTE-based cellular network. The system bandwidth B_{SYS} is composed of N_{SC} subcarriers grouped in N_{PRB} Physical Resource Blocks (PRBs). In LTE/LTE-A, a PRB is the minimum allocable resource unit in frequency domain. It is composed of 12 contiguous subcarriers each of them spaced 15 kHz. In time domain, the Transmission Time Interval (TTI) is 1 ms and it contains 14 OFDM symbols. The first 3 symbols are devoted to the PDCCH as it is illustrated in Figure 1. The total available power per cell is $P_{\text{max}}^{\text{Cell}}$. The conclusions obtained in this study are independent of the value of N_{PRB} and hence, more or less PRBs would just shift absolute values. The cellular network, composed of L trisectorial cells, provides service to a coverage area divided in A small area elements (pixels). Given

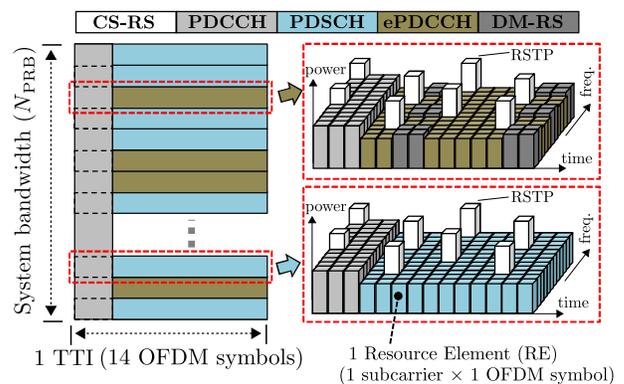


Fig. 1. Structure of the PDCCH and ePDCCH in LTE/LTE-A.

the small granularity used in this study, it is reasonable to assume that within each pixel the average received power and hence, average SINR are constant. Average SINR values \bar{S} are computed based on the average Reference Signals Received Power (RSRP). In LTE and LTE-A, *cell-specific* Reference Signals (CS-RS) are embedded into the system bandwidth to allow for channel estimation, synchronization and cell selection procedures [10]. Due to their importance, CS-RSs are the highest powered components within the downlink signal and they are transmitted with constant power within each cell as it is shown in Figure 1. Thus, the vector $\mathbf{p}_{\text{RSTP}} \in \mathbb{R}^L$ represents the Reference Signals Transmit Power (RSTP) of each cell. The matrix $\mathbf{R}_{\text{RSRP}} \in \mathbb{R}^{A \times L}$ corresponds to the average RSRP at each pixel with respect to each transmitter and it is obtained according to:

$$\mathbf{R}_{\text{RSRP}} = \mathbf{G} \cdot \text{diag}(\mathbf{p}_{\text{RSTP}}) \quad (1)$$

The matrix $\mathbf{G} \in \mathbb{R}^{A \times L}$ contains the Long Term Channel Gain (LTCG) of each pixel with respect to each transmitter. LTCG includes propagation losses, large scale fading and antenna gains. The pixel a (a^{th} row in \mathbf{R}_{RSRP}) is served by the cell l^* from which it receives the highest RSRP, thus:

$$l^* = \underset{l \in \{1, 2, \dots, L\}}{\text{argmax}} \mathbf{R}_{\text{RSRP}}(a, l) \quad (2)$$

Therefore, the binary matrices $\mathbf{S}, \mathbf{S}^c \in \{0, 1\}^{A \times L}$ indicate the coverage of each cell. If a is served by l^* , then $\mathbf{S}(a, l^*) = 1$ and the rest of the a^{th} row is 0. \mathbf{S}^c is the binary complement of \mathbf{S} . It is assumed, without loss of generality, that the power allocated to RSs is the same for all cells and hence, the cell coverage pattern depends on local propagation conditions.

B. Description of the ePDCCH

As it was indicated previously, in order to allow frequency domain ICIC, the ePDCCH is based on FDM as it is illustrated in Figure 1. Note that additional Demodulation Reference Signals (DM-RS) are inserted within the ePDCCH to allow for user-specific beamforming and spatial diversity. Thus, each serving cell can configure a User Equipment (UE) with one or more ePDCCH PRB sets, i.e., a set of contiguous PRBs devoted to allocate the ePDCCH. This user-specific allocation

is transmitted to UE by means of higher layers signaling. The exact position and amount of resources devoted to the ePDCCH can be changed dynamically and it depends on aspects such as system bandwidth, required control capacity and location of the ePDCCH in neighbor cells. Details about the resource allocation control mechanism for the ePDCCH in LTE-A, i.e., how to localize and index the eCCEs within the PRBs carrying the ePDCCH, can be found in [10].

The information transmitted over the ePDCCH includes downlink (and uplink) scheduling grants, power control commands and data required to decode and demodulate OFDM symbols in the downlink (encode and modulate in the uplink) [10]. Given the importance of such information, a target Block Error Rate (BLER) of 1% is pursued for the ePDCCH. Therefore, different ALs in which one or more eCCEs can be grouped have been defined. In this manner, several coding rates provide the required reliability. For a given UE, the selection of the appropriate AL depends on the reported SINR for the subband in which its ePDCCH is allocated. Thus, a user i is assigned with the AL x if its SINR in the subband carrying the ePDCCH is greater than the target SINR of that AL for a BLER of 1% S_x^T . In this study, the focus is precisely on improving the radio quality (SINR) of the ePDCCH by means of SFR, a static ICIC technique described in the next subsection. Other degrees of freedom to enhance the performance of the ePDCCH are dynamic control resource allocation schemes and efficient design of users' search space; examples of these approaches include [8] and [13], respectively.

C. Soft Frequency Reuse

Broadly speaking, the main goal of any ICIC strategy is to enhance the radio channel quality of cell edge users, a well known issue in OFDMA-based cellular technologies such as LTE/LTE-A [15]. As such, SFR accomplishes this target by classifying users in Exteriors (\mathcal{E}) or Interiors (\mathcal{I}) according to their average channel quality (based on CS-RSs and expressed in terms of SINR) and then, applying different power levels to each group in order to reduce the amount of Intercell Interference (ICI) received by cell edge users, thus increasing their SINR. The operation of SFR is illustrated in Figure 2. In order to accomplish such target, a classification threshold S_{TH} must be defined either globally in the network or locally at each cell. This figure has a great impact on the performance of SFR (see [16] and [17]) as it determines the amount of users in each class. Similarly, the bandwidth and power allocated to each group is controlled by means of the parameters β and α respectively. Although SFR proved its effectiveness as an ICIC technique in the context of OFDMA technologies ([18] and [19] are representative examples) its usage was mainly focused on data channels due to the rigid structure of the PDCCH (time-multiplexed). However, as it was introduced previously, the flexible design (based on FDM) of the ePDCCH in LTE-A opens new possibilities from the perspective of ICIC; SFR is certainly an interesting option that is investigated in this paper.

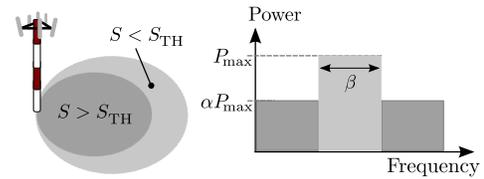


Fig. 2. Operational principle of SFR.

III. RESEARCH FRAMEWORK

A. Multiobjective Problem Formulation

This work investigates the advantages of applying SFR to control channels in LTE-A. However, it is well known that the enhancements in terms of cell edge performance achieved by ICIC techniques are usually obtained at expenses of spectral efficiency losses [16]. For this reason, it is desirable to have a complete picture of the tradeoff existing among conflicting criteria such as spectral efficiency, cell edge performance, energy consumption, etc. Moreover, the problem studied herein adds another interesting perspective: the impact of allocating resources (normally employed for data) to control channels on the overall system performance. Therefore, in order to provide such *visibility*, the performance assessment is based on the joint optimization of the following metrics:

- 1) Maximization of the **average cell capacity** (f_1 [Mbps]): A metric proportional to the system spectral efficiency.
- 2) Maximization of the **capacity of the worst 5% of the coverage area** (f_2 [Mbps]): This indicator indicates the capacity associated to cell edge areas and hence, it is a measure of cell edge performance.
- 3) Minimization of the **average eCCE consumption** (f_3 [eCCE]): This metric reflects the impact of ICI on the radio quality associated to the ePDCCH. It indicates the average consumption of eCCEs per cell.
- 4) Minimization of the **worst eCCE consumption** (f_4 [eCCE]): It corresponds to the average eCCE consumption in the worst cell of the system, i.e., the most interfered cell.
- 5) Maximization of **ePDCCH resources** (f_5 [PRB]): This metric quantifies how many resources are devoted to the ePDCCH. Thus, the maximization of this metric implies more capacity for the control channels. However, it is worth saying that this objective is in conflict with the capacity associated to data channels f_1 .
- 6) Minimization of the **normalized energy consumption** (f_6 [·]): Indicates the energy consumption in the system.

Thus, in order to achieve simultaneous optimization of the previous conflicting objectives, the problem under consideration, i.e., optimization of SFR for the ePDCCH, has been addressed as a multiobjective optimization task. Such problem has been formulated as follows:

$$\begin{aligned} & \text{minimize } \mathbf{f}(\mathbf{x}) \\ & \text{subject to: } \mathbf{x}(l) = x^l \in [x_{\min}^l, x_{\max}^l] \quad \forall l \quad (3) \\ & \mathbf{f}(\mathbf{x}) = [-f_1(\mathbf{x}) \quad -f_2(\mathbf{x}) \quad f_3(\mathbf{x}) \quad f_4(\mathbf{x}) \quad -f_5(\mathbf{x}) \quad f_6(\mathbf{x})] \quad (4) \end{aligned}$$

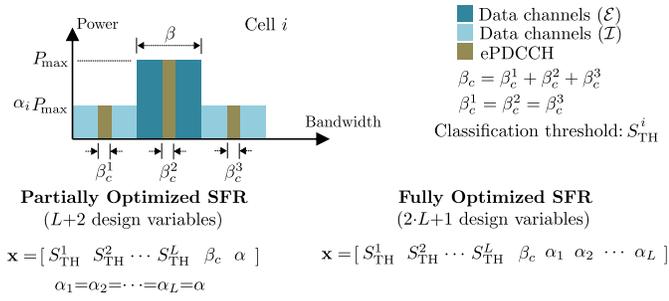


Fig. 3. SFR-based optimization models.

The vectors \mathbf{x} and \mathbf{f} contain the optimization (design) variables and objective function values respectively. The parameters x_{\min}^l and x_{\max}^l are the bounds of the l^{th} design variable. Thus, in order to provide a flexible framework, the optimization of these performance indicators is done by tuning the operational parameters of SFR: α , and S_{TH} (see Figure 2). Two different optimization models are proposed: *Partially Optimized SFR* (POS) and *Fully Optimized SFR* (FOS). Both models are described in the following points:

- 1) **Partially Optimized SFR (POS)**: In this scheme, there are L local design variables (classification thresholds) optimized at cell level plus 2 additional network-wide design variables (β_c and α) that are applied globally in the network, i.e., the same value for all cells. The reason for selecting classification threshold S_{TH}^i as local design variables is twofold: first, the performance of SFR is highly sensitive to this parameter with the advantage that varying this parameter locally has no effect on neighbor cells (for a common value of α) and, second, its usage was demonstrated to be effective in SFR optimization for realistic/irregular deployments [20]. In [20], a multiobjective approach is also employed; however it is focussed exclusively on data channels. The parameters β_c and α determine how many resources are devoted to the ePDCCCH and the power ratio between exterior and interior users respectively. As it can be seen in Figure 3, the resources allocated to the ePDCCCH (controlled by β_c) are distributed between the bandwidth portions of each class of user (\mathcal{E} and \mathcal{I}). The bandwidth sharing coefficient β is kept as an input and its value is set to the maximum avoiding overlapping between cell edge subbands, thus $\beta = 1/3$. Thus, the optimization framework employed by this model corresponds to the following mapping: $\mathbf{x} \in \mathbb{R}^{L+2} \rightarrow \mathbf{f} \in \mathbb{R}^6$. The design target of this scheme is achieving a competitive optimization level while keeping the computational complexity as low as possible; that was the reason why in this model α is defined as a network-wide design variable.
- 2) **Fully Optimized SFR (FOS)**: This scheme is similar to the previous scheme with the only difference that the power ratio α is optimized locally at each cell in order to attain a higher optimization level although at expense of

Function PreliminaryComp(\cdot)

input : $\mathbf{G}, \mathbf{S}, \mathbf{S}^c, \Psi'_j, \eta, \mathbf{v}_\varphi, \text{PRSTP}$
output: $\mathbf{G}_j, \mathbf{S}_j, \mathbf{S}_j^c, \Psi'_j$

```

// STEP 1: Average SINR (based on CS-RS);
1  $\Psi' = [(\mathbf{S} \odot \mathbf{G}) \cdot \text{PRSTP}] \oslash [[(\mathbf{S}^c \odot \mathbf{G}) \cdot \text{PRSTP}] \oplus \eta];$ 
// STEP 2: Azimuth classification;
2  $\mathbf{t} \leftarrow \text{AzimuthClass}(\mathbf{v}_\varphi, \mathbf{S});$ 
// STEP 3: Segmentation;
3 for each  $j \in \mathcal{J}$  do
4    $\{\mathbf{G}_j, \mathbf{S}_j, \mathbf{S}_j^c, \Psi'_j\} \leftarrow \text{Segmentation}(\mathbf{t}, j, \mathbf{G}, \mathbf{S}, \mathbf{S}^c, \Psi');$ 
5 end
    
```

 Fig. 4. Preliminary computations required for evaluating the objective functions: $f_i, i = 1, 2, \dots, 6$.

additional computational cost. This model corresponds to the following mapping: $\mathbf{x} \in \mathbb{R}^{(2L)+1} \rightarrow \mathbf{f} \in \mathbb{R}^6$

B. Evaluation of Objective Functions

In order to evaluate the objective functions, a preliminary set of computations need to be performed as indicated in Figure 4. Therefore, the objective functions are evaluated by means of the pseudo-code shown in Figure 5. Both pseudo-codes are explained in the following points. Note that the optimization model POS is indeed a particular case of the scheme FOS, when $\alpha_1 = \alpha_2 = \dots = \alpha_L = \alpha$. Thus, the process of evaluating the objective function values is described for the most general case, i.e., the model FOS.

- **PreliminaryComp()**: The first step (line 1) corresponds to the computation of average SINR values based on RSRP. These figures are used to classify the pixels as \mathcal{E} or \mathcal{I} . Note that \odot, \oslash and \oplus indicate Hadamard (point-wise) operations and η corresponds to the noise power. Then, Function **AzimuthClass()** classifies cells according to their azimuth φ (stored in $\mathbf{v}_\varphi \in \mathbb{R}^L$) as follows:

$$j = \begin{cases} 0 & 0^\circ \leq \varphi < 120^\circ \\ 1 & 120^\circ \leq \varphi < 240^\circ \\ 2 & 240^\circ \leq \varphi < 360^\circ \end{cases} \quad (5)$$

Pixels are also classified according to the type of serving cell (see (2)). Thus, cells and pixels belong to one of the sets \mathcal{A}_j with $j \in \mathcal{J} = \{0, 1, 2\}$. This classification is stored in the vector $\mathbf{t} \in \mathbb{N}^A$, where each element (representing one pixel) indicates the type of serving cell it belongs to. Next, Function **Segmentation()** (line 4) pulls out from the matrices $\mathbf{G}, \mathbf{S}, \mathbf{S}^c$ and Ψ' the rows whose corresponding value in the vector \mathbf{t} is equal to $j, \forall j \in \mathcal{J}$. In other words, once instructions 3-5 are executed, each one of these matrices is segmented in $|\mathcal{J}|$ submatrices ($\mathbf{S}_j, \mathbf{S}_j^c, \mathbf{G}_j$ and Ψ'_j) each of them having L columns, but a different number of rows and so:

$$\mathbf{S}_j, \mathbf{S}_j^c, \mathbf{G}_j, \Psi'_j \in \mathbb{R}^{|\mathcal{A}_j| \times L} \quad \forall j \in \mathcal{J}$$

- **ObjFunc()**: Once Function **PreliminaryComp()** is executed, and having fixed $\mathbf{G}_j, \mathbf{S}_j, \mathbf{S}_j^c, \Psi'_j \forall j \in \mathcal{J}$, the

Function ObjFunc(**x**)

```

input : x, Gj, Sj, Sjc, Ψj',
output: f
// Classification thresholds;
1 STH ← x(1 : L);
// ePDCCCH bandwidth;
2 βc ← x(L + 1);
// Power ratios: vα(l) = αl;
3 vα ← x(L + 2 : 2 · L + 1);
// Classification of pixels:  $\mathcal{E}$  or  $\mathcal{I}$ ;
4 for each j ∈  $\mathcal{J}$  do
5   | Cj ← Class(Sj, STH, Ψj');
6 end
// Relative coverage per cell;
7 Φ ← RelCov(S1, ..., S|J|, C1, ..., C|J|);
// Power matrices;
8 Pser =  $\begin{bmatrix} P_{\max} & P_{\max} & \dots & P_{\max} \\ \alpha_1 P_{\max} & \alpha_2 P_{\max} & \dots & \alpha_L P_{\max} \end{bmatrix}^T$ ;
9 for each s = 0 : ((L/3) - 1) do
10  | Pintbase(s) =  $\begin{bmatrix} P_{\max} & \alpha_{(3s)} P_{\max} & \alpha_{(3s)} P_{\max} \\ \alpha_{(3s+1)} P_{\max} & P_{\max} & P_{\max} \\ P_{\max} & \alpha_{(3s+1)} P_{\max} & \alpha_{(3s+1)} P_{\max} \\ \alpha_{(3s+2)} P_{\max} & \alpha_{(3s+2)} P_{\max} & P_{\max} \\ \alpha_{(3s+2)} P_{\max} & P_{\max} & \alpha_{(3s+2)} P_{\max} \end{bmatrix}^T$ ;
11 end
12 Pint = [Pintbase(0)T Pintbase(1)T ... Pintbase((L/3) - 1)T]T;
// Bandwidth matrix;
13 B ← Bsys · [ (1/3 - (1/3 · βc)) ((2/3) - (2/3 · βc)) ];
// Reset accumulators;
14 f1 ← 0, f2 ← 0, r̂ ← [ ];
// Resize pγ (pointers to per-cell SINR values);
15 Resize(pγ, (2 · L));
// For each group of pixels (for each j ∈  $\mathcal{J}$ );
16 for each j ∈  $\mathcal{J}$  do
17  | P̄int ← Pint(: , 2j : 2j + 1);
18  | Δj ← [((Sj ⊙ Gj) · Pser) ⊙ [ [(Sjc ⊙ Gj) · P̄int] ⊕ η ];
19  | UpdatePerCellSINR(pγ, Δj);
20  | Ψj ← Δj ⊙ Cj;
21  | Λj ← LinkPer(Ψj);
22  | f1 ← f1 + [B · [(ΛjT · Sj) ⊙ Φ]] · 1;
23  | r ← [ [Sj · (ΦT · diag(B))] ⊙ Λj] · 1;
24  | r̂ ← [r̂ rT];
25 end
26 f2 ← CapPerc(r̂);
27 f5 ← βc · NPRB;
28 f6 ← 1 - ((2/3) · (1 - α));
// For each cell l;
29 for each l = 1 : L do
30  | fε ← CDF(pγ(2 · l - 1)); fI ← CDF(pγ(2 · l) + 1);
31  | pε0 ← (1 - fε(S00)); pε1 ← (fε(S00) - fε(S10));
32  | pε2 ← (fε(S10) - fε(S20)); pε3 ← fε(S20);
33  | pI0 ← (1 - fI(S00)); pI1 ← (fI(S00) - fI(S10));
34  | pI2 ← (fI(S10) - fI(S20)); pI3 ← fI(S20);
35  | Γεl ← (pε0 · AL0) + (pε1 · AL1) + (pε2 · AL2) + (pε3 · AL3);
36  | ΓIl ← (pI0 · AL0) + (pI1 · AL1) + (pI2 · AL2) + (pI3 · AL3);
37  | με ← (1/Φ(0, l)); μI ← (1/Φ(1, l));
38  | Γ(l) ← [(με/(με + μI)) · Γεl] + [(μI/(με + μI)) · ΓIl];
39 end
40 f3 ← mean(Γ);
41 f4 ← max(Γ);
42 f ← [ -f1/L -f2 f3 f4 -f5 f6 ];

```

Fig. 5. Pseudo code corresponding to the evaluation of the objective functions considered in this study.

vector **f** (objective function values) depends exclusively on **x**, a potential solution or SFR configuration. First, Function Class() (line 5) calculates the binary matrices

$\mathbf{C}_j \in \mathbb{R}^{|\mathcal{A}_j| \times 2} \forall j \in \mathcal{J}$ indicating the class (\mathcal{E} or \mathcal{I}) to which each pixel belongs to. The value ‘1’ in column 0 or 1 indicates the pixel belongs to \mathcal{E} or \mathcal{I} respectively. Next, Function RelCov() (line 7) computes the matrix $\Phi \in \mathbb{R}^{2 \times L}$. While the columns of Φ correspond to cells, the rows 0 and 1 are the inverse of the number of pixels classified as \mathcal{E} and \mathcal{I} (at each cell) respectively. As SFR assigns specific portions of the system bandwidth to each group of pixels (\mathcal{E} and \mathcal{I}), Φ is required to guarantee bandwidth proportionality and hence, calculate true average values. Instructions from line 8 to 12 create the matrices corresponding to the particular power allocation specified by **x**. In line 13, the bandwidth allocated to each class of users is calculated as a function of β_c . In line 15, a vector of $2 \cdot L$ pointers is created. They point to vectors storing the SINR distribution of each class (\mathcal{E} and \mathcal{I}) at each cell. The loop comprising the lines 16 to 25 determines the capacity in bps associated to each pixel; once this loop is executed, the vector $\hat{\mathbf{r}} \in \mathbb{R}^A$ contains such information. Function UpdatePerCellSINR() updates the statistic of the SINR per class/cell by using the SINR values stored in the matrix $\Delta_j \in \mathbb{R}^{|\mathcal{A}_j| \times 2}$. Function LinkPer() computes for every single element of $\Psi_j \in \mathbb{R}^{|\mathcal{A}_j| \times 2}$ a non-decreasing function of the SINR. Thus LinkPer(z) = $\text{Log}_2(1 + z)$ [bps/Hz], i.e., the Shannon bound. Finally, the metrics related to the consumption of eCCEs (f_3 and f_4) are obtained in a per-cell basis by means of the instructions in the loop comprising the lines 29 to 39. To be precise, the Cumulative Density Function (CDF) of each class (\mathcal{E} and \mathcal{I}) at each cell is estimated by means of Function CDF() using the information pointed from \mathbf{p}_γ . Note that $\mathbf{p}_\gamma(2 \cdot l - 1)$ and $\mathbf{p}_\gamma(2 \cdot l)$ point to the vectors containing the SINR values of the classes \mathcal{E} and \mathcal{I} (at the l^{th} cell) respectively. Once the CDFs are obtained, the instructions from line 31 to 38 determine the average eCCE consumption at each cell taking into account the relevant information of each AL. To be precise, for the AL x , the number of eCCEs and target SINR are specified by AL_x and S_x^T respectively. Recall that S_x^T represents the target SINR of the AL_x for a BLER of 1%.

C. Multiobjective Evolutionary Optimization

In the light of the previous subsection, where objective functions were defined, it is clear that the *domain* (search space) created by the optimization variables is a n -dimensional space, where $n = L + 2$ and $n = 2 \cdot L + 1$ for POS and FOS respectively. Given that the objective space (or *image* defined by such objectives is not only highly non-linear, non-convex, but also full of discontinuities and local optima [21], traditional optimization approaches such as Simplex [22] can not be employed as they are susceptible to be trapped in local minima. Other techniques such as Sequential Quadratic Programming based methods [23] require convexity (a very strong assumption in this context) to guarantee convergence.

Summarizing, the problem of interest requires of an optimization tool fulfilling the following set of features:

- It must be able to find good solutions by efficiently exploring the search space.
- It should operate efficiently with multiple criteria and a large number of design variables.
- It should not have strong requirements on objective functions such as linearity, convexity, continuity or differentiability.

Multiobjective evolutionary algorithms (MOEAs) [24] fulfill the previous requirements and hence, their usage in this context has been investigated. MOEAs are population-based metaheuristics that simulate the process of natural evolution. In MOEAs, a population of individuals (candidate solutions) is iteratively modified by means of two basic principles: selection and variation. While selection tries to imitate the battle for reproduction among living beings, variation mimics their inherent ability of creating new (better adapted) individuals through recombination and mutation. A well-known MOEA has been selected for this study: The Non-dominated Sorting Genetic Algorithm II (NSGA-II) [25]. The reason is that this algorithm provides means to accomplish desired features such as: elitism, fast convergence and good distribution of solutions. Interested readers are referred to [26] for further details. Finally, recall that multiobjective optimization has an additional advantage: the solution is not only one single configuration but a set of so many Pareto Efficient solutions. A solution \mathbf{x}_1 is preferred to (dominates in the Pareto sense) another solution \mathbf{x}_2 , ($\mathbf{x}_1 \succ \mathbf{x}_2$), if \mathbf{x}_1 is better than \mathbf{x}_2 in at least one criterion and no worse with respect to the remaining ones. The set of nondominated solutions \mathcal{X}^* is called Pareto Front.

IV. NUMERICAL RESULTS

In order to perform numerical evaluations, a realistic cellular deployment covering the city of Vienna and its surroundings has been considered. This deployment represents the system model described in Subsection II-A. Actual parameter values are shown in Table I. To be precise, the cellular layout is composed of 60 tri-sectorial cells. The evaluation area corresponds to a urban subarea of $2.75 \times 2.625 \text{ km}^2$. The digital elevation model and cell parameters have been obtained from the MORANS initiative [27]. The propagation model is the COST 231-Walfish-Ikegami. Figure 6 shows the cellular layout and the resulting propagation pattern for one site as reference. In addition, the CDFs of \bar{S} corresponding to the coverage of each cell are also shown. It can be observed how much different those CDFs are. It can be noticed that areas in which the average SINR $\bar{S} < 0 \text{ dB}$ vary from 5 to 70% of the total coverage at different cells. These *differences* are exploited by the proposed framework in order to fit SFR to the scenario under consideration and hence, maximize its benefits.

The list of parameters employed for NSGA-II together with the rest of parameters employed in numerical evaluations are shown in Table I. In general, the performance (convergence) of

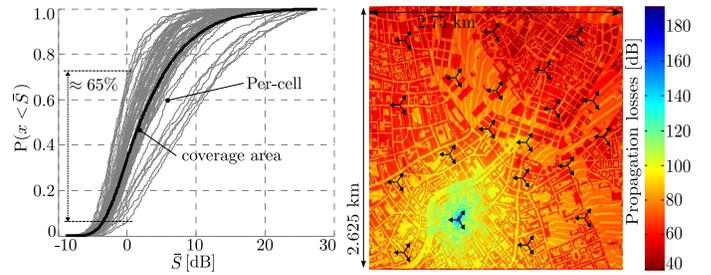


Fig. 6. Realistic deployment used in numerical evaluations. Left: Per-cell and global SINR statistic. Right: one single site propagation losses.

TABLE I
EVALUATION SETTING

Parameter	Value
NSGA-II: Population size	300
NSGA-II: Max number of generations	3000
NSGA-II: Crossover probability	1.0
NSGA-II: Mutation probability	$1/(L+1)$
NSGA-II: Design variables type	Binary coded
System model: [$A \ L \ \eta$]	[288750 60 -125 dBm/15kHz]
System model: Pixel resolution	$5 \times 5 \text{ m}^2$
System model: [$B_{\text{SYS}} \ N_{\text{SC}} \ N_{\text{PRB}}$]	[5.40 MHz 360 30]
System model: [$P_{\text{Cell}}^{\text{max}} \ \beta$]	[43.0 dBm 1/3]
System model: [$P_{\text{max}} \ \text{RSTP}$]	[17.4 18.5] dBm/15kHz
System model: [$AL_0 \ AL_1 \ AL_2 \ AL_3$]	[1 2 4 8]
System model: [$S_0^T \ S_1^T \ S_2^T \ S_3^T$]	[9.25 2.50 -0.50 -2.50] dB [28]
MO problem: [$\beta_{c,\text{min}} \ \beta_{c,\text{max}}$]	[0.10 0.30]
MO problem: [$\alpha_{l,\text{min}} \ \alpha_{l,\text{max}}$]	[0.15 0.60]
MO problem: [$S_{\text{TH},\text{min}}^l \ S_{\text{TH},\text{max}}^l$]	[-3.00 6.00] dB

NSGA-II depends on its operational parameters. The following points are practical calibration guidelines:

- **Population size:** It is widely accepted that populations larger than 100 individuals only provide marginal gains and the same global convergence is obtained [29]. However, in some cases such as the problem herein, larger populations will have the benefit of getting more solutions or network settings at expenses of computational cost.
- **Termination criterion:** The execution of NSGA-II finishes when the improvement of each objective function is less than 0.001% after a block of 40 generations.
- **Genetic operators:** Crossover and mutation are important in genetic algorithms to preserve elitism and achieve good diversity respectively. Crossover and mutation rates (probabilities) follow the recommendations given in [25].

For comparison, several benchmarks have been considered. These references include the important case of Full Frequency Reuse (FFR) and baseline designs of SFR according to the bandwidth proportionality criterion (see [15] and [30]), i.e., schemes in which optimization is not available and hence, the parameters α , β and S_{TH} are selected according to the SINR statistic (observed in the whole coverage area) and applied globally. The configurations of these benchmarks together with their corresponding objective function values are shown in

TABLE II
 BENCHMARKS

Ref	Type	S_{TH} [dB]	α	β	β_c	f_1 [Mbps]	f_2 [Mbps]	f_3 [eCCE]	f_4 [eCCE]	f_5 [PRB]	f_6 [·]
x_{FFR}^1	FFR	N/A	N/A	N/A	0.10	8.38	7.54	4.25	6.26	3.00	1.00
x_{FFR}^2	FFR	N/A	N/A	N/A	0.20	7.45	6.70	4.25	6.26	6.00	1.00
x_{FFR}^3	FFR	N/A	N/A	N/A	0.30	6.52	5.87	4.25	6.26	9.00	1.00
x_{SFR}^1	SFR	0.00	0.40	1/3	0.10	7.51	8.11	4.08	5.67	3.00	0.60
x_{SFR}^2	SFR	0.00	0.40	1/3	0.20	6.50	7.03	4.08	5.67	6.00	0.60
x_{SFR}^3	SFR	0.00	0.40	1/3	0.30	5.84	6.31	4.08	5.67	9.00	0.60

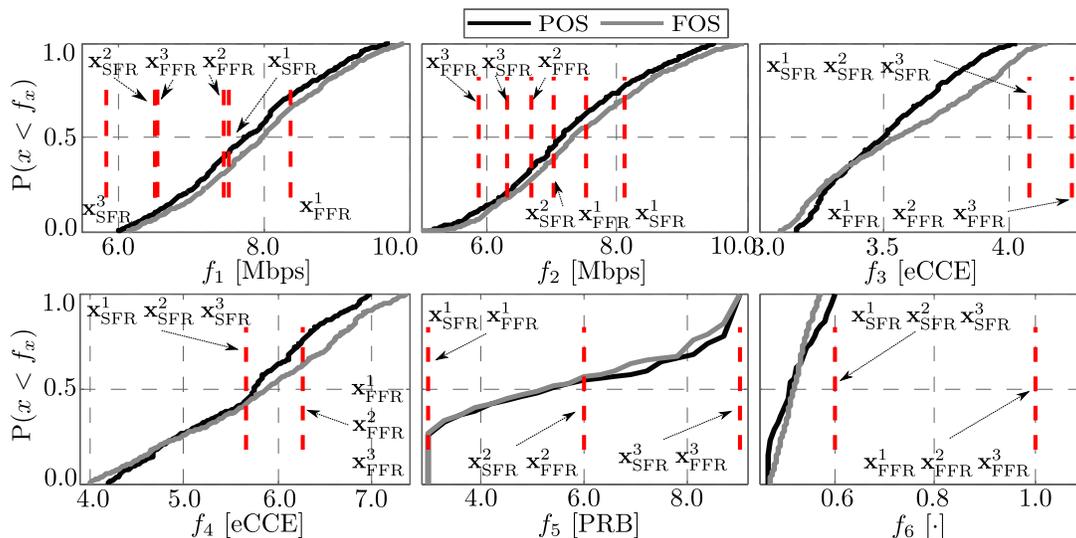


Fig. 7. Overall performance obtained by means of the optimization frameworks POS and FOS.

Table II. The statistic of the different performance metrics corresponding to the sets of nondominated solutions of the models POS and FOS, \mathcal{X}_{POS}^* and \mathcal{X}_{FOS}^* respectively, is shown in Figure 7. In the figure, the performance of the benchmarks is also plotted (in red). These results indicate, on the one hand, that the SFR settings obtained through the proposed schemes are able to offer different tradeoffs among f_1 , f_2 , f_4 and f_5 while, on the other hand, the performance in terms of f_3 and f_6 is always enhanced with respect to the benchmarks. This means that by applying the solutions in the sets \mathcal{X}_{POS}^* and \mathcal{X}_{FOS}^* it is possible to select different tradeoffs among the performance metrics such as spectral efficiency vs. cell edge performance (f_1 vs. f_2) or spectral efficiency vs. resources allocated to control channels (f_1 vs. f_5), but always achieving gains in terms of eCCE and energy consumption (f_3 and f_6). Note that the performance obtained by means of the model FOS is slightly better than the model POS. This is especially noticeable in terms of f_1 and f_2 . Moreover, the smallest values of f_3 and f_4 are also attained through elements in \mathcal{X}_{FOS}^* , which is expected given the higher number of optimization variables used by this model. However, focusing on each benchmark separately, Figure 8 shows the gains that can be achieved with respect to each benchmark and performance metric. Such gains are computed based on the subsets of solutions dominating

each benchmark in the Pareto sense and hence, no loss is expected. The cardinality of these subsets for each model is presented in Table III. Clearly, the proposed frameworks always succeed in finding non-empty sets of SFR configurations outperforming each reference scheme. The gains shown in Figure 8 make evident that significant improvement are obtained with respect to the important case of Full Frequency Reuse. Both subfigures include an example indicating how they must be read. For instance, focusing on Figure 8b and x_{FFR}^2 , it has been obtained that the energy consumption (f_6) can be decreased up to 53% without losses in terms of any other performance metric. The proposed schemes outperform all the benchmarks achieving gains of 20% or more in at least one of the objective functions.

These results confirm the effectiveness of the optimization models presented herein as they are able not only to **achieve effective ICIC for the ePDCCH** (this is evident given the gains in terms of f_3 and f_4) but also to **improve spectral efficiency and cell edge performance** (f_1 and f_2). This is quite interesting if one takes into account that some of the SFR settings found so far, accomplish the previous and, in addition, increase the amount of resources available for the ePDCCH.

In order to provide additional insights on the tradeoff between the performance metrics, 2D profiles are presented

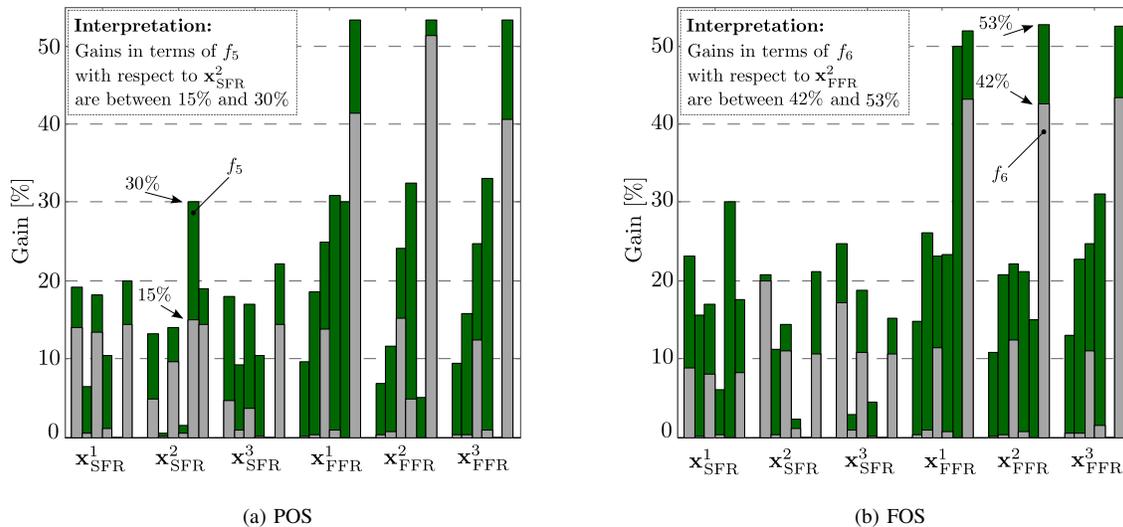


Fig. 8. Gains achieved by means of the proposed optimization models.

in Figure 9. 2D profiles are generated by projecting the Pareto Front onto different planes in order to obtain alternative representations providing better insights about the tradeoff between any pair of objective functions. The figure includes 3 different planes: f_1 - f_2 , f_1 - f_5 and f_3 - f_4 . While in the two first cases, it is notorious the conflicting nature of each pair of metrics, note how the last case, f_3 - f_4 , suggests that SFR settings attaining better average eCCE consumption are also able to minimize the consumption of control resources in the worst cell of the system, which is not evident from the analogy that could be made a priori with f_1 - f_2 . Thus, as a conclusion, while the amount of resources allocated to the ePDCCH f_5 presents a tradeoff with the spectral efficiency f_1 and hence f_2 , the average consumption of control resources is basically linked to the ICIC mechanism employed to protect the ePDCCH. In practice, it will also depend on system load, and in particular, on the amount of low-rate services such as VoIP that tend to heavily load control channels.

Finally, to close the section, Figure 10 illustrates the convergence pattern of both optimization frameworks in terms of the hypervolume of the Pareto front \mathcal{L} [31]. It can be seen that initially the convergence of FOS is slower than POS. This is due to the higher number of design variables that need to be adjusted in FOS. However, after a certain number of generations (approx. 500), FOS features better convergence, explaining so the differences in performance previously found. However, it is worth mentioning that both schemes are valid since there are cases in which the computational cost is a limiting factor, for instance, in very large scale scenarios, and hence a 3% in convergence can be traded by an interesting saving of 26% in processing that can be achieved by means of POS.

 TABLE III
 CARDINALITY OF THE SUBSETS DOMINATING EACH BENCHMARK.

Model	x_{FFR}^1	x_{FFR}^2	x_{FFR}^3	x_{SFR}^1	x_{SFR}^2	x_{SFR}^3
POS	36	3	22	8	2	10
FOS	47	7	29	13	3	10

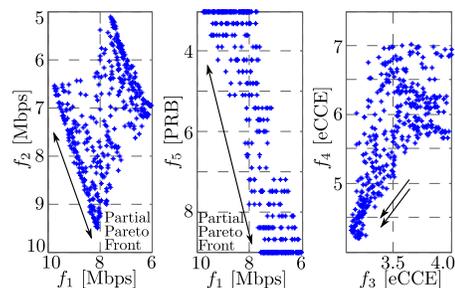


Fig. 9. 2D representations of the resulting Pareto Front: POS model.

V. CONCLUSIONS AND FUTURE WORK

A novel multiobjective algorithm has been proposed for SFR optimization aiming at improving the performance of the ePDCCH in LTE-A networks. The optimization frameworks presented herein have been designed to be applied in realistic/irregular macro/micro cellular deployments, for which only (commonly available) propagation information is required. The proposed schemes succeed in finding good quality SFR settings achieving significant gains with respect to baseline designs and important reference schemes such as full frequency reuse. It was found that, by means of the proposed framework, the overall performance of the ePDCCH can be increased **without penalizing** neither spectral efficiency nor cell edge performance. Indeed, results indicate that the effective intercell interference coordination achieved by the

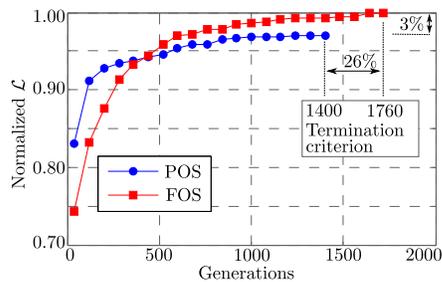


Fig. 10. Convergence pattern of the optimization models: POS and FOS.

proposed schemes results in significant gains in terms of average eCCE consumption (f_3) and transmitted power (f_6). The gains with respect to the important case of full frequency reuse range from 12% to 30% and 11% to 25% for POS and FOS respectively while reducing the transmitted power at least 40% in any of these cases.

Finally, as future research item, adaptive mechanisms able to cope with heterogeneous/irregular traffic distributions both in time and space are going to be investigated. Also, more realistic traffic models (real-time and elastic services) are expected to be included into our framework.

ACKNOWLEDGMENT

This work has been funded through the project TEC2011-27723-C02-01 (Spanish Industry Ministry) and the European Regional Development Fund (ERDF).

REFERENCES

- [1] Mobile web watch 2012, Accenture, 2012, Available online at: www.accenture.com, [retrieved: Feb, 2013].
- [2] S. Sesia, I. Toufik, and M. Baker, *The UMTS Long Term Evolution: From Theory to Practice*, 2nd ed. John Wiley & Sons, 2011.
- [3] Status of the LTE ecosystem, Global mobile Suppliers Association (GSA), 2013, Available online at: www.gsacom.com/news/, [retrieved: Jan, 2013].
- [4] S. Ye, S. Wong, and C. Worrall, "Enhanced physical downlink control channel in LTE advanced release 11," *IEEE Communications Magazine*, vol. 51, no. 2, pp. 82–89, February 2013.
- [5] Group Radio Access Network (GRAN), Technical Specifications: 36.211 (Physical Channels and Modulation), 36.213 (Physical Layer Procedures), 3GPP, 2008, Release 8.
- [6] J. Puttonen, H.-H. Puupponen, K. Aho, T. Henttonen, and M. Moisio, "Impact of control channel limitations on the LTE VoIP capacity," in *Networks (ICN)*, 2010 Ninth Int. Conference on, Apr 2010, Muires, France.
- [7] D. González, M. García-Lozano, S. Ruiz, and J. Olmos, "On the role of downlink control information in the provision of QoS for NRT services in LTE," in *Vehicular Technology Conference (VTC Spring)*, 2012 IEEE 75th, May 2012, pp. 1–5, Yokohama, Japan.
- [8] Y. Yan, A. Li, A. Harada, and H. Kayama, "Enhanced downlink control channel resource allocation algorithm for cross-carrier scheduling in LTE-Advanced carrier aggregation system," in *Vehicular Technology Conference (VTC Spring)*, 2011 IEEE 73rd, May 2011, pp. 1–5, Budapest, Hungary.
- [9] Y. Zhu, A. Li, and A. Harada, "Novel method to improve control channel reliability in LTE-Advanced heterogeneous network," in *Vehicular Technology Conference (VTC Fall)*, 2012 IEEE 76th, Sept. 2012, pp. 1–5, Québec City, Canada.
- [10] Group Radio Access Network (GRAN), Technical Specifications: 36.211 (Physical Channels and Modulation), 36.213 (Physical Layer Procedures), 3GPP, 2012, Release 11.
- [11] Group Radio Access Network, TS 22.146: Multimedia Broadcast/Multicast Service, 3GPP, Jun 2008, v8.4.0.
- [12] M. Einhaus, C. Wengerter, J. Ohlhorst, and S. Feng, "Performance study of an enhanced downlink control channel design for LTE," in *Vehicular Technology Conference (VTC Spring)*, 2012 IEEE 75th, 2012, pp. 1–5, Yokohama, Japan.
- [13] W. Yi, Z. Hua, and W. Jianming, "The search space design for enhanced downlink control channel in LTE-Advanced system," in *Wireless Communications and Mobile Computing Conference (IWCMC)*, 2012 8th International, 2012, pp. 322–326, Venice, Italy.
- [14] Huawei, R1-050507: Soft Frequency Reuse Scheme for UTRAN LTE, 3GPP, May 2005, TSG RAN WG1 Meeting #41: Athens, Greece.
- [15] D. González, M. García-Lozano, S. Ruiz, and J. Olmos, "On the need for dynamic downlink intercell interference coordination for realistic LTE deployments," *Wireless Communications and Mobile Computing*, pp. 1–26, 2012, John Wiley & Sons, Ltd. DOI: 10.1002/wcm.2191.
- [16] T. Novlan, R. Ganti, A. Ghosh, and J. Andrews, "Analytical Evaluation of Fractional Frequency Reuse for OFDMA Cellular Networks," *Wireless Communications*, *IEEE Transactions on*, no. 99, pp. 1–12, Sep 2011.
- [17] D. González G., M. García-Lozano, S. Ruiz, and J. Olmos, "An analytical view of static intercell interference coordination techniques in OFDMA networks," in *IEEE Wireless Communications and Networking Conference (WCNC)*, Apr 2012, pp. 1–6, Paris, France.
- [18] D. Jia, G. Wu, S. Li, G. Li, and X. Zhu, "Dynamic soft-frequency reuse with inter-cell coordination in OFDMA networks," in *Computer Communications and Networks (ICCCN)*, Proceedings of 20th International Conference on, Aug 2011, pp. 1–6, Hawaii, United States.
- [19] L. Chen and D. Yuan, "Soft Frequency Reuse in Large Networks with Irregular Cell Pattern: How Much Gain to Expect?" in *Personal, Indoor and Mobile Radio Communications (PIMRC)*, IEEE 20th International Symposium on, 2009, pp. 1467–1471, Tokyo, Japan.
- [20] D. Gonzalez G., M. Garcia-Lozano, S. Ruiz Boque, and D. Lee, "Optimization of soft frequency reuse for irregular LTE macrocellular networks," *Wireless Communications*, *IEEE Transactions on*, pp. 1–14, 2013.
- [21] T. Weise, *Global Optimization Algorithms - Theory and Application*, 2nd ed. Self-Published, Jun. 26, 2009, available online at <http://www.it-weise.de/>, [retrieved: Mar, 2013].
- [22] D. Gale, "Linear programming and the Simplex method," *Notices of the AMS*, vol. 54, no. 3, pp. 364–369, March 2007.
- [23] Gill, P. E. and Wong, E., "Sequential Quadratic Programming Methods," Department of Mathematics, University of California, San Diego, La Jolla, CA, Tech. Rep. NA-10-03, Aug 2010.
- [24] C. A. Coello, G. B. Lamont, and D. A. Van Veldhuizen, *Evolutionary Algorithms for Solving Multi-Objective Problems*, 2nd ed. Springer: Genetic and Evolutionary Computation Series, 2007.
- [25] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *Evolutionary Computation*, *IEEE Transactions on*, vol. 6, no. 2, pp. 182–197, Apr 2002.
- [26] J. Jürgen Branke, K. Deb, K. Miettinen, and R. E. Slowiński, *Multiobjective Optimization: Interactive and Evolutionary Approaches*. Springer: Lecture Notes in Comp. Science 5252, 470p, 2008.
- [27] Verdone, R. *et al*, "MORANS White Paper - Update," COST 273, Greece, Tech. Rep. available as TD(04)062, Jan. 26–28, 2004.
- [28] HTC, R1-122331: eREG and eCCE Definitions for ePDCCH, 3GPP, May 2012, TSG RAN WG1 Meeting #69: Prague, Czech Republic.
- [29] J. C. Spall, *Introduction to stochastic search and optimization*, 1st ed. Wiley-Interscience, 2003.
- [30] D. González G., M. García-Lozano, S. Ruiz, and J. Olmos, "Static Inter-Cell Interference Coordination Techniques for LTE Networks: A Fair Performance Assessment," *Lecture Notes in Computer Science*, vol. 6235, no. 68, pp. 211–222, Sep 2010, Springer Berlin / Heidelberg.
- [31] E. Zitzler and L. Thiele, "Multiobjective optimization using evolutionary algorithms - A comparative case study," in *Parallel Problem Solving from Nature V*. Springer, 1998, pp. 292–301.

Modeling and Simulation of Radio Signals Attenuation Using Informed Virtual Geographic Environments (IVGE)

Mehdi Mekni

University of Minnesota
Crookston, Minnesota, United States
mmekni@umn.edu

Abstract—A radio communication system is a complex dynamic phenomenon where transmitter and receiver antennas are constantly constrained by the physical environment in which they are deployed. In the real world, radio transmissions are subject to propagation effects which deeply affect the received signals because of geographic and environmental characteristics (foliage and vegetation, buildings, mountains and hills, etc.). Multi-Agent Geo-Simulation aims to simulate such phenomena involving a large number of autonomous situated actors (implemented as software agents) evolving and interacting within a representation of the physical environment. Using a geo-computation approach, we propose to use an Informed Virtual Geographic Environment (IVGE) along with MAGS paradigm. In addition, we propose a multi-agent prototype to analyze the attenuation effect due to the radio signal's traversal between antennas (simulated as software agents) through terrain shape, vegetation area, and buildings using a 3D line-of-sight computation technique.

Keywords—Line of Sight; Excess attenuation; Vegetation and Foliage; Radio Propagation; Informed Virtual Geographic Environments

I. INTRODUCTION

Rapid advances in wireless communications have made mobile data applications a high-growth area of development. So far, most applications only focus on geographic data collection and access using geographic information systems. However, many emergency management applications need such geographic data in order to ensure that field workers and command center operators collaborate under acceptable operating conditions. Under emergency conditions, emergency systems need to quickly establish an *ad hoc*, ground-level network of radio stations mounted on temporary command centers, vehicles, or temporary masts, interacting with moving field operators using mobile devices.

Radio Frequency (RF) communication does have some limitations that must be considered. The maximum line-of-sight range between two shoulder-height devices is limited to 12km considering the curvature of the earth but not considering refraction of radio waves. The actual range may be considerably less depending on transmission power and receiver sensitivity, and the radio signal can be attenuated or

degraded due to obstruction resulting from its interactions with features on its transmission path. In an urban environment, possible obstructions include buildings, trees, and bridges (Figure 1). The difficulty of evaluating an ad hoc radio network with hundreds of nodes and various levels of mobility operating in a complex geographic environment (i.e., rugged terrain, dense foliage, buildings, etc.) motivated us to use our IVGE model to analyze communication attenuation. In order to analyze the communication attenuation in real geographic environments, we propose to use the Multi-Agent Geo-Simulation (MAGS) approach involving an enriched description of the geographic environment representation to precisely compute the radio transmission's attenuation. The radio transmission is computed using the line of sight between two points located in the IVGE.

During the last decade, the Multi-Agent Geo-Simulation (MAGS) approach has attracted a growing interest from researchers and practitioners to simulate phenomena in a variety of domains including traffic simulation, crowd simulation, urban dynamics, and changes of land use and cover, to name a few [1]. Such approaches are used to study phenomena (i.e., car traffic, mobile robots, mobile networks, crowd behaviours, etc.) involving a large number of simulated actors (implemented as software agents) of various kinds evolving in, and interacting with, an explicit description of the geographic environment called Virtual Geographic Environment (VGE). Nevertheless, simulating such autonomous situated agents remains a particularly difficult issue, because it involves several different research domains: geographic environment modeling, spatial cognition and reasoning, situation-based behaviours, etc. When examining situated agents in a VGE, whether for gaming or simulations purposes, one of the first questions that must be answered is how to represent the world in which agents navigate. Since a geographic environment may be complex and large-scale, the creation of a VGE is difficult and needs large quantities of geometrical data describing the environment characteristics (terrain elevation, location of objects and agents, etc.) as well as semantic information that qualifies space (building, road, park, etc.). Current approaches usually consider the environment as a monolithic structure,

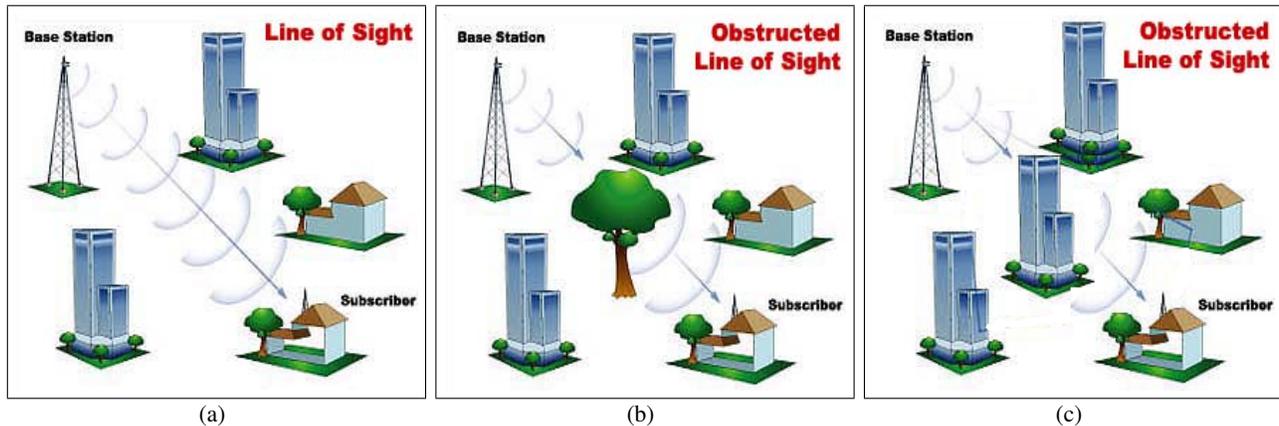


Figure 1: Radio signal propagation: (a) an obstruction-free propagation; (b) propagation obstructed by vegetation and foliage; and (c) propagation obstructed by buildings.

which considerably limits the way that large-scale, real world geographic environments and agent's spatial reasoning capabilities are handled.

In this project, our goal is to analyze the radio communication attenuation in complex, dynamic and large-scale geographic environments. In order to achieve such a goal, a geographic environment model should precisely represent geographic features. It should also integrate several semantic notions characterizing these geographic features. Since we deal with large-scale geographic environments, it would be appreciable to have a VGE organized efficiently in order to reduce the spatial computation algorithms such as line-of-sight algorithms. There is also a need for autonomous situated agents representing antennas, either transmitters or receivers which are able to communicate using radio signal propagation in presence of both *static* and *dynamic* obstacles located in the VGE. Static obstacles correspond to areas that affect radio signal propagation such as walls, fences, trees, rivers, etc. Static obstacles also include obstructions resulting from terrain elevation (mountains, hills, etc.). Dynamic obstacles correspond to moving or stationary antennas which may interfere with the radio communication system.

In this paper, we propose a novel approach to model and simulate radio communication networks using multi-agent systems evolving in and interacting with a virtual geographic environments. The rest of the paper is organized as follows: in Section II, we provide a brief overview on related work in the field of environment representation and analysis of radio communications. Section III presents our methodology for the creation of informed virtual geographic environments. Section IV highlights the unique properties of our IVGE model which easily and efficiently enable spatial reasoning algorithms and geometrical computations such as line-of-sight computation. Section V details the radio communication analysis tool. Finally, Section VI concludes

and presents our perspectives.

II. RELATED WORKS

In this section we provide a brief overview of prior works related to *environment representation*, and *analysis of radio communications* in virtual environments.

A. Environment Representation

Virtual environments and spatial representations have been used in several application domains. For example, Thalmann *et al.* proposed a virtual scene for virtual humans representing a part of a city for graphic animation purposes [2]. Donikian *et al.* proposed a modelling system which is able to produce a multi-level data-base of virtual urban environments devoted to driving simulations [3]. More recently, Shao *et al.* proposed a virtual environment representing the New York City's Pennsylvania Train Station populated by autonomous virtual pedestrians in order to simulate the movement of people [4]. However, since the focus of these approaches is computer animation and virtual reality, the virtual environment usually plays the role of a simple background scene in which agents mainly deal with geometric characteristics. Indeed, the description of the virtual environment is often limited to the geometric level, though it should also contain topological and semantic information for other types of applications. Therefore, most interactions between agents and the environment are usually simple, only permitting to plan a path in a 2D or 3D world with respect to free space and obstacle regions.

B. Analysis of Radio Communications

Jiancheng *et al.* proposed a prediction of mobile radio propagation by regression analysis of signal measurements, investigated path loss characteristics of mobile radio signals

in urban and rural areas [5]. The results were used to evaluate the accuracy of Okumura-Hata and Lee prediction models [6]. In his perspectives on the effects of harmattan on radio frequency waves, Dajab computed harmattan dust particle densities in air and observed that attenuation due to dust particles increases as harmattan intensity increases [6]. Eyo *et al.* researched into microwave signal attenuation in harmattan weather along Calabar-Akampkpa terrestrial line of sight microwave link [7]. Using measured results in conjunction with some meteorological data they were able to deduce the mean signal level, fog attenuation, fade depth and fade margin of the link.

Most works achieved towards on radio propagation modeling and simulation do not consider the geographic environments in which the communication system is located (See Figure 2 for an illustration of a typical radio communication system). Signal specialists currently have limited capabilities for predicting radio performance in complex geographic environments that may involve jammers and may have combinations of open terrain and obstructed locations [8]. Foliated areas, buildings, or terrain may cause obstructions to the radio line-of-sight path [9]. Therefore, there is a need for a communication analysis tool that helps fill this void [10]. We propose to use geo-computation techniques in order to build a tool for the analysis of radio communications attenuation involving a geometrically-accurate and semantically-informed virtual geographic environment model. This tool is an easy to use, stand-alone, GUI-based application that runs on a PC and provides a rich set of functionalities to aid the user to compute the total path loss of a given operational scenario that is directly coupled to an operational area using reliable GIS data.

This tool enables the user to plan radio deployments and determine link connectivity using actual radio parameters, taking into account the presence of obstacles, while accounting for the excess attenuation due to terrain, foliage (vegetation), and building obstructions. To compute the total path loss, our tool uses the following parameters: height of transmitter antenna (meters), height of receive antenna (meters), transmitter antenna position (x, y, z), receiver antenna position (x, y, z), and frequency of operation (GHz).

III. COMPUTATION OF IVGE DATA

In this section, we present our automated approach to compute the IVGE data directly from vector GIS data. This approach is based on four stages which are detailed in this section (Figure 3): *input data selection*, *spatial decomposition*, *maps unification*, and finally the generation of the *informed topologic graph*.

GIS Input Data Selection: The first step of our approach consists of selecting the different vector data sets

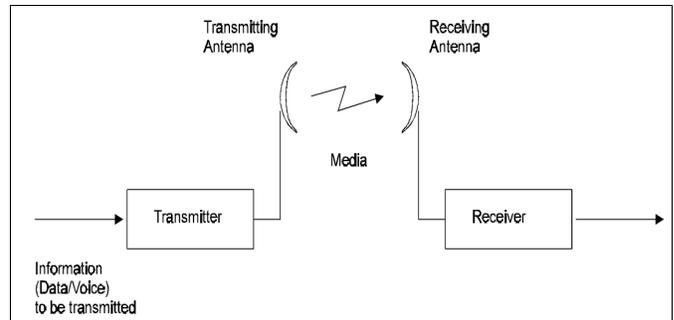


Figure 2: Typical Radio Communication System

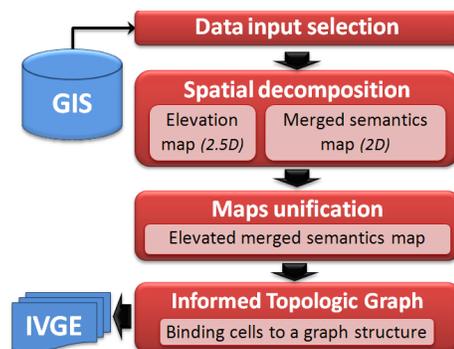


Figure 3: The four stages to obtain an IVGE from GIS data. All the stages are automatic but the first.

which are used to build the IVGE. The only restriction concerning these data sets is that they must respect the same scale. The input data can be organized into two categories. First, *elevation layers* contain geographical marks indicating absolute terrain elevations. As we consider 2.5D IVGE, a given coordinate cannot have two different elevations, making it impossible to represent tunnels for example. Multiple elevation layers can be specified, and if this limitation is respected, the model can merge them automatically. Second, *semantic layers* are used to qualify various types of data in space. Each layer indicates the physical or virtual limits of a given set of features with identical semantics in the geographic environment, such as roads or buildings. The limits can overlap between two layers, and our model is able to merge the information.

Spatial Decomposition: The second step consists of obtaining an exact spatial decomposition of the input data into cells. This process is entirely automatic using Delaunay triangulation, and can be divided into two parts in relation to the previous phase. First, an elevation map is computed, corresponding to the triangulation of the elevation layers. All the elevation points of the layers are injected into a 2D triangulation, the elevation being considered as an attribute of each node. This process produces an environment subdivision composed of connected triangles. Such a subdivision provides information about coplanar areas: the

elevation of any point inside a triangle can be deduced by using the elevation of the three original data points to form a plane. Second, a merged semantics map is computed, corresponding to a constrained triangulation of the semantic layers. Indeed, each segment of a semantic layer is injected as a constraint which keeps track of the original semantic data by using an additional attribute for each semantic layer. The obtained map is then a constrained triangulation merging all input semantics. Each constraint represents as many semantics as the number of input layers containing it.

Merging Elevation and Semantics Layers: The third step to obtain our IVGE consists of unifying the two maps previously obtained. This phase can be depicted as mapping the 2D merged semantic map onto the 2.5D elevation map in order to obtain the final 2.5D elevated merged semantics map. First, preprocessing is carried out on the merged semantics map in order to preserve the elevation precision inside the unified map. Indeed, all the points of the elevation map are injected into the merged semantics triangulation, creating new triangles. This first process can be dropped if the elevation precision is not important. Then, a second process elevates the merged semantics map. The elevation of each merged semantics point P is computed by retrieving the corresponding triangle T inside the elevation map, i.e., the triangle whose 2D projection contains the coordinates of P . Once T is obtained, the elevation is simply computed by projecting P on the plane defined by T using the Z axis. When P is outside the convex hull of the elevation map then no triangle can be found and the elevation cannot be directly deduced. In this case, we use the average height of the points of the convex hull which are visible from P .

Informed Topologic Graph: The resulting unified map now contains all the semantic information of the input layers, along with the elevation information. This map can be used as an *Informed Topologic Graph* (ITG), where each node corresponds to the map's triangles, and each arc corresponds to the adjacency relations between these triangles. Then, common graph algorithms can be applied to this topological graph, and graph traversal algorithms in particular. One of these algorithms retrieves the node, and therefore the triangle, corresponding to given 2D coordinates. Once this node is obtained, it is possible to extract the data corresponding to the position, such as the elevation from the 2.5D triangle, and the semantics from its additional attributes. Several other algorithms can be applied, such as path planning or graph abstraction, but they are out of the scope of this paper and will not be detailed here.

IV. PROPERTIES OF INFORMED VGE

The subdivision of space into convex cells allows us to preserve the original geometric definition of the geographic environment, unlike the grid-based representations

that discretise the environment. Furthermore, the proposed data reorganization produces triangles that feature good properties: convexity which facilitates the geometric calculations; support of heterogeneous geometric constraints (points, segments, polygons); Since each constraint is linked to its nearest neighbor, it is easy to compute the widths of the bottlenecks in the virtual geographic environments. The width computation corresponds to the minimum of borders' width that are not qualified as obstacle (Figure 4(a)).

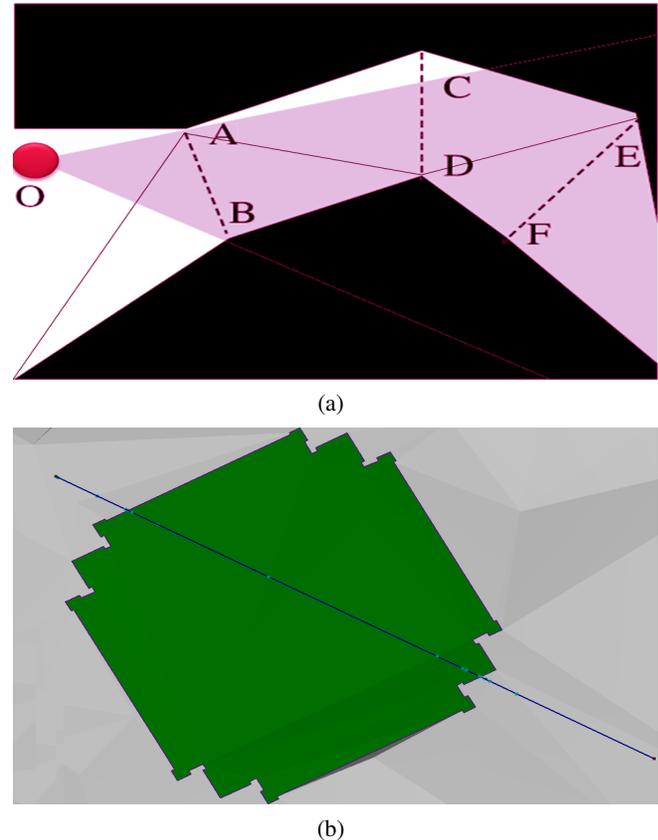


Figure 4: (a) Computation of the widths of the bottlenecks; (b) Computation of the ray tracing.

The subdivision of space into convex cells also allows us to extract an informed topologic graph of the environment featuring relatively few nodes compared to grid-type representations. Additionally, the triangulation is not dependent on a fixed spatial scale for the environment, but only on its complexity (number of constrained segments). It should also be emphasized that curved geometries will produce a lot of triangles since they are represented by a large number of constrained segments. However, since the produced triangulation is represented as a graph, it is possible to abstract it in order to reduce the number of elements. All these properties are of interest to address the issue of line-of-sight computation.

A. Line of Sight's Computation

The spatial subdivision provides a structure of convex cells which facilitates and accelerates the calculation of ray tracing in three dimensions. We define the radius α using the following information: the position of the origin p ; the direction vector \vec{d} ; and the maximum distance considered. Let $Get_{free}(Cell)$ and $Get_{constrained}(Cell)$ be two functions returning respectively the list of free (S_{free}) and constrained (S_{const}) borders bounding the convex cell $Cell$. Let $N(Cell, b)$ be a function returning the normal vector to the border b which belongs to the cell $Cell$ and directed towards the inside of the cell. Finally, let us note $\wp(\beta)$ the $2X2$ rotation matrix of \vec{d} . The test checking if there is an intersection between the ray and the border b , which links the vertices I and J , is performed using the following expression:

$$(\alpha \leq 0) \wedge (\beta \times \gamma) \leq 0 \quad (1)$$

The parameters α , β , and γ , in equation 1, are computed as detailed in equations 2, 3 and 4.

$$\alpha = \vec{d} \cdot N(Cell, b) \quad (2)$$

$$\beta = \left(\frac{I+J}{2} - p \right) \quad (3)$$

$$\gamma = \wp\left(\frac{\pi}{2}\right) \cdot (I - p) \times \left(\left(\frac{I+J}{2} - p \right) \times \left(\frac{\pi}{2} \cdot (J - p) \right) \right) \quad (4)$$

The line of sight computation algorithm proceeds as follows:

- **Step 1:** the cell $Cell$ containing the source of the line of sight vector \vec{LoS} is determined.
- **Step 2:** an intersection test is performed between \vec{LoS} and each border b of $Cell$.
- **Step 3:** compute $S_{free}(Cell)$ using $Get_{free}(Cell)$ and $S_{const}(Cell)$ using $Get_{constrained}(Cell)$.
- **Step 4:** if no intersection is found with borders from $S_{free}(Cell)$, then \vec{LoS} must intersect with a border from $S_{const}(Cell)$.
- **Step 5:** the border b is pushed back to the list of borders crossed by \vec{LoS} .
- **Step 6:** the cell $Cell$ is pushed back to the list of cells crossed by \vec{LoS} .
- **Step 7:** the cell sharing the border b which intersects with \vec{LoS} becomes the current cell. Proceed to Step 2.

The line of sight algorithm, due to its low computational cost, can be extensively used in MAGS involving a large number of agents evolving in a complex IVGE.

V. ANALYSIS OF RADIO COMMUNICATION

Planning communications links requires the ability to assess the performance of each link in the presence of a number of degrading factors. In addition to the normal free space signal attenuation loss, other losses reduce the signal level. These additional losses can be caused by obstructions such as buildings and vegetation (foliage). Analyzing obstruction losses can be difficult because of the geometric, topologic, and semantic characteristics of the geographic environment and the need for path loss models.

We use a ray-tracing approach which determines the path that a radio signal takes to arrive at the receiver's position from a given transmitter within our 3D IVGE. The ray-tracer implemented in the tool uses optimized algorithms for detecting direct (i.e., line of sight) paths (Figure 4(b)). An analytic ray-tracing technique [11] is used rather than the approximate technique, for example, as used in [12]. For the analytical approach, the transmitter and receiver are each modelled as infinitesimally small points such that paths are computed precisely and cannot be duplicated or missed as sometimes could happen in the approximate approach. Obviously, the analytical technique is more precise and reliable. Our approach computes the total source-to-destination path length and then determines whether the vector defined by the source and destination points (locations in the IVGE) passes through an obstruction area. Doing so, it is able to compute the total path loss between transmitter and receiver antennas. One of three cases may occur: 1) *Obstruction-Free Path*: Vector does not penetrate any obstruction; 2) *Obstruction Block Penetration*: Vector penetrates one or more foliage obstruction blocks and/or one or more building obstruction blocks; and 3) *Ground Penetration*: Vector penetrates the ground (earth) once or several times.

The computations performed by our tool, to quantify the path attenuation for each of the three cases defined above, are based on the following well-established mathematical models described in [13].

Plane-Earth Attenuation Model: Let L_p be the path attenuation using the plane-Earth model (dB):

$$L_p = 40\text{Log}(D) - 20\text{Log}(H_t) - 20\text{Log}(H_r) \quad (5)$$

where D is the total source-to-destination path length (meters), and H_t and H_r are the heights of the transmitter and receiver antennae above ground level, respectively (meters).

Free-Space Attenuation Model: Let L_{fs} be the path attenuation using the Free-Space model (dB):

$$L_{fs} = 32.45 + 20\text{Log}(D) + 20\text{Log}(f) \quad (6)$$

where D is the total source-to-destination path length (meters) and f is the RF frequency (GHz).

Obstruction Block Penetration Model: Let L_B be the path attenuation term due to propagation through a building obstruction block (dB) [14]:

$$L_B = K_1(0.6)^f + K_2 D_B \quad (7)$$

where f is the RF frequency (GHz), K_1 is a constant used to map the first expression above to building penetration data reported in [13]. $K_1 = 35$; K_2 is a constant to account for the attenuation (per meter) of the signal within the building. $K_2 = 1$ (dB/m); and D_B is the distance that the signal propagates through the building (meters). The first term in Equation 7 accounts for the penetration into and out of the building by the signal and was derived from data reported in [13] using a regression analysis technique. The second term in the equation above accounts for the attenuation through the building and is based on data reported by Willassen in [15].



(a)



(b)

Figure 5: Simulation of radio communications' attenuation; yellow lines correspond to obstacle-free line-of-sight radio signal propagation; red lines correspond to obstructed line-of-sight; (1) represents the transmitter antenna implemented using the agent paradigm; (2) an example of a plane-earth obstruction; (3) an example of a block-penetration obstruction.

Foliage obstruction Attenuation Model: If the penetration is through a foliage obstruction block, the tool computes an excess path attenuation term called L_f using the Wiessberger model [16] as follows:

$$L_f = 1.33f^{0.284}.D_f^{0.588}, 14 < D_f \quad (8)$$

$$L_f = 0.45f^{0.284}.D_f^{1.0}, 0 \leq D_f \leq 14m \quad (9)$$

where D_f is the distance that the signal propagates through the foliage obstruction (meters) and f is the RF frequency (GHz). The term "excess attenuation" refers to the additional attenuation above the basic transmission loss, for a given path length, in the absence of foliage. Our analysis approach applies equations 5 to 9 to calculate the basic transmission loss for the link. Thus, the total path attenuation for the obstruction penetration case called L_{total} is calculated as follows:

$$L_{total} = \max(L_p, L_{fs})_{Def f} + \text{sum}(L_f) + \text{sum}(L_B) \quad (10)$$

where $\max(L_p, L_{fs})_{Def f}$ means that this term is calculated at $Def f$; $Def f$ is the effective path length over which the *Plane-Earth* and *Free-Space* path attenuation model is applied and is equal to the total path length minus the sum of the building obstruction block path length segments or $\text{sum}(DB)$; $\text{sum}(L_f)$ is the sum of the excess attenuation terms in dB due to signal propagation through the foliage obstruction block(s) (dB); and $\text{sum}(L_B)$ is the sum of the path attenuation terms due to propagation through the building obstruction block(s) (dB).

Figure 5 illustrates the agent-based simulation tool that we developed in order to implement our approach to analyze the radio signal attenuation in informed virtual geographic environments. This figure presents a snapshot of the simulation at time t_0 with an agent representing a transmitter antenna and several agents representing receiver antennae. Red lines in Figure 5(a) highlight the strength of the radio signal attenuation with comparison to an arbitrary user-defined threshold Δ : (red) for severe signal attenuation, (green) for acceptable signal attenuation. Moreover, Figure 6 shows how to collect the list of cells crossed by the radio signal propagation path.

VI. DISCUSSION AND CONCLUSION

Using reliable GIS data along with the line of sight algorithm (ray tracing feature) provided by our IVGE allows the system to compute the exact locations of intersections occurring between the radio signal propagation path and the terrain shape (Figure 5). Our IVGE also allows us to collect the list of cells crossed by the radio signal propagation path and to determine which analytic model to apply in order to precisely compute the path loss.

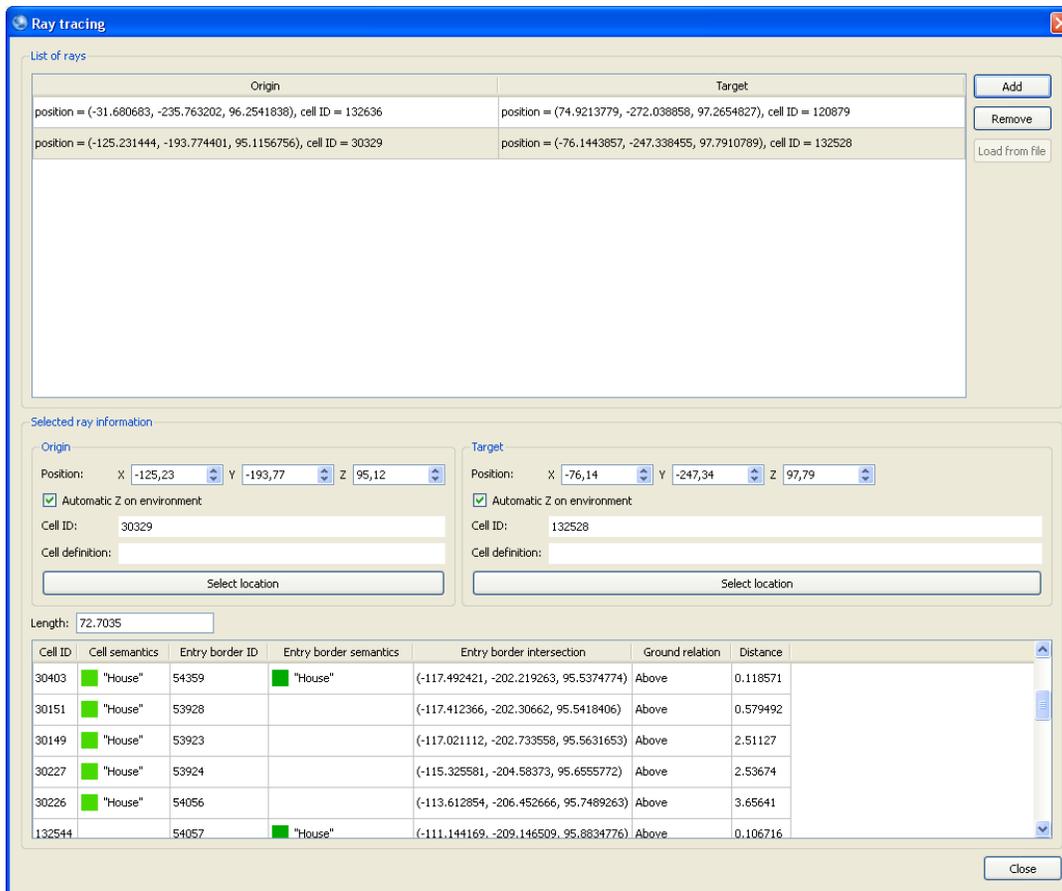


Figure 6: The graphic user-interface for the computation of the radio signal line-of-sight; Cells and borders.

We have shown a tool that leverages the enriched description of the IVGE and computes the radio signal attenuation due to buildings, foliage and field obstructions. However, other phenomena can also degrade the radio signal transmission. Examples of such phenomena include transmitter power, receiver sensitivity, and radio signal’s absorption, reflection, and scattering from interaction with features on or near its transmission path. The proposed tool should be easily extended to take into account absorption, reflection and scattering phenomena when computing the radio signal attenuation.

In the future, we propose to extend our tool in order to integrate an advanced ray-tracing process [17] which combines both the geometric optics and Keller’s [18] geometric theory of diffraction (GTD). Moreover, we propose to include the uniform theory of diffraction (UTD) [19] extension to GTD which removes the inaccuracies close to the incident and reflection boundaries.

To conclude, our geometrically-precise and semantically-enhanced IVGE enables us to provide wireless network planners with a tool for the analysis of the communications’

attenuation. In contrast with mathematical models which only approximate the radio signal attenuation based on a coarse-grained qualification of the geographic environment: *urban*, *suburban* and *rural*, we compute more precisely the radio signal propagation path and qualify obstructions in order to apply the appropriate analytical model.

ACKNOWLEDGEMENT

This research was supported in part by the Grant in Aid provided by the University of Minnesota. The author would like to thank the reviewers for their valuable comments.

REFERENCES

- [1] I. Benenson and P. Torrens, *Geosimulation: Automata-Based Modeling of Urban Phenomena*. John Wiley and Sons Inc., 2004.
- [2] N. Farenc, R. Boulic, and D. Thalmann, “An informed environment dedicated to the simulation of virtual humans in urban context,” in *Computer Graphics Forum (Eurographics ’99)*, P. Brunet and R. Scopigno, Eds., vol. 18(3). The Eurographics Association and Blackwell Publishers, 1999, pp. 309–318.

- [3] S. Donikian and S. Paris, "Towards embodied and situated virtual humans," in *Motion in Games*, 2008, pp. 51–62.
- [4] W. Shao and D. Terzopoulos, "Environmental modeling for autonomous virtual pedestrians," *Digital Human Modeling for Design and Engineering Symposium*, 2005.
- [5] S. Jiancheng, Z. Taiyi, and L. Feng, "Nonlinear prediction of mobile-radio fading channel using recurrent least squares support vector machines and embedding phase space," in *International Conference on Communications, Circuits and Systems, ICCAS'04*, vol. 1, 2004, pp. 282–286.
- [6] D. D. Dajab, "Perspectives on the effects of harmattan on radio frequency waves," *Journal of Applied Sciences Research*, vol. 11, no. 2, pp. 1014–1018, 2006.
- [7] O. Eyo, A. Menkiti, and S. Udo, "Microwave signal attenuation in harmattan weather along calabar-akampkpa line-of-sight link," *Turkish Journal of Physics*, vol. 27, pp. 153–160, 2003.
- [8] P. Graniero, "A spatial analysis of gps availability and radio data transmission reliability for real-time, wireless gis updating in urban environments," MEMF Lab, University of Windsor, Dept. of Earth Sciences, Tech. Rep., 2004.
- [9] P. Graniero and H. Miller, "Real-time, wireless field data acquisition for spatial data infrastructures," in *GeoTec Event 2003*, Vancouver BC, Canada, March 2003.
- [10] N. Savage, D. Ndzi, A. Seville, E. Vilar, and J. Austin, "Radio wave propagation through vegetation: Factors influencing signal attenuation," *Radio Science*, vol. 38, no. 5, pp. n/a–n/a, 2003.
- [11] G. Li, S. Yang, Y. Liu, and Z. Nie, "A variable step length hybrid approach for electromagnetic ray tracing in ionosphere," *Electromagnetics*, vol. 27, no. 6, pp. 331–340, 2007.
- [12] S. Seidal and T. Rappaport, "Site-specific propagation prediction for wireless in-building personal communication system design," *IEEE Transactions on Vehicular Technology*, vol. 43, no. 4, pp. 879–891, 1994.
- [13] G. Comparetto, J. Schwartz, N. Schult, and J. Marshall, "A communications analysis tool set that accounts for the attenuation due to foliage, buildings, and ground effects," in *MILCOM 2003: Proceedings of the Military Communications Conference*, vol. 1, no. 1, Boston, MA, USA, October 2003, pp. 1 – 5.
- [14] J. Schwartz, G. Comparetto, and J. Marshall, "The communications resource planning tool," in *Military Communications Conference, 2003. MILCOM 2003. IEEE*, vol. 1, no. 1, Boston, MA, USA, October 2003, pp. 227–230.
- [15] S. Y. Willassen, "A method for implementing mobile station location in GSM," Master's thesis, Norwegian University of Techniques and Sciences, 1998.
- [16] M. Weissberger, "An initial critical summary of models for predicting the attenuation of radio waves by trees," in *Final Report Electromagnetic Compatibility Analysis Center, Annapolis, MD.*, Annapolis, Md, July 1982.
- [17] R. Matschek, "A geometrical optics and uniform theory of diffraction based ray tracing optimisation by a genetic algorithm," *Comptes rendus de physique*, vol. 6, no. 6, pp. 595–603, July 2005.
- [18] J. Keller, "Geometric theory of diffraction," *Journal of the Optical Society of America*, vol. 52, pp. 162–130, 1962.
- [19] R. Kouyoujian and P. Pathak, "A uniform geometric theory of diffraction for an edge in perfectly conducting surface," in *IEEE*, vol. 62, 1974, pp. 1448–1461.

A New Physical Cell Identifier Structure in Femtocell Networks

Robert Bestak

Department of Telecommunication Engineering
Czech Technical University in Prague
Prague, Czech Republic
robert.bestak@fel.cvut.cz

Abstract—In densely populated areas, a quite high number Femto Access Points (FAPs) can be deployed and operate under one macro base station. Due to limited number of available Physical Cell Identifiers (PCI), these identifiers have to be inevitably reused in the network. However, the repetition of identifiers under one macrocell results in confusion events at the macro base station. To deal with the confusions and scarcity of identifiers in a highly dense femtocell environment, we propose dividing a macrocell into smaller logical regions, called clusters, and enhance the PCI structure about the cluster identifier. Additionally, we discuss the assignment process of FAPs into particular clusters and analyze possible solutions how to implement the new cluster identifier in today's mobile networks.

Keywords: LTE/LTE-A, femtocells, synchronisation signals, cell identifier, implementation.

I. INTRODUCTION

The wireless traffic has significantly grown over the past decade and there is no indication this growth will slow down. Rather contrary, due to new range of applications, massive increase of connected devices and extended usage of video oriented applications, there are predictions that the traffic growth trend will also continue in the next decade [1]. Handling this traffic in an affordable and sustainable way is one of major challenges of emerging wireless communication systems such as Long Term Evolution (LTE) and LTE-Advanced (LTE-A). A cost-effective mean to manage this growth represents small-cell/local-area deployments (i.e., femtocells, picocells, or metrocells). Network densification and thereby bringing network nodes physically closer to the users enables handling two shortages of present mobile systems: capacity and indoor coverage. The low-power nodes make possible to provide high traffic capacity and high user throughput locally, e.g., in indoor and outdoor hotspots. Hereafter, we focus on the femtocell concept.

Femtocells can be seen as small cells covered by inexpensive, low-power base stations that are, in general, deployed by customers themselves. These base stations are denoted as Femto Access Point (FAP). FAPs are connected to mobile operator's networks using either a wired, or a wireless backhaul; e.g., ADSLs, optics, WiFi, or nowadays even satellite links [2]. The FAPs can be configured to operate either in open access, or closed access, or hybrid access [3]. Closed access cells are accessible to a set of

users, whereas in open access cells any user can get access to the network. Finally, in case of hybrid access, the priority is given to owner of the FAP but additional selected users can be accepted as well.

Compare to the present macro/micro base stations, the number of FAPs within a mobile network can be huge. Therefore to enable smooth, simple and mass deployment of femtocells, self-organizing concept has to be employed when deploying femtocells. Customers cannot be expected to have whatever knowledge how to install/configure FAPs, all configurations have to be done automatically by FAPs themselves, or with a network assistance.

As in case of conventional macrocell base stations (MBSs), among others, a newly introduced FAP needs to be assigned a cell identifier that unambiguously identifies the cell in the given area. In LTE/LTE-A systems, this (macro/femto) cell identifier is called Physical Cell Identifier (PCI) and the number of identifiers is limited to 504 PCIs [4]. The PCI pool is shared by all types of cells in the network (femto, pico, metro, macro, etc.). The PCI is derived from LTE/LTE-A physical layer signals known as Primary Synchronization and Secondary Synchronization Signal (PSS, SSS, [5]).

A PCI should be selected in such a manner to avoid collision (i.e., a PCI is unique in the area that the cell covers, see FAP_{A,B} in Fig.1) and confusion (i.e., a cell has no neighboring cells that have identical PCI, see FAP_E and FAP_{C,D} in Fig.1) events [6]. Notice that these events can occur as from the point of femtocell level (collision: FAP_A - FAP_B, or confusion: FAP_C - FAP_D) as from the point of macrocell level (collision: MBS - FAP_{A,B}, or confusion: MBS - FAP_{C,D,F}).

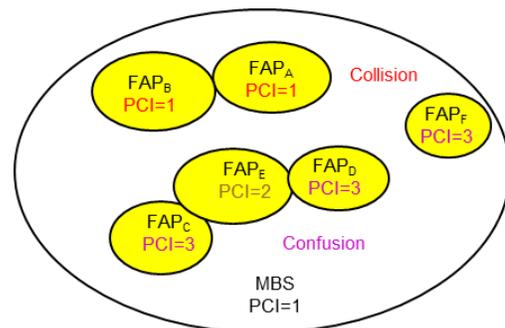


Figure 1. Example of collision and confusion events in a macrocell.

In view of the fact that FAPs are usually in possession of customers, a PCI assignment method has to be automatic and adaptive to manage FAPs movement; in case a customer decides changing the FAP location. In highly populated metropolises, a quite high number of femtocells can operate under coverage of one macro base station, e.g., Paris population density is about 21 thousands per km² [7]. Due to the limited number of PCIs, the PCIs have to be inevitably reused in the network. Thus, confusion events from the point of a macro base station can occur (MBS - FAP_{C,D,F} in Fig.1). To deal with both confusion events and lack of identifiers, we propose: i) increasing number of PCIs, ii) dividing a superior macrocell into several smaller regions denoted as FAP clusters, and iii) enhancing the PCI structure about FAP cluster sub-identifier. Additionally, we discuss possible ways of FAP cluster identifier implementation in the LTE/LTE-A mobile networks.

The rest of this paper is organized as follows. The next section provides brief overview of related works and our motivation. Section III details the proposed clustering concept. Possible implementation of femtocell cluster in mobile networks is presented in section IV. Finally, major findings are concluded in the last section.

II. RELATED WORKS AND BACKGROUND

In conventional macrocell deployment, several PCI assignment methods have been investigated last couple of years [8]. These methods can be in general classified as distributed (e.g., random, or radio scanning methods) and centralized (e.g., network planning approach, or use of specific network entity). Authors usually employ a graph coloring approach to cope with the collision/confusion problems [9-11]. However, the femtocell deployment has certain specificities compare to the macrocell one. These include: i) larger number of base stations, ii) unstable position of base stations in time, and iii) a base station placement is not really managed and controlled by operators.

To handle femtocell collision/confusion events, authors in [12] combine the radio scanning and random methods. The former method helps a newly introduced FAP to identify neighboring PCIs that are in the second phase omitted when randomly choosing a PCI. In the femtocell scenarios, researchers mainly focus on how to distribute PCIs among different type of cells. For example in [13], authors discuss how to split the PCI range among different FAP access types, i.e., open, closed and hybrid. The sharing of PCIs between macro base station and FAPs are discussed in [14], where an automatic PCI allocation method is modified according to the PCI utilization and the scale of base stations. New dynamic reservation schemes are introduced in [15]. The authors present several reserving types and each type corresponds to a different femtocell PCI pool. The transition among the types depends on the number of deployed FAPs, or the number of occurring PCI confusion events. However, in all previously mention works, the number of considered femtocells is relatively low, the number FAP remains below 503 femtocells (one PCI is dedicated to the MBS).

To the best of our knowledge, a research work regarding the macrocell-femtocell collision/confusion issue in the dense FAP environment has not been carried out. To illustrate the problem, let's consider a simple scenario where we randomly distribute femtocells within a macrocell; the standard uniform distribution is used to generate FAP positions. An example of FAP distribution in a macrocell is illustrated in Fig. 2. In this scenario, we would like to just demonstrate the problem of high number of femtocells per macrocell; femtocell related issues such as cell interferences, attenuation of signals, etc. are not taken into account.

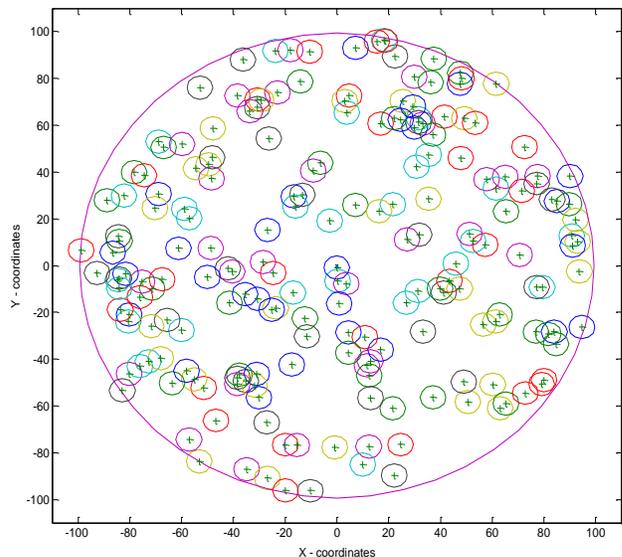


Figure 2. Example of FAPs placement in macrocell.

Fig. 3 shows number of femtocells per macrocell for different macrocell radius (denoted as R_m in the figure) and parameter k that is given as:

$$k = \frac{\sum_i^N S_i^{femto}}{S^{macro}} \quad (1)$$

where S^{Macro} denotes the macrocell surface, S_i^{Femto} is surface of i -th femtocell, and N is number of femtocells per macrocell. In this analysis, we consider femtocells to be circles with radius (R_f), which is same for all femtocells. Thus, the equation (1) can be simplified and rewritten as:

$$k = N \cdot r^2 \quad (2)$$

where r represents ratio R_f/R_m . For $k = 1$, the sum of all femtocell surfaces theoretically entirely fulfilled the macrocell surface, i.e., femtocells do not overlap. However, femtocells can and do overlap and therefore the macrocell surface is not entirely fulfilled by femtocells for $k = 1$. Fig. 4, subsequently, shows number of femtocells per macrocell for different femtocell radius. Results illustrate that for certain combinations of R_f and R_m , we can rather quickly begin missing PCI (the LTE max. value of PCI is indicated in both

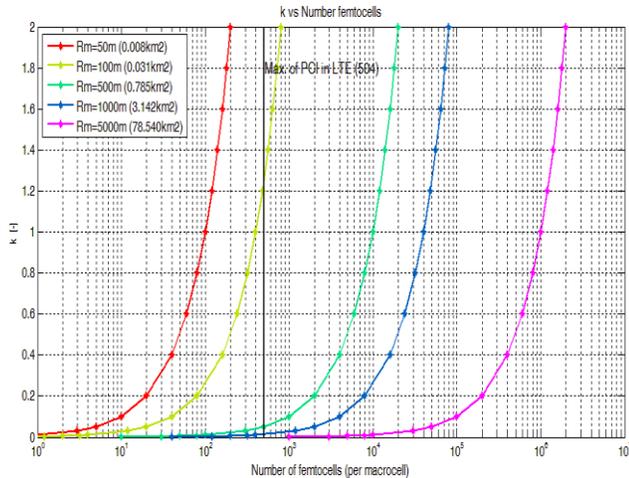


Figure 3. Number of femtocells for various macrocell radius (R_m), $R_f = 5m$.

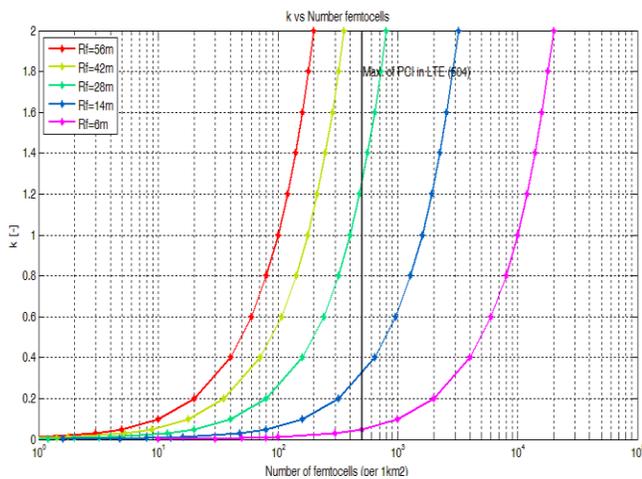


Figure 4. Number of femtocell for various femtocell radius (R_f), $R_m = 564m$.

figures by vertical black lines). For example, already in case of about 35% of macrocell surface is covered by femtocells (Fig. 4, $R_f = 14m$), we run out of unique PCIs.

In [16], two deployment models are considered for the femtocell environment: i) suburban model, and ii) dense-urban model. In case of the dense-urban model, there are assumed 6928 households per km^2 . Thus, if we consider $1km^2$ macrocell and 10% FAP penetration (that gives 693 households equipped with FAPs), we would need to repeat 190 PCIs within the given macrocell.

The previous analysis shows that even if the PCIs are ideally distributed and there are no collision/confusions at the femtocell level, we start to have macrocell-femtocell confusions since PCIs has to be repeated in the given area.

III. FAP CLUSTER CONCEPT

As discussed in the previous section, even in case of ideal PCI assignment as the number of FAPs in the given macrocell exceeds 503, the PCIs have to be inevitably

repeated and confusion events at the macrocell-femtocell level takes place. To deal with this issue, we propose: a) increasing PCI range, and b) introducing FAP clusters. The FAP cluster concept can be described as follows.

A superior macrocell (or microcell) is divided into several smaller logical regions, called FAP clusters, where a FAP cluster is formed by set of neighboring FAPs. A FAP is part of a FAP cluster if the FAP can hear (detect) at least one FAP of the cluster, i.e., a FAP cluster contains at least two FAPs. Thus, a macrocell is composed of at least one (see next), or several FAP clusters (see Fig 5).

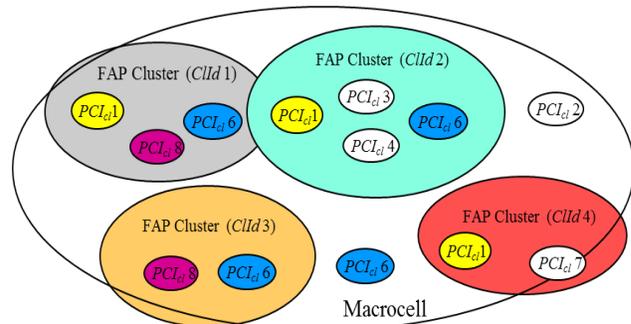


Figure 5. FAP cluster concept.

Each cluster within a macrocell is assigned a unique FAP cluster identifier, called as Cld . Within a FAP cluster, a FAP is assigned, if possible, collision/confusion free PCI, denoted as PCI_{cl} . Notice that any kind of PCIs assignment methods can be used here to avoid collision/confusion events. The same PCI_{cl} can be subsequently repeated in other FAP clusters (see for example $PCI_{cl} = 1$ in Fig. 5). The suggested PCI structure can be illustrated in Fig. 6.

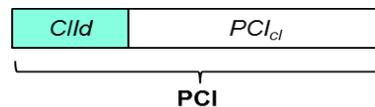


Figure 6. Enhanced PCI structure.

All isolated FAPs, i.e., a FAP that cannot detect any other FAP in its vicinity, form a FAP cluster called isolated FAP cluster (see for example $PCI_{cl} = 2$, or 6 in Fig. 5). Therefore, there is always at least one FAP cluster within a macrocell. In the next, we abbreviate the isolated femto access point as FAP_{is} and the isolated FAP cluster as FAP_{ISCL} . The FAP_{ISCL} is assigned a predefined, default, value of Cld that is known by all MBSs and FAPs in the network, e.g., $Cld = 0$.

Based on the FAP density evolution, the number of FAP clusters can vary in time and from macrocell to macrocell. A MBS knows the current number of FAP clusters in its macrocell and informs all FAPs under its coverage about the Cld values. To do that, MBS/FAP broadcast channels, or femtocell backhaul can be employed to distribute this information. Notice that if there are only FAP_{is} in the macrocell, the Cld value does not have to be distributed as

the $Clid$ is known, which reduces signaling overhead in the network. In other words, the today's way of PCIs assignment, without the FAP cluster concept, can be seen as if $Clid = 0$. Thus, the FAP cluster approach can be gradually introduced in today's networks.

When introducing a new FAP (FAP_{new}), the following three scenarios can occur; the scenarios and steps are summarized in the flow chart in Fig. 7:

a) *Joining an isolated cluster*. Not detecting any FAP, the FAP_{new} selects the predefined value of $Clid$ for the FAP_{ISCL} . This scenario will mainly occur in the initial phase of FAPs deployment when the number of FAPs within a macrocell is small. As stated above, this scenario can be also seen as today's approach without the FAP clusters.

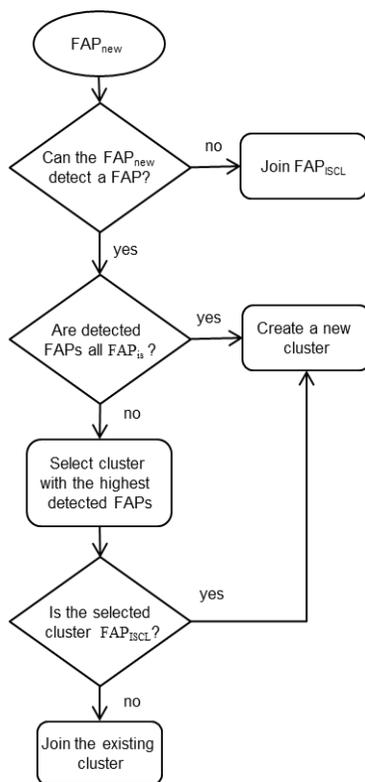


Figure 7. Steps when introducing a new FAP within a macrocell.

b) *Joining an existing cluster*. Detecting only FAP_{is} belonging to one, or more FAP clusters, the FAP_{new} selects and joins a cluster in which detects the highest number of FAPs. This solution enables keeping FAP clusters compact otherwise clusters would become mixed with each other.

c) *Creating a new cluster*. Detecting one, or more FAPs belonging to one, or more FAP clusters, the FAP_{new} selects and joins a cluster (including the FAP_{ISCL}) in which detects the highest number of FAPs. If the joined cluster corresponds to FAP_{ISCL} , the FAP_{new} initiates creation of a new FAP cluster that will be composed of all detected FAP_{is}

(they update their $Clid$). Thus, the number of FAP_{ISCL} is kept as low as possible within the macrocell. The FAP cluster creation is done with the network/MBS assistance and results in updating of $Clid$ value that is spread out in the macrocell.

IV. IMPLEMENTATION OF FAP CLUSTER IDENTIFIER

To implement the FAP cluster concept in the present LTE/LTE-A mobile networks, there can be considered three approaches denoted as: a) sub-identifier, b) new identifier, and c) alternate transmissions.

A. Sub-identifier

The simplest and easiest way to implement the FAP cluster identifier, $Clid$, is to dedicate from the existing PCIs range certain PCIs to the FAP cluster identifiers. This is a similar approach like propositions in [13-15] where authors suggest splitting PCIs range among different type of cells (MBSs/FAPs, open/closed/hybrid). However, the small number of PCIs still remains a limiting factor, i.e., the confusion events at macrocell level remain. Therefore, this solution is only reasonable in a case of relatively low density femtocell environment where the PCI range reduction can be accepted.

B. New identifier

In this case, the $Clid$ is considered to be a new independent part of the present PCI structure (see Fig. 6).

The LTE radio frame, 10ms long, is divided into 10 sub-frames of 1 ms duration. Each sub-frame is further split in two 0.5ms slots. A slot contains seven, resp. six, OFDM symbols in case of normal cyclic prefix, resp. extended cyclic prefix. In LTE-FDD mode, the PSS is embedded in the last OFDM symbol and the SSS in the second last OFDM symbol, of the sub-frame 0 and 5 in each radio frame [5]. The PCI is obtained by decoding PSS in the first phase (which represents one of three possible cell identities) and then decoding the SSS (which represents one of 168 cell-group-identities).

In the frequency domain, 6 Resource Blocks (RB) around the DC subcarrier are reserved for transmission of the synchronization signals; an RB is formed by 12 subcarriers with a subcarrier spacing of 15 kHz.

Both synchronization signals, PSS and SSS, are based on Zadoff-Chu (ZC) sequences [5]. Since the PSS uses only 62 of the 72 reserved subcarriers (6RBs x 180 kHz), the required length of ZC is 63 (62 subcarriers and 1 un-used DC subcarrier, which is punctured). Thus, the ZC sequence is padded with five zeros at the edges. The same sequence is repeated in both PSS symbols.

Similar to the PSS, the SSS has 62 non-zero elements ZC sequence. The 62 elements are an interleaved concatenation of two length-31 sequences, each taking 31 different values. Contrary to the PSS, the SSS sequence in sub-frame 0 and 5 are different from each other.

The FAP cluster identifier can be implemented by extending the PSS and SSS signals and called Extended Synchronization Signals (ESS), as it is proposed in [17].

Within the ESS proposal, the authors propose replacing the padding zeros by additional information elements. These 20 zero-elements can be used for different purposes, in our case for implementing the FAP cluster identifier (*Clid*). In comparisons with the solution A (sub-identifier), this implementation solution makes possible to increase and flexibly adapt PCIs range based on evolution of FAP density in a macrocell. For example, allowing 8 FAP clusters would extend the PCI range up to 4031 (theoretically) non-confusion FAPs.

C. Alternate transmissions

The third way of implementing FAP clusters could be either periodically scrambling the PSS with a *Clid* specific scrambling sequence, or periodically alternate transmissions of PSS and *Clid* in the sub-frame 0 and 5. The switching period between PSS and *Clid* would have to be sufficiently long to allow a FAP/terminal to correctly decode the PSS. In the first step a FAP/terminal would decode the PSS and in the second step the cluster identifier. A drawback of this solution represents the extra time the FAP/terminal needs to decode the PSS and *Clid*.

V. CONCLUSIONS

In this paper, we discuss an approach of assigning physical cell identifier to FAPs in highly dense femtocell environment. We show that in such scenario even if the PCIs are ideally distributed among FAPs and there are no collisions/confusions from the point of view of femtocells, once the number of FAPs under a macrocell exceeds the PCI range, the PCIs need to be repeated and we begin to have confusion events at macrocell-femtocell level.

To deal with this problem, we propose a part of increasing number of PCI to divide a macrocell into smaller logical regions, that we denote FAP clusters, and to enhance the current PCI structure about the cluster identifier. Based on the FAP density evolution in the macrocell, the number of FAP clusters can be progressively adjusted. Finally, three possible implementation approaches of the FAP cluster identifier within today's LTE/LTE-A mobile networks are outlined and discussed.

The main aim of this paper has been to describe the principle of the proposed FAP cluster concept and its possible implementation and deployment in LTE/LTE-A mobile networks. In the next step, we plan to evaluate performance of the clustering method via simulation by taking into account different scenarios and parameter settings. In addition to that, we also intend to investigate more sophisticated methods that would take into consideration several aspects before deciding, which FAP cluster the newly introduced FAP should join; e.g., positions of detected FAPs, number of FAP clusters in the macrocell, or number of FAPs in the macrocell.

ACKNOWLEDGMENT

This research work was supported by the Grant Agency of the Czech Technical University in Prague, grant no. SGS13/199/OHK3/3T/13.

REFERENCES

- [1] Qualcomm, "The 1000x mobile data challenge" Oct. 2012.
- [2] Informa telecoms & media, "Satellite backhaul for rural small cells", 2012.
- [3] A. Golaup, M. Mustapha, and L. B. Patanapongpibul, "Femtocell access control strategy in UMTS and LTE" IEEE Communication Magazine, Sep. 2009, vol. 47, no. 9, pp. 117–123, doi. 10.1109/MCOM.2009.5277464
- [4] 3GPP TS 36.211, "Physical Channels and Modulation", Dec. 2012.
- [5] E. Dahlman, S. Parkvall, and J. Skold, "4G LTE/LTE-Advanced for Mobile Broadband", Academic Press, 2011.
- [6] 3GPP TS 36.902, "Self-configuring and self-optimizing network (SON) use cases and solutions", Mar. 2011.
- [7] Institut National de la Statistique et des Etudes Economiques, "Commune: Paris (75056) – Theme : Evolution et structure de la population", 2009.
- [8] 3GPP R3-080812, "Solution(s) to the 36.902s Automated Configuration of Physical Cell Identity Use Case". Mar. 2008.
- [9] Y. Liu, W. Li, H. Zhang, and W. Lu, "Graph based automatic centralized PCI assignment in LTE. 2010 IEEE Symposium on Computers and Communications (ISCC), Jun. 2010, pp. 919-921, doi:10.1109/ISCC.2010.5546812.
- [10] T. Bandh, G. Carle, and H. Sanneck, "Graph Coloring Based Physical-Cell-ID Assignment for LTE Networks," ACM International Conference on Wireless Communications and Mobile Computing (IWCMC), Jun. 2009, pp 116-120, doi: 10.1145/1582379.1582406.
- [11] F. Ahmed, O. Tirkkonen, M. Peltomäki, J-M. Koljonen, C-H. Yu, and M. Alava, "Distributed Graph Coloring for Self-Organization in LTE Networks," Journal of Electrical and Computer Engineering, vol. 2010, Article ID 402831, 10 pages, 2010. doi:10.1155/2010/402831.
- [12] Y. Liu, W. Li, H. Zhang, and L. Yu, "Distributed PCI assignment in LTE based on consultation mechanism", 2010 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM), Sep. 2010, doi: 10.1109/WICOM.2010.5601210.
- [13] 3GPP R1-082762, "LS reply on CSG cell identification", 2008.
- [14] T. Wu, L. Rui, A. Xiong, and S. Guo, "An Automation PCI Allocation Method for eNodeB and Home eNodeB Cell", 2010 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM), Sep. 2010, doi: 10.1109/WICOM.2010.5600764.
- [15] P. Lee, J. Jeong, N. Saxena, and J. Shin, "Dynamic Reservation Scheme of Physical Cell Identity for 3GPP LTE Femtocell Systems", Journal of Information Processing Systems, Dec 2009, s, Vol.5, No.4, pp 207-220, 10.3745/JIPS.2009.5.4.207.
- [16] Femto Forum, "Interference management in UMTS femtocells", 2008.
- [17] A. H. Zahran, "Extended Synchronization Signals for eliminating PCI confusion in heterogeneous LTE", Wireless Communications and Networking Conference (WCNC), Apr. 2012, pp. 2588-2592, doi: 10.1109/WCNC.2012.6214236.

A New Bio-Crypto System Approach to Enhance Security of 4G Mobile Networks

<p>Saeed Ashry ITIDA Cairo, Egypt sashry@itida.gov.eg</p>	<p>Abdulatif Elkouny Ain Shames University Cairo, Egypt aalkouny@gmail.com</p>	<p>Hesham Elbadawy NTI Cairo, Egypt heshamelbadawy@ieee.org</p>	<p>Salwa Elramly Ain Shames University Cairo, Egypt sramlye@netscape.net</p>	<p>Ahmed ElSherbini ITU Geneva, Switzerland sherbini@mcit.gov.eg</p>
---	--	---	--	--

Abstract— This research paper examines the vulnerable security elements that MAC management messages in WiMAX initial network entry faces. The paper proposes an innovative hybrid approach to resolve such vulnerability problem and establish secret communication channels via insecure domains. The proposed protocol is based on Bio-crypto systems and to improve current security level of authentication and Key Exchange between the Subscriber Station SS and the Base Station BS. AES-BDK system is defined as the integration of Advanced Encryption Standard and Biometric Digital Key. Within this integration, AES is used for the encryption process for initial ranging Request/Response messages RNG_REQ/RSP, while BDK generates the secret key.

Keywords- initial ranging; MAC Management Messages; key exchange; key generation; bio-crypto systems.

I. INTRODUCTION

Computer security experienced a significant and grew rapidly over the past few years. Worldwide Interoperability for Microwave Access (WiMAX) based on IEEE 802.16e-2005 for Mobile WiMAX [1]. Bolsters a large number of enhanced security features compared to the fixed IEEE 802.16-2004 [2]. based on security schemes. The initial network entry is the most critical process in WiMAX networks as it is the foremost step to establish network connection while carrying multiple parameters such as performance factors and security context between the Base Station (BS) and the Subscriber Station (SS) determined in the middle of this process. However, security schemes are utilized and applied only to normal data traffic after the initial network entry process, but not to control messages during the initial network entry. BS and SS communication in the initial network entry is susceptible to forgery. As a result, there can be various security vulnerabilities especially unauthenticated messages and unencrypted management communication that uncover important management data. As a consequence, no common key to generate message digest is the source of difficulty to authenticate these types of messages. According to literature

survey, little pertinent papers discuss the security issues of messages in this process whilst most of these messages are very short [3]. Thus, it is an area worthy of interest. Due to this fact, a tradeoff between security and effectiveness is being considered in this paper through a new approach of enhancing MAC management message security during network entry initialization. The proposed scheme is applicable to Point to – Multi- Point (PMP), while it could be scalable to multiple zones depending on the security channels between different Trusted Third Parties (TTP).

The paper proposes a hybrid protocol relying on Bio-crypto system. The biometric-crypto system is a combination of biometrics and cryptography and is considered a very promising technique [3]. Bio-cryptography is an emerging technology that inherits the advantages of both and provides strong means of protection against attacks targeting MAC Management messages: Ranging Request/Response (RNG_REQ/RSP) messages.

The objective of biometric data is to provide privacy, cryptographically- secure authentication of human users, and non-repudiation by using biometric template of fingerprint. Fingerprint minutiae matrix is produced to generate the Biometric Digital Key (BDK). To guarantee the security of the key, the proposed protocol adds a new message to the standard control message. BS uses stored biometric templates in the Trusted Third Party (TTP) to fulfill the authentication key exchange protocol banking on biometric system to authenticate user messages. The produced key is used in Advanced Encryption Standard (AES) to encrypt user messages. Bio-Crypto system establishes a secure channel between SS and BS and vice versa.

The remaining of this paper is structured along these lines: Section II describes the initialization of WiMAX network in accordance with amendment IEEE 802.16e, while Section III shows cases of the security vulnerabilities of MAC management messages. Section IV discusses the previous work. Section V introduces the projected approach, while Section VI shows the experimental results. Finally, Section VII concludes this paper.

II. NETWORK ENTRY INITIALIZATION

When a SS initializes some synchronization, parameter adjustments take place to establish a robust and well-fitted link between SS and BS. The performed initial tasks, when a SS turns on in the network, include, but not limited to, channel acquirement, PHY (Physical Layer) synchronization, channel descriptors identification and interpretation. After PHY synchronization, MAC (Medium Access Control Layer) can detect and identify MAC management messages such as DCD (Downlink Channel Descriptors), UCD (Uplink Channel Descriptors), DL-MAP (Downlink Map) and UL-MAP (Uplink Map). MAC management messages determine the contention slots to initial ranging on the UL sub-frame. These messages are transmitted by BS on the first available DL burst. Different SS and BS messages are exchanged to stabilize signal power, frequency and time offset [4].

The initial ranging adjusts transmission until an adequate profile is achieved for the link, including synchronization issues. After network entry, when SS is already transmitting data, takes place, the periodic ranging, which guarantees that transmission, will react to channel changes as well as maintain link quality. Periodic ranging exchanges MAC-PDU (Packet Data Unit) messages between BS and SS, while initial ranging uses both contention slots on the UL sub-frame and MAC-PDU messages. The SS will transmit a ranging request message (RNG-REQ) in one of the initial ranging slots. If collision occurs, the SS uses a truncated exponential back-off. This message is described in [1]. The BS could receive or not the RNG-REQ. If the SS doesn't receive an answer from BS, it retransmits the RNG-REQ message after a timeout. If the BS successfully received the RNG-REQ, it would return a RNG-RSP message using the initial ranging CID (Connection Identification) [5]. A ranging response message (RNG-RSP) is sent in one of the available DL bursts containing the CIDs of the Basic and Primary connections to this SS. The RNG-RSP could contain power, frequency and time corrections to the SS. This message is described in [1].

III. SECURITY VULNERABILITIES OF MAC MANAGEMENT MESSAGES

With a reference to IEEE 802.16-e amendment, the security schemes are only applied to standard data traffic after the initial network entry process not to control messages during the initial network entry. The initial network entry is considered the main process in Mobile WiMAX network. It is the first gateway to set up a connection to Mobile WiMAX. Consequently, numerous physical parameters, factors of performance, and security contexts between SS and BS are set during the process. In Figure 1 the initial network entry in normal process is illustrated.

A. RNG_REQ/RSP vulnerabilities

This section studies vulnerabilities contained by RNG_REQ/RSP in the initial network process. The ranging Request (RNG_REQ) message is transmitted by an SS requesting and seeking a network join. The message integrates the SS's existence meanwhile acts as a request for transmission timing, power, frequency and burst profile information. The BS responds to the SS request using a Ranging Response (RNG_RSP) message. This message is composed of important information, such as ranging status, time adjusts information and power adjusts information. Nonetheless, the RNG_RSP message is neither encrypted nor authenticated, it holds no definite status. An attacker would manipulate such a leak and execute an attack using DoS because SS's action could be conducted by any validly formatted RNG/RSP that addresses to it [6][7].

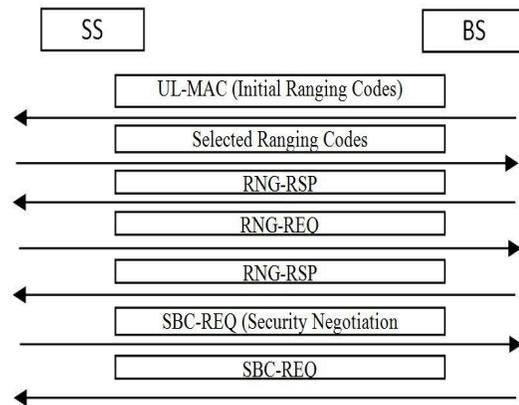


Figure 1. Normal Initial Network Entry Procedure.

B. Rouge BS

Other distinguished attacks on wireless networks include man-in-the-middle (MITM) attack, namely, Rouge BS. The attacker could intercept messages during the process of communication establishment or a public key exchange. The attacker then retransmits them, counterfeiting the contained information in the messages, so that the two original parties still seem to be contacting with each other [8].

IV. PREVIOUS SOLUTIONS

This work is inspired from a number of preceding works concerning key exchange during initial network entry. Some messages in the initial network entry inject some vulnerability such as RNG-REQ/RSP. T. Shon 'et al.' [9]. Introduces a solution to secure these messages counting on the Diffie-Hellman (DH) key agreement. De facto, it can secure the message during initial network entry process and boost the security grade; however it is still liable to man-in-the-middle attack. T. Han 'et al.' [10]. proposed an alternation to DH key agreement protocol to fit into mobile WiMAX relying on the presumption that every MS possesses its own International Subscriber Station Identity (ISSI) from which the MS could generate its Temporary ISSI

(TSSI) and the authors postulate that the legitimate BS has the hash value $H(TSSI)$. Then they launch their protocol which is based on five steps to avert the former attack. T. Shon ‘et al.’ [10], has a concern on the previous solution, as it could cause additional overhead when distributing initial DH random number. The perspective of this paper contrives that this solution does not clarify the source of private and public key ($p&q$) calculations, in addition to time limitation for $p&q$ calculation. Furthermore, they have to disseminate TSSI to the entire neighbor BS. Hence, there is no security for ISSI. M. Rahman ‘et al.’ [11], utilizes DH key agreement so as to secure the initial network entry and encrypt the initial management communication. Their adjustment on the DH is carried out to dismiss man-in-the-middle breach by employing cryptographic sealing function. T. Shon ‘et al.’ [9], are debating over this approach explicating it as obscure and providing no clear vision of how the random number is generated and distributed to others to use DH scheme. Consequently, those authors suggested a modernistic altered DH scheme [11]. The schema can be performed using hash authentication as well as one of the ranging codes as a prime number seed. Furthermore, it uses hash authentication when SS singles out one of ranging codes RC_i which consists of $A1$ and $A2$, then transmits only part of RC_i ($A1$ or $A2$) and hash value of RC_i . When BS receives a segment of RC_i ($A1$) and the hashed value $H(RC_i)$, BS obtains $A2$ from ranging code pool using $A1$, then authenticates MS throughout received hash value verification [12]. Therefore, they could produce a prime number “p” from the chosen ranging code. In a subsequent phase, SS generates the other global variables “g” and public/private key pair then sends them to BS. If the received key and variables are verified, BS sends its public key to MS. Thence, BS and SS commence sharing DH global variables and public key with each other at the hand of initial network entry ranging process. It is unclear whether the protocol steps could counteract the bogus BS from sending RC_i or not.

V. PROPOSED PROTOCOL

This paper suggests an approach of securing the RNG_REQ/RSP messages. The model is composed of SS, BS, and TTP server to manage the authentication and key exchange between SS and BS. The protocol assumes that the channel between BS and Trusted Third Party (TTP) is secured, as illustrated in Figure 2. The protocol uses Advanced Encryption Standard (AES) as the most popular algorithm used in symmetric key cryptography with a combination of Biometric data (e.g fingerprint, iris, etc.) to generate the BDk.

A. Setting up a SS User Device

The SS user firstly captures its biometric data, then registers itself to the TTP, and asks the system make its biometric template (i.e., fingerprint minutiae matrix), which will be stored on both SS device and TTP server.

B. Biometric Template Protection

The TTP saves biometric template of SS’s secured. Biometric data is stored on protected storage. The confidentiality of biometric template can be assured by implementing permutation module using obscure scheme. Firstly, TTP generates fake minutiae set and inserts it to the user’s biometric template. Secondly, for hiding private key, polynomial for real minutiae set and polynomial for fake minutiae set are established. Finally, the protected biometric template is made by combining these results. It consists of minutiae’s (location, angle, result) value set [13].

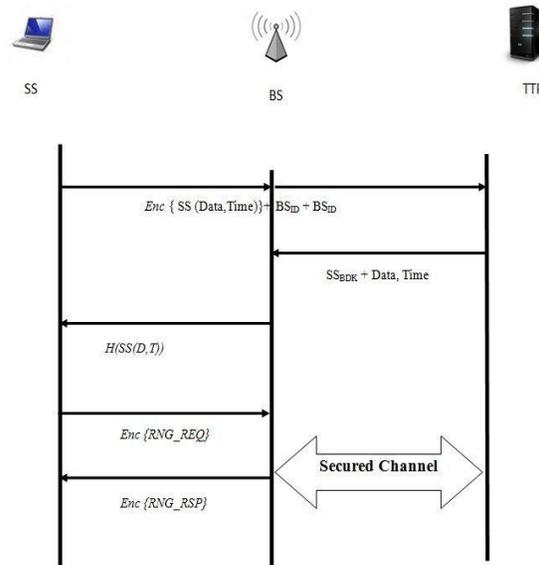


Figure 2. Proposed secure initial network entry using AES-BDK

C. Biometric Key Generation

The protocol proposed Biometric Digital Key with uniqueness, randomness property that cannot be falsified. The SS holds a stored biometric template on its device [13][14]. (i.e., fingerprint minutiae). The fingerprint minutiae matrix acted as the seed data in Secure Hash Algorithm (SHA). The seed data length is selected considering the length of key lengths of 128 bits, 256 bits, and 512 bits. Then SS starts to generate the initial BDk. The initial BDk is formed based on minutiae points. For extracting minutia points from fingerprint preprocessing phases should done, Image enhancement, Binarize, ridge ending and ridge bifurcation. Figure 3 clarifies key generation from minutiae. Further to strengthen the security of the key and to randomize the BDk between SS and BS, the seed data will be based on Data and Time.

D. Encryption and Decryption phases

Advanced Encryption Standard (AES) is used in this protocol as a block cipher which is the most popular algorithm used in symmetric key cryptography; see Figure 4 that clarifies encryption and decryption process using BDk. As AES supports key sizes of 128 bits, 256 bits, and recently

512 bits. Security increases with the larger key sizes as well as the number of rounds (10, 12, and 14, respectively). Therefore, the complexity of AES encryption and decryption also grows. The proposed protocol is evaluated with the different key lengths.

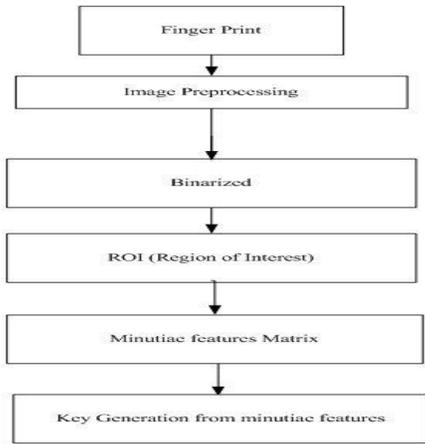


Figure 3. Key Generation from Minutiae Feature.

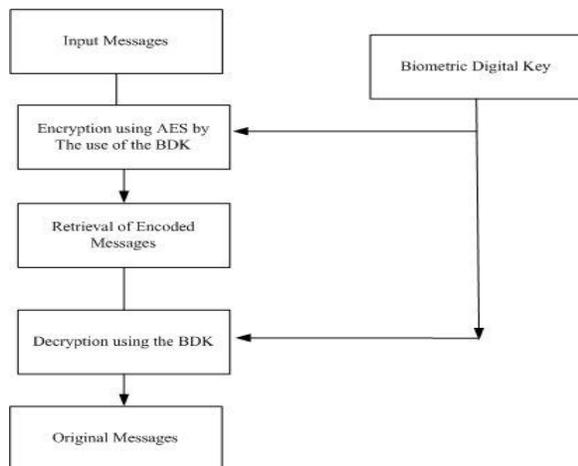


Figure 4. Encrypt and Decrypt using BDK.

E. Protocol Steps

The following steps describe the proposed protocol:

- SS starts sending a new message including Data and Time encrypted by the initial BDK to the BS Enc {Data, Time} + BS_{ID} + SS_{ID}.
- BS receives the Enc message, and then delivers it to TTP.
- TTP starts searching on its Database by SS_{ID}, if TTP verifies that it is a real SS, and then the TTP retrieves the stored biometric template of SS starting to generate the SS initial BDK, and Dec the message.
- TTP pass the SS initial BDK, the Data and Time to the BS.

- BS sends the hash value of SS (Data and Time) $H(D, T)$, as a challenge message to the SS.
- SS starts sending the Enc {RNG_REQ} to the BS using the new BDK.
- SS starts sending the Enc {RNG_REQ} to the BS using the new BDK.
- BS starts Dec (RNG_REQ).
- BS sends the Enc (RNG_RSP) to the SS.
- SS starts Dec (RNG_RSP).

VI. EXPERIMENTAL RESULTS

A prototype implementation of MAC frame structure has been developed. PC specifications are CPU Intel Core 2 Duo CPU 2GHZ, RAM 4GB, O/S Windows 7 SP2.

The physical layer of model system uses Orthogonal Frequency Division Multiplex (OFDM) with basic OFDM parameters and with 20 MHz bandwidth TDD mode while the frame length is set to 10 ms, modulation code QPSK [4]. The scenario frame is illustrated in Figure 5. The illustrated protocol, in the previous section, deployed MATLAB script for AES encryption, decryption phases, and key generation phase conjointly with Secure Hash Algorithm (SHA), to produce various key lengths.

CASIA fingerprint database version 5[16]. is used to produce biometric data. To extract biometric templates, 1000 fingerprint samples were taken from this database. The average total time of reading fingerprint image and enhancement process is 3.7 sec. The average time process of generating the BDK is 0.0003808 sec. The consuming time of encryption and decryption processes for the proposed protocol is presented in Table 1 in msec.

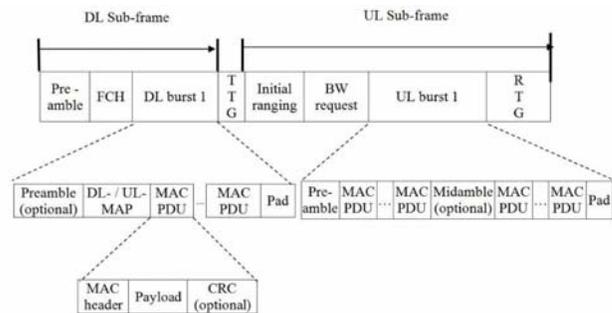


Figure 5. Example of MAC frame structure scenario.

This performance comparison is made in terms of time consumed for encryption and decryption of AES-128 bits, AES-256 bits and AES-512 bits. The model employed measures time at the start of encryption process and its end time while the same process is repeated for decryption. The evaluation results show that due to the increase in the key length used in AES-512 bits, and consequently, the number of rounds for both encryption and decryption boosts.

But, the encryption and decryption procedures become more complex thereby degrading the speed of the 512 bit AES algorithm. Figures 6 and 7 show the consumed time of Encrypted/Decrypted REG_REQ and RNG_RSP messages using various key lengths, respectively.

TABLE I. ENCRYPTION/DECRYPTION TIME

	Enc (RNG_REQ)	Dec (RNG_REQ)	Enc (RNG_RSP)	Dec (RNG_RSP)
AES-128	3.23	3.5575	5.1414	6.491
AES-256	3.4005	3.8486	5.3823	6.8198
AES-512	4.05	4.4035	5.9293	7.3486

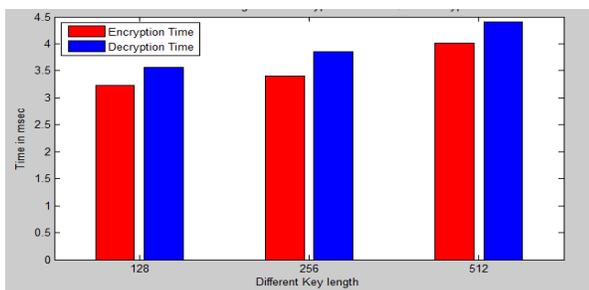


Figure 6. Time of Encrypt/Decrypt RNG_REQ with different key lengths.

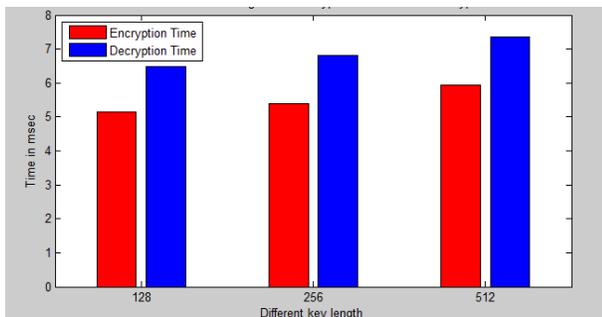


Figure 7. Time of Encrypt/Decrypt RNG_RSP with different key lengths.

VIII. CONCLUSION AND FUTURE WORK

The contributions of this paper are meant to enhance a security of PMP initial network entry. This paper proposed a new authentication key exchange protocol that enables SS and BS exchange parameters securely. The generated BDK from fingerprint minutiae is utilized to encipher data messages thereafter more securely derived and produced. Consequently, this key is not liable to be cracked easily [17]. The approach compared AES-256 bits and AES-512 bits in terms of time taken for encryption and decryption. Thus there is a tradeoff between speed and security. The proposed approach substantially considers the items of

privacy, authentication and non-repudiation, to reduce the probability of some threats such as eavesdropping, MITM attacks.

In the future, further efforts will be focused to find the solution for Mesh network to guarantee the security of the network entry. As the PKI authentication uses ECC and is an area worth of more research and analysis, a Mesh Biometric Certificate X.509 V.3 with ECC is an important component of the network security system.

ACKNOWLEDGMENT

I would like to express my gratitude to Yasser Hashem for the useful comments, remarks and engagement through the learning process of this research paper. Also, I like to thank AbdelRahman El-Mawawi for his great efforts in proof reading and reviewing this research.

REFERENCES

- [1] IEEE 802.16e-2005, 'IEEE standard for local and Metropolitan Area Networks-part 16; Air Interface for Fixed Broadband Wireless Access Systems', IEEE press, 2005.
- [2] IEEE 802.16-2004, 'IEEE standard for local and Metropolitan Area Networks-part 16; Air Interface for Fixed Broadband Wireless Access Systems', IEEE press, 2004.
- [3] P. Stavroulakis and M. Stamp, Handbook of Information and Communication Security., Ch.7, BioCryptography, pp. 129-157, Springer,2010.
- [4] C. Hoymann. "Analysis and performance evaluation of the OFDM-based metropolitan area network IEEE 802.16", pp. 341-363, June 2005.
- [5] H. Hernadnes, F. Quinelato, and A. Alberti. "A Simplified Simulation Model for Initial Ranging in IEEE 802.16d WirelessMAN-OFDM", [Online] Available: http://www.iiis.org/CDs2010/CD2010SCI/CCCT_2010/Abstract.asp?myurl=TB807EH.pdf/06.2013
- [6] T.Shon and W.Choi, "An Analysis of Mobile WiMAX Security: Vulnerabilities and Solutions", Lecture Notes in Computer Science, vol. 4658, August 2007, pp. 88-97.
- [7] T. Nguyen. "A Survey of WiMAX Security Threats", Aug. 2009. [Online] Available: <http://www1.cse.wustl.edu/~jain/cse57109/ftp/wimax2.pdf/05.2013>.
- [8] P. Rengaraju, C. Lung, Yi Qu, and A. Srinivasan, "Analysis on Mobile WiMAX Security", IEEE TIC-STH, Information Assurance in Security and Privacy, Toronto, Ontario: Canada, Sept. 2009, pp. 27-29.
- [9] T.Han, N.Zhang, K.Liu, B.Tang and Y.Liu, "Analysis of Mobile WiMAX Security: Vulnerabilities and Solutions", Proc. of 5th int. Conf. on Mobile Ad Hoc and Sensor Syst., Germany, Aug. 2008, pp.828-833.
- [10] A. Deininger, S. Kiyomoto, J. Kurihara, and T. Tanaka. " Security Vulnerabilities and Solutions in Mobile WiMAX", International Journal of Computer Science and Network Security (IJCSNS), Vol. 7, No. 11, Novemeber, 2007.
- [11] M. Rahman, Mir Md. S. Kowsar, "WiMAX Security Analysis and Enhancement", proc. of 12th Int. Conf. on Comput. and Inform. Tech (ICCI), Dhaka, Dec.2009, pp. 679 - 684, doi:10.1109/ICCIT.2009.5407321.
- [12] T. Shon, B. Koo, J. Park, and H. Chang, "Novel Approach to Enhance Mobile WiMAX", EURASIP J. on Wireless Communication and Networking, Republic of Korea, 2010, pp.1-11.
- [13] Y. Chung and K. Moon, "Biometric Certificate Based Biometric Digital Key Generation with Protection Mechanism", IEEE comput.

- Soc. Frontiers in the Convergence of Biometric and Inform. Technologies, Jeju City, 2007, pp. 709-714. doi:10.1109/FBIT.2007.151.
- [14] P. Arul and A. Shanmugam "Generate a Key for AES using Biometric for VOIP Network Security", Journal of Theoretical and Applied Information Technology, Vo.5, No.2, 2005, pp. 107- 112.
- [15] H. Lee and T. Kwon "Biometric Digital Key Mechanisms for Telebiometric Authentication Based on Biometric Certificate", pp. 428-437, 2007.
- [16] "Biometrics Ideal Test" from <http://iris.idealtest.org/findTotalDbByMode.do?mode=Fingerprint/6.2013>
- [17] J. Kim, S. Hong, and B. Preneel "Related-key rectangle attacks on reduced AES-192 and AES-256," In FSE'07, volume 4593 of LNCS, Springer, 2007, pp. 225-241.

The Advantage of Moving Nodes in Formations in MANETs and M2ANETs

Abdullah Alshehri, John DeDourek, Przemyslaw Pohec

Faculty of Computer Science
University of New Brunswick
Fredericton, Canada

E-mail: {j0lp6, dedourek, pohec}@unb.ca

Abstract—MANETs are self-configuring infrastructure-less mobile networks characterized by dynamically changing topology and intermittent connectivity. Performance of MANETs depends on the movement pattern of their nodes. We propose, and simulate in network simulator NS-2, a novel group mobility model, a towing formation, and show its advantage in controlling the movement of mobile nodes in a MANET. In the towing formation model, one of the nodes within each formation acts as a leader, with the other nodes, called followers, trailing closely. We show the impact of the number of towing formations in a MANET network, the number of nodes in each towing formation, and the distance between the nodes in each towing formation on the data traffic through a MANET. The greater the number of formations and the greater the size of each formation, the better the throughput.

Keywords—MANET; group mobility; formations; NS-2.

I. INTRODUCTION

In recent years, we observed a heightened interest in mobile networks for use in emergency situations and disaster recovery. Ideally, such networks would have a limited reliance on (existing) fixed infrastructure, be easy to set up and resilient to disruptions resulting from hazardous environments and infrastructure breakdown.

The general model of a Mobile Ad hoc Network (MANET) fits very well the emergency deployment scenarios. MANET is a network consisting of mobile autonomous nodes that forward the data among themselves without, as opposed to cellular networks, the need for any fixed infrastructure [19]. Nodes are relatively inexpensive and therefore can be used in large numbers if necessary, to facilitate communication. The new M2ANET model [4] proposes to use a large number of mobile nodes to create a “cloud” of routing nodes, a mobile medium, to facilitate communication between designated senders and receivers. In M2ANET, when two stations cannot communicate directly, a large number of routing nodes is dispersed between them and acts as the medium forwarding and routing messages.

In many situations (i.e., military communications, emergency relief, search and rescue) members participating in the effort are organized into teams, often with a designated leader: an officer leading a platoon of soldiers, a coordinator giving directions to a rescue team, etc. Given the available resources, it may be advantageous (and economical) to

organize the members into teams. The leaders may possess special qualities, skills, intelligence and equipment that distinguish them from the rest of the team.

If mobile networks were to be organized on some of the same principles, we might designate some mobile network nodes as leaders. We may then equip these nodes with special facilities like advance control and guidance with GPS etc., and perhaps even a powered motion facility (e.g. an engine), while the other nodes may be left more limited in these functions and may only need, for instance, to be guided or even to be towed by a leader. Such an approach would simplify the control (i.e., decision made at a node, which way to move next) of movement in a MANET with a large number of nodes: one would only need to guide (a limited number of) leaders while the followers would simply tag along. In case of simple physical towing, neither a guidance system on a towed node, nor any type of communication between the leader and the followers would be required. One example of such a scenario where the leader is responsible for the direction of movement of another object is aerotowing a glider by a powered plane; another application is a multiple decoy system towed behind a warship [21].

In this paper, we propose a group mobility strategy based on the above scenario where some mobile nodes can independently choose their movements, while the other nodes are limited in their movements, and limited to following the leaders (designated nodes). In MANETs, small changes to the movements of nodes can lead to changes in network topology and therefore affect the performance characteristics, such as throughput. We study the impact of having the nodes moving in formations on the performance of a MANET running the AODV routing protocol [14].

In Section II, we present background on MANETs. In Section III, dynamic mobility control strategy for MANET with formations is proposed based on group behavior. Experiments with different group sizes and lengths are described, including simulation set up, results and analysis in Sections IV and V. Finally, we present the conclusion and future work, in Section VI and Section VII, respectively.

II. STATE OF THE ART

MANET is a kind of wireless network that consists of a group of mobile nodes that communicate with each other without relying on fixed infrastructure [19]. The Mobile Medium Ad hoc Network (M2ANET) is a kind of a MANET where the mobile nodes are divided into two

categories: nodes that forward the data only and cannot be a source or the final destination of any transmission, and the nodes that can originate and receive data [4]. The forwarding nodes form a cloud of mobile medium that enables communication between transmitting nodes. The data communication through the mobile medium is affected by the properties of the medium: node density, speed, and especially the movement patterns, etc.

The mobility models have been developed to represent the node distribution and the movements [5][6][7]. They capture the properties of node location, time-varying node speed, and the distinct behaviors of the nodes, for example following different movement paths. These statistical properties of network connectivity are studied and identified with the mobility model [8][9][11][12][13].

The closest to our towing group mobility model is the Reference Point Group Mobility (RPGM) proposed by Hong et al. [9]. In our towing model, the nodes in a group follow the group leader, while in RPGM it is the “logical center” that determines the group motion behavior.

III. THE STRATEGY: MOVING IN FORMATIONS

A. Preliminary

We assume that all of the nodes have the same communication range R . The nodes inside the range are called neighbors, and two or more neighbors can communicate. Each node has its location, which is simply denoted as $L(v)$. The location information can be maintained using Global Positioning System (GPS), or by using an inertial guidance system. In the experiments we initially place the clusters of nodes belonging to formations randomly at different locations in area $n \times n$ (Fig. 1). Within each cluster one node is arbitrarily designated as a leader (node with lowest index assigned during the simulation is used). We test the communication between one source, S , and one destination, D , in every scenario where both S and D are stationary and positioned at the opposites sides of the experimental area. The number of relay (forwarding) mobile nodes depends on the scenario.

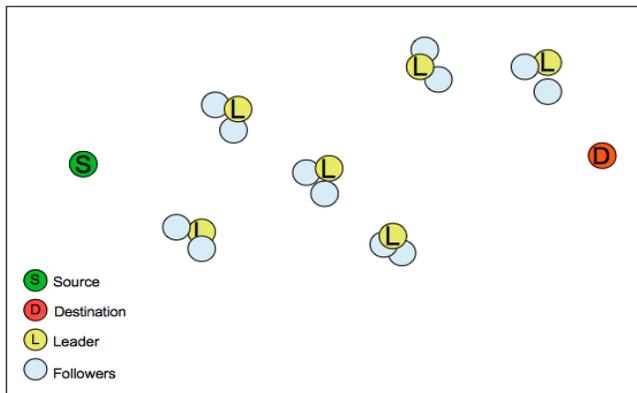


Figure 1. The initial placement of the leaders and the followers.

B. The Movement of the Leader and the Followers in the Towing Group Mobility Model

We propose a simple model for moving the nodes in a formation: each designated leader moves in a random direction following the random way point mobility model [20], while the followers retrace the path of the leader following it at a predetermined distance. The direction chosen by the leader not only defines the motion of the group leader itself, but it also provides the general motion trend of the whole formation.

In the first set of experiments, the distance between the followers ranges from 100 - 110 meters. In the second set, the distance was 150 - 210 m. In the simulation we assume that the leader L randomly chooses the direction of the next move, and sends information to the followers N including coordinates and movements. The movement of the followers is directly affected by the movement of its group leader, where every group member eventually lands on the same X and Y coordinates earlier visited by the leader. Every follower, N , sends directional information to the next $N+1$ in the same formation F and follows L , etc. When the movement of groups of nodes is viewed as an animation it looks like if the leader was towing a number of nodes behind it (Fig. 2).

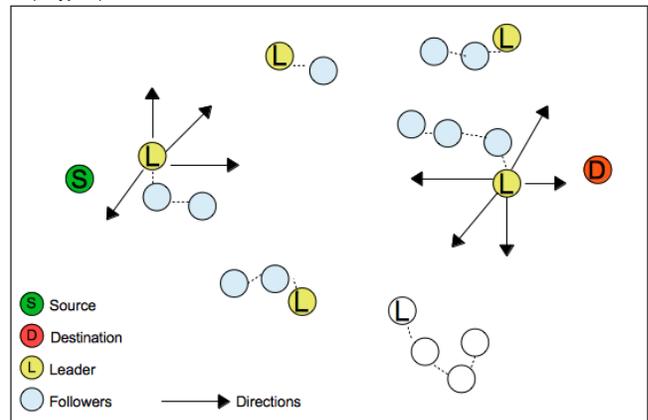


Figure 2. Movement and structure of formations of different lengths.

IV. SIMULATION

The towing group mobility model introduced in the previous section was implemented and simulated in NS-2. We used AWK scripts to analyze the trace files produced by the simulator.

A. Simulation Environment

In each case the network consisted of a different number of nodes roaming in a 1000 x 1000 meters square with a reflecting boundary. The NS-2.34 default settings for 802.11 legacy mode (Distributed Coordination Function DFC and 2Mbps links) are used. The transmission range is 250m. The data is generated at the source node at rate of 1 Mbps. Every packet has a size of 512 bytes. The buffer size at each node is 50 packets. Data packets are generated following a constant bit rate (CBR) process. The experiments we transmit packets

from one source to one destination for 900 seconds, and measure the effective throughput with increasing mobility range and density. The other assumptions are listed below and shown in Table 1.

- The source node, S, is stationary, and is located in a specific location on the far left.
- The destination node, D, is stationary, and is located in a specific location on the far right (see Fig. 1).
- The distance between S and D is 1000 meters.
- Nodes are being generated randomly at random locations as clusters.
- The intermediate nodes are moving at a constant speed of 10m/s.
- The distribution of the intermediate nodes is divided into a number of formations with different lengths (2, 3, or 4) clustered as an initial placement, F1, F2,..Fn (Fig. 2).
- A leader, L, from each formation is automatically elected.

We simulated a number of groups of nodes, each with a leader (L) and one to three followers behind it. All of the followers follow their leaders' X and Y coordinates. Each leader with its followers is called a formation. Fig. 3 shows the screen shot of the node movements for a number of formations each with two followers, all simulated in NS-2 and visualized using NAM animation tool, e.g., nodes n2, n3, and n4 form a formation of 3 nodes (1 leader + 2 followers).



Figure 3. Formations of three nodes; NS2 simulation.

TABLE I. SIMULATION PARAMETERS

Parameters	
Simulator	NS-2.34
Channel Type	Channel / Wireless Channel
Network Interface Type	Phy/WirelessPhy
Mac Type	Mac/802.11
Radio-Propagation Type	Propagation/Two-ray ground
Interface Queue Type	Queue/Drop Tail
Link Layer Type	LL
Antenna	Antenna/Omni Antenna
Maximum Packet in ifq	50
Area (n * n)	1000 x 1000
Source Type	(UDP) CBR
Simulation Time	900 sec
Routing Protocol	AODV

B. Performance Metric

The network throughput is a metric used to calculate the amount of data transmitted from source to destination in a specific period of time (in bits per second). It can be calculated as:

$$\gamma = \frac{\text{no. of bits received by node D}}{\text{Observation time}} \tag{1}$$

V. RESULTS AND ANALYSIS

The experimental results show the average throughput for the data transmitted from the source node S to the destination node D. The results are averaged over three experiments and plotted with the standard deviation shown as an error bar.

A. Case 1: Different Number of Formations

In Fig. 4, we show the performance of a MANET when a different number of formations of different lengths are used. The number of formations is varied from 6 to 10, and the length of formations is varied from 1 to 4. The graph shows that, for a given number of formations, the longer formations perform better. It should be noted that longer formations use larger number of nodes in total in the experiment: 10 formations of length 1 require only 10 nodes, while 10 formations of 4 would use the total of 40 nodes (4 leaders and 36 followers). Also, increasing the length of a formation we keep the number of leaders the same and add followers only, which may be an advantage in some applications.

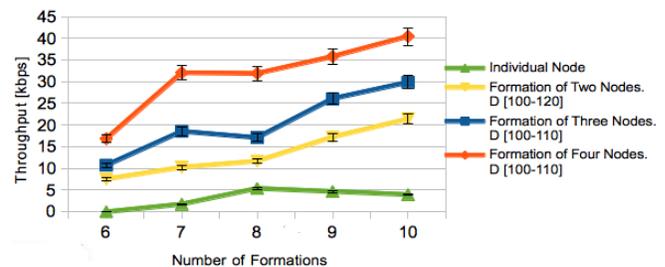


Figure 4. Throughput vs. number of formations.

B. Case 2: Different Number of Nodes

We tested the scenarios where a given number of mobile nodes (24, 36, 48, and 60) was equally divided into a number of formations with predetermined length (2, 3, and 4). For example, with 36 nodes we had either 18 formations of two nodes, or 12 formations of three nodes. We also show the effectiveness of increasing the default distance between the

nodes in a formation in our group mobility by running two sets of experiments one with an average of 100m - 110m and the other with 150m - 210m, as illustrated in Fig. 5.

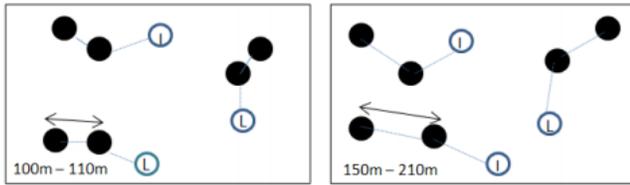


Figure 5. The distance between follower nodes

For a formation of two nodes (the leader and one follower), increasing the distance between the nodes improves the performance, but only for the cases with 36 and 48 mobile nodes (Fig. 6). Formations of three always work better when the followers keep back at a larger distance (Fig. 7). However, Fig. 8 shows the opposite: the three followers following at a short distance give slightly better throughput compared with the scenario when nodes follow at a longer distance. One could stipulate that only for short formations, the performance is improved with the increased distance between the nodes.

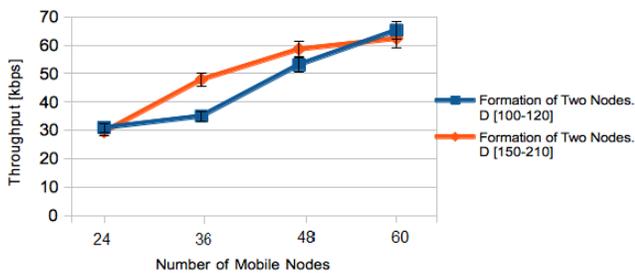


Figure 6. Throughput with length two formations.

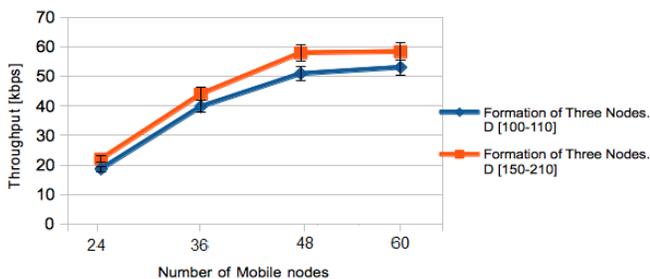


Figure 7. Throughput with length three formations.

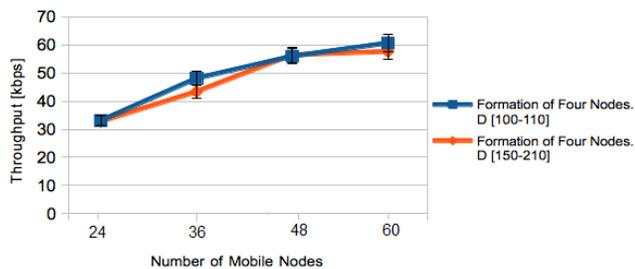


Figure 8. Throughput with length four formations.

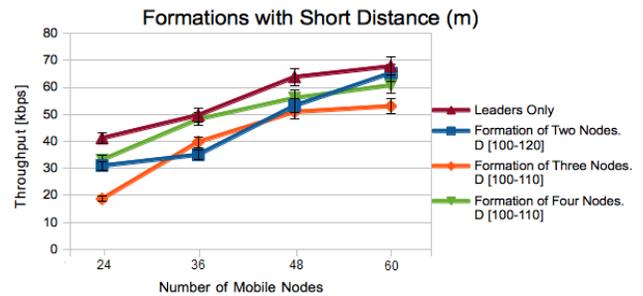


Figure 9. Throughput for short distance between the nodes for different formations.

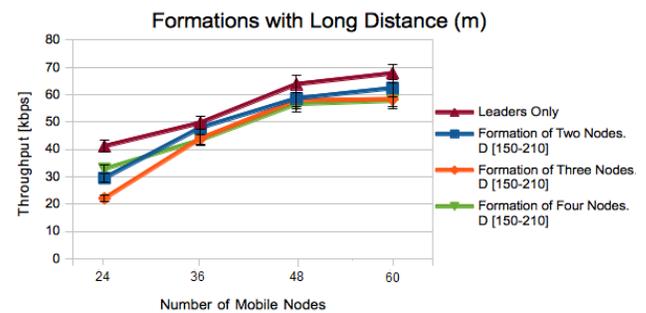


Figure 10. Throughput for long distance between the nodes for different formations.

In all the comparisons, the formation of one node (the leader only) consistently gave the best throughput (Figs. 9 and 10).

For all other cases, if we set the distance separating the nodes in the towing formation to be large and analyze the role of the formation size, we see consistent results for all the cases with a larger distance between the nodes in a formation: for the cases with following distance in the 150-210m range we notice the shorter the formation the better the throughput for all experiments except for the case with the smallest number of nodes (24 in Fig. 10). We also noticed that there was no performance difference when formations of three or four nodes were used in the experiments with 36 and more nodes (Fig. 10).

For all the cases where the distance separating the nodes in the towing formation is small, the preference for any particular length of the formation is not so clear cut. Moving nodes individually (formation of one) is still always the best, formation of four is always significantly better than the formation of three, and the relative performance of the formation of two varies and seems to depend on the number of nodes.

VI. CONCLUSION AND FUTURE WORK

In this paper, we presented a group mobility mode for moving mobile nodes in formations. All the available nodes in a network are divided into groups and each group forms a formation and moves independently from the others. We proposed a mobility model for moving the nodes in a formation called “towing”. In the towing formation model

one of the nodes within a formation acts as a leader, with the other nodes, now called followers, trailing closely.

The towing formation model was tested by simulation. We used the recorded throughput for the CBR traffic between designated nodes as the main performance metric. We observed that the larger the total number of mobile nodes the better the throughput. Also, the larger the number of formations, the higher the throughput. Consequently, for a given total number of nodes, using shorter formations results in more formations being created, and leads to a better performance (observed at larger distances separating the nodes in a formation). In the case of a fixed total number of mobile nodes, increasing the number of formations by shortening the formations all the way to only one node, i.e., moving nodes individually always gives the best performance. For short formations, increasing the distance between the followers improves the performance.

It should be noted that the fact that the best throughput was recorded when moving nodes individually does not negate the results of this research. As stated in Section I, using the towing formation implies having two types of nodes: the leaders and the followers. For example, when the number of available leaders is limited, one can still improve the performance of a MANET by increasing the number of followers. As a part of future work, we would like to benefit of adding more nodes to a network and the tradeoffs between adding the leaders and the followers. We would like to investigate forming formations automatically from nodes that are not prearranged into clusters. The (direction of) movement of the towing formations could be optimized to take a maximum advantage of the guaranteed connectivity between the nodes within the cluster to extend the coverage area to the maximum. The random movement in free space could be replaced with a more realistic model based on the actual maps and street layout.

ACKNOWLEDGMENT

We would like to thank the Ministry of Higher Education of Saudi Arabia including the Saudi Cultural Bureau in Canada for their support.

REFERENCES

- [1] M. K. J. Kumar and R.S. Rajesh. "Performance analysis of MANET routing protocols in different mobility models." *IJCSNS International Journal of Computer Science and Network Security*, vol. 9, no. 2, Feb 2009, pp. 22-29.
- [2] C. M. Cordeiro and D. P. Agrawal, "Mobile ad hoc networking." in *Tutorial/Short Course in 20th Brazilian Symposium on Computer Networks*, May 2002, pp. 125-186.
- [3] D. Goldenberg, J. Lin, A. S. Morse, B. Rosen, and Y. R. Yang. "Towards Mobility as a Network Control Primitive," in *Proc. of the 5th ACM MobiHoc*, 2004, pp. 163-174.
- [4] J. DeDoutre and P. Pochee, "M2ANET: a Mobile Medium Ad Hoc Network," *Wireless Sensor Networks: Theory and Practice*, WSN 2011, Paris, France, Feb. 2011, pp. 1-4.
- [5] X. Chen, Z. Jiang, and J. Wu, "Mobility control schemes with quick convergence in wireless sensor networks," *Parallel and Distributed Processing*, IEEE International Symposium on, April 2008, pp. 1-7, doi: 10.1109/IPDPS.2008.4536119.
- [6] J. Wu and F. Dai, "Mobility control and its applications in mobile ad hoc networks," *Network*, IEEE, vol.18, no.4, July-Aug. 2004, pp. 30-35, doi: 10.1109/MNET.2004.1316759.
- [7] F. Bai and A. Helmy, *A Survey of Mobility Modeling and Analysis in Wireless Ad Hoc Networks, A Survey of Mobility Models in Wireless Adhoc Networks*, Kluwer academic Publishers, University of Southern California, U.S.A., pp. 2-32, 2004.
- [8] A. Jadbabaie, L. Jie, and A.S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *Automatic Control*, IEEE Transactions on, vol. 48, no. 6, June 2003, pp. 988-1001, doi: 10.1109/TAC.2003.812781
- [9] X. Hong, M. Gerla, G. Pei, and C.-C. Chiang. "A Group Mobility Model for Ad Hoc Wireless Networks." *Proc. ACM Int'l Workshop Modeling, Analysis, and Simulation of Wireless and Mobile Systems (MSWiM)*, 1999, pp. 53-60.
- [10] J. Hortelano, M. Nacher, J.-C. Cano, C. Calafate, and P. Manzoni. "Evaluating the goodness of MANET's performance results obtained with the ns-2 simulator," in *NSTools'07*, Nantes, France, 2007, pp. 1-7.
- [11] S. Gunasekaran and N. Nagarajan, "An Improved Realistic Group Mobility Model for MANET Based On Unified Relationship Matrix," *Advance Computing Conference*, 2009. IACC 2009. IEEE International, March 2009, pp. 1270-1274, doi: 10.1109/IADCC.2009.480919.
- [12] B. Malarkodi, P. Gopal, and B. Venkataramani, "Performance Evaluation of Adhoc Networks with Different Multicast Routing Protocols and Mobility Models," *Advances in Recent Technologies in Communication and Computing*, 2009. ARTCom '09. International Conference on, Oct. 2009, pp. 81-84, doi:10.1109/ARTCom.2009.29.
- [13] K. Blakely and B. Lowekamp, "A structured group mobility model for the simulation of mobile ad hoc networks," in *Int. Conf. Mob.Comp. Netw., Proc. 2nd Int. Worksh. Mob. Manag. Wirel. Acc. Protoc.*, Philadelphia, USA, 2004, pp. 111-118.
- [14] D. O. Jorg, "Performance comparison of MANET routing Protocols in different network", *Institute of Computer Science and Applied Mathematics, Computer Networks and Distributed Systems, University of Berne, Switzerland, Computer Science Project Report*, February 2004.
- [15] H. A. Lagar-Cavilla, G. Baron, T. E. Hart, L. Litty and E. de Lara, "On the robustness of simple indoor MANET simulation models," *Ad hoc & sensor wireless networks*, 2007, pp. 321-354.
- [16] H. Akcan, V. Kriakov, H. Bronnimann and A. Delis. "GPS-Free Node Localization in Mobile Wireless Sensor Networks." *Proc. Fifth Int'l ACM Workshop Data Eng. for Wireless and Mobile Access (MobiDE '06)*, June 2006, pp. 35-42.
- [17] L.A. Latiff and N. Fisal, "Routing Protocols in Wireless Mobile Ad Hoc Network" - a review, *Communications, The 9th Asia-Pacific Conference on*, 2003, pp. 600-604.
- [18] E. Royer and C. Toh, "A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Networks". *IEEE Personal Communications* 6 (2), April 1999, pp. 46-55.
- [19] C. E. Perkins, *Ad Hoc Networking*, New York, Addison-Wesley, 2001.
- [20] T. Issariyakul and E. Hossain, *Introduction to Network Simulator NS2*, Springer, 2009.
- [21] N. Friedman, *The Naval Institute Guide to World Naval Weapon Systems*, Naval Institute Press, 2006

BER Performance of BICM-coded Cooperative Networks with Selection Decode-and-Forward Relaying over Nakagami- m Fading Channels

Tsang -Wei Yu

Dept. of Electrical and Computer Engineering
National Chiao Tung University
Hsinchu, Taiwan
toshiba.cm93g@nctu.edu.tw

Wern-Ho Sheen

Dept. of Communication and Information Engineering
Chaoyang University of Technology
Taichung, Taiwan
whsheen@cyut.edu.tw

Abstract—Bit-error-rate (BER) performance of coded cooperative networks with selection decode-and-forward (S-DF) relaying is analyzed over the Nakagami- m fading channels. Previously, in the literature, BER analysis was done only for the un-coded network, where an impractical symbol-based forwarding is employed. In this paper, we analyze the bit-interleaved code modulation (BICM) coded cooperative network with an easy-to-implement packet-based forwarding. In particular, a closed-form BER expression is derived for the S-DF with source re-transmission (S-DF/RT) relaying, where the source re-transmits the packet to the destination on behave of a relay if it fails to decode. The accuracy of the proposed analysis is confirmed by computer simulations.

Keywords—cooperative communication; bit-interleaved coded modulation; selection decode-and-forward.

I. INTRODUCTION

Cooperative communication has recently emerged as a promising technique to combat multi-path fading in wireless systems, thanks to its ability to provide spatial diversity for the size-limited mobile terminals [1][2]. It exploits the broadcast nature of wireless communications by allowing intermediate nodes, called relays, to overhear the packet transmitted from the source and forward it to the destination. One of the commonly employed relaying methods is the fixed decode-and-forward (F-DF) [2][3], where relays decode the received packet, re-encode and forward it to the destination. However, as was shown in [2], F-DF fails to provide full diversity because the received packet is always forwarded by a relay even if it is decoded erroneously. Selection DF (S-DF) is a relaying method proposed in [2] to overcome the shortcoming of F-DF, where a relay forwards the overheard packet only when it is decoded correctly. In this way, a full diversity can be achieved.

Performance analysis of the S-DF relaying has been a topic of extensive research [2][4]-[10]. Analyses were done over the Rayleigh fading channels from the aspects of capacity [4], outage probability [2][5] and symbol-error-rate (SER) [6], respectively. Very recently, analysis has been extended to the Nakagami- m fading channels in [7]-[10] for the un-coded network. In particular, in [7], SER was analyzed for a single-relay network under the correlated and uncorrelated channels, and exact SER was provided in [8] for multiple-relay networks. In [9], a closed-form expression for the moment generating function of the received signal-

to-noise ratio (SNR) at the destination was derived, and it was used to evaluate SER, outage probability and channel capacity. Lastly, in [10], SER and diversity order were investigated for the networks with inter-relay links.

In [6]-[10], the analyses were focused on the un-coded network using a symbol-based forwarding in which symbols are detected separately at a relay, and only the correct symbols are forwarded to the destination. Unfortunately, the symbol-based forwarding is not practical in real systems because whether a particular symbol is detected correctly or not is not known at the relay. In addition, a huge signaling overhead is needed for notifying exactly which symbols are forwarded.

In this paper, we analyze the bit-interleaved coded modulation (BICM) [11][12] coded cooperative network over the Nakagami- m fading channels, using a packet-based forwarding which can be easily implemented by using a cyclic redundancy check (CRC). BICM has been extensively applied in real systems. To our best of knowledge, this work is the first attempt to do the performance analysis for such a system. Only the S-DF with source re-transmission (S-DF/RT), where the source re-transmits the packet to the destination on behave of a relay if it fails to decode [2], is treated explicitly here; the S-DF with source idle in which source stays idle in the case of relay decoding failure can be viewed as a special case of S-DF/RT. A closed-form BER expression is derived, and simulation results are given to confirm the accuracy of the proposed analysis.

This paper is organized as follows. Section II describes the system models. BER performance is analyzed in Section III with numerical results presented in Section IV. Finally, conclusions are given in Section V.

II. SYSTEM MODELS

We consider the cooperative relaying network with one source, R relays and one destination which are indexed by $0, 1, \dots, R$ and $R+1$, respectively. Each node is equipped with one antenna, and relays operate in the half-duplexing manner implying that they cannot transmit and receive simultaneously. In the S-DF/RT relaying [2], transmission of a packet is divided into two phases; at phase-I, the source broadcasts a packet to relays and the destination, and, at phase-II, relays forward the received packet over orthogonal channels to the destination if it is decoded correctly (with a CRC). In the case of decoding failure at a relay, on the other hand, the relay keeps silent, and the source retransmits the

packet on the relay's behave over the orthogonal channel allotted to that relay. For convenience, the orthogonal channel allotted to relay j at phase-II will be denoted as orthogonal-channel j .

A. Channel Model

A frequency-nonselctive fading channel is considered. Define $h_{0,j}(k)$, $j=1,\dots,R+1$ the channel gain between the source and node j at the k -th channel use, $h_{i,R+1}(k)$ $i=1,\dots,R$ the channel gain between relay i and the destination, and $h_{0,R+1}^{(j)}(k)$, $j=1,\dots,R$ the channel gain between the source and the destination at phase-II that uses orthogonal-channel j for retransmission. The channels are assumed to be mutually independent, and under the assumption of a symbol inter-leaver with a depth larger than the channel coherent time, the channel gains of a channel are independent and identically distributed (i.i.d.) over different time index k .

The general Nakagami- m fading model is adopted in this study, with the probability density function (pdf)

$$p(h) = \frac{2m^m h^{2m-1}}{\Gamma(m)\Omega^m} \exp\left(-\frac{mh^2}{\Omega}\right) \quad (1)$$

to characterize the channel gain h [12], where m is the shaping factor (assumed to be a positive integer), Ω is the average power of the channel, and $\Gamma(\cdot)$ is the Gamma function. Perfect channel-state-information (CSI) will be assumed available at all receiving nodes.

B. Bit-interleaved Coded Modulation

BICM is employed at all nodes. At the source, an information bit sequence \mathbf{b} of length K is encoded into a coded sequence \mathbf{c} of length N . After interleaving, the sequence is partitioned successively into groups of l bits, called the labels, which are then mapped to a sequence of complex symbols $\{x(k) \in \mathcal{X}\}$ for transmission according to a signal mapper μ and a signal constellation \mathcal{X} .

At phase-I, the received signals at relays and the destination at time k are given by

$$y_{0,j}(k) = h_{0,j}(k)\sqrt{P_0}x(k) + \omega_{0,j}(k), \quad j=1,2,\dots,R+1, \quad (2)$$

where P_0 is the source transmit power, and $\omega_{0,j}(k)$ is the AWGN (additive white Gaussian noise) at node j . All noises are modeled as i.i.d. zero-mean, circularly-symmetric complex Gaussian random variables with variance $N_0/2$ per dimension. Upon receiving $y_{0,j}(k)$, relay j calculates the maximum log-likelihood ratio (LLR) for the i -th bit of the k -th symbol according to

$$\min_{x \in \mathcal{X}_b^i} \frac{|y_{0,j}(k) - h_{0,j}(k)\sqrt{P_0}x|^2}{N_0} - \min_{x \in \mathcal{X}_b^i} \frac{|y_{0,j}(k) - h_{0,j}(k)\sqrt{P_0}x|^2}{N_0}, \quad (3)$$

where \mathcal{X}_b^i is the subset of signal points in \mathcal{X} with the binary value b at the i -th position of the label. The LLRs of the coded sequence are then de-interleaved and decoded. The Max-log MAP (maximum a posteriori probability) decoder [16] is employed at all receiving nodes

In the S-DF/RT relaying, relay j forwards the received packet to the destination if \mathbf{b} is decoded correctly. Otherwise, it notifies the source to re-transmit the packet over orthogonal-channel j . Define $\Theta \subseteq \{1,\dots,R\}$ as the set of active relays which have decoded successfully at phase-I and $\bar{\Theta}$ as its complement set. Then, at phase-II, the signals received at the destination can be expressed by

$$y_{j,R+1}(k) = h_{j,R+1}(k)\sqrt{P_j}x(k) + \omega_{j,R+1}(k), \quad j \in \Theta, \quad (4)$$

and

$$y_{0,R+1}^{(j)}(k) = h_{0,R+1}^{(j)}(k)\sqrt{P_j}x(k) + \omega_{0,R+1}^{(j)}(k), \quad j \in \bar{\Theta}, \quad (5)$$

where P_j is the transmit power over orthogonal-channel j , and $\omega_{j,R+1}(k)$ and $\omega_{0,R+1}^{(j)}(k)$ are the corresponding AWGNs. For $j \in \bar{\Theta}$, P_j may assume different values depending on whether the orthogonal channels are implemented in the time, frequency or code domain. The received signals at phase-I and phase-II are combined and decoded jointly at the destination.

The max-log likelihood ratio for the i -th bit of the k -th symbol at the destination is evaluated by

$$\min_{x \in \mathcal{X}_b^i} \sum_{j=0}^R \frac{|\tilde{y}_{j,R+1}(k) - \tilde{h}_{j,R+1}(k)\sqrt{P_j}x|^2}{N_0} - \min_{x \in \mathcal{X}_b^i} \sum_{j=0}^R \frac{|\tilde{y}_{j,R+1}(k) - \tilde{h}_{j,R+1}(k)\sqrt{P_j}x|^2}{N_0}, \quad (6)$$

where $\tilde{y}_{j,R+1}(k) \doteq \begin{cases} y_{j,R+1}(k), & \text{if } j=0 \text{ or } j \in \Theta \\ y_{0,R+1}^{(j)}(k), & \text{if } j \in \bar{\Theta} \end{cases}$, and

$$\tilde{h}_{j,R+1}(k) \doteq \begin{cases} h_{j,R+1}(k), & \text{if } j=0 \text{ or } j \in \Theta \\ h_{0,R+1}^{(j)}(k), & \text{if } j \in \bar{\Theta} \end{cases}. \quad \text{The max-log}$$

likelihood ratios of the coded sequence are then de-interleaved and passed to the decoder. For notation simplicity, $\Theta' \doteq \Theta \cup \{0\}$ is used in the rest of the paper.

III. BER ANALYSIS

Let $p_{b,R+1}^{\text{RT}}$ denote the BER at the destination using the S-DF/RT relaying. Under the packet-based forwarding, the BER is given by

$$p_{b,R+1}^{\text{RT}} = \sum_{\Theta \in \{1,2,\dots,R\}} p_{b,R+1}^{\text{RT}}(\Theta) \prod_{j \in \Theta} (1 - p_{f,j}) \prod_{j \in \bar{\Theta}} p_{f,j}, \quad (7)$$

where $p_{b,R+1}^{\text{RT}}(\Theta)$ is the BER at the destination given Θ , and $p_{f,j}$ is the packet-error-rate (PER) at relay j . In what follows, $p_{b,R+1}^{\text{RT}}(\Theta)$ is analyzed first, followed by the

analysis of $p_{f,j}$. Recall that in [6]-[10] SER at the destination was analyzed for the un-coded relaying system under an impractical symbol-based forwarding.

Using the assumptions of ideal interleaving and symmetrization in [12], $p_{b,R+1}^{\text{RT}}(\Theta)$ can be estimated by

$$p_{b,R+1}^{\text{RT}}(\Theta) \approx \sum_{d_h=d_f}^N w_l(d_h) f_{\text{ex}}^{\text{RT}}(d_h, \Theta), \quad (8)$$

where $w_l(d_h)$ is the total information bits of the error events with Hamming weight d_h divided by K , d_f is the free distance of the code, and $f_{\text{ex}}^{\text{RT}}(d_h, \Theta)$ is the expurgated upper bound of the pair-wise error probability (PEP) between two coded sequences with Hamming distance d_h .

As is shown in [14], $f_{\text{ex}}^{\text{RT}}(d_h, \Theta)$ can be expressed as

$$f_{\text{ex}}^{\text{RT}}(d_h, \Theta) = \frac{1}{2\pi j} \int_{s_0-j\infty}^{s_0+j\infty} \prod_{k=1}^{d_h} \left[\frac{1}{l2^l} \sum_{i=1}^l \sum_{b=0}^1 \sum_{x(k) \in \mathcal{X}_b^i} \prod_{j=0}^R \Phi_{\Delta_j(x(k), \hat{z}(k))}(s) \right] ds \quad (9)$$

where $j = \sqrt{-1}$,

$$\Phi_{\Delta_j(x(k), \hat{z}(k))}(s) = E_{\tilde{h}_{j,R+1}(k)} \left[\exp \left[(-s + s^2) \frac{P_j}{N_0} \tilde{h}_{j,R+1}^2(k) |x(k) - \hat{z}(k)|^2 \right] \right] \quad (10)$$

is the moment generating function of the metric difference

$$\Delta_j(x(k), \hat{z}(k)) \doteq E_{\tilde{h}_{j,R+1}(k)} \left[\begin{array}{l} \log p(\tilde{y}_{j,R+1}(k) | x(k), \tilde{h}_{j,R+1}(k)) \\ -\log p(\tilde{y}_{j,R+1}(k) | z(k), \tilde{h}_{j,R+1}(k)) \end{array} \right], \quad (11)$$

and $\hat{z}(k)$ is the nearest neighbor of $x(k)$ in \mathcal{X}_b^i .

To evaluate $f_{\text{ex}}^{\text{RT}}(d_h, \Theta)$, firstly, $\Phi_{\Delta_j(x(k), \hat{z}(k))}(s)$ can be derived as [14]

$$\Phi_{\Delta_j(x(k), \hat{z}(k))}(s) = \left(1 - (-s + s^2) \frac{\tilde{\Omega}_{j,R+1} P_j |x(k) - \hat{z}(k)|^2}{\tilde{m}_{j,R+1} N_0} \right)^{-\tilde{m}_{j,R+1}} \quad (12)$$

with the region of convergence (ROC)

$$\frac{1}{2} - \sqrt{\frac{\tilde{m}_{j,R+1} N_0}{\tilde{\Omega}_{j,R+1} P_j |x(k) - \hat{z}(k)|^2} + \frac{1}{4}} < \text{Re}\{s\} < \frac{1}{2} + \sqrt{\frac{\tilde{m}_{j,R+1} N_0}{\tilde{\Omega}_{j,R+1} P_j |x(k) - \hat{z}(k)|^2} + \frac{1}{4}}, \quad (13)$$

where

$$(\tilde{m}_{j,R+1}, \tilde{\Omega}_{j,R+1}) = \begin{cases} (m_{j,R+1}, \Omega_{j,R+1}), & j \in \Theta' \\ (m_{0,R+1}, \Omega_{0,R+1}), & j \in \bar{\Theta} \end{cases}, \quad (14)$$

and $m_{i,j}$ and $\Omega_{i,j}$ are the shaping factor and the average power of the channel between node i and j , respectively. Since the saddle point 0.5 always lies in the ROC, the

integration in (9) can be evaluated efficiently along with the vertical line of $s = 0.5 + jt$ [15]. Using this, (9) becomes

$$f_{\text{ex}}^{\text{RT}}(d_h, \Theta) = \frac{1}{4\pi} \int_{-\infty}^{\infty} \left[\frac{1}{l2^l} \sum_{i=1}^l \sum_{b=0}^1 \sum_{x \in \mathcal{X}_b^i} \prod_{j=0}^R \left(1 + \left(t^2 + \frac{1}{4} \right) \frac{\tilde{\Omega}_{j,R+1} P_j |x - \hat{z}|^2}{\tilde{m}_{j,R+1} N_0} \right)^{-\tilde{m}_{j,R+1}} \right]^{d_h} \left(t^2 + \frac{1}{4} \right)^{-1} dt, \quad (15)$$

where the time index k has been dropped because $\tilde{m}_{j,R+1}$ and $\tilde{\Omega}_{j,R+1}$ are the same for all k . In addition, (15) only contains the real part because the imaginary part of the integral in (9) is an odd function of t .

Secondly, some of (x, \hat{z}) pairs in (15) have the same squared Euclidean distance $|x - \hat{z}|^2$ and can be grouped together. By doing so, (15) is rewritten as

$$f_{\text{ex}}^{\text{RT}}(d_h, \Theta) = \frac{1}{4\pi} \int_{-\infty}^{\infty} \left[\sum_{i=1}^M C_i \prod_{j=0}^R \left(1 + \left(t^2 + \frac{1}{4} \right) \frac{\tilde{\Omega}_{j,R+1} P_j D_i}{\tilde{m}_{j,R+1} N_0} \right)^{-\tilde{m}_{j,R+1}} \right]^{d_h} \left(t^2 + \frac{1}{4} \right)^{-1} dt, \quad (16)$$

where $D_i = |x - \hat{z}|^2$ is a squared Euclidean distance, C_i is the number of (x, \hat{z}) pairs with $|x - \hat{z}|^2 = D_i$ over $l2^l$, and M is the number of distinct D_i 's. the examples, M , C_i and D_i for QPSK, 16-QAM and 64-QAM with Gray mapping, can be found in [14].

Lastly, from [14], a close-form expression of (16) is obtained as

$$f_{\text{ex}}^{\text{RT}}(d_h, \Theta) = \frac{1}{4} \sum_{\substack{u_1, u_2, \dots, u_M \\ d_h = u_1 + \dots + u_M}} \frac{d_h!}{u_1! u_2! \dots u_M!} \prod_{i=1}^M \left[C_i \prod_{j=0}^R \left(A_{i,j} - \frac{1}{4} \right)^{B_{i,j}} \right] \cdot \left(\sum_{i=1}^M \sum_{j=0}^R \sum_{k=1}^{B_{i,j}} E_{i,j,k} \frac{(2k-3)!!}{(2k-2)!!} A_{i,j}^{-\frac{2k-1}{2}} + E' \right), \quad (17)$$

$$\text{where } A_{i,j} = \frac{1}{4} + \frac{\tilde{m}_{j,R+1} N_0}{\tilde{\Omega}_{j,R+1} P_j D_i}, \quad B_{i,j} = u_i \tilde{m}_{j,R+1},$$

$$(2k+1)!! = 1 \cdot 3 \cdot 5 \cdots (2k+1), \quad (2k)!! = 2 \cdot 4 \cdot 6 \cdots (2k),$$

$$E_{i,j,k} = \frac{1}{(B_{i,j} - k)! ds^{B_{i,j} - k}} \left((A_{i,j} + s)^{B_{i,j}} G(s) \right) \Big|_{s=A_{i,j}}, \quad (18)$$

$$E' = \left(\frac{1}{4} + s \right) G(s) \Big|_{s=-\frac{1}{4}}, \quad (19)$$

and $G(s) = \prod_{i=1}^M \prod_{j=0}^R (A_{i,j} + s)^{-B_{i,j}} \left(s + \frac{1}{4} \right)^{-1}$. With (17), the

BER $p_{b,R+1}^{\text{RT}}(\Theta)$ is now estimated by

$$\hat{p}_{b,R+1}^{\text{RT}}(\Theta) = \sum_{d_h=d_f}^N w_l(d_h) f_{\text{ex}}^{\text{RT}}(d_h, \Theta). \quad (20)$$

On the other hand, the PER at relay j , $p_{f,j}$, can be approximated by

$$\hat{p}_{f,j} = \begin{cases} 1, & \hat{p}_{b,j} \geq 1 \\ 1 - (1 - \hat{p}_{b,j})^K, & \text{otherwise} \end{cases}, \quad (21)$$

where $\hat{p}_{b,j}$ is an estimation of the BER $p_{b,j}$ at relay j . Following the same steps as in (8)-(20), $\hat{p}_{b,j}$ is given by

$$\hat{p}_{b,j} = \sum_{d_h=d_f}^N w_i(d_h) \frac{1}{4} \cdot \sum_{u_1, u_2, \dots, u_M} \frac{d_h!}{u_1! u_2! \dots u_M!} \prod_{i=1}^M \left[C_i^{u_i} \left(\tilde{A}_{i,j} - \frac{1}{4} \right)^{\tilde{B}_{i,j}} \right], \quad (22)$$

$$\cdot \left(\sum_{i=1}^M \sum_{k=1}^{\tilde{B}_{i,j}} \tilde{E}_{i,j,k} \frac{(2k-3)!!}{(2k-2)!!} \tilde{A}_{i,j}^{-\frac{2k-1}{2}} + \tilde{E}' \right)$$

where $\tilde{A}_{i,j} = \frac{1}{4} + \frac{m_{0,j} N_0}{\Omega_{0,j} P_0 D_i}$, $\tilde{B}_{i,j} = u_i m_{0,j}$, and $\tilde{E}_{i,j,k}$ and \tilde{E}' are obtained as in (18) and (19) with $A_{i,j}$, $B_{i,j}$ and $G(s)$ replaced by $\tilde{A}_{i,j}$, $\tilde{B}_{i,j}$, and $\tilde{G}_i(s) \triangleq \prod_{i=1}^M (\tilde{A}_{i,j} + s)^{-\tilde{B}_{i,j}} \left(s + \frac{1}{4} \right)^{-1}$, respectively. Using $\hat{p}_{b,R+1}^{\text{RT}}(\Theta)$ and $\hat{p}_{f,j}$, $p_{b,R+1}^{\text{RT}}$ now can be estimated by

$$\hat{p}_{b,R+1}^{\text{RT}} \doteq \sum_{\Theta \in \{1,2,\dots,R\}} \hat{p}_{b,R+1}^{\text{RT}}(\Theta) \prod_{j \in \Theta} (1 - \hat{p}_{f,j}) \prod_{j \in \bar{\Theta}} \hat{p}_{f,j}. \quad (23)$$

IV. SIMULATION RESULTS

In this section, simulation results are given to verify the accuracy of the proposed BER analysis. In all the simulations, the half-rate convolutional code with the generator matrix $(1 + D + D^2 + D^3 + D^6, 1 + D^2 + D^3 + D^5 + D^6)$ is employed, an S-random interleaver with depth 20 is used to break to correlation between bits in a label, and the Gray mapping is used for all QAM constellations. In addition, we assume $P_0 = P_1 = \dots = P_R = P$ for simplicity, where $P = E_b \cdot R_c \cdot l$ with the bit energy E_b and the channel code rate $R_c = 0.5$. The considered network configurations are summarized in Table I.

Fig. 1 shows the analytical and simulation results of $p_{b,2}^{\text{RT}}(\{1\})$, $p_{b,2}^{\text{RT}}(\emptyset)$, $p_{b,1}$ and $p_{f,1}$ for Network-1 (3-node), which is used to illustrate the accuracy of the analyses given

TABLE I. THE CONSIDERED NETWORK CONFIGURATIONS

Networks ^o	R^o	S-D link ^o	S-R link(s) ^o	R-D link(s) ^o
Network-1	1 ^o	$m_{0,2} = \Omega_{0,2} = 1^o$	$m_{0,1} = \Omega_{0,1} = 4^o$	$m_{1,2} = \Omega_{1,2} = 2^o$
Network-2 ^o	1 ^o	$m_{0,2} = \Omega_{0,2} = 1^o$	$m_{0,1} = \Omega_{0,1} = n^o$	$m_{1,2} = \Omega_{1,2} = 2^o$
Network-3 ^o	2 ^o	$m_{0,2} = \Omega_{0,2} = 1^o$	$m_{0,1} = \Omega_{0,1} = 4,^o$ $m_{0,2} = \Omega_{0,2} = 3^o$	$m_{1,3} = \Omega_{1,3} = 1,^o$ $m_{2,3} = \Omega_{2,3} = 2^o$

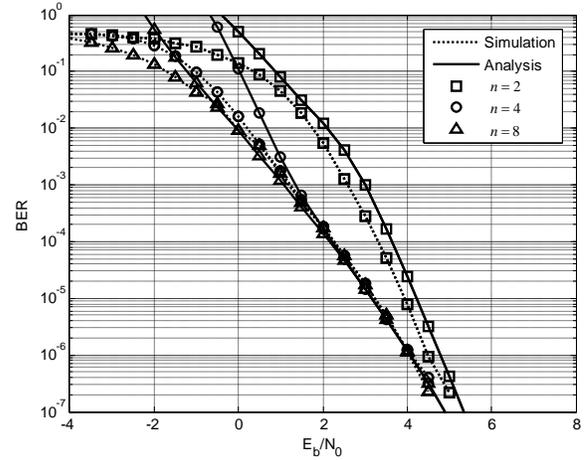


Figure 1. Simulation and analytical results of $p_{b,2}^{\text{RT}}(\{1\})$, $p_{b,2}^{\text{RT}}(\emptyset)$,

$p_{b,1}$ and $p_{f,1}$.

in (20), (21) and (22). Recall that in the 3-node network, the BER at the destination is given by (see (7))

$$p_{b,2}^{\text{RT}} = p_{b,2}^{\text{RT}}(\{1\})(1 - p_{f,1}) + p_{b,2}^{\text{RT}}(\emptyset)p_{f,1}. \quad (24)$$

As can be seen, the analytical results obtained in (20) and (22) predict the BERs very accurately at performance of interest, say $\text{BER} = 10^{-5}$. The PER prediction obtained in (21), on the other hand, is a bit mismatch with the true one, but the difference is only around 0.5 dB at $\text{PER} = 10^{-3}$.

Fig. 2 shows the simulation and analytical results of $p_{b,2}^{\text{RT}}$ for Network-2, which is a 3-node network as well but with a varying channel condition on the S-R channel. Clearly, the analysis in (23) predicts the BER performance at the destination very accurately at $\text{BER} = 10^{-5}$ for all the cases of $n = 2, 4$ and 8. The slight difference between the analytical and simulation results for $n = 2$ is due to that in this case $p_{b,2}^{\text{RT}}(\emptyset)p_{f,1}$ in (24) plays a more prominent role in determining $p_{b,2}^{\text{RT}}$, and there is a bit error in predicting $p_{f,1}$ as is shown in Fig. 2. In the typical scenarios of applying relay stations, the S-R channel often has a good channel condition; for example there is light-of-sight between the source and relay, and that results in a better BER prediction by the proposed analysis because in this case $p_{b,2}^{\text{RT}}$ is well approximated by $p_{b,2}^{\text{RT}}(\{1\})$ and which can be predicted accurately. (see the cases of $n = 4$ and 8 in Fig. 2.)

Fig. 3 shows the simulation and analytical results for Network-3 (4 nodes) for the S-DF/RT with different constellations. As is shown, the predictions are quite accurate; a less than 0.3 dB error at BER of 10^{-5} is observed.

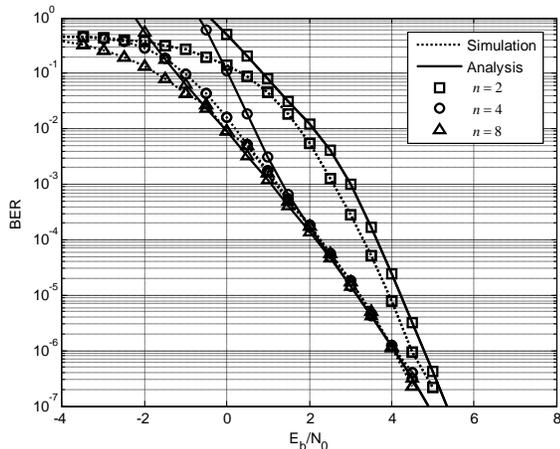


Figure 2. Simulation and analytical results of $p_{b,2}^{RT}$ for Network-2 with $n = 2, 4, \text{ and } 8$.

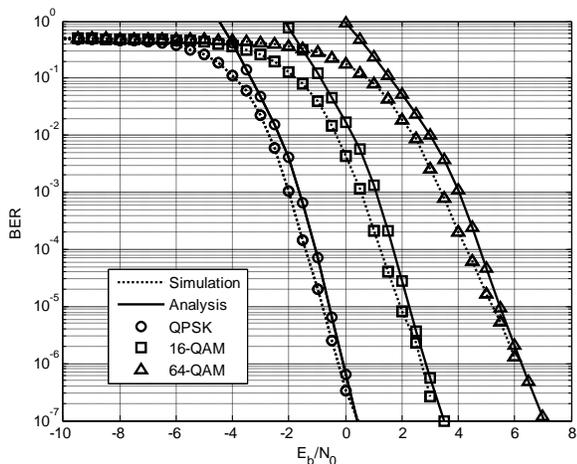


Figure 3. Simulation and analytical results of $p_{b,3}^{RT}$ for Network-3.

V. CONCLUSIONS

This paper investigates the performance of BICM-coded cooperative networks with the S-DF/RT relaying over the Nakagami- m fading channels. In particular, the BER performance is analyzed with a packet-based forwarding which can be implemented easily with CRC. In the literature, performance has been reported only for the un-coded system that uses an impractical symbol-based forwarding. A closed-form expression is derived and verified by computer simulations. The numerical results show that the prediction error is less than 0.5 dB at $\text{BER} = 10^{-5}$.

REFERENCES

[1] J. N. Laneman and G. W. Wornell, "Distributed space-time coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. on Inform. Theory*, vol. 49,

Oct. 2003, pp. 2415-2425.

[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 Inform. Theory*, vol. 50, Dec. 2004, pp. 3062-3080.

[3] Y. G. Kim and N. C. Beaulieu, "Exact closed-form solutions for the BEP of decode-and-forward cooperative systems in Nakagami- m fading channels," *IEEE Trans. on Commun.*, vol. 59, Sept. 2011, pp. 2355-2361.

[4] S. Lee, M. Han, and D. Hong, "Average SNR and ergodic capacity analysis for opportunistic DF relaying with outage over Rayleigh fading channels," *IEEE Trans. on Wireless Commun.*, vol. 8, June 2009, pp. 2807-2812.

[5] Y. Zhao, R. Adve, and T. J. Lim, "Outage probability at arbitrary SNR with cooperative diversity," *IEEE Commun. Letter*, vol. 9, Aug. 2005, pp. 700-702.

[6] W. Su, A. K. Sadek, and R. J. K. Liu, "SER performance analysis and optimum power allocation for decode-and-forward cooperation protocol in wireless networks," *Proc. IEEE WCNC'05*, IEEE Press, Mar. 2005, pp. 984-989.

[7] Y. Lee and M.-H. Tsai, "Performance of decode-and-forward cooperative communications over Nakagami- m fading channels," *IEEE Trans. on Vehicular Tech.*, vol. 58, March 2009, pp. 1218-1228.

[8] S. Amara, H. Boujemaa, and N. Hamdi, "SEP of cooperative systems using amplify and forward or decode and forward relaying over Nakagami- m fading channels," *Proc. IEEE Signals, Circuits and Systems (SCS) 2009*, IEEE Press, Nov. 2009, pp. 1-5.

[9] S. S. Ikki and M. H. Ahmed, "Performance analysis of multi-branch decode-and-forward cooperative diversity networks over Nakagami- m fading channels," *Proc. IEEE ICC 2009*, IEEE Press, June 2009, pp. 1-6.

[10] F. Xu, F.C.M. Lau, D.-W. Yue, and S.F. Hau, "Error rate and diversity order of multinode cooperative communications in dissimilar Nakagami fading channels," *IET Commun.*, vol. 3, Dec. 2009, pp. 1843-1850.

[11] E. Zehavi, "Eight-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, pp. 873-884, May 1992.

[12] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. on Inform. Theory*, vol. 44, May 1998, pp. 937-946.

[13] M. Nakagami, "The m -distribution, a general formula of intensity distribution of rapid fading," in *Statistical Methods in Radio Wave Propagation*, W. G. Hoffman, ed. Oxford, U.K.: Pergamon, 1960.

[14] T. W. Yu, "Performance Analysis and Power Allocation for BICM-Coded Cooperative Relaying Networks," Ph.D. Dissertation, Dept. of Elect. & Comp. Eng., National Chiao Tung Univ., Hsinchu, Taiwan, July 2012.

[15] C. W. Helstrom, *Elements of Signal Detection and Estimation*. New Jersey: PTR Prentice Hall, 1995.

[16] P. Robertson, E. Villebrun, and P. Hoeher, "A comparison of optimal and sub-optimal decoding algorithms in the log domain," *Proc. IEEE ICC'95*, IEEE Press, Jun. 1995, pp. 1009-1013.

High-Level Energy Saving Strategies for Mobile Location-Based Services on Android Devices

Wolfgang Narzt

Department of Business Informatics – Software Engineering
Johannes Kepler University Linz
Linz, Austria
wolfgang.narzt@jku.at

Abstract— The use of mobile location-based services (LBS) on modern smartphones, especially continuous utilization of GPS, significantly reduces the operating times of batteries. Considering the fact that the exhausting exploitation of power for determining the current location is unnecessary for various situations (e.g., when the device is not being moved), several high-level energy saving strategies can be developed in order to extend the batteries' operating times. This paper sketches the architecture for high-level energy saving strategies for LBS on Android devices encapsulated in a social-web application ("Spotnick") and illustrates their positive results.

Keywords - *Energy Saving Strategies; Location-Based Services; Android; Spotnick.*

I. INTRODUCTION

A manifold of location-based services (LBS) utilizing GPS, cellular network triangulation, or WLAN SSID mapping have evolved as more or less beneficial apps available for modern smartphones. Built-in positioning technology for location determination in the mobile devices is considered state-of-the-art and supports its consumers to find a path from A to B, recognize things around one's own position, record your training routes, or exchange current whereabouts with your friends. LBS are therefore considered an indispensable paradigm in mobile computing environments.

Nevertheless, continuous use of LBS still faces a considerable drawback in terms of batteries' operating times, which are likely to be reduced to less than 10% compared to the operating times in standby mode or to a maximum of 12 hours depending on the type of device (please, refer to the exact measurements in the following sections). In addition, the devices generate substantial heat, which is not always perceived as pleasing. In most cases, the exhausting use of GPS is the main reason for draining the battery. There is evidence on this statement due to several experiments, which we have conducted in order to figure out the energy guzzlers in LBS (see Section V). Naturally, the transmission of packets via the cellular network or WLAN also impacts operating times. The reduction of data transmission is not subject of this survey, though.

So, our main research issue for mobile LBS was to find a way to extend the batteries' operating times of smartphones for continuous active operation (i.e., in foreground and in

background mode). Of course, there cannot be a general answer or method to this issue because LBS strongly differ in terms of required accuracy and/or latency of reaction times. If a service e.g., claims permanent use of GPS due to exact location-based examinations, there is hardly any chance to develop a strategy for energy saving. Thus, we define a family of LBS applications with constraints to be fulfilled in order to be suitable to apply the energy saving strategies, which we propose in this paper. The constraints are defined as follows:

(i) The service is continuously operational, i.e., the position of the device must be determined both in foreground and in background mode. (ii) When in foreground mode, the best possible position of one's own device is required. (iii) The service dynamically offers location-based points of interest (POIs), which have to be triggered, i.e., their content displayed, an alarm set off, an external service initiated, etc., when the device reaches spatial proximity to them. (iv) An observation mechanism is included within a network of participating clients, i.e., the service shows the current whereabouts of other floating devices. (v) When in observation mode (presuming that there is no permanent monitoring station), the best possible positions of the monitored devices are required.

This means, that within a distributed LBS environment an exact position is only needed when the application is in foreground, observed by another client or close to POIs. For the remaining time (which represents the major part during operation in many cases, e.g., when the device is in background mode, not moving, not observed, or generally in a static position in the office or at night times, etc.), less energy consuming positioning methods (e.g., cellular network triangulation) could be used in order to extend the batteries' operating times. Hence, we propose to avoid utilization of GPS as often as possible and therefore present strategies for Android devices for implementing this paradigm (with the drawback of less accurate position determination in background mode and latency for the reactivation of GPS when required). The strategies are encapsulated in a social-web app named "Spotnick", which is available on Google Play and on the App Store, and provides the measuring results included in this paper.

The paper is structured as follows: Section 2 deals with selected points of state-of-the-art methods and technology. Section 3 gives an insight into the proposed energy saving

strategies. Section 4 sketches the test scenarios and measuring metrics. Section 5 provides figures and measured results and finally, Section 6 concludes the paper and prospects future work.

II. RELATED WORK

Investigations concerning energy consumption for location-based services on mobile devices have been the focal point of research in various scientific and industry labs [2][3][8][12]. Bareth and Kupper [5] e.g., have confirmed our experiments and derived findings about energy consumption of the different positioning technologies available on smartphones (see Table I), where they have measured the energy consumption for GPS, WiFi-, and cell-based positioning techniques.

TABLE I. PROPERTIES OF POSITIONING TECHNIQUES [5]

Technology	Accuracy	Precision	Energy
<i>A-GPS</i>	10m	95%	6.616Ws
<i>WiFi</i>	50m	90%	2.852Ws
<i>Cell-Id</i>	5km	65%	1.013Ws

Generally, there are three ways to face the draining problems: (1) directly on a hardware-related level, where ameliorations are achieved e.g., by considering Kalman filters [4], (2) by several software-based and application-related high-level strategies utilizing available SDK functions, software-architectures and framework [1][6][7][9] and (3) with hybrid approaches using additional (less energy consuming) sensor technology (e.g., accelerometers) and context-based movement detection algorithms as a substitute or an extension of functionality [11][13].

As the strategy described in this paper can be classified to approach (2), let us find out how similar methodologies work: Kjaergaard et al. [10] e.g., have developed a context-based model which tries to estimate and predict system conditions and movements in order to calculate a schedule for utilization of GPS. In a way, our approach can be compared to that, trying to schedule different positioning techniques due to system states (e.g., device is not moving). Although, the authors have succeeded in significantly saving energy using their high-level strategies using a Nokia N95, we try to give a more up-to-date approach using a less complex algorithm and utilizing enhanced positioning techniques on an Android device.

III. ENERGY SAVING STRATEGIES

The energy saving strategies for our approach are settled on a high application-oriented level, cleverly using, not using and switching between system functions and operations provided by standard SDKs. They are not based on hardware-related optimization attempts.

In principle, the strategy aims at avoiding usage of GPS whenever applicable. Instead, it uses cellular- or WiFi-based positioning or region monitoring (a technique which utilizes the WLAN and Cell-Id handover functions of the phone in order to detect a potential movement) and continuously tries to switch between those operation modes according to

particular events. Thus, the architecture results in a Finite State Machine (FSM) and has to take decisions for state conversions between those states built upon the following properties and its values:

1. *Application mode*: the app is either in *foreground mode (f)* (the user looks at it) or in *background mode (b)* (the app is idle).
2. *Movement*: the user/device is either *moving (m)* or *stationary (s)* (not moving).
3. *Observation mode*: the device is either *observed (o)* or *unobserved (u)*.
4. *POIs nearby*: points of interest for a device/user are either *near (n)* (i.e., within a distinct radius) or *distant (d)*.

Thus, there are four properties having two values each, which makes 16 different states, strictly speaking (see Table II). Most of them can be combined, though, because it is irrelevant for the strategy if the device is moving, observed or POIs are near or not when the app is in foreground mode (i.e., states 1 to 8 can be combined to one). In this state, the most accurate position is required. Also states 9 to 11 can be combined to one state, in which GPS is likely to be turned on, because the device is either being observed while in background mode (states 9 and 10) or a POI is near, the approach to it must be exactly determined (state 11). From a technical point of view, those two combinations of states 1 to 8 and 9 to 11 can be merged, once more, for they require the best possible position as a common characteristic. State 12 is unique: the app is in background mode, the device is moving, but not observed or any POIs near. This means that nobody takes notice of the app at that moment. Positioning must continue, though (by an inaccurate technique), in order to reactivate GPS at spatial proximity to a POI (which exactly requires to be triggered, transferred, etc.; for a deeper insight into location-triggered processes see [14]) within the inaccuracy range of the used positioning method. Finally, states 13 to 16 can again be combined to one state where region monitoring is used in order to recover from an idle mode where the app is in the background and not moving (observation and POIs are irrelevant at that state).

TABLE II. DIFFERENT STATES OF THE FSM

No.	App. mode	Move ment	Obs. mode	POIs near	No.	App. mode	Move ment	Obs. mode	POIs near
1.	<i>f</i>	<i>m</i>	<i>o</i>	<i>n</i>	9.	<i>b</i>	<i>m</i>	<i>o</i>	<i>n</i>
2.	<i>f</i>	<i>m</i>	<i>o</i>	<i>d</i>	10.	<i>b</i>	<i>m</i>	<i>o</i>	<i>d</i>
3.	<i>f</i>	<i>m</i>	<i>u</i>	<i>n</i>	11.	<i>b</i>	<i>m</i>	<i>u</i>	<i>n</i>
4.	<i>f</i>	<i>m</i>	<i>u</i>	<i>d</i>	12.	<i>b</i>	<i>m</i>	<i>u</i>	<i>d</i>
5.	<i>f</i>	<i>s</i>	<i>o</i>	<i>n</i>	13.	<i>b</i>	<i>s</i>	<i>o</i>	<i>n</i>
6.	<i>f</i>	<i>s</i>	<i>o</i>	<i>d</i>	14.	<i>b</i>	<i>s</i>	<i>o</i>	<i>d</i>
7.	<i>f</i>	<i>s</i>	<i>u</i>	<i>n</i>	15.	<i>b</i>	<i>s</i>	<i>u</i>	<i>n</i>
8.	<i>f</i>	<i>s</i>	<i>u</i>	<i>d</i>	16.	<i>b</i>	<i>s</i>	<i>u</i>	<i>d</i>

The consolidation of states reduces the complexity of the FSM to three states. Its state transitions are extractable from Table II, which is a simplified depiction of the concept, though; the actual implementation considers several details, which cannot be discussed here due to space limitations except the following (see Figure 1):

When transiting to a state requiring high accuracy, we do not immediately turn on the GPS provider of the Android platform. We first check the accuracy value of the network provider (which both combines Cell- and WiFi-based positioning on Android systems). If this value is good (i.e., just a few meters inaccuracy, which refers to a WiFi-based determination), then there is no need to activate GPS, e.g., when residing within a building where GPS is unavailable, anyway. Only a high inaccuracy value triggers the GPS provider (see Figure 1).

To turn GPS off again (and switch back to network positioning) several requirements have to comply: While moving, no other client must be observing and no POI must be near (i.e., within the inaccuracy radius of the network provider position). However, deactivation of GPS can also lead to a passive state where region monitoring is enabled (i.e., the app is completely idle until a system signal wakes it up again due to a network handover event among WLAN access points or cellular antennas). In order to detect a non-moving device, we have defined a distance threshold ϵ , which may not be crossed, and a time interval t , which must be exceeded until we assume a stationary state where the device is not moving (see Figure 1).

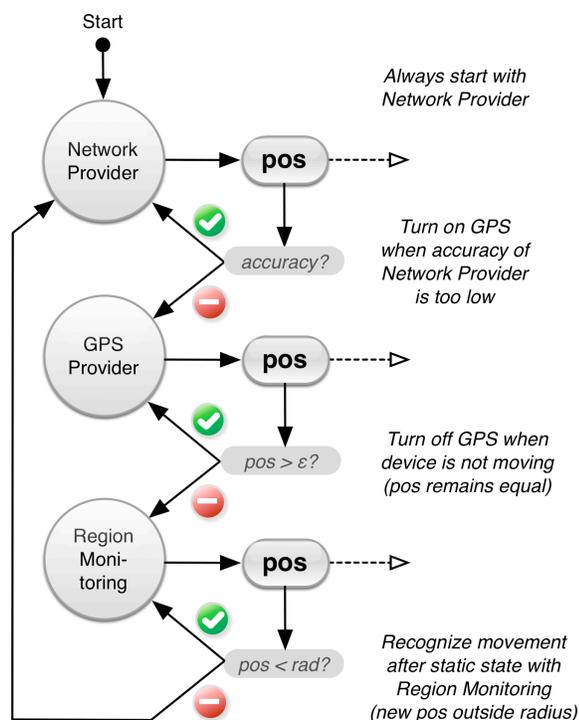


Figure 1. Flow Chart of Energy Saving Strategy (Simplified Excerpt)

The FSM as described above including several technical and strategic tricks as indicated briefly but omitted in detail have been implemented as an *intent service* in the core of a social-web application named Spotnick. Spotnick comprises a friend finder component for a user's Facebook friends with the opportunity of mutually perceiving their current whereabouts with just a few seconds delay. When looking at

the app, users are able to view their own position and those of their observed friends with maximum accuracy (depending on the devices' positioning capabilities). In addition, a user can trigger so called *spots*, i.e., self-set POIs that actively announce the arrival of a user on his Facebook wall when he reaches spatial proximity to it. Thus, Spotnick requires best possible accuracy when the app is in foreground mode, when observing friends or at regional closeness to spots. For the remaining time of usage (which we presume represents the major part of a day for a human being, e.g., when he/she is in his/her office or at night times), less accurate positioning methods comply in order to keep the app functional but not wasting energy.

IV. METRICS AND TEST SCENARIOS

In order to benchmark the energy saving qualities of our proposed methods using our app Spotnick, we have repetitively conducted a series of tests, which were organized as follows: All tests have been carried out sequentially on the same device, an HTC Desire HD with Android 2.3.5 as the operating system. All tests have used the same mini-app, which has recorded time stamps, WGS84 coordinates (if available) and the battery status via the Android SDK in regular time intervals.

Although, we have not been able to provide exact equal test conditions for the particular tests due to cellular network fluctuations, arbitrary WLAN availability, or the speed or duration of our trips while moving through the city traffic, we have at least tried to provide similar test conditions: Naturally, we have always used the same phone settings (i.e., WLAN on, Bluetooth off, UMTS preferred but EDGE allowed, no phone calls, no messages, no Internet surfing, no other apps running, etc.) and have performed daily recurring procedures, i.e., driven the same route from home to the office, which is about 20km distance or approximately half an hour of driving time (depending on the traffic situation) twice a day (bidirectional). At home and in the office, the app has been used in foreground for a minute to see where the other clients are residing. While driving, our device has been observed by another app for a minute and also been used in foreground for a minute. Thus, every possible status permutation has been created at least once (app in foreground, in background, device moving, not moving, observed, not observed, while driving, when stationary, etc.).

With these prerequisites, four test series have been conducted (using the mini-app for data recording):

1. *Standby Mode Test*: This test should examine the practical operating time when no app is running. It serves as a reference.
2. *Full GPS Test*: This test should examine the practical operating time when GPS is operational all the time.
3. *Full GPS and Data Transmission Test*: Same as test number 2, including the transmission of a simple HTTP data package every minute (simulating the transmission of position data for monitoring mode).
4. *Spotnick Test*: This test executed the app Spotnick, which contained the energy saving strategies described in the previous section.

V. RESULTS

According to the four test scenarios the measured results are as follows: Table III lists the parameters and typical results for one of the conducted *Standby Mode Tests*. The test was started on Jan. 31, 2013 and lasted for nearly 4 days until the battery was drained. All location services were turned off (i.e., no GPS and no cellular or WLAN-based positioning, however WLAN was turned on) and there was no data transmission during the test phase. We measured an average decline of 1,1% battery power per hour and took these values as a reference for the subsequent tests (i.e., the operating times are referred to as 100%).

TABLE III. MEASUREMENT RESULTS STANDBY MODE

HTC Desire HD, Android 2.3.5			
<i>Start Date</i>	Jan 31, 2013	<i>Duration</i>	3d 21h 36min
<i>Location Services</i>	off	<i>Avg. Decline</i>	1,1 % p.h.
<i>Data Transmission</i>	none, WLAN on	<i>Rel. Duration</i>	100,0 %

Figure 2 additionally illustrates the decline during the test phase. Please note, that the tests were not started at midnight – it is just the scale that starts at 0:00 for an easier comparison to the remaining tests. We do not recognize a clear linear decline of battery power, which we guess is due to varying environmental conditions, e.g., (wasted) attempts of the smartphone to connect to WLAN. However, we have not investigated these divergences in detail.

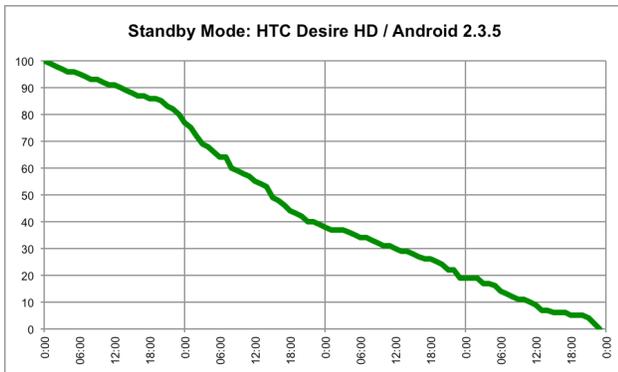


Figure 2. Battery Consumption in Standby Mode

Test number 2 was the *Full GPS Test*. Table IV lists the parameters and results for a typical test run carried out on Feb. 4, 2013. Continuous use of GPS and network-based location determination (i.e., cellular triangulation and WLAN mappings) reduce the operating time down to approx. 9 hours (which is only 9,7% of the *Standby Mode Test*) with an average decline of 11% battery power per hour.

TABLE IV. MEASUREMENT RESULTS CONTINUOUS GPS

HTC Desire HD, Android 2.3.5			
<i>Start Date</i>	Feb 4, 2013	<i>Duration</i>	9h 07min
<i>Location Services</i>	GPS+Network	<i>Avg. Decline</i>	11,0 % p.h.
<i>Data Transmission</i>	none, WLAN on	<i>Rel. Duration</i>	9,7 %

Figure 3 illustrates this drastic decline of battery operating times. For a better impression of the difference the original scale in this graph is the same as in Figure 2.

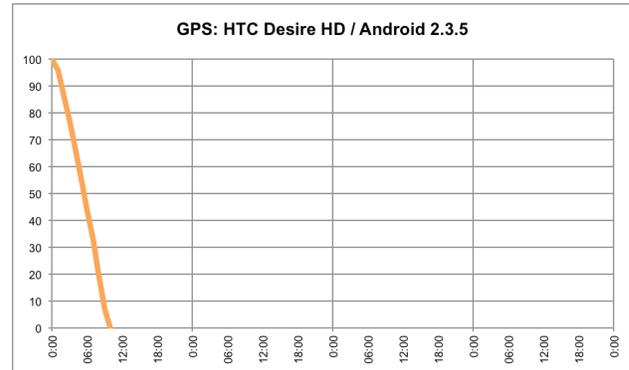


Figure 3. Battery Consumption with continuous GPS

However, it still goes worse. If we additionally transmit small HTTP packets in regular time intervals (i.e., every minute in this example) containing a timestamp, the current position and accuracy of the device’s positioning technique such that other observing clients may perceive the current whereabouts of this device (*Full GPS and Data Transmission Test*, see Table V and Figure 4) then we measure an operating time of only 7 hours or 7,5% compared to the *Standby Mode Test*. The battery loses power by 14,3% per hour!

TABLE V. MEASUREMENT RESULTS GPS+HTTP REQUESTS

HTC Desire HD, Android 2.3.5			
<i>Start Date</i>	Feb 6, 2013	<i>Duration</i>	7h 00min
<i>Location Services</i>	GPS+Network	<i>Avg. Decline</i>	14,3 % p.h.
<i>Data Transmission</i>	Every Minute	<i>Rel. Duration</i>	7,5 %

The measured results of tests 2 and 3 tell their own tale. It is obvious that continuous utilization of GPS (and beyond that coupled with regular data transmission) decisively decrease batteries’ operating times and therefore the attractiveness of location-based services for mobile devices continuously using GPS.

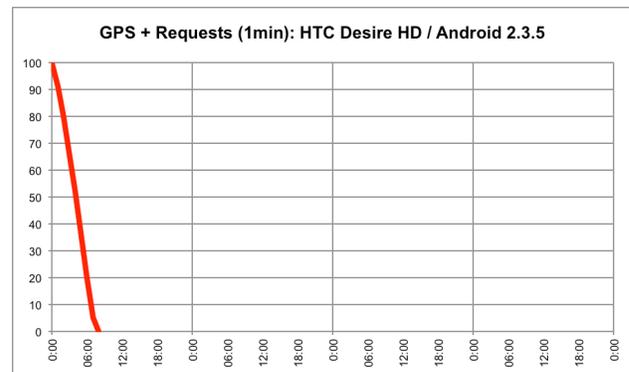


Figure 4. Battery Consumption for continuous GPS usage and HTTP Requests every Minute

Hence, the results for our proposed high-level energy saving strategies presented in section III are vital (regarding the *Spotnick Test*). They are consolidated in Table VI. The test lasted 2 ½ days, which is 65,5% compared to the *Standby Mode Test* with an average decline of 1,6% of battery power per hour.

TABLE VI. MEASUREMENT RESULTS SPOTNICK

HTC Desire HD, Android 2.3.5			
<i>Start Date</i>	Feb 17, 2013	<i>Duration</i>	2d 13h 16min
<i>Location Services</i>	Spotnick Strategy	<i>Avg. Decline</i>	1,6 % p.h.
<i>Data Transmission</i>	On Pos. Update	<i>Rel. Duration</i>	65,5 %

Figure 5 shows the corresponding graph. Considering the descent of the curve it is not clearly recognizable when the device was moving, stationary or observed, which applies to the fact that GPS was only used on demand (i.e., when the application was in foreground mode or the device had been observed – both situations occurred seldom during the tests and are therefore not evident in the graph).

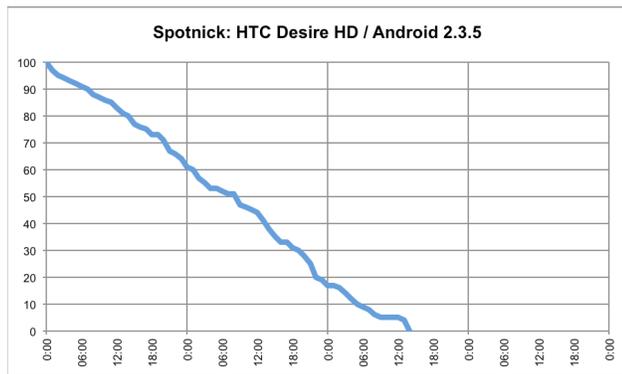


Figure 5. Battery Consumption using Spotnick

Considering the accuracy of the calculated positions, we refer to the successive two maps in Figures 6 and 7. Figure 6 shows a clipping of the daily way from our office at the University of Linz (starting in the top right corner of the map) to a home location (beyond the left edge). The red line marks the detailed path measured by GPS (recorded at test number 2). It represents the most accurate path we can get from the used mobile device (HTC Desire HD). The blue line marks the path from the Spotnick app including all energy saving mechanisms presented in section III. In the urban areas (in the right half of the map) the recorded positions are clearly distinguishable from the GPS points, however, for this use-case still accurate enough to recognize an unambiguous movement along a particular city district. In the rural areas (in the left half of the map) the recorded positions were less in amount (because farther apart) and less accurate (because of rural cellular positioning which is imprecise due to distant antennas). Nevertheless, also here an unambiguous movement is recognizable for possible observers for the short period of time until the moving device is informed about being monitored and therefore turns on GPS with a slight latency.

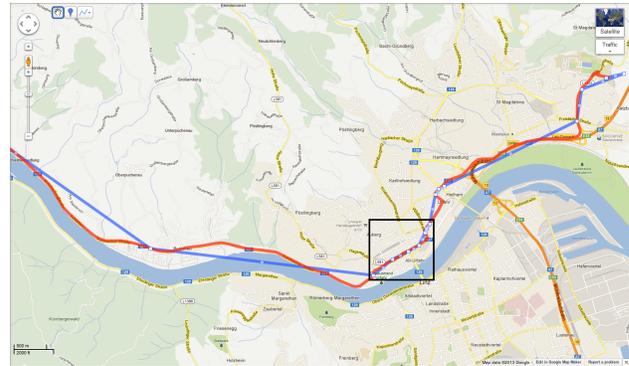


Figure 6. Accuracy Comparison GPS (red) vs. Spotnick (blue)

Figure 7 shows the switching on of GPS for a short period of time when a device is being monitored (see the blue line, clipped part out of Figure 6). Coming from the top right corner in this map, GPS has been turned on after the junction where the blue line diagonally crosses a block of buildings.

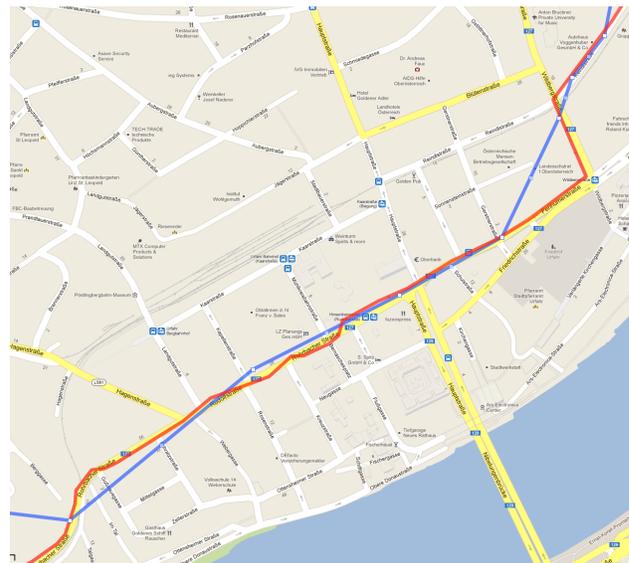


Figure 7. Detailed Accuracy Comparison GPS (red) vs. Spotnick (blue)

Please note, that the blue line (even when GPS is on) is built upon just a few measured points, which is due to a minimum distance to be covered and a minimum period of time to be elapsed for a position to be recorded and transferred. At the very left edge of Figure 7 the observation phase has ended and GPS turned off again, which immediately results in a clear deviation in the successive points.

VI. SUMMARY, CONCLUSION AND FUTURE WORK

Continuous utilization of GPS in LBS for modern smartphones expedites the process of draining batteries. In order to extend operating times we have developed high-level strategies for saving energy. In particular, we avoid usage of GPS whenever possible and use region monitoring and network-based positioning when a device is not moved,

e.g., at night times or while residing in the office. As a consequence, accuracy is diminished in certain situations (e.g., while moving with the app in background) and we have to accept latency for the reactivation of GPS whenever needed (e.g., when being monitored by other devices). Nevertheless, we succeeded in saving energy while guaranteeing a full functional service regarding location-based operations. Figure 8 summarizes the results of our conducted tests with a theoretical optimum when the device is in standby mode (green line) and a worst result with GPS engaged and regular data transmission (red line). In between is our energy saving strategy with Spotnick (blue line), which indicates a clear improvement compared to the exhaustive utilization of GPS.

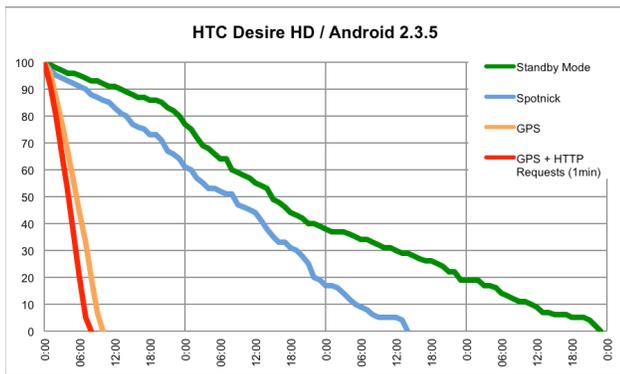


Figure 8. Battery Consumption HTC Desire HD with Android 2.3.5

In principle, these strategies also apply to further devices and operating systems. We have implemented them for iOS devices, as well. The technical details are different, because Apple restricts access to hardware-related functions and to system-related operations like process execution in background mode. However, the basic concept is similar, and also the results seem promising. Although, investigations on energy saving strategies for iOS devices are still underway, the first version of Spotnick including these strategies is available for iPhones in the App Store. Presuming comparable success for iOS, our proposed high-level energy saving strategies could be regarded as code of practice for mobile LBS, in general.

REFERENCES

- [1] S. Gabriel, C. Maciocco, C. Tai, and A. Min, "A EE-AOC: Energy Efficient Always-On-Connectivity Architecture," Proceedings of the 8th International IARIA Conference of Wireless and Mobile Communications, ICWMC 2012, ISBN: 978-1-61208-203-5, Italy 2012, pp. 110-117.
- [2] A. Carroll, "An Analysis of Power Consumption in a Smartphone," Proceedings of the 2010 USENIX Annual Technical Conference, Boston, MA, USA, USENIX Association Berkeley, CA, USA, 2010, pp. 21-21.
- [3] S. J. Barbeau, M. A. Labrador, A. Perez, P. Winters, N. Georggi, D. Aguilar, and R. Perez, "Dynamic Management of Real-Time Location Data on GPS-enabled Mobile Phones," Proceedings of the 2nd International Conference on Mobile Ubiquitous Computing, Systems, Services, and Technologies, (UbiComm 2008), Valencia, Spain, 2008, pp. 343 – 348.
- [4] I. M. Taylor and M. A. Labrador, "Improving the Energy Consumption in Mobile Phones by Filtering Noisy GPS Fixes with Modified Kalman Filters," Proceedings of IEEE Wireless Communications and Networking Conference, March 28-31, 2011, pp. 2006-2011.
- [5] U. Bareth and A. Kupper, "Energy-Efficient Position Tracking in Proactive Location-Based Services for Smartphone Environments," 2011 IEEE 35th Annual Computer Software and Applications Conference, July 2011, pp. 516–521,.
- [6] T. O. Oshin, S. Poslad, and A. Ma, "Improving the Energy-Efficiency of GPS Based Location Sensing Smartphone Applications," 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications, June 2012, pp. 1698–1705.
- [7] Z. Zhuang, D. T. R., and L. Altos, "Improving Energy Efficiency of Location Sensing on Smartphones," in Proceedings of the 8th international conference on Mobile systems, applications, and services, ISBN: 978-1-60558-985-5, 2010, pp. 315–330.
- [8] N. Balasubramanian, A. Balasubramanian, and A. Venkataramani, "Energy consumption in mobile phones: a measurement study and implications for network applications," Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference IMC 2009, New York, NY, USA, ISBN: 978-1-60558-771-4, 2009, pp. 280-293.
- [9] T. Farrell, R. Lange, and K. Rothermel, "Energy-efficient Tracking of Mobile Objects with Early Distance-based Reporting," Fourth Annual International Conference on Mobile and Ubiquitous Systems: Networking Services, MobiQuitous 2007, pp. 1-8.
- [10] M. Kjaergaard, J. Langdal, T. Godsk, and T. Toftkjaer, "EnTracked: energy-efficient robust position tracking for mobile devices," Proceedings of the 7th international conference on Mobile systems, applications, and services, MobiSys 2009, New York, NY, USA, ISBN: 978-1-60558-566-6, pp. 221-234.
- [11] A. Rahmati and L. Zhong, "Context-for-wireless: context-sensitive energy-efficient wireless data transfer," Proceedings of the 5th international conference on Mobile systems, applications and services, MobiSys 2007, New York, NY, USA, ISBN: 978-1-59593-614-1, pp. 165-178.
- [12] L. Wang and J. Manner, "Energy Consumption Analysis of WLAN, 2G and 3G interfaces," Green Computing and Communications (GreenCom), 2010 IEEE/ACM Int'l Conference on Int'l Conference on Cyber, Physical and Social Computing (CPSCom), 2010, pp. 300-307.
- [13] M. Youssef, M. Yosef, and M. El-Derini, "GAC: Energy-Efficient Hybrid GPS-Accelerometer-Compass GSM Localization," IEEE Global Telecommunications Conference (GLOBECOM 2010), ISSN: 1930-529X, Print ISBN: 978-1-4244-5636-92010, pp. 1-5.
- [14] W. Narzt and H. Schmitzberger, "Location-triggered code execution — dismissing displays and keypads for mobile interaction," in UAHCI '09: Proceedings of the 5th International Conference Universal Access in Human-Computer Interaction. Part II. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 374–383.

Efficient Mobility Management in 6LoWPAN Wireless Sensor Networks

Abdelwahed Berguiga, Habib Youssef

PRINCE Research Unit, ISITCom, Hammam Sousse

University of Sousse, Tunisia

abdelwahed.berguiga@infcom.rnu.tn, habib.youssef@fsm.rnu.tn

Abstract—Wireless Sensor Networks (WSNs) represent one of the most exciting today's research areas. Their utility ranges from causal sensing and data collection to mission critical applications such as battle field control, medical assistance, and natural disaster forecasting. With all the constraints related to WSNs, and with the urgent need to provide connectivity between sensor networks and the Internet, a new IETF Working Group, namely 6LoWPAN, was created. This group opens up a wealth of application opportunities as well as a number of difficult challenges. One of these challenges is the efficient support of sensor mobility. This paper proposes a mobility management scheme for IPv6 over Low power Wireless Personal Area Network (6LoWPAN) sensor nodes. We propose FPMIPv6-S (Fast handover Proxy Mobile IPv6 for Sensor networks), an improved version of the protocol PMIPv6, a protocol recently adopted as an RFC to support mobility in IPv6 based networks. We present a performance comparison of FPMIPv6-S with PMIPv6, a recently proposed modification to PMIPv6 using an analytical approach. Performance results show that FPMIPv6-S exhibits a significantly lower number of messages exchanged and handoff latency, thus extending the network lifetime.

Keywords—Wireless Sensor Networks; 6LoWPAN; FPMIPv6-S; PMIPv6; Binding update cost; Packet delivery cost.

I. INTRODUCTION

The emergence of low cost technologies in wireless communication has enabled the development of small wireless sensors. Characterized by their low cost, low power consumption and their implementation in different applications, wireless sensors have attracted attention in the academic field as well as in industry. Indeed, wireless sensors can be used in various applications such as remote control, health, and military applications. The distinguishing characteristic of sensors is their limited resources (memory and processing) and autonomous, but usually limited power supply.

Although it is a part of the domain of ad hoc networks, sensor networks and ad hoc networks differ in many ways. A WSN is composed of simple nodes. Applications require a few bytes sent periodically or on demand, generally triggered by an external event. Each node can be either a source or a destination of information. A sensor network may consist of hundreds of mobile nodes which can move from one place to another. These mobile nodes can communicate with each other or with external networks such as the Internet. So far, sensor networks use non-IP protocols such as the ZigBee 802.15.4 protocol stack [1]. Besides the fact that the ZigBee protocol is incompatible with the Internet Protocol, it introduces several constraints such as resource usage, energy consumptions, limited bandwidth, etc.

With all the constraints related to WSNs, and with the urgent need to provide connectivity between sensor networks and the Internet, a new IETF Working Group, namely 6LoWPAN [2], was created. The 6LoWPAN working group proposed two RFCs: RFC 4919 [2], which provides an overview, assumptions, problems, and goals for improving IPv6 over low power wireless PAN; and RFC 4944 [3], which specifies the transmission of IPv6 packets over IEEE 802.15.4 protocols. The integration of IPv6 in LoWPAN is expected to facilitate the introduction of new applications such as information sharing with other networks as well as the ability to provide cell-based mobility so that a session can be maintained while the device moves.

Indeed, providing IP connectivity to mobile 6LoWPAN devices may suggest that the devices can use traditional mobility related IP protocols such as Mobile IPv6 (MIPv6) [4], Hierarchical Mobile IPv6 (HMIPv6) [5], and Proxy Mobile IPv6 (PMIPv6) [6]. However, mobility using traditional mobility related IP protocols poses several challenges. First, the signaling overhead triggered by a sensor move can deplete sensor battery. Second, mobility always results in energy consumption due to the messages exchanged to synchronize with the new channel and also degrades the performance of the network due to packet loss. In addition, the connection discontinuity during handover can increase the delay and packet loss. Further, 6LoWPAN nodes are not equipped with a mobility protocol. Special proxy agents such as 6LoWPAN gateways are responsible for maintaining connectivity between 6LoWPAN mobile node and the Internet. The 6LoWPAN gateway is responsible for managing mobility-related functions on behalf of the 6LoWPAN sensor devices.

The objective of this work is to introduce a new IP-based related mobility protocol named Fast handover PMIPv6 for Sensor networks (FPMIPv6-S). We analyze and compare the binding update cost and packet delivery cost of our proposed scheme with PMIPv6. This performance evaluation shows the most suited protocol to manage mobility in a wireless sensor network.

The remainder of the paper is organized as follows. Section II reviews background data related to IEEE 802.15.4 and mobility in WSNs. Section III presents an overview of Fast handover for PMIPv6. In Section IV, the proposed mobility scheme is described and discussed. Section V evaluates the location update cost and packet delivery cost using analytical model. Experimental results are given in Section VI. Finally, conclusions and future work are given in Section VII.

II. RELATED WORKS

Many proposals have been reported for wireless sensor networks to increase the lifetime of mobile sensors and thus improve their performance. The idea presented by David Kiyoshi Goldenberg et al. [7] is to have the nodes move to predetermined positions that minimize energy consumption to transmit the sensed data toward a static sink. Howard et al. [8] proposed an algorithm to be implemented on sensor guided by mobile robots. The main objective of this algorithm is to maximize the coverage area in order to increase the line-of-sight between robots. Rajesh Rao et al. [9] proposed distributed algorithms for mobile sensors. In this work, the proposed algorithms enable the sensor to move to new areas to optimize the transmitting power needed to send collected data toward the static sink. The new position of sensors is determined by distributed simulated annealing algorithms.

Chatzigiannakis and Nikolettseas [10] explore the possibility of using a small number of mobile coordinators for efficient communication between any pair of nodes in the network. These nodes act as mobile relays and they carry packets from source to destination. Indeed, the packets are exchanged when the source node and the relay are in the radio vicinity of each other. Then, the mobile relay forwards packets to the sink. With this approach, the energy required to transmit a packet from a node to the sink is reduced significantly. However, the problem of energy has been replaced by the delay because the sensor has to wait for the mobile relay to pass nearby. Also, this solution is feasible only in the so-called delay tolerant networks.

Most of the above proposed approaches do not take into account the connectivity between IEEE 802.15.4 based-network and existing IP-based infrastructure. Further, the majority of the proposed solutions provide mobility within the same IEEE 802.15.4 network. However, with the introduction of the *Internet of Things* (IoT), a wireless sensor network can be composed of several sub-networks (or clusters) that are interconnected via an existing transport network, e.g., the Internet. Communication across subnets is likely to be multi-hops.

Interconnecting a Wireless Personal Area Network (WPAN) such as a WSN to the Internet presents several challenging tasks. To overcome this problem, a 6LoWPAN protocol has been proposed to insure internetworking between ZigBee/802.15.4 and TCP/IP. Given the limited packet size of 802.15.4, the 6LoWPAN protocol describes methods and assumptions for improving IPv6 over low power wireless PAN. Providing IP connectivity to mobile 6LoWPAN devices may suggest that the devices can use traditional mobility related IP protocols such as Mobile IPv6 (MIPv6), Hierarchical Mobile IPv6 (HMIPv6), and Proxy Mobile IPv6 (PMIPv6). Unfortunately, deploying HMIPv6 and MIPv6 in the sensor networks requires the modification of the protocol stack of the mobile nodes. However, with sensors characterized by limited resources in terms of processing capacity and memory, it is difficult to modify the structure of their protocol stack. The PMIPv6 protocol provides network-based mobility management. With PMIPv6, a Mobile Node (MN) is exempt from participating in any mobility-related signaling. In fact, all messages signalling are supported by the proxy mobility agents namely Local Mobility Anchor (LMA) and Mobility

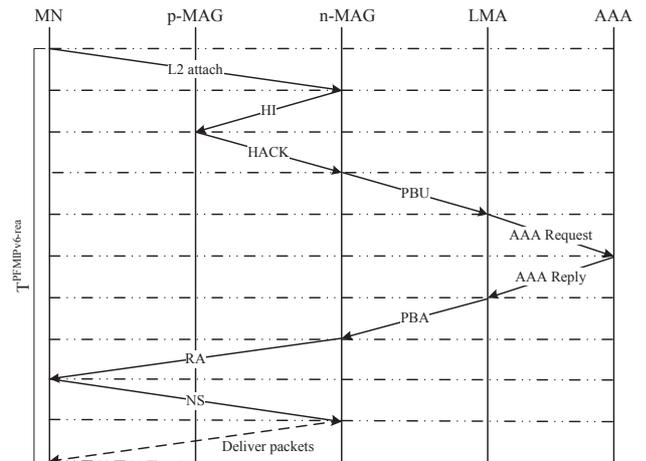


Fig. 1. PFMIPv6 handover procedure

Access Gateway (MAG). The main idea of PMIPv6 is that the MN is not involved in any IP layer mobility-related signaling. The problem with PMIPv6 is that all data traffic between the mobile and correspondent nodes is forwarded through the LMA. This presents two main drawbacks:

- 1) Besides its mobility management role, the LMA will get overburden with data forwarding task between the communicating nodes.
- 2) Packets delivery delay will get increased because of the imposed transfer through the LMA.

Some efforts have recently been reported to adapt this protocol for WSNs. Islam et al. [11] have proposed a new scheme called Sensor Proxy Mobile IPv6 (SPMIPv6). SPMIPv6 is an adaptation of the PMIPv6 standard. However, SPMIPv6 inherits the centralized approach of PMIPv6 which has some limitations such as non-optimized routing path, single point of failure, packet loss, and long handoff latency which result degradation of mobility performance. In order to reduce packet losses during the handover latency, MIPSHOP working group [12] has standardized the Fast Handover for Proxy Mobile IPv6 (PFMIPv6). However, PFMIPv6 reduces only packet loss by creating, during the handover, a bi-directional tunnel between the previous MAG that is currently serving the MN and the new MAG to which the MN is most likely to move. However, handover latency is longer than PMIPv6, because of increased signalling during PFMIPv6 handover. Also, PFMIPv6 requires that a mobile node is capable of reporting lower-layer information, as stated in [12], which may also require a modification of the mobile node.

III. OVERVIEW OF FAST HANDOVERS FOR PMIPv6

The overall PFMIPv6 signaling flow includes two modes of handover: the first mode is reactive and the second is predictive. For simplicity, the terms MAG, LMA, and MN will be used here-in-after to unify the terms among PFMIPv6. Each step of the first mode, as shown in Figure 1, is described as follows [12]:

- 1) When a MN detects that a handover is imminent, it reports its identifier (MN ID) and the previous Access

- Point Identifier (AP ID) to the new MAG (n-MAG) to which the MN is about to move.
- 2) Based on its neighbor local table, n-MAG can determine the previous MAG (p-MAG) address according to the tuples (AP-ID, MAG) of the context information. Then, the new MAG initiates the exchange of the handoff information by sending the Handover Initiate (HI) message to the previous MAG.
 - 3) The p-MAG sends a Handover Acknowledge (HACK) message back to the n-MAG. The HACK message contains MN related context such as MN-HNP, and the LMA address that is currently serving the MN.
 - 4) Once the n-MAG receives the HACK message, it establishes a bi-directional tunnel with the p-MAG and packets destined for the MN are forwarded from p-MAG to the n-MAG over this tunnel. This kind of routing may increase the end-to-end delay.
 - 5) After that, n-MAG sends a Proxy Binding Update (PBU) message to the MN's LMA, whose address is provided in HACK message, to update the MN's new location.
 - 6) The LMA sends a request message to the Authentication Authorization Accounting (AAA) server for the MN's access authentication.
 - 7) After successful authentication, the LMA will reply by sending back a Proxy Binding Acknowledgment (PBA) message. In addition, the LMA establishes a bidirectional tunnel with the n-MAG. From this time, the packets to/from the MN go through the n-MAG instead of the p-MAG. In addition, n-MAG sends a RtrAdv message to the MN to advertise MN-HNP as the hosted on-link prefix.
 - 8) Finally, after completing the address configuration procedure the MN is able to use the new address to continue sending/receiving data to/from corresponding node (CN).

However, although in PMIPv6 the MN is exempt from any participation in any mobility related signaling, additional HI/HACK messages are required to be exchanged between p-MAG and n-MAG requesting buffered packets. Nevertheless, it is not defined based on which criteria and when n-MAG has to request the buffered packets from p-MAG. Also, another problem that may arise is a neighbor queue overflow at new MAG, as well as the long handoff latency, and non-optimized communication path.

IV. PROPOSED SCHEME (FPMIPv6-S)

A. FPMIPv6-S protocol architecture

In this section, we present our proposed scheme: Fast handover Proxy Mobile IPv6 for Sensor networks (FPMIPv6-S). FPMIPv6-S is an enhanced architecture of PMIPv6 and its main objective is to reduce the handover latency of MN while moving and changing the attachment point to the new network. The FPMIPv6-S architecture consists of the following entities:

- **Sensor Mobile Access Gateway (SMAG):** Sits on the network border and it acts like an access gateway router between MN and external network. Its main role is to detect the MN's movement and perform the

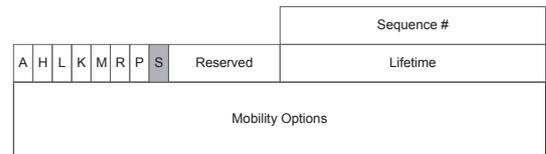


Fig. 2. FPMIPv6-S SBU Message Format

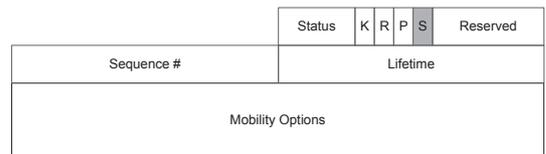


Fig. 3. FPMIPv6-S SBA Message Format

mobility-related signaling with SLMA on behalf of the MN. Packets sent/received to/from the MN are routed through a tunnel created between the SMAG that is currently serving the 6LoWPAN MN and the SMAG whose the corresponding node is attached.

- **Sensor Local Mobility Anchor (SLMA):** Acts like a Home Agent in MIP. SLMA is responsible of maintaining reachability with MN while it moves within a PMIPv6 domain.
- **6LoWPAN Mobile Node (MN):** A device that changes its point of attachment from one network to another. The MN may change its location without changing its home address.

B. Message format for FPMIPv6-S

This section gives the message formats exchanged to perform the binding, and communication processes in the sensor PMIP domain. These messages include Sensor Binding Update (SBU), Sensor Binding Acknowledgement (SBA), Localized Routing Initiation (LRI), and Localized Routing Acknowledgement (LRA).

For binding query operation in FPMIPv6-S, we define the two new messages, SBU and SBA, by adding the 'S' flag bit into the existing PBU and PBA messages of PMIPv6, respectively, as shown in Figures 2 and 3. The SBU and SBA messages are exchanged between SMAGs and SLMA to update the current location of the 6LoWPAN MN. The definition and description of the other flags are beyond the scope of this study and are described in [6].

To perform the communication process, SLMA and SMAGs exchange localized routing (LR) messages to request local forwarding for a pair 6LoWPAN MN-CN which locates the MAG who's the CN is attached for delivery data packets. The key idea of the LRI and LRA messages is to introduce the route optimization process and reduce end-to-end delay [13]. Hence, all data packets can travel between 6LoWPAN MN and CN through a tunnel that is created between SMAGs without being intercepted by the SLMA. Therefore, this allows better routing of data packets between the 6LoWPAN MN and CN which results in the reduction of network load and end-to-end delivery delay.

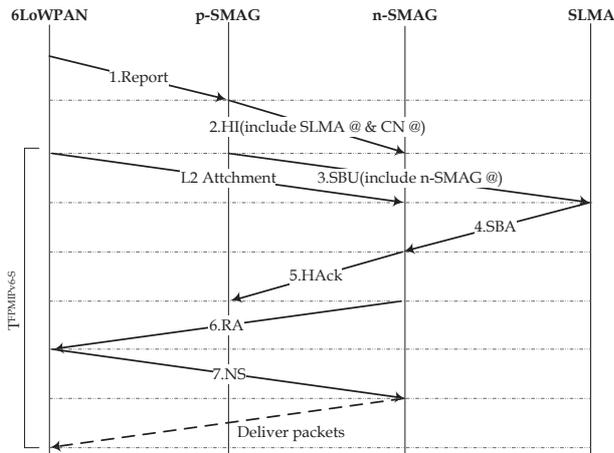


Fig. 4. FPMIPv6-S handover procedure

The procedure of sending data packets from 6LoWPAN MN to CN can be described as follows. First, given that the communicating MN and CN are anchored to the same SLMA, this latter initiates LR by sending two separate LRI messages to the two SMAGs. Each LRI message contains the IP address of the counterpart SMAG. When the SMAGs receive the LRI, each SMAG creates a local forwarding entry and a bi-directional tunnel is established between the two SMAGs such that all data packets, for which the destination is CN, are sent from the 6LoWPAN MN over this tunnel without being intercepted by the SLMA.

C. Message Flow in FPMIPv6-S

The steps of the sequence diagram of the message flow are shown in Figure 4. There are seven steps:

- **Step 1 and 2:** When a MN detects that it is entering a new sensor network, it sends a Report message to the previous SMAG (p-SMAG) that is currently serving the MN. The Report message contains the MN Identifier (MN ID) and the New Access Point Identifier (New AP ID). The p-SMAG sends a Handover Initiate (HI) message to the new SMAG (n-SMAG). The HI message MUST include both SLMA and CN addresses.
- **Step 3:** The p-SMAG sends the SBU message to the SLMA on behalf of the n-SMAG. The SBU message contains the default information like the PBU message on PFMIPv6 plus the n-SMAG address. From this moment, n-SMAG will wait for a SBA message from SLMA.
- **Step 4:** In response to the SBU, SLMA which includes both the authentication (AAA) and network information sends back a Sensor Binding Acknowledge (SBA) message.
- **Step 5:** Once n-SMAG receives SBA, it replies by sending HAcK to p-SMAG, configures the required routing information needed to reach 6LoWPAN MN,

TABLE I. Performance evaluation related parameters

Parameter	Description
T_{x-y}	Transmission cost of a packet between nodes x and y
P_C	Processing cost of node C for binding update or lookup
N_{MAG}	Number of MAGs in the PMIP domain
$N_{MN/MAG}$	Number of active MNs per MAG
C_{x-y}	Hop count between nodes x and y
S_{ctrl}	Size of control packet (byte)
S_{data}	Size of data packet (byte)
a	Unit cost of binding update with LMA
b	Unit cost of lookup for MN at LMA
t	Unit transmission cost of a packet per hop (wired link)
k	Unit transmission cost of a packet per hop (wireless link)

and registers the requesting 6LoWPAN MN in its BUL table.

- **Step 6:** n-SMAG sends a RtrAdv message to 6LoWPAN MN. When the 6LoWPAN MN receives this RtrAdv message, it will configure its IP address using either a stateful or stateless address configuration.
- **Step 7:** Once the MN performs address configuration, it sends a Neighbor Solicitation (NS) message to n-SMAG, and accordingly the MN is connected to the n-SMAG. From this moment, the sensor node is able to communicate with the CN through SMAGs.

V. PERFORMANCE EVALUATIONS

In this section, we analyze the performance of PFMIPv6 and FPMIPv6-S in terms of binding update cost (BUC) and packet delivery cost (PDC). We define the total cost (C_{total}) as the sum of binding update cost and packet delivery cost, i.e.,

$$C_{total} = BUC + PDC \quad (1)$$

Table I gives notations used in this work [14] [15] [16]. For simplicity, we consider that all costs are symmetric, i.e., $T_{x-y} = T_{y-x}$.

A. PFMIPv6 cost analysis

The binding update process of PFMIPv6 is performed as follows: When MN enters a new MAG (n-MAG) region, it performs the channel scanning which corresponds to the handover latency at L2 (T_{MAG-MN}). Then, there is an exchange of HI and HAcK messages between the new MAG and the previous MAG (p-MAG). This operation takes $2T_{MAG-MAG}$.

After that, n-MAG performs the binding update operations by exchanging the PBU and PBA control messages with LMA which takes $2T_{MAG-LMA} + P_{LMA}$. The n-MAG, on receiving the PBA message, sends back a Router Advertisement (RA) message to the MN which takes T_{MN-MAG} . After receiving RA and configuring its IP address using either a stateful or stateless address configuration, the MN sends a neighbor solicitation message to n-SMAG and performs the Layer 2 attachment which corresponds to $2T_{MN-MAG}$.

We noted that the authentication process to be done by LMA via Authentication, Authorized and Accounting server (T_{AAA}) will be simply expressed as the delay to exchange

TABLE II. Parameter values

Parameter	Default value
N_{MAG}	20
$N_{MN/MAG}$	200
$C_{MAG-LMA}$	5
$C_{MAG-MAG}$	$\sqrt{N_{MAG}}$
S_{ctrl}	50 bytes
S_{data}	1024 bytes
a	3
b	2
t	2
k	4

messages between LMA and AAA server [17]. Hence, the binding update cost can be expressed as follows:

$$\begin{aligned}
 BUC^{PFMIPv6} &= S_{ctrl}(4T_{MN-MAG} + 2T_{MAG-LMA} \\
 &\quad + 2T_{MAG-MAG} + T_{AAA}) + P_{LMA} \\
 &= S_{ctrl}(4kC_{MN-MAG} + 2tC_{MAG-LMA} \\
 &\quad + 2tC_{MAG-MAG} + T_{AAA}) \\
 &\quad + a \log(N_{MAG} \times N_{MN/MAG})
 \end{aligned} \quad (2)$$

It is assumed that the processing cost for binding update with LMA (P_{LMA}) is proportional to the total number of active MNs in the LMA domain ($N_{MAG} \times N_{MN/MAG}$) in the log scale by using a tree-based data structure to implement the database. Therefore, the processing cost at the LMA can be expressed as follows:

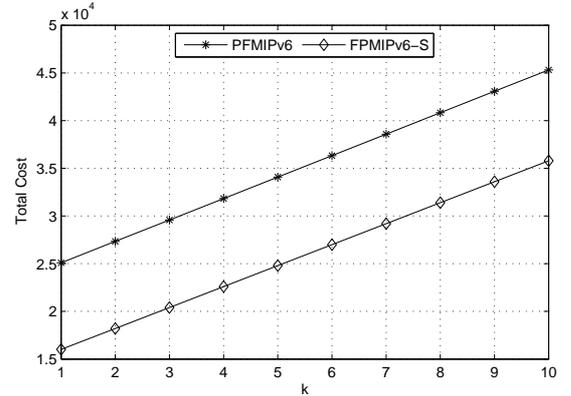
$$P_{LMA} = a \log(N_{MAG} \times N_{MN/MAG}) \quad (3)$$

In PFMIPv6, the packet delivery process consists of sending data packet from MN to CN. First, a packet is sent from the MN to the LMA via its MAG, which is equal to $T_{MN-MAG} + T_{MAG-LMA}$. Then, the LMA will look for CN address in its binding cache which requires P_{LMA} . Finally, LMA sends the packet to MAG of CN ($T_{MAG-LMA}$), and further to CN (T_{MAG-CN}). Therefore, the PDC for PFMIPv6 can be expressed as follows:

$$\begin{aligned}
 PDC^{PFMIPv6} &= S_{data}(T_{MN-MAG} + 2T_{MAG-LMA} \\
 &\quad + T_{MAG-CN}) + P_{LMA} \\
 &= S_{data}(kC_{MN-MAG} + 2tC_{MAG-LMA} \\
 &\quad + kC_{MAG-CN}) \\
 &\quad + b \log(N_{MAG} \times N_{MN/MAG})
 \end{aligned} \quad (4)$$

B. FPMIPv6-S cost analysis

As we have already mentioned in Section IV-C, in FPMIPv6-S the p-SMAG sends SBU message to the SLMA on behalf of the n-SMAG, which corresponds to $T_{SMAG-SLMA}$. On receiving the SBU, SLMA will perform the needed authentication and registration processes which take $2P_{SLMA}$. After that, SLMA replies by sending back the SBA that contains the MN's home network prefix ($T_{SLMA-SMAG}$). Once the n-SMAG gets SBU message, it will directly reply by sending HAcK to the p-SMAG and RA to the MN, which takes $T_{SMAG-SMAG} + T_{SMAG-MN}$. After receiving RA and configuring its IP address using either a stateful or stateless


 Fig. 5. Total cost versus k

address configuration, the MN sends a neighbor solicitation message to n-SMAG and performs the Layer 2 attachment which corresponds to $2T_{MN-SMAG}$.

$$\begin{aligned}
 BUC^{FPMIPv6-S} &= S_{ctrl}(3T_{MN-SMAG} + 2T_{SMAG-SLMA} \\
 &\quad + 2T_{SMAG-SMAG}) + 2P_{SLMA} \\
 &= S_{ctrl}(3kC_{MN-SMAG} \\
 &\quad + 2tC_{SMAG-SLMA} + 2tC_{SMAG-SMAG}) \\
 &\quad + 2a \log(N_{SMAG} \times N_{MN/SMAG})
 \end{aligned} \quad (5)$$

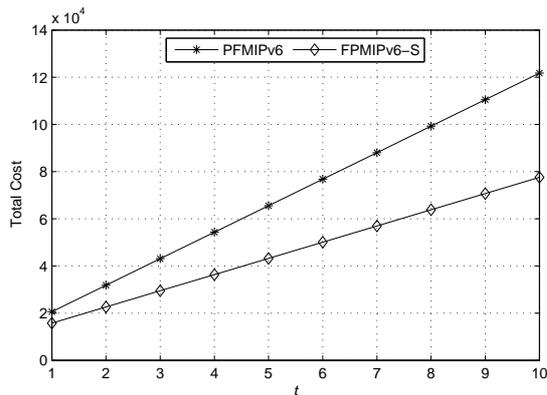
We note that, from Figure 4, the handover delay is equal to the period from the moment that 6LoWPAN MN starts L2 attachment to the moment that 6LoWPAN MN receives the first packet from n-SMAG.

In the case of data packet delivery, we analyze the communication between two MNs that belong to different MAGs in the same domain. As we mentioned earlier, when the MN is attached to the n-SMAG, SLMA exchanges localized routing messages with n-SMAG and p-SMAG to request local forwarding for a pair 6LoWPAN MN-CN. When the SMAGs receive the LRI, each MAG creates a local forwarding entry and a bi-directional tunnel is established between two SMAGs such that all data packets, in which the destination is CN, are sent from the MN over this tunnel. Accordingly, the packet delivery cost for FPMIPv6-S can be expressed as follows:

$$\begin{aligned}
 PDC^{FPMIPv6-S} &= S_{data}(T_{MN-SMAG} + T_{SMAG-SMAG} \\
 &\quad + T_{SMAG-CN}) + S_{ctrl}4T_{SMAG-SLMA} \\
 &= S_{data}(kC_{MN-SMAG} \\
 &\quad + tC_{SMAG-SMAG} + kC_{SMAG-CN}) \\
 &\quad + S_{ctrl}4tC_{SMAG-SLMA}
 \end{aligned} \quad (6)$$

VI. NUMERICAL RESULTS

In this section, we show numerical results based on the analysis derived in the previous section. To simplify our analysis, we only focus on analyzing the handover latency within a same domain. The other possible scenarios [18] for interdomain movement are not considered in the present document, and they are set as future work. The parameter

Fig. 6. Total cost versus t

values are taken from [14] [15] [16]. They are shown in Table II.

Figure 5 shows the total cost with respect to the wireless link delay. In this analysis, we have changed the parameter (k) and setting all other parameters to their default values. We see that the total cost for the two protocols increases linearly with the increment in the wireless link delay. The proposed scheme FPMIPv6-S gives better performance than PFMIPv6 scheme. This is due to the protocol properties. Indeed, as shown in Section IV-C, the handover latency is reduced since the L2 attachment is not taken into account during the exchange of Sensor Binding Update/Acknowledge messages.

Figure 6 shows the impact of wired link delay on total cost. For all of the mobility protocols, it can be observed that the total cost considerably increases as the wired link delay increases. FPMIPv6-S results in a lower total cost latency than PFMIPv6. As mentioned earlier with PFMIPv6, when a MN wants to send a data packet to a CN, all data packets get routed to the LMA through a tunnel between the MAG and the LMA. Then, LMA forwards the packet to the destination MAG. Thus, when the data packet size is increased, the cost to send it is also increased. Also, this can create the triangle routing problem. However, with FPMIPv6-S, we don't need to send a data packet to SLMA. Indeed, control messages must be exchanged between SMAGs and SLMA. As a result, a tunnel between SMAGs is established, and all packets exchanged between 6LoWPAN mobile and correspondent nodes are tunneled through this optimized routing path.

Figure 7 illustrates the variation in the total cost as the distance between the SMAGs and SLMA (MAGs and LMA) is changed. From the Figure, it is clear that the total cost increases for all the mobility protocols schemes. However, when the hop count is larger than 3 hops, PFMIPv6 has higher total cost than FPMIPv6-S. This is because sending a data packet from MN to CN must include intermediate nodes such as SLMA (LMA). However, with FPMIPv6-S, only control packets must be exchanged between SMAGs and SLMA to find the CN's location.

VII. CONCLUSION AND FUTURE WORK

In this paper, we studied the problem of mobility management for 6LoWPAN mobile sensor nodes. We presented

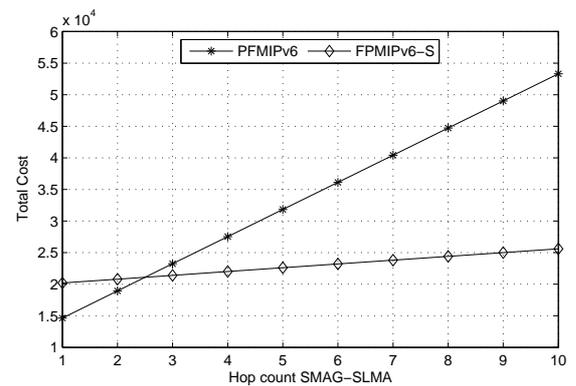


Fig. 7. Total cost versus hop count SMAG-SLMA

PFMIPv6-S, a new fast proxy-based mobility management protocol, which is an improved version of PFMIPv6. We have conducted a comparative analytical study for two mobile protocols, PFMIPv6 and FPMIPv6-S. We compared the total cost which is expressed as the sum of binding update cost and packet delivery cost. The performance analysis and the numerical results presented in this work shows that our proposal significantly outperforms PFMIPv6. We are in the process of building NS3 simulation models of PFMIPv6 and FPMIPv6-S protocols in order to validate our analytical results and to perform a more thorough evaluation.

REFERENCES

- [1] Z. Alliance, "Zigbee specification," tech. rep., <http://www.zigbee.org>, 2007.
- [2] N. Kushalnagar, G. Montenegro, and C. Schumacher, "IPv6 over Low-Power Wireless Personal Area Networks (6LoWPANs): Overview, Assumptions, Problem Statement, and Goals." RFC 4919, Aug. 2007.
- [3] G. Montenegro, N. Kushalnagar, J. Hui, and D. Culler, "Transmission of IPv6 Packets over IEEE 802.15.4 Networks." RFC 4944, Sept. 2007.
- [4] D. Johnson, C. Perkins, and J. Arkko, "Mobility Support in IPv6." RFC 3775, June 2004.
- [5] H. Soliman, C. Castelluccia, K. ElMalki, and L. Bellier, "Hierarchical Mobile IPv6 (HMIPv6) Mobility Management." RFC 5380, Oct. 2008.
- [6] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury, and B. Patil, "Proxy Mobile IPv6." RFC 5213, Aug. 2008.
- [7] D. K. Goldenberg, J. Lin, A. S. Morse, B. E. Rosen, and Y. R. Yang, "Towards mobility as a network control primitive," *MobiHoc*, 2004, pp. 163–174.
- [8] A. Howard, M. J. Matorić, and G. S. Sukhatme, "An incremental self-deployment algorithm for mobile sensor networks," *Auton. Robots*, vol. 13, Sept. 2002, pp. 113–126.
- [9] R. Rao and G. Kesidis, "Purposeful mobility for relaying and surveillance in mobile ad hoc sensor networks," *IEEE Educational Activities Department*, vol. 3, July 2004, pp. 225–232.
- [10] I. Chatzigiannakis and S. E. Nikolettseas, "An adaptive compulsory protocol for basic communication in highly changing ad hoc mobile networks," *IEEE Computer Society*, 2002, pp. 193–202.
- [11] M. M. Islam and E.-N. Huh, "Sensor proxy mobile ipv6 (spmipv6) a novel scheme for mobility supported ip-wsns," *Sensors*, vol. 11, no. 2, 2011, pp. 1865–1887.
- [12] H. Yokota, K. Chowdhury, R. Koodli, B. Patil, and F. Xia, "Fast Handovers for Proxy Mobile IPv6." RFC 5949, Sept. 2010.
- [13] S. Krishnan, R. Koodli, P. Loureiro, Q. Wu, and A. Dutta, "Localized Routing for Proxy Mobile IPv6." RFC 6705, Sept. 2012.

- [14] H. Jung, M. Gohar, J. I. Kim, and S. J. Koh, "Distributed mobility control in proxy mobile ipv6 networks," *IEICE Transactions*, vol. 94, no. 8, 2011, pp. 2216–2224.
- [15] Y.-H. Choi and T.-M. Chung, "Using correspondent information for route optimization scheme on proxy mobile ipv6," *Journal of Networks*, Vol. 5, no. 8, 2010, pp. 984–989.
- [16] A. Jabir, S. Subramaniam, Z. Ahmad, and N. A. W. Hamid, "A cluster based proxy mobile ipv6 for ip-wsns," *EURASIP J. Wireless Comm. and Networking*, vol. 2012, no. 1, 2012, p. 173.
- [17] K.-S. Kong, W. Lee, Y.-H. Han, M.-K. Shin, and H. You, "Mobility management for all-IP mobile networks: mobile IPv6 vs. proxy mobile IPv6," *IEEE Wireless Communications*, Apr. 2008, pp. 36–45.
- [18] G. Giarretta, "Interactions between Proxy Mobile IPv6 (PMIPv6) and Mobile IPv6 (MIPv6): Scenarios and Related Issues." RFC 6612, May 2012.

Mobile Ad-Hoc Search and Retrieval in the iTrust over Wi-Fi Direct Network

Isaí Michel Lombera, Louise E. Moser, Peter M. Melliar-Smith, Yung-Ting Chuang

Department of Electrical and Computer Engineering

University of California, Santa Barbara

Santa Barbara, USA

imichel@ece.ucsb.edu,moser@ece.ucsb.edu,pmms@ece.ucsb.edu,ytchuang@ece.ucsb.edu

Abstract—iTrust over Wi-Fi Direct is a peer-to-peer information publication, search and retrieval system for mobile ad-hoc networks. In this paper, we present the fundamental design of iTrust over Wi-Fi Direct. Next, we describe the iTrust over Wi-Fi Direct API and components as implemented on the Android platform for mobile devices, and show how user applications can easily interface with the API to gain P2P functionality. Then, we present the iTrust over Wi-Fi Direct networking model, and the interactions between the Android and Linux stacks. iTrust over Wi-Fi Direct addresses the need for sharing of personal information and simultaneously prevents a third party from censoring information and preventing dissemination of information. It enables users with Wi-Fi Direct enabled mobile devices to publish, search for, and retrieve information.

Keywords-Android; publication, search and retrieval; mobile ad-hoc network; peer-to-peer network; Wi-Fi Direct

I. INTRODUCTION

In our increasingly wireless and mobile world, the typical mobile phone user has come to expect ubiquitous access to public and private information. The traditional information search engines, such as those of Google, Yahoo! and Bing, have transitioned from mouse pointer desktop computers to touch screen mobile devices; thus, searching for public information is relatively easy, and continues to improve. To a certain extent, private information can also be easily indexed and searched; e.g., searching for pictures taken from a camera phone, on that *same* phone, is relatively straightforward even if the graphical user interface is difficult to use. However, in the same way that public Web sites enrich the user's search experience (i.e., increase the number and quality of relevant search hits), private stores of information from which to search also enhance the user's search experience. The gap between searching numerous public Web sites and searching only an individual device has, until recently, not been addressed.

The recent resurgence of personal search includes examples such as Facebook using facial recognition to tag friends in uploaded pictures, or Google+ using *circles* to suggest products/advertisements to the user. In both cases, the approach of those companies has been to have all users or participants upload personal information to a central

information repository and then have the users access the centralized repository to search for personal information of their friends (or circles in the case of Google+). As long as each user has no objection to submitting personal information to a third party, the centralized search approach works; the individual users (both first and second parties) benefit by having a third party perform the search functions for them, and the third party benefits by extracting and selling the personal information of the other two parties (e.g., advertising and data mining).

The centralized search approach does not work well when the third party has no incentive to enable the sharing of information, or even worse when the third party has an incentive to restrict or censor the sharing of information. Consider the two cases of politics and economics. We have seen, in the recent past and in the present, political upheavals in Egypt, Tunisia and Syria, where the government has no incentive to enable the sharing of information (such as pictures, video, protest information, etc.) among its citizens. In fact, it is in the best interest of the government to censor and restrict the dissemination of non-government sanctioned information. In the economic case, consider the example of many buyers and sellers in a concentrated area, such as weekend shoppers at a shopping mall, swap meet or bazaar. It is in the best interest of each buyer to share information with other buyers by comparing the products and the prices of the sellers (perhaps through pictures or short text messages); it is in the best interest of each seller not to allow buyers to compare such information. In both cases, individuals benefit from sharing personal information, and the third party either has no incentive to enable sharing or has an incentive to restrict sharing.

To address the need for sharing of personal information and simultaneously to prevent a third party from censoring information or preventing the dissemination of information, we created iTrust over Wi-Fi Direct. The iTrust over Wi-Fi Direct system enables users with Wi-Fi Direct enabled mobile devices to publish, search for, and retrieve information among themselves. Wi-Fi Direct [20] is a relatively new wireless technology, based on IEEE 802.11, that enables devices to form a peer-to-peer (P2P) network without the

need for a third intermediary device, such as an access point. Wi-Fi P2P is the previous name of Wi-Fi Direct; we use both terms interchangeably in this paper.

In the rest of this paper, we describe the fundamental design of iTrust over Wi-Fi Direct (Section II), the application programming interface (API) and components as implemented on the Android platform (Section III), the associated networking model (Section IV), related work (Section V), and conclusions and future work (Section VI).

II. ITRUST FUNDAMENTAL DESIGN

The iTrust over Wi-Fi Direct network consists of peers that form a mobile ad-hoc network. Multiple iTrust over Wi-Fi Direct networks may exist simultaneously, and a peer may join any such network over time. Peers in the same network are said to be in the same *membership*, although the peers do not *all* need to be within range of each other.

Figure 1 illustrates how information is published, searched for, and retrieved in the iTrust over Wi-Fi Direct network. Any peer with information to share (which we call a *source*) generates metadata describing that information and distributes that metadata to a subset of the membership chosen at random (1). A peer requesting information (which we call a *requester*) distributes a query to a subset of the membership chosen at random (2). In the distribution of both the metadata and the query, a peer that receives the message may *relay* the message to yet another subset of the membership chosen at random. Message flooding is prevented, but is not addressed here. When a peer finds a match between the metadata it holds and a query it receives, we say that an *encounter* occurs (3). The peer with the match sends a message to the requester, which identifies the source holding the desired information (4). The requester then directly fetches the information from the source (5).

In previous work, we described iTrust over HTTP [9] and iTrust over SMS [10]. We have established that the probability of a match is high even if the metadata and the queries are distributed to relatively few peers [7]. Moreover, we have developed mechanisms that prevent malicious peers from censoring information or disrupting the dissemination of information [2].

III. ITRUST OVER WI-FI DIRECT

The iTrust over Wi-Fi Direct system is implemented in Android, and is compatible with version 4.1 (and above) of the mobile platform. The choice of Android was made for a variety of reasons including previous experience and hardware compatibility. The previously implemented iTrust over SMS system enables decentralized publication, search and retrieval of information between mobile devices, as long as those devices are SMS-capable (data are transmitted over SMS). However, there are circumstances in which the centralized SMS store-and-forward model can be shut down by a third party; therefore, we developed iTrust over Wi-Fi

Direct as the natural technological progression that gives iTrust a completely decentralized and robust method of information transfer. At the time of this writing, Android is the only mobile platform that has hardware support for Wi-Fi Direct; neither Apple's iOS nor Microsoft's Windows Phone supports Wi-Fi Direct. It remains to be seen whether the new Firefox operating system (also known as B2G) will include support for Wi-Fi Direct.

Figure 2 illustrates the iTrust over Wi-Fi Direct API and components, their relationships to both user/programmer applications and the underlying Android/Linux mobile platform. We discuss in detail each of the blocks in the diagram and their relationships and interactions below. The center iTrust over Wi-Fi Direct blocks are the most pertinent, but we also discuss the two surrounding user and operating system blocks.

A. Application

The app block at the left of Figure 2 is not strictly part of iTrust over Wi-Fi Direct but is, instead, a placeholder for the user and program code that interfaces with iTrust over Wi-Fi Direct. In the previously implemented versions of iTrust, namely iTrust over SMS and to a much lesser extent iTrust over HTTP, the application had only minimal interaction with the components of iTrust, which remains true in iTrust over Wi-Fi Direct. The application has access to only the signal parser and the node core, which decouples the application and the components; doing so creates a clear separation of tasks and ensures that any application can easily add iTrust network functionality to existing modes of communication.

For example, a simple instant message text chat application was written for iTrust over SMS [10], to demonstrate the easy way in which any application can add the publication, search and retrieval functionality of iTrust. In particular, we showed how iTrust may be added to existing chat applications such as AIM, Jabber, XMPP, etc. iTrust over Wi-Fi Direct retains this same functionality; any application written for iTrust over SMS can use iTrust over Wi-Fi Direct without any additional function calls. The increased transmission rate available to Wi-Fi Direct, compared to SMS, enables a broader range of applications, including: file transfer, picture sharing, music sharing, text document collaboration, etc.

B. Signal Parser

The signal parser in iTrust over Wi-Fi Direct is similar to the signal parser in iTrust over SMS, but with several enhancements. The most extensive enhancement is the peer management protocol required to enable mobile ad-hoc functionality in the mobile device. Peer management is essential to ensuring that a moving device, or peer, can remain connected to other nearby devices and maintain the network connection required for publishing, searching for,

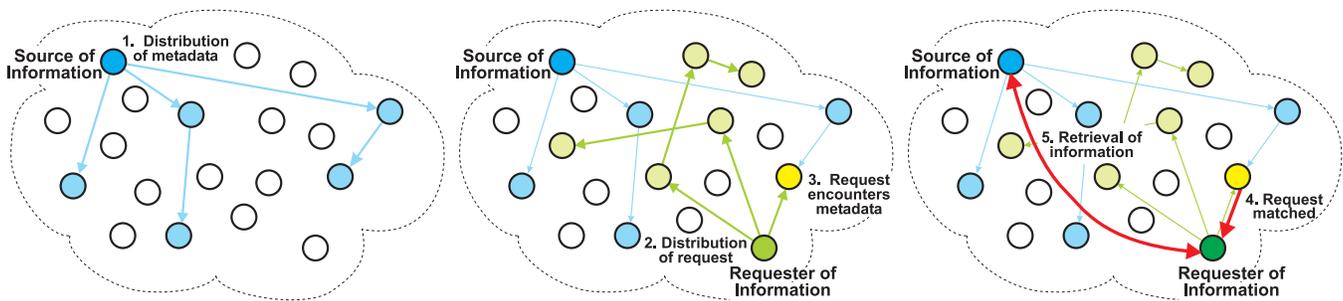


Figure 1. Information publication, search and retrieval in the iTrust over Wi-Fi Direct network.

and retrieving information. The peer management code is quite extensive and is described in detail elsewhere [11] and, thus, is not described here; however, it is important to note that the signal parser is responsible for reading the peer management message types and properly informing the node core of which actions to take regarding peer connection states. Apart from these additions, the signal parser retains the iTrust over SMS tasks for decoding incoming messages and appropriately informing the node core of which actions to take.

C. Node Core

The node core is an integral part of iTrust over Wi-Fi Direct, and handles all program accounting and system state information. It retains the functionality found in the iTrust over SMS node core, but adds a substantial amount of Wi-Fi P2P adapter information that is required for proper operation such as: self peer state, Wi-Fi P2P adapter state, and nearby peer states. For example, a peer must extract its own MAC address from the Android platform to self identify to other peers when joining the iTrust membership (this functionality is explained in Section IV). Also, the Wi-Fi P2P hardware adapter state is *separate* from the normal Wi-Fi hardware adapter state, i.e., Android makes a distinction between the normal Wi-Fi connectivity to an access point and Wi-Fi P2P connectivity directly to another peer. The two adapters must be managed separately, and the node core performs these functions, with the help of the Wi-Fi P2P service component described in Section III-E.

The node core serves the fundamental iTrust functions of: node management (not P2P related), metadata generation (keyword creation) and distribution (JSON import/export), query distribution and matching (encounters), message relaying, and message formatting (protocol finite state machine and logic control). Furthermore, the node core is the only component that interacts with the database.

D. Database Adapter

The database (DB) adapter in iTrust over Wi-Fi Direct is structurally similar to that in iTrust over SMS, and is not discussed here in detail. The singular enhancement is the enlargement of the node table (the database table that

holds information on other peers in the iTrust membership). Specifically, whereas a particular peer is assumed to be always connected over SMS (because it is a store-and-forward bearer of information), the mobile ad-hoc nature of Wi-Fi Direct does not allow this same assumption to be made. The node table has three vital additions: a name identifier to identify the peer, a Boolean field to specify whether the peer is *in range* (physically within radio distance), and a Boolean field to store the connection state of the peer. The database adapter simply stores this information, and has no logic to process the information. The node core and the Wi-Fi P2P service component act on the peer information.

E. Wi-Fi P2P Service Component

The Wi-Fi P2P service component is the centerpiece of iTrust over Wi-Fi Direct, and effectively merges the fundamental iTrust network logic with the twin goals of: sending and receiving messages, and handling Wi-Fi Direct network connections.

This component is a daemon that is separate from the application (and the other components) and that services incoming and outgoing messages through separate threads. Specifically, in Android terminology, the Wi-Fi P2P service component is a *started* Service object that is invoked with an Intent object near the beginning of application execution; once created, it remains active indefinitely (until the device is powered off). Because of the relatively aggressive memory management in Android, the Wi-Fi P2P service component may be torn down by the Android memory manager if another application requires more memory; in this case, it is automatically restarted when more memory becomes available (in practice, a delay of one to two seconds).

At start-up time, the Wi-Fi P2P service component creates an Inbox thread to listen for incoming messages (described in Section III-G) and acts as an intermediary between the Inbox thread and the signal parser (all the incoming messages must be parsed by the signal parser). When the node core needs to send a message, the Wi-Fi P2P service component creates an Outbox thread (described in Section III-H).

The handling of network connections is the other important function of the Wi-Fi P2P service component; indeed, it is responsible for starting and maintaining all Wi-Fi Direct

functionality. Again, the peer management details are outside the scope of this paper; however, it is necessary to outline the primary steps required to transmit information between peers in the network.

First, the Wi-Fi P2P service component must check the device's system settings and request permission to control the Wi-Fi P2P hardware adapter. If permission is granted, the device immediately announces itself to the network and begins searching for peers. Second, new peers are detected, and a connection is attempted if the peer is deemed available for a connection (in Android parlance, the `onPeersAvailable` interface is implemented). At this point, the node core is informed of the peer changes, and all peer database entries are updated to reflect the current in-range status. Third, if the invited peer is successfully connected, the node core is informed of the new connection details, and the connection status of the newly connected peer is stored in the database (in Android parlance, the `onConnectionInfoAvailable` interface is implemented). Normally, Android requires the user to accept each connection invitation between nodes manually by tapping a consent dialog window on the screen; the device that initiates the connection request must wait until the invited device explicitly agrees to connect. However, iTrust over Wi-Fi Direct simplifies this task by automatically accepting (in addition to automatically making) device invitations; thus, the application and the individual using the device do not have to confirm every connection request (including re-connections after accidental disconnects). Fourth, the device begins device negotiation to manage peers within the iTrust over Wi-Fi Direct membership.

Although the Wi-Fi P2P service component is responsible for the majority of all Wi-Fi Direct related functionality, it cannot by itself service all incoming requests from Android. To handle all communication, it off-loads a majority of the event handling to the Wi-Fi P2P broadcast receiver.

F. Wi-Fi P2P Broadcast Receiver

The Wi-Fi P2P service component operates in the background to service the primary Wi-Fi Direct functions; however, Android requires that a specific component listen to and monitor the system for state changes. The Wi-Fi P2P broadcast receiver is similar in function to an interrupt service handler or even a handler manager that continuously receives messages broadcast by Android.

In most cases, the Wi-Fi P2P broadcast receiver simply passes on the messages to the Wi-Fi P2P service component. The Wi-Fi P2P broadcast receiver listens for four main actions emitted by Android: the state change, the peer change, the connection change, and the device change actions.

A state change action is simply a Wi-Fi P2P hardware adapter power settings status; listening to this state enables the Wi-Fi P2P broadcast receiver to know whether the Wi-Fi P2P adapter is functioning correctly (and, indirectly,

whether the Wi-Fi P2P broadcast receiver has permission to access the device). The peer change action occurs when a peer is in the range of the device, or a peer has left the range of the device; this event is passed on to the Wi-Fi P2P service component for further processing. A connection change action occurs when the device attempts to establish a connection to a particular peer, which simply means that *something* has happened in relation to a connection attempt – there is no guarantee that the connection actually succeeded. This event is also passed on to the Wi-Fi P2P service component for further processing; e.g., a successful connection triggers device negotiation to manage the peers. Finally, a device change action signals that the current state of the device has been altered in some meaningful way. This event is primarily useful for the peer management algorithms of iTrust over Wi-Fi Direct and, thus, is outside the scope of this paper; however, it does have an important role for the other components. When a device change action is triggered, it means that the Wi-Fi P2P hardware adapter has changed state and thus can (and should) be read. Reading the state of the Wi-Fi P2P adapter at this point *guarantees* that a valid self MAC address can be extracted from Android.

G. Inbox Thread

The Inbox thread is responsible for reading all incoming messages; it is a common Java network server socket that simply listens/waits for an incoming client socket connection. When a client connection is made, the message is buffered and passed on (through the common Java *handler* object) to the Wi-Fi P2P service component. The same thread is maintained throughout the life of the Wi-Fi P2P service component; if the Wi-Fi P2P service component is killed and restarted, the Inbox thread is restarted.

H. Outbox Thread

The Outbox thread is responsible for sending all outgoing messages; it is a common Java network client socket that connects to the destination peer's Inbox thread to send the message. Unlike the Inbox thread, the Outbox thread is created to send a specific message; the Outbox thread is created on demand by the Wi-Fi P2P service component, sends its message, and then dies. Doing so conserves resources; also, the probabilistic distribution of messages in the iTrust network means that the *next* message sent probably has a different destination peer. Thus, there is little reason to keep the connection open for later immediate use.

I. Android/Linux

The Android/Linux block in Figure 2 represents the Android mobile platform, and is combined because the Linux kernel provides the functionality for the Android user space. Importantly, interaction with Linux is required to implement all parts of iTrust over Wi-Fi Direct; e.g., the Inbox and Outbox threads use Linux and not Android to setup sockets

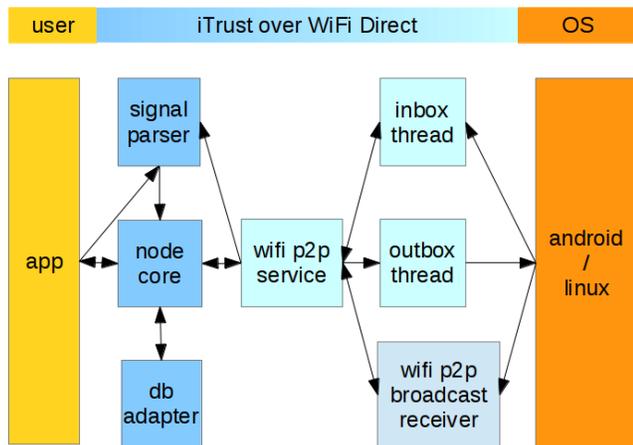


Figure 2. The iTrust over Wi-Fi Direct API and components.

and transfer information. However, unlike traditional Linux systems where the kernel dominates network functionality, Android plays an important role in controlling the Wi-Fi P2P network adapter (in addition to iTrust logic non-specific to network-related functions). We elaborate this interaction between the Android user space and the Linux kernel in creating and using the Wi-Fi Direct connection below.

IV. ITRUST OVER WI-FI DIRECT NETWORKING MODEL

To understand iTrust over Wi-Fi Direct as implemented on the Android platform more fully, it is necessary to understand not only the API and components but also the associated networking model. Figure 3 illustrates the iTrust over Wi-Fi Direct networking model by rearranging the components of Figure 2 along two axes. Horizontally, the components are separated into the portion of the Android platform to which they pertain: the Android user space on the left and the non-Android portions (mostly the Linux kernel) on the right. The Android user space and the Linux operating system are actually tightly intertwined within the Android platform but, for simplicity, we say that the Android and Linux network stacks exist in parallel. Vertically, components are arranged from top to bottom according to the layers of the Internet protocol stack: Application, Transport, Internet, and Link layers. The Physical layer is not shown.

The remainder of this section starts with some general observations, and then explains the placement of the iTrust over Wi-Fi Direct components in their respective layers, beginning with the Link layer and ending with the Application layer.

First, we observe that the iTrust over Wi-Fi Direct components are placed within the layer with which they interact (operating system or user space code). Second, most components individually interact within only a single layer, but there are several components that span multiple layers, e.g., the signal parser is entirely within the Application layer but the Wi-Fi P2P broadcast receiver spans both the Transport

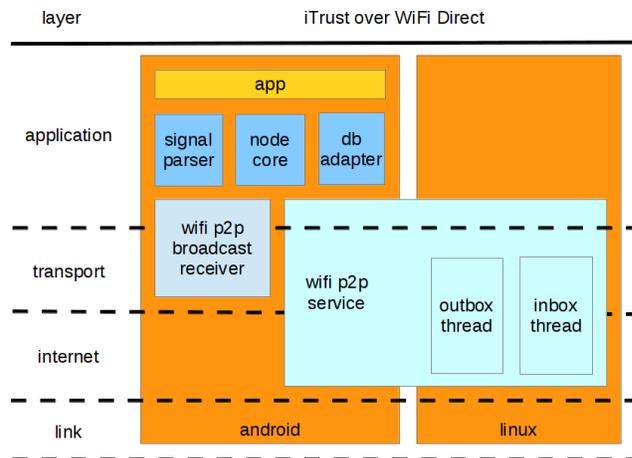


Figure 3. The iTrust over Wi-Fi Direct networking model.

and the Application layers. Third, a single component might span both network stacks, e.g., the Wi-Fi P2P service component is so complex that it spans both the Android and the Linux network stacks. Finally, although the app is shown in the Android network stack, the user/programmer may choose to use the Linux network stack for purposes not related to iTrust over Wi-Fi Direct. The iTrust over Wi-Fi components are fully contained, and do not require the user to interface directly with a network adapter (indeed, they transparently handle all networking internally); the programmer may of course add more functionality to the application.

A. Link Layer

The Link layer does not contain any iTrust over Wi-Fi Direct components. However, it is necessary to understand the interactions and differences between Android and Linux within the Link layer.

Using standard Wi-Fi with an access point, the Linux network stack plays a dominant role in providing network access at the Link layer; Android plays only a small role which, for the most part, is insignificant. Similar to the role that Linux plays on portable laptop computers, the Linux kernel provides the Link layer, whereas the user space code (GNU utilities, KDE/GNOME window managers, etc.) plays a minor role. Indeed, this organization allows any mobile device, to which the Linux kernel is ported, to gain network access easily.

However, Wi-Fi Direct is different, and Android plays a much more significant role in the Link layer. Wi-Fi Direct requires more support from the Android user space compared to standard Wi-Fi. The differences are vast enough that the entire Wi-Fi Direct access libraries are entirely within the Android Java namespaces. For example, all Wi-Fi Direct code is within the *android.net.wifi.p2p.** namespace, instead of the standard Java namespaces in *java.net.**. Consequently, the Linux kernel is not necessarily aware of the Wi-Fi P2P network adapter, and the traditional ways of accessing the

network state, hardware adapter, etc. do not work. Because no other mobile platform provides Wi-Fi Direct support, we do not know whether this organization is a specific Android platform design decision or whether the nature of the protocol does not allow for tight kernel integration.

The remaining sections have more examples of the individualistic nature of Wi-Fi Direct.

B. Internet Layer

The Wi-Fi P2P service component spans both the Android and the Linux network stacks at the Internet layer: the Android side mostly accounts for the Wi-Fi Direct functionality, whereas the Linux side is relatively minor.

Most Wi-Fi P2P service component functionality occurs within the Internet layer on the Android stack. Here, the Wi-Fi Direct hardware adapter is probed and powered on; peers are discovered; and connections are made to peers. During the connection phase, there is a pseudo P2P negotiation between peers to exchange basic network information including: MAC address, unique device name, and adapter configurations. Once a connection is made, the Wi-Fi Direct hardware adapter state changes and is internally saved within Android (specifically within an Android *Intent* object). In summary, the Wi-Fi P2P service component within the Internet layer deals with finding and connecting to peers; once a connection is made, control is passed up to the Transport layer (and eventually to the Application layer).

The Linux network stack plays a minor role in the Internet layer; instead of establishing a network, it mostly off-loads this functionality to the Android stack. However, because the Linux network stack is not involved with peer connections taking place on the Android stack, it does not have MAC address information. Although finding MAC addresses of other peers in the network is *not* part of the TCP/IP model, it is often used by a device for finding *its own* MAC address. (Note that Wi-Fi Direct identifies a peer by its MAC address.) Because the Linux stack is not aware of its own MAC address (because the Wi-Fi P2P adapter is controlled entirely by the adjacent Android stack), it cannot play any major role in network functions in higher network layers.

Trivially, the Outbox and Inbox threads have access to the Linux network stack in the Internet layer (the functionality is very similar to traditional client/server sockets). However, most of the socket functionality is in the Transport layer, as discussed in the next section.

C. Transport Layer

The Transport layer contains four components: the Wi-Fi P2P broadcast receiver, the Wi-Fi P2P service component, the Outbox thread, and the Inbox thread.

The Android stack within the Transport layer contains a portion of the two Wi-Fi P2P components. The Wi-Fi P2P broadcast receiver has the critical functionality of interfacing with the Wi-Fi P2P service component in the Transport layer.

Specifically, the Wi-Fi P2P broadcast receiver passes control from its Application layer side to its Transport layer side and passes on event states to the Wi-Fi P2P service component.

The important task of the device's reading its own MAC address is also performed in the Transport layer between the Wi-Fi P2P service component and the Wi-Fi P2P broadcast receiver. Once the Wi-Fi P2P service component establishes a connection with a peer (in the aforementioned Internet layer on the Android stack), an event is triggered within Android (device changed action); this event is caught by the Wi-Fi P2P broadcast receiver. At this point, the Wi-Fi P2P broadcast receiver parses the Android Intent object (previously set by the Wi-Fi P2P service component in the Internet layer) and successfully extracts a valid MAC address associated with the Wi-Fi Direct adapter. The device now can distinguish itself from the other peers, and can share its MAC address with other peers; iTrust over Wi-Fi Direct can then establish and automatically maintain a persistent network connection with peers. Importantly, the timing of the Intent object parsing is crucial: it *must* be done immediately after the device changed action is received and *before* any other event is received; otherwise, there is no guarantee that the MAC address exists or is readable.

As mentioned in Section IV-B, the Linux stack has no major role in Wi-Fi Direct and, for this reason, the traditional Transport layer utilities provided by Unix are not useful. Specifically, the Unix address resolution protocol (ARP) tables cannot be used and, thus, IP addresses for sending/receiving files cannot be queried from Linux. Instead, we provide this functionality in the iTrust over Wi-Fi Direct peer management system [11].

The Wi-Fi P2P service component plays a much more passive role in the Transport layer than in the Internet layer; the component simply passes messages between the Application layer portion (described in Section IV-D) and across to the Outbox/Inbox threads in the Linux stack in the Transport layer. Trivially, the Wi-Fi P2P service component creates the persistent Inbox thread and the on-demand Outbox thread. Recall that both the Inbox thread and the Outbox thread exist only within Linux and are not part of Android and, thus, exist only in the Linux stack.

For example, when the application sends a request or metadata message, the node core sends the message to the Application layer portion of the Wi-Fi P2P service component, the message is passed down from the Application layer to the Transport layer (still within the Wi-Fi P2P service component), then transfers from the Android stack to the Linux stack, and finally is passed on to the Outbox thread. When a message is received, Linux informs the Inbox thread in the Transport layer; the message is passed on to the Wi-Fi P2P service component (within the Transport layer from the Linux stack to the Android stack), passed up from the Transport layer to the Application layer within the Wi-Fi P2P service component, and finally passed on to the signal

parser.

D. Application Layer

The Application layer is relatively simple compared to the other network layers. Most of the iTrust over Wi-Fi Direct components exist in the Application layer, but are not network related. The signal parser, node core and DB adapter components have no networking functions; indeed, the Wi-Fi P2P service component and the Wi-Fi P2P broadcast receiver are specifically charged with managing all network access in iTrust. The only (minor) exceptions are that the node core sends outgoing messages to the Wi-Fi P2P service component, and the signal parser receives incoming messages from the Wi-Fi P2P service component. All of these component interactions occur in the Application layer on the Android stack.

The Wi-Fi P2P broadcast receiver, within the Application layer, interfaces directly with Android to capture events. In effect, it reads the reactions of the Wi-Fi Direct network adapter and relays the information to the Wi-Fi P2P service component.

The Wi-Fi P2P service component spans the Android and Linux stacks within the Application layer. Apart from the small but important roles of relaying messages between threads and the iTrust logic components, the Wi-Fi P2P service component handles the complex task of managing peer connections within the Application layer. The details of peer management, such as how peer IP addresses are assigned, how peers join the membership, and how connections are repaired, can be found in [11].

V. RELATED WORK

Several researchers [12], [16], [19] have provided methodologies, surveys and comparisons of distributed search for peer-to-peer networks. The structured approach requires the nodes to be organized in an overlay network based on distributed hash tables, trees, rings, etc. The unstructured approach uses randomization, and requires the nodes to find each other by exchanging messages. The iTrust system uses the unstructured approach, which is particularly appropriate for mobile ad-hoc networks.

The Commotion Wireless project [15] aims to ensure that communication cannot be controlled or cut off by authoritarian regimes, which is also one of the goals of the iTrust project. Their device-as-an-infrastructure distributed communication platform integrates Wi-Fi enabled mobile phones, computers and other personal devices to create a metro-scale network that supports local peer-to-peer communication, as well as local-to-Internet communication.

Thomas and Robble [17] have created a mobile ad-hoc network for disaster and emergency relief, using the Wi-Fi chips in Android phones, allowing them to connect directly without using cellular networks, like iTrust over Wi-Fi Direct. Their Smart Phone Ad-Hoc Networks (SPAN)

project reconfigures the onboard Wi-Fi chip of a smartphone to act as a Wi-Fi router to nearby similarly configured smartphones. SPAN intercepts communication at the Global Handset Proxy, so that typical applications, such as e-mail, Twitter, etc., still work. In contrast, our implementation of iTrust for mobile ad-hoc networks uses Wi-Fi Direct, which Android now supports.

The Serval project [3] is developing a wireless ad-hoc mobile phone platform, named Serval BatPhone. The project targets rural and remote populations, disaster and emergency relief, and governments that disable the Internet or the cellular network. The team chose to use Wi-Fi ad-hoc mode in the ISM2400 band and Android phones. At the time, Android phones did not support Wi-Fi Direct, so they had to manipulate the Wi-Fi hardware on the Android phones. Our implementation of iTrust for mobile ad-hoc networks uses Wi-Fi Direct, which Android now supports.

Meroni et al. [8] describe an opportunistic platform for Android-based devices using Wi-Fi in a mobile ad-hoc network. Their platform is intended to address concerns of scalability, flexibility and bandwidth in cellular networks by supporting local peer-to-peer communication between nodes. It enables peers to query for information and receive responses locally and, thus, to save network bandwidth, particularly when that information is large.

Motta and Pasquale [13] describe a JXTA middleware architecture for peer-to-peer networks, which exploits the features of mobile devices and optimizes mobile resources. They apply the JXTA middleware to a search infrastructure for structured peer-to-peer networks that uses resource indexing based on distributed hash tables. The iTrust over Wi-Fi Direct system for publication, search and retrieval uses an unstructured approach, which is more appropriate for mobile ad-hoc networks.

The Mobile Agent Peer-To-Peer (MAP2P) system [5] supports mobile devices in a Gnutella [4] file-sharing network using mobile agents. A mobile agent attaches itself to the peer-to-peer network, and acts as a proxy for the mobile device. The iTrust system has a lower message cost than Gnutella and, thus, a lower message cost than MAP2P.

The 7DS system [14] supports information sharing among peers in a mobile ad-hoc network. It uses multicasting of queries together with a multi-hop flooding algorithm. In contrast, the iTrust system forwards messages selectively to nodes based on a relay probability that limits the number of nodes to which the metadata and the requests are distributed to about $2\sqrt{n}$ nodes, where n is the number of nodes in the membership [7]. Moreover, iTrust does not relay metadata or requests that a node has seen previously.

The Distributed Mobile Search Service [6] broadcasts query results locally and forwards them over several hops. It is based on a distributed index, stored in a local index cache on each mobile device, that contains keywords and corresponding document identifiers. iTrust over Wi-Fi Direct

likewise maintains a distributed index, with metadata and corresponding node addresses and resource ids stored on the mobile devices. However, iTrust distributes metadata and the corresponding node addresses and resource ids first, rather than on receipt of the query results.

Tiago et al. [18] describe a system for mobile search in social networks based on the Drupal content site management system. Their system is fully distributed, is based on the network of social links formed from the mobile phone's address book, and exploits the independence of nodes. iTrust over Wi-Fi Direct is not based on the links provided by the mobile phone's address book but, rather, on the nodes within a node's neighborhood.

VI. CONCLUSION AND FUTURE WORK

The iTrust over Wi-Fi Direct system is a peer-to-peer information publication, search and retrieval system that operates over mobile ad-hoc networks. In this paper, we have described the iTrust over Wi-Fi Direct API and components as implemented on the Android platform for mobile devices. User applications can easily interface with the API to gain P2P functionality. We have presented the iTrust over Wi-Fi Direct networking model, and have described the interactions between the Android and Linux stacks. In the future, we plan to experiment with the iTrust over Wi-Fi Direct implementation and to evaluate its performance. We also plan to add an easy-to-use graphical user interface and to release an app that is useable even by computer novices.

ACKNOWLEDGMENT

This research was supported in part by U.S. National Science Foundation grant number NSF CNS 10-16193.

REFERENCES

- [1] Android Developers, Android 4.0 Platform Highlights, <http://developer.android.com/sdk/android-4.0-highlights.html> [retrieved: May 15, 2013]
- [2] Y. T. Chuang, I. Michel Lombera, P. M. Melliar-Smith, and L. E. Moser, "Protecting the iTrust information retrieval network against malicious attacks," *Journal of Computing Science and Engineering* 6(3), Sep. 2012, pp. 179–192.
- [3] P. Gardner-Stephen, "The Serval project: Practical wireless ad-hoc mobile telecommunications," http://developer.servalproject.org/site/docs/2011/Serval_Introduction.html [retrieved: May 15, 2013]
- [4] Gnutella, <http://www.gnu.org/philosophy/gnutella.html> [retrieved: May 15, 2013]
- [5] H. Hu, B. Thai, and A. Seneviratne, "Supporting mobile devices in the Gnutella file sharing network with mobile agents," *Proc. 8th IEEE Symposium on Computers and Communications*, Kemer-Antalya, Turkey, Jul. 2003, pp. 1035–1040.
- [6] C. Lindemann and O. P. Waldhorst, "A distributed search service for peer-to-peer file sharing in mobile applications," *Proc. Second International Conference on Peer-to-Peer Computing*, Linköping, Sweden, Sep. 2002, pp. 73–80.
- [7] P. M. Melliar-Smith, L. E. Moser, I. Michel Lombera, and Y. T. Chuang, "iTrust: Trustworthy information publication, search and retrieval," *Proc. 13th International Conference on Distributed Computing and Networking*, Hong Kong, China, Jan. 2012, LNCS 7129, Springer, pp. 351–366.
- [8] P. Meroni, E. Pagani, G. P. Rossi, and L. Valerio, "An opportunistic platform for Android-based mobile devices," *Proc. Second International Workshop on Mobile Opportunistic Networking*, Pisa, Italy, Feb. 2010, pp. 191–193.
- [9] I. Michel Lombera, Y. T. Chuang, P. M. Melliar-Smith, and L. E. Moser, "Trustworthy distribution and retrieval of information over HTTP and the Internet," *Proc. Third IARIA International Conference on the Evolving Internet*, Luxembourg City, Luxembourg, Jun. 2011, pp. 7–13.
- [10] I. Michel Lombera, L. E. Moser, P. M. Melliar-Smith, and Y. T. Chuang, "Decentralized search and retrieval for mobile networks using SMS," *Proc. IEEE 8th International Conference on Wireless and Mobile Computing, Networking and Communications*, Barcelona, Spain, Oct. 2012, pp. 134–141.
- [11] I. Michel Lombera, L. E. Moser, P. M. Melliar-Smith, and Y. T. Chuang, "Peer management for iTrust over Wi-Fi Direct," *Proc. International Symposium on Wireless Personal Multimedia Communications*, Atlantic City, NJ, Jun. 2013.
- [12] J. Mischke and B. Stiller, "A methodology for the design of distributed search in P2P middleware," *IEEE Network* 18(1), 2004, pp. 30–37.
- [13] R. Motta and J. Pasquale, "Wireless P2P: Problem or opportunity," *Proc. Second IARIA Conference on Advances in P2P Systems*, Florence, Italy, Oct. 2010, pp. 32–37.
- [14] M. Papadopouli and H. Schulzrinne, "Effects of power conservation, wireless coverage and cooperation on data dissemination among mobile devices," *Proc. ACM Symposium on Mobile Ad Hoc Networking and Computing*, Long Beach, CA, 2001, pp. 117–127.
- [15] P2P Foundation, <http://p2pfoundation.net/Commotion> [retrieved: May 15, 2013]
- [16] J. Risson and T. Moors, "Survey of research towards robust peer-to-peer networks: Search methods," *Tech. Rep. UNSW-EE-P2P-1-1*, Univ. New South Wales, Sep. 2007, RFC 4981, <http://tools.ietf.org/html/rfc4981> [retrieved: May 15, 2013]
- [17] J. Thomas and J. Robble, "Off grid communication with Android: Meshing the mobile world," *Proc. IEEE Conference on Technologies for Homeland Security*, Waltham, MA, Nov. 2012, pp. 401–405.
- [18] P. Tiago, N. Kotiainen, M. Vapa, H. Kokkinen, and J. K. Nurminen, "Mobile search – Social network search using mobile devices," *Proc. 5th IEEE Consumer Communications and Networking Conference*, Las Vegas, NV, Jan. 2008, pp. 1201–1205.
- [19] D. Tsoumakos and N. Roussopoulos, "A comparison of peer-to-peer search methods," *Proc. Sixth International Workshop on the Web and Databases*, San Diego, CA, Jun. 2003, pp. 61–66.
- [20] Wi-Fi Alliance, Wi-Fi Direct, <http://www.wi-fi.org/discover-and-learn/wi-fi-direct> [retrieved: May 15, 2013]

Proposal to Reduce the Computational Cost of Wireless Network Selection Process

Francirley Resendes Borges Costa and Claudio de Castro Monteiro
 Network Computer Group - GREDES
 IFTO Federal Institute of Education, Science and Technology of Tocantins
 Palmas, Brasil
 fborges@ifto.edu.br, ccm@ifto.edu.br

Abstract—The network selection process has been widely studied in recent years as a result of technological evolutions which has been presented to the user with ease of access involving issues related to mobility and ubiquity of access to data. Thus, several proposals have already been presented, always showing a lot of variables that need to be considered for a good selection. However, these solutions presents a computational cost that grows with the complexity of the technique used, the number of variables involved and the ways of collecting these variables. The most common solutions, and still used in smart-phones and tablets, are those based on the signal level measured at the interface of the mobile, due there is no incremental computing cost to devices, maintaining these solutions as feasible. Therefore, this paper proposes a new metrics based on a mapping between the variables of network QoS (delay, jitter and packet loss) and the signal level, presenting, through experimental tests, prediction equations these variables based only on signal level measured at the device interface. The validation of the metrics is presented using a simple network selection algorithm, operating in an environment with two independent and overlapped WLANs and no distribution system, comparing the results obtained with the proposed metric and WLAN-First algorithm, present in most of smart-phones and tablets today.

Keywords-WLAN; RSSI; QoS

I. INTRODUCTION

Nowadays, it is common the access to wireless networks, being those, the most diverse kinds. It is also increasing the popularization of devices, like smart-phones and laptops. Besides providing this technology with affordable prices, it offers an enormous mobility because the user can move it easily, keeping always best connected to the networks.

The access to multimedia content, like videos, songs and others, through the Internet is an example of services consumed by this increasing demand of users. These users do not want only to keep best connected, but also, to enjoy of a connection with quality of services.

To define what would be the best network for devices that use wireless networks, it is necessary several requirements analysis, what generate high computational cost. The network selection needs take in account the quality of experience of the users during a connection, without generate incremental computational cost. However, today, this is a big problem.

It is noticed that there is a challenge when we speak of network selection. Efficient Techniques to choose a better connection for these multiple devices are still are searched for researchers. This study looks for a solution with low

computational cost, but efficient that could solve this challenge. Considering that the computational cost to measure the signal level information is minimum, the study seeks, through experiments, to relate such parameter with the QoS network, proposing in this way, a metric that can be used on the mobile devices as an alternative to the algorithm WLAN-First currently used.

The article is organized the following way. In Section II, we present some theoretical references for the work understanding. In Section III, we present the proposal, determining the objectives, beside of where and how the processing will be done. In Section IV, we present the “test bed” and methodology used for the realization of the study, like equipment and statistical studies about the data collected. In Section V, we present the results of the study and the models of network QoS prediction from the variables studied. In Section VI, we show the final considerations and some future work proposed.

II. RELATED WORKS

During the bibliographic survey done for the realization of this research, we found assignments [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11] that show the signal level, measured on the mobile devices as a parameter to select networks.

Some of them use methods of artificial intelligence, by combining many networks variables (including signal level), determining which is the best network in a available group [3], [4]. Other authors use methods of decision making (MADM) to realize the same task [5], [6]. Other works present solutions for handover prediction based only in the signal level [7], [8]. Nevertheless, there is still a problem to solve in network selection scenario: How to establish what is possible to be done considering the computational cost added to the mobile device? Considering that the most efficient solutions for the networks selection has been focused on mobiles [1] and [2], the increasing of computational cost must be taken into account, in order to the applications can be feasible.

This way, even knowing that the level of the signal measured in a wireless interface of a mobile device, it not represent, in most cases, the quality of the service provided by a network [9] [10], this work was motivated by the fact that to measure the power of the signal on wireless interfaces, a device does not have increasing computational cost. This way, the study was conducted on an experimental basis, with a purpose to verify traces that could determine a relationship between the main network QoS variables (delay, jitter, and loss) and the signal intensity (RSSI) measured on the device [11].

III. PROPOSAL

Normally, modern equipments of wireless access, have the possibility to access and/or migrate between various networks at any moment. However, there is a diversity of parameters that influences the networks selection, mainly, when the focus is QoS. Due to complexity of this process, variables and context to consider, such techniques have not become yet, a pattern adopted by the industry, that continues to insert on its mobile devices, algorithms of network selection, based on methods/criteria not efficient [12].

Through statistic methods, this work seeks to predict the value of the other variables that affect the network quality. Thus, the primary objective of this work is to propose a metrics to be used in a mechanism of network selection that presents low computational cost (resulting also in energy saving). The hypothesis is based mainly on the relationship between signal level and QoS variables (delay, jitter and packet loss), that affect directly the QoS of the networks in general.

This way, a practical scenario was mounted for measuring the involved variables and a collector algorithm was proposed and validated with an implementation. This algorithm can be seen in Figure 1.

A. Sample Collector

The first stage of the work was to develop a software to make the automatized collect of the necessary variables for the studies. In this case, we develop a program in C, receiving as parameters: i) the destination IP address (Host); ii) the quantity of ICMP requests; iii) the output interface, and iv) the number of iterations that it will be realized in each point. The number of iterations indicates how many times the group of requests will be executed at the point's collect.

After the execution of the routines, a file with the following data for each iteration is generated:

- Average delay: variable that stores the mean delay of the requisitions;
- Average jitter: variable that stores the mean jitter of the requisitions;
- Quantity of packet loss: variable that stores the quantity of requisitions ICMP that did not achieve its destination;
- Average network signal: variable that stores mean level of the wireless network signal over iteration.

When the program executes the ICMP requests, it creates a new process, in order to verify and store the signal level, until the end of every the set of ICMP requests. The algorithm follows the steps illustrated in Figure 1.

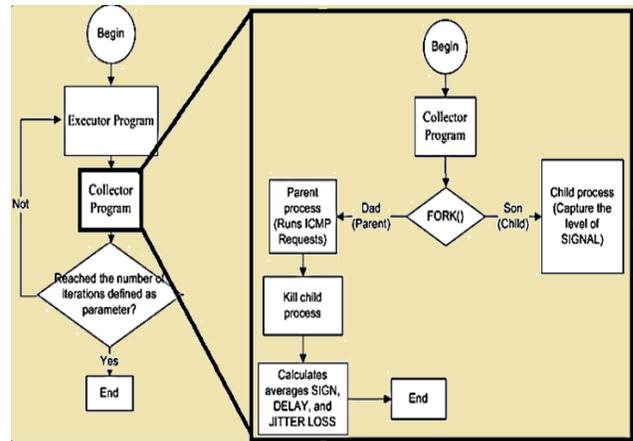


Figure 1. Collector algorithm.

In Figure 1, it is noticed that the algorithm uses two programs, one named **executor** and the other named **collector**. The first makes the call of the second according to the number of iterations defined via parameter. On the other hand, the second, divides its processing into two processes where one realizes the ICMP requests and the other stores the signal level. When the father process finishes its execution, it finishes the process son itself, calculating the averages of the collected values. The program verifies if the number of iterations was achieved. If yes, the program is finished; otherwise, the executor program calls the collector until it achieve the number of iterations established.

B. Environment Used

We define ten points, at ten meters from each other one. Every points were defined as illustrated in Figure 2. The set of points and the distance between each one of them, were defined following a straight line from the wireless network (Access Point – AP).



Figure 2. Collection Points

The coordinates of the points were collected using a GPS (Global Position System) data test. The location of each point is shown on Table I.

TABLE I. COORDINATES

POINT	COORDINATES	
0	10°11'58.95"S	48°18'48.13"W
1	10°11'58.60"S	48°18'48.14"W
2	10°11'58.28"S	48°18'48.15"W
3	10°11'57.96"S	48°18'48.15"W
4	10°11'57.62"S	48°18'48.14"W
5	10°11'57.31"S	48°18'48.15"W
6	10°11'56.99"S	48°18'48.15"W
7	10°11'56.65"S	48°18'48.15"W
8	10°11'56.30"S	48°18'48.15"W
9	10°11'55.95"S	48°18'48.15"W

We try to avoid obstacles like walls, uphill, downhill, among others. The point 0 (zero), showed in Figure 2, is where the AP is located, but also, it is the first point's collect.

IV. METHODOLOGY

The equipments used were two notebooks model ASUS EEE PC 1001PX, considering that one had the Access Point function and the other had the client station function. The wireless devices on the notebooks are ATHEROS AR9285 model. Every the tests were realized, exclusively, with these equipments. The operating system used was LINUX UBUNTU version 11.04. The wireless network model used was the Ethernet 802.11g.

Then, with the collector program, five iterations of six thousand requests were performed at three point's collect, constituting the pre-sample of data. These parameters were chosen randomly, using the statistical calculations to define the ideal set of the collect.

The studies were realized to define how many iterations and how many requests per iteration should be done at each point's collect. Through the normal distribution, we can calculate the size of set of tests.

First, it was necessary to define the number of ICMP requisitions for each iteration. For this, through the pre-sample of data, cited previously, and with only the data of one iteration of each one of the three points, we obtain the following results, shown in the Tables II and III:

TABLE II. STANDARD DEVIATIONS AND QUANTITY OF LOSS OF THE COLLECTION POINTS

Variable	Point 1	Point 2	Point 3
Packet loss (dgram)	74	102	91
Delay (ms)	16,61	11,56	19,8
Signal level (db)	1,94	1,49	1,11
Jitter (ms)	23,11	15,55	27,51

a. dgram = Datagrams;
 b. ms = Miliseconds;
 c. db = Decibels.

TABLE III. IDEAL QUANTITIES PER VARIABLE REQUISITIONS

Variable	Point 1	Point 2	Point 3
Packet loss (dgram)	384	784	476
Delay (ms)	5270	2806	4568
Signal level (db)	3136	2401	1739
Jitter (ms)	6147	5270	9604

In Table II, there are the results from the standard deviations of the points 1, 2 and 3. In Table III, the quantities of requests necessary for each iteration are presented, by considering the observation of each variable analyzed. The biggest value found was of the jitter analysis (9604); therefore, the value adapted for the number of the test was rounded to 10,000 requests. With these data it was possible to estimate the Confidence Intervals (C.I.).

In previous studies, we notice that even collecting the data at the same point, some factors can change the values of the variables studied, therefore, it was necessary to estimate how many iterations should be done at each point. To define the number of ideal iterations, we used the same method previously described. The results are exposed in Table IV.

TABLE IV. IDEAL QUANTITIES OF ITERATIONS PER VARIABLE.

Variable	Averages (AVG)	Standard Deviation	Error	Ideal Sample
Delay	3,98477	0,826	0,194	69,575
Signal	62,07 %	0,998	0,235	69,347
Jitter	4,13384	1,040	0,250	66,515
Packet loss	130,2	122,328	29	68,355

In this case, we use the averages of the five iterations, instead of one. Another information that also must be quoted, is the error considered. The value of the error was defined through diverse considerations to enable the feasibility of the study. Some example of this is: i) the time for collecting; ii) climate factors and iii) the capacity of the equipment's battery.

With the data analysis, we can observe the values in Table IV, are around to seventy (70). Therefore, we use this value as the ideal number of necessary iterations at each collection point. After the collect, we calculate the confidence interval, with a significance level of 95%. These values may be visualized in Table V. The delay information and jitter are presented in milliseconds (ms), while the signal, in decibels (db) and loss, in datagrams (dgram).

TABLE V. CONFIDENCE INTERVALS

P[i]	Variable	Standard Deviation	avg	C.I. α=95%	Min Margin	Max Margin
0	Signal (db)	0,01520852	69,99	0,004	69,992	69,999
	Delay (ms)	6,18539727	5,989	1,449	4,540	7,438
	Jitter (ms)	2,62774534	2,828	0,616	2,212	3,444
	Loss (dgram)	7,76973211	5,471	1,820	3,651	7,292
1	Signal (db)	1,51335227	42,94	0,355	42,586	43,295
	Delay (ms)	5,45189219	4,790	1,277	3,513	6,067
	Jitter (ms)	2,09261122	1,769	0,490	1,279	2,259
	Loss (dgram)	67,6242572	59,05	15,842	43,215	74,899
2	Signal (db)	1,91760654	33,40	0,449	32,952	33,850
	Delay (ms)	60,1989292	37,96	14,102	23,866	52,071
	Jitter (ms)	5,82958986	7,275	1,366	5,909	8,641
	Loss (dgram)	69,2890859	436,5	16,232	420,29	452,76
3	Signal (db)	0,85534434	39,41	0,200	39,213	39,614
	Delay (ms)	4,39566103	3,565	1,030	2,535	4,594
	Jitter (ms)	1,58620896	1,197	0,372	0,825	1,568
	Loss (dgram)	63,0608566	73,22	14,773	58,456	88,001
4	Signal (db)	1,22242838	37,44	0,286	37,156	37,729
	Delay (ms)	208,923603	131,2	48,943	82,308	180,19
	Jitter (ms)	21,9951686	18,47	5,153	13,326	23,631
	Loss (dgram)	81,6578087	200,7	19,129	181,59	219,85

5	Signal (db)	1,09617497	42,84	0,257	42,587	43,101
	Delay (ms)	32,0866230	19,27	7,517	11,759	26,793
	Jitter (ms)	4,88242644	6,525	1,144	5,381	7,669
	Loss (dgram)	47,5931711	98,18	11,149	87,037	109,33
6	Signal (db)	0,93811564	30,20	0,220	29,983	30,422
	Delay (ms)	2405,75533	682,7	563,57	119,16	1246,3
	Jitter (ms)	25,8061138	21,77	6,045	15,724	27,815
	Loss (dgram)	647,759259	469,9	151,74	318,22	621,71
7	Signal (db)	1,11200051	39,01	0,260	38,749	39,270
	Delay (ms)	33,9789006	16,58	7,960	8,628	24,548
	Jitter (ms)	5,42724717	4,583	1,271	3,312	5,855
	Loss (dgram)	79,0995925	119,5	18,530	100,99	138,05
8	Signal (db)	1,40141661	31,61	0,328	31,287	31,944
	Delay (ms)	144,401208	58,39	33,827	24,565	92,220
	Jitter (ms)	12,8705671	9,183	3,015	6,168	12,198
	Loss (dgram)	241,156497	323,9	56,493	267,42	380,4
9	Signal (db)	0,46778293	21,27	0,110	21,161	21,380
	Delay (ms)	7983,11352	7736	1870,1	5866,7	9607
	Jitter (ms)	85,1955070	136,2	19,95	116,30	156,21
	Loss (dgram)	2392,03947	3477	560,3	2916,7	4037,4

P[i] = Pont in Position i, from 0 to 9.

V. RESULTS

With the ideal collect finished, the next step was to seek the QoS prediction models, through of relationship between the variables studied. For this, we use techniques of linear regression (multiple and simple); therefore the prediction model was determined like illustrated on the equation 1.

$$Y = b_0 + b_1.x_1 + b_2.x_2 + b_3.x_3 \tag{1}$$

where:

- Y = Dependent variable;
- b₀ = Linear adjustment coefficient;
- b_{1,2,3} = Coefficients of the variables 1, 2 e 3.
- x_{1,2,3} = Average values of the independent variables 1, 2 e 3.

In this work, we study, always, four variables: signal level, delay, jitter and packet loss, with one of them is considered as dependent variable and the other three are considered as independent variable. In this context, four models of prediction were found, but only three of them have shown coefficient of determination to validate the prediction. These models are shown in Table VI.

TABLE VI. MODELS OF PREDICTION.

Depend variable	Prediction model equation	(R ²) coefficients of determination
Delay	$Y = 320,5642 - 26,928925x_s + 47,1463341x_j + 0,0127093x_p$	0,8822654429
Jitter	$Y = 23,3186 + 0,0144121124x_a - 0,381432x_s + 0,002428774x_p$	0,9843574233
Loss	$Y = 584,8607 + 0,3679054x_a - 1,12599x_j - 9,5240063x_s$	0,9739415504

a. X_a = Average delay;
 b. X_j = Average jitter;
 c. X_p = Average loss;
 d. X_s = Average signal.

R² indicates how much of the variance of Y can be represented by the variance of the independent variables. The values of R² are in the interval between 0 to 1, where 1 indicates that model is more accurate. For example, in Table VI,

the variance of the depend variable delay presents values around 88.23%, considering the group of independent variables: signal, jitter and packet loss.

In real environments, these variables cannot exist separately. It is not possible to observe jitter without observing delay, besides of the loss of datagrams being inherent to network environments. Beside the multiple regression model is the most efficient to predict values, analyzing more than one variable, it is not indicated in this case. In this sense, due the need to analyze the impacts caused by each one of the QoS variables, we use for analysis, the models of simple regression, considering only one variable at each time, in order to verify the relationship with the signal level variable.

A. QoS Models of Prediction with Simple Regressions.

The models of simple regressions relate only two variables, therefore, with the value of one of them, it is possible to predict the value of other one. This models are used, in this work, to obtain a model of prediction.

We use the collected data in some simple regression models and we notice that both exponential and potential models has been goods correlation coefficients. The Figures 3, 4 and 5 illustrate a graphic analysis of the data, showing relationship between the variables: loss, delay and jitter to the signal level. Tables VII and VIII present the coefficients of determination found.

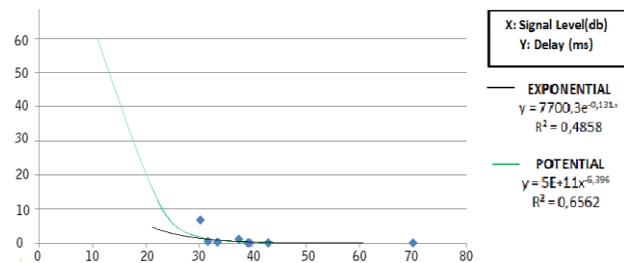


Figure 3. Relation signal level and delay.

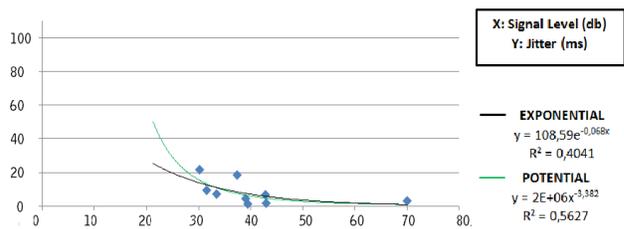


Figure 4. Relation signal level and jitter.

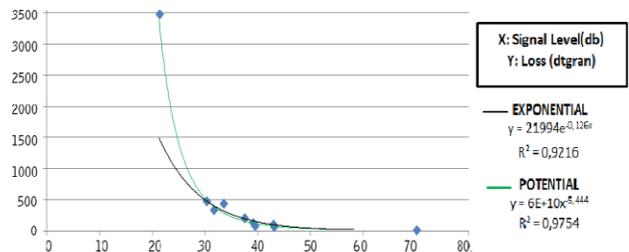


Figure 5. Relation signal level and loss.

TABLE VII. MODELS OF PREDICTION WITH EXPONENTIAL REGRESSION.

Depend variables	Prediction model equation	(R ²) coefficients of determination
Delay	$y = 7700,3e^{-0,131x}$	0,4858
Jitter	$y = 108,59e^{-0,068x}$	0,4041
Packet loss	$y = 21994e^{-0,126x}$	0,9216

TABLE VIII. PREDICTION MODELS WITH POTENTIAL REGRESSION.

Depend variables	Prediction model equation	(R ²) coefficients of determination
Delay	$y = 5e+11x^{-6,396}$	0,6562
Jitter	$y = 2e+6x^{-3,382}$	0,5627
Packet loss	$y = 6e+10x^{-5,444}$	0,9754

We notice, through Figure 5 and Tables VII and VIII, that the relationship between signal level and packet loss is the most explained by the regression models. In the exponential and potential models, the determination coefficients (R²) show more than 92.16% and 97.54% respectively, when predicting the values of the variable Y (loss). In other cases, Figures 3 and 4 and also through Tables VII and VIII, for both delay and jitter, the better results are obtained for the potential regression, respectively. The coefficients of determination achieve around 65.62% and 56.27%.

The purpose of the study was achieved in potential model, because only with the value of the variable signal level was has been possible to predict the other 3 variables, independently.

B. Validation of the proposed models

To realize the validation of the models found for the prediction of each QoS variable, in function of the measured signal level, it was mounted an environment with two overlay WLANs, without distribution system. In a mobile device, model SAMSUNG P6200, it was implemented a simple algorithm based on the models presented in this work.

We use the described environment to transmit a video sequence known as highway, containing 2000 frames in the QCIF format, using a video server based on the VLC software. Beside this, we consider the threshold of 20 packets loss to be used by algorithm to choice between the networks.

By considering always a level of trust of 95%, we realize 10 groups of tests, each one with 100 transmissions of this video sequence, from the server to the mobile. During the transmissions, we capture, using ICMP requests, from the mobile to the gateway of the two WLAN networks available, the values of the datagrams loss. During every transmissions, the mobile was submitted to a movement that allowed a large variations at the signal level measured on each network, aiming to test of the model found. The results can be seen in Figures 6 and 7.

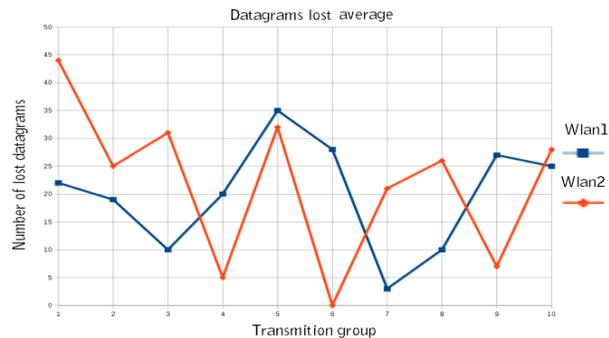


Figure 6. Average Datagram Loss Measured by the Mobile on WLAN1 and WLAN2.

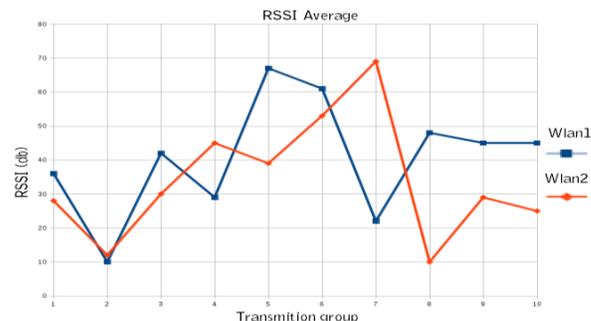


Figure 7. RSSI measured on the Mobile of WLAN1 and WLAN2

As we can see in Figures 6 and 7, the algorithm of network selection, proposed based on the presented model, did the change of network whenever the datagrams loss was above the threshold, providing the user with a better experience in the reception of services on its own devices.

An example can be seen at the point referring to the sixth group of transmissions, where the average value of RSSI for WLAN1 was 60db and the loss for the same network was of 28 datagrams. Meanwhile, for WLAN2, RSSI was lower, 53db and the loss of datagrams was null. In this case, while WLAN-First didn't show sensibility, indicating to the mobile that it should remain in the same network (WLAN1), the algorithm based in the our proposal, had recommended to the mobile, the change of network to WLAN2. Even this network having a smaller RSSI, the proposed model for datagrams loss showed to be efficient, suggesting the change to a network with less loss. Meanwhile, using the traditional algorithm, native of the device (WLAN-First), the mobile remained in its origin network from the beginning until the end of the video sequence transmissions, not taking into consideration the losses of the network.

Similar results were found using the prediction models proposed for delay and jitter, characterizing that the proposed models can be used together or separated, depending of the traffic type considered by the network selection.

We notice that the results of this work do not increase the computational cost to the device, demonstrated by the battery consumption, if compared to the WLAN-First algorithm, as in Figure 8.

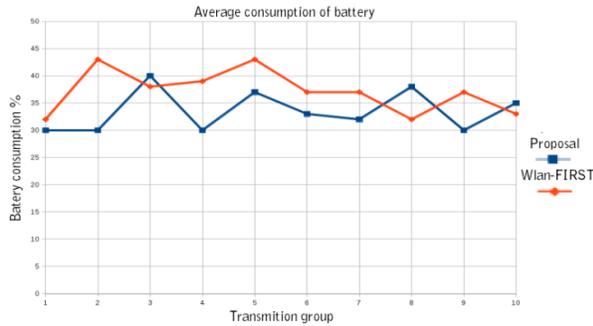


Figure 8. Percentage of consumed battery

In Figure 8, it is possible to observe that the battery consumption of our proposal did not change, if compared to WLAN-First algorithm. Considering that WLAN-First adopts the strategy of keeping the mobile always connected to a WLAN, without considering the QoS network, we can affirm that our proposal achieved better results, basically with the same battery consumption.

VI. CONCLUSION AND FUTURE WORKS

As we presented before, there are no efficient computational network selection techniques available. This study presents satisfactory results, showing that there is a relationship between the variables studied. Table VI shows satisfactory results, reaching coefficients of determination in the interval between 0 and 1.

With simple regressions, it was possible to determine R2 to each one QoS variables, noticing, that, in 97.54% of the cases, it is possible to predict the loss of datagrams on a network, through of the signal level measured. Even without having an expressive relationship between other QoS variables, delay and jitter present coefficients, in order to show that there is a relationship with the signal level.

We can notice, then, that it is possible to achieve success in the development a technique of network selection with low computational cost and efficient. Even that any data has presented values with a low relationship, it is possible to determine some variables, that affect directly the QoS of a network observing the signal level only.

However, the proposed environment shows only a portion of the possible environments, some aspects need to be better investigated and yet improved, facts that is being studied.

A study mixing both simple and multiple regressions can result in more accurate values. Besides this, the proposed model here can be expanded and generalized to other test beds, using mathematical tools more elaborated that can verify other non-linear relationship between the variables.

REFERENCES

[1] C. C. Monteiro, "Um ambiente para apoio à integração de redes sem fio heterogêneas", Doctoral Thesis. Submitted to the Department of Electrical Engineering. University of Brasília, Brasília, June 2012.

[2] V. M. Rios, "Seleção de redes sem fio baseada em técnicas de apoio à decisão.", Master's thesis submitted to the Department of Electrical Engineering. University of Brasília, Brasília, June 2012.

[3] R. B. Ali and P. Samuel, "On the Impact of Soft Vertical Handoff on Optimal Voice Admission Control in PCF-Based WLANs Loosely Coupled to 3G Networks", IEEE Transactions on Wireless Communications 8, pp. 1356–1365, March 2009.

[4] S. Das, "et al", "IEEE 802.21: Media Independent Handover: Features, Applicability, and Realization", IEEE Communications Magazine 47 (1), pp. 112–120, January 2009.

[5] A. De La Oliva, "et al", "An Overview of IEEE 802.21: Media-Independent Handover Services", Wireless Communications, IEEE 15 (4), pp. 96–103, August 2008.

[6] M. Kim, T. Moon, and S. Cho, "A study on IEEE 802.21 MIH Frameworks in Heterogeneous Wireless Networks", 11th International Conference on Advanced Communication Technology-Volume 1, IEEE, pp. 242–246, South Korea, February 2009.

[7] C. Lim, D. Kim, O. Song and C. Choi, "Share: Seamless Handover Architecture for 3G-WLAN Roaming Environment", Wireless Networks 15, pp. 353–363, April 2009.

[8] P. Machan, S. Serwin, J. Wozniak, "Performance of Mobility Support Mechanisms in a Heterogeneous UMTS and IEEE 802.11 Network Offered Under the IEEE 802.21 Standard", 1st International Conference on Information Technology, IEEE, pp. 1–4, Nice, May 2008.

[9] E. Stevens-Navarro and V. W. S. Wong, "Resource Sharing in an Integrated Wireless Cellular/WLAN System", Canadian Conference on Electrical and Computer Engineering, CCECE, 2007, pp. 631–634, Vancouver, April 2007.

[10] Y. Wang, H. Lo, Y. Li, and W. Lee, "Seamless Handover with Buffer Prediction for Wireless Networks Based on IEEE 802.21" International Conference on Information Networking, LNCS, 2008, pp. 1–5, Tokio, January 2008.

[11] Y. Wei, Y. Hu, and J. Song, "Network Selection Strategy in Heterogeneous Multi-Access Environment", The Journal of China Universities of Posts and Telecommunications 14, pp. 16–20, China, October 2007.

[12] F. F. Silva, "Jornalismo live streaming: tempo real, mobilidade e espaço urbano", VI National Meeting of Journalism Researchers, pp. 1–4, Universidade Metodista, São Paulo, November, 2008 .

Challenges of Operationalizing PACS on Cloud Over Wireless Networks

Ricardo Ferraro de Souza, Carlos Becker Westphall, Daniel Ricardo dos Santos, Carla Merkle Westphall
 Networks and Management Laboratory
 Federal University of Santa Catarina
 Florianopolis, Brazil

ricardoferraro@hotmail.com, westphal@inf.ufsc.br, danielrs@inf.ufsc.br, carlamw@inf.ufsc.br

Abstract— Clinics and hospitals are acquiring more technological resources to help providing a faster and more precise diagnostic, with the goal of making it more dynamic and effective. This is pushing health institutions to search for more modern equipment, with greater technological features. Besides last generation equipment, another problem faced by these institutions is enabling the connection of physicians to a Picture Archive and Communication Systems (PACS) from anywhere. With the use of communication resources increasingly present in everyday life, like Wireless-Fidelity (Wi-Fi), third generation of mobile telecommunications technology (3G), fourth generation of mobile telecommunications technology (4G), Worldwide Interoperability for Microwave Access (WiMax) and other wireless networks that allow the connection of mobile devices, it becomes easier and cheaper to provide quality medical services at a distance. Diagnoses that needed a doctor to be present, for instance, can now be performed from anywhere, provided there is an Internet connection. Cloud-based PACS is shown to be efficient for archiving medical images, allowing access to exams and reports from anywhere, over wireless networks, regardless of the platform used for access.

Keywords - Cloud Computing; PACS; Healthcare Systems; Wireless Networks.

I. INTRODUCTION

Medicine is being perfected through the use of innovative technological solutions in various equipment with diverse applications, such as image processing, blood analysis, surgical assistance and distance patient control.

Inside hospitals and clinics for diagnostic imaging, it is usual to find PACS. PACS have the goal of managing the storage and exhibition of medical images. Through workstations, doctors have access to the PACS system where they manipulate images independently of their physical location.

Specialized physicians achieve diagnosis through the analysis of images or the reading of reports. But these physicians are not always present where the exam was performed, especially in cases where the participation of a second doctor is necessary or in the case of training for resident physicians. The involvement of these doctors can happen with telemedicine.

Telemedicine comprehends the offering of services related to health care in cases where distance is a critical factor; such services are performed by health professionals using communication and information technologies for the interchange of information valid for diagnostics, prevention

and treatment of illnesses and the continual education of health service providers, as well as research [1].

The practice of telemedicine is only made possible because of significant advances in communication systems. The possibility of connectivity to the World Wide Web from mobile devices, a constantly evolving technology, allows patients to obtain adequate medical care in less-favored regions, where there are no doctors or wired Internet connection available. Wi-Fi Networks, 3G, 4G, WiMax and other wireless networks are being constantly improved with higher data transmission rates, allowing access to content not explored before, which aim to improve, simplify and complement the services related to patient care and make them more efficient.

Cloud computing is currently the main theme of a lot of research in information technology. The possibility of sharing resources through clusters, virtualization and the ease of access to information attracts more and more information technology researchers. This technology is also a powerful tool to promote the homogenization or virtualization of space [2].

Images from radiological exams are used in clinics and hospitals for medical diagnosis. The inter-relation among clinics, hospitals and radiology departments are increasingly dependent of the accessibility of these images, from anywhere inside or outside of the health care unit [3].

The idea is to use cloud as a model for applications being delivered as services over the Internet. Cloud services are built in such a way that if a machine fails, the system resets, in order to prevent the service to crash or that the contractor knows that there was some kind of problem. Cloud computing enables the growth of processing and storage infrastructure for hospitals and clinics without causing much impact. Thus cloud based PACS enables medical activity from anywhere using computers or devices connected to the Internet.

This paper is organized as follows. Section 2 discusses the interoperability of medical equipment and wireless networks. Section 3 discusses the fundamentals of cloud computing. Section 4 talks about PACS and cloud-based PACS. Section 5 cites the performance of wireless networks for medical image transmission. Section 6 approaches wireless security in medical environments. Section 7 shows a scenario where medical images may be accessed by remote devices. Section 8 is the conclusion.

II. INTEROPERABILITY CHALLENGES

Health charts, medical and laboratory reports, medical images and prescribed medicine are some of the items in a

medical record, and those records are becoming more and more complex. Physicians in hospitals and clinics need a flexible resource that allows them accessing information and history for each patient, because they work and meet patients in several places; they need to be frequently following exams and giving support to several people.

Medical records occupy a great storage space and the management of these data is a challenging task for hospitals and clinics [4]. To solve those issues, these organizations invest large amounts of money in infrastructure for communication, processing and storage of exams. Inside this infrastructure are equipment for ultrasonography, MRI, CT scans and radiography. Figure 1 shows the sending of medical images to the PACS.

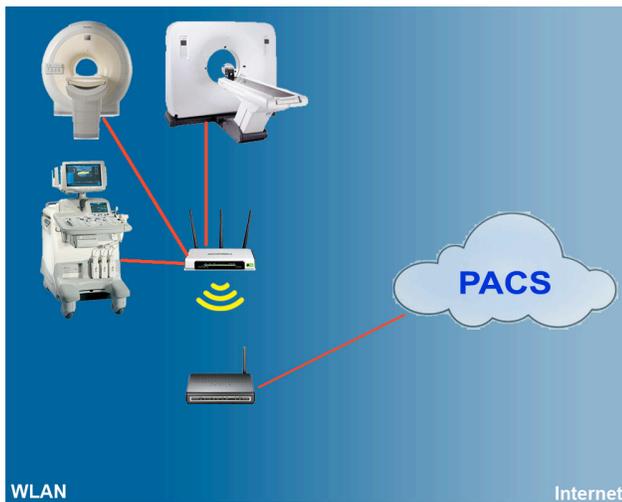


Figure 1 - Sending DICOM images to the cloud-based PACS

The modalities, as known equipment's, send images using Digital Image and Communication in Medicine (DICOM), a digital standard to store and transmit medical images, over a Wireless Local Area Network (WLAN) to a PACS located on a cloud. Communication from the medical equipment to the WLAN is achieved by using a device that connects to a Local Area Network (LAN), which in turn accesses the PACS that is on the cloud using the Internet.

III. CLOUD COMPUTING

The cloud computing approach enables the growth of processing and storage infrastructure for hospitals and clinics without causing much impact. Internet access and computing devices are available in most places, creating new opportunities to share and use online resources. A great number of Internet features and services like e-mail and storage are used daily as a kind of commodity. Patients are continuously being monitored undisturbed during their daily activities [5].

Cloud services are built in such a way that if a machine fails, the system readjusts, in order to prevent service crashes or the consumer even knowing there was some kind of problem.

The National Institute of Standards and Technology (NIST) defines cloud computing as "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [6].

It could be said that cloud computing is the result of the union of computational paradigms such as virtualization, service level agreements and grid computing, aimed at providing on-demand, service-based business models of utility computing [7].

Cloud computing can be classified into three models [6][8]:

- Software as a Service (SaaS): the customer uses the provided software, without being able to control the infrastructure. It can be accessed through a web browser or directly through the interface of a program;
- Platform as a Service (PaaS): the customer has control over the deployed applications and settings of the environment hosting applications in the cloud;
- Infrastructure as a Service (IaaS): the customer can deploy and run arbitrary software, and has control over operating systems, storage and deployed applications.

Clouds can be deployed in four ways and security policies depend on the business process, the type of information and desired levels of service:

- Private Clouds are managed by the company itself or by third parties, this cloud is accessed only by an organization;
- Public Clouds are available to the general public;
- Community Clouds are shared by companies with interests in common; and
- Hybrid Clouds are a composition between two or more implementations of clouds.

IV. CLOUD-BASED PACS

The idea of managing the storage of digital images from different modalities into a central database, emerged in the mid 80s. This database should be a file system that manages the storage and answers queries for images and related clinical data [9].

The system has a central database was called PACS. PACS can acquire, transmit, store and display information from medical images [10]. Acquire images directly from the modalities; store these images in their central database, so the hospitals and clinics no longer need to file radiological films reducing costs by helping in the preservation of nature and besides; PACS can transmit and display medical images through stations job. Workstations allow manipulate and process medical images.

PACS systems allow consultations through texts linked to images, however consultations through attributes of the image itself can facilitate diagnosis [11].

A cloud-based PACS solution must grant access to the file server from any place or platform [12]. A traditional PACS server consists of the following components: DICOM repository system and database [13]. The object repository calls an infrastructure with storage capacity to support all

DICOM exams. The database module supports the DICOM Information Model, which contains metadata information related to patients, the series of examinations and images. When PACS exams are received, the images are stored in the DICOM repository and the database is updated with elements drawn from examination.

For doctors, the PACS cloud allows access to historical images and the selection of "key images", which are images that have a supposed variation in normal patterns. Some files emphasize putting medical examination and opinion on the case available, so that other doctors can use it as a parameter or continue researching the subject.

V. NETWORK PERFORMANCE CHALLENGES

Being available on the cloud, medical data may be processed by an intelligent or distributed system and sent to a medical team for analysis [14].

Applications that access PACS allow doctors to perform diagnosis from any physical location using only a mobile device. These devices must have a DICOM application that allows the downloading of images using wireless networks, and the manipulation of a series of images.

Large data transfer rates, such as gigabits per second, will be available in common places, supporting applications that involve the synchronization of local device data with the cloud or the pre-caching of data for posterior use [15].

Wi-Fi or 802.11 networks are present in places such as airports, bus stations, cafes, shopping malls, squares and universities, among others. Those networks can achieve transmission rates of up to 300 Mbps at varying frequencies, allowing adequate access to the images of tests that are in the cloud.

3G networks appear as an alternative to users of mobile phones and tablets, especially. These networks reach long distances and support data and voice transmission, with rates of up to 10 Mbps. The efficiency of access to medical images on the cloud will vary based on the data transmission rate being used, but those networks allow the physicians to access images on distant places and even on the move.

4G networks, which are totally IP-based, are still used in few countries, but may achieve speeds of up to 100 Mbps in movement and 5 Gbps while still, making the access of pictures on a cloud-based PACS, through mobile devices as well as through computers, similar to the access over wired networks in CAT 5e and CAT 6 standards.

WiMax networks were developed to support Wireless Metropolitan Area Networks, using the 802.16 protocol, allowing rates beyond 1 Gbps, which is higher than the other technologies, except 4G, and again enabling the visualization of medical images to be as fast as over a wired network.

VI. SECURITY CHALLENGES

The security challenges in telemedicine and cloud-based PACS can be divided in three main categories: data transmission, data storage and data access.

Currently, medical data is generated in many different equipment to be later stored and processed [16], and all of these data must be transmitted over different networks, so

that those who need it may have ubiquitous and time-independent access to it.

During transmission, private data will probably transit through different types of networks, each with its specific security design. Medical data can be generated in equipment such as heart rate and blood pressure monitors and transit over a Wireless Body Area Network (WBAN), then be transmitted to a cloud server over a WLAN to be later accessed by a doctor via a 4G WWAN. All of the devices used in transmission and storage must ensure security properties such as availability, confidentiality, integrity and privacy [17]. Among the main threats to private data are misuse, malicious disclosure or modification [18].

The issues in secure data transmission and storage are solved by using cryptography in every communication, as well as keeping private and sensitive data stored in cryptographic devices at all times, while it is not being processed. The use of cryptography, however, brings the challenges of key management in distributed environments.

All sensitive and private data stored or in transit must be encrypted, and the access points, gateways and other communication systems must provide support for it. This includes every data that contains medical information or that may be used to identify patients and doctors.

Secure data access is not entirely solved by using cryptography. To allow efficient access to medical data, good Identity and Access Management procedures are necessary. The use of Federated Identity Management is extremely useful to allow the collaboration between doctors in different institutions and to enable ubiquitous access to important data.

Another primary concern in the manipulation of private data is the logging of every access to it, allowing posterior audit and possible incident response.

Accessing patients' private data in mobile devices brings many facilities, but it also presents problems, since this kind of device is easily lost or stolen, and even if all important data are stored on a cloud, these devices may contain access credentials or cached data, so they must also be encrypted.

A. Federated Identity

The concept of federated identity is used to provide reliability between the medical and health institutions and implement secure access. Users are represented by identities, which are a "representation of an entity (or group of entities) in the form of one or more elements of information (attributes) that enable the entity to be recognized only within a context" [19].

The credentials informed by users are managed by an Identity Management System (IMS), which uses Single Sign-On (SSO), and ensures the security of users' information through a set of functions and capabilities called Identity Management (IDM).

The IMS performs several functions, the most important being:

- Provisioning: the practice of providing identities within an organization;
- Authentication: making sure that the user that makes the access is really who he says he is;

- Authorization: the process responsible for allowing or denying access requests in different parts of the system or operations;
- Federation: a group of organizations or Services Provide (SPs) which provide a circle of trust in order to allow the sharing of information through digital identities [20].

Entities can be classified into three types: the User, the Identity Provider (IDP), responsible for issuing credentials and issuing and managing user identities, and the Service Provider (SP), which is the entity that provides services to users based on their attributes.

B. Shibboleth

Shibboleth [21] [22] is an open-source Identity Management System developed by Internet2 that enables the construction of identity federations and the use of SSO. It also allows the sharing of attributes using the Security Assertions Markup Language (SAML) standard. It is widely used in academic and industrial scenarios to deploy both Identity Providers as well as Service Providers.

Shibboleth is comprised of three main components, the identity provider, responsible for managing user identities and authentication; the service provider, responsible for controlling the access of users to resources; and the Where Are You From? (WAYF), responsible for allowing users to choose their desired identity provider during authentication.

Users authenticate with their organizational credentials. The identity provider then passes minimal identity information, in the form of attributes predefined in an agreement, to enable authorization decisions by the service providers.

Federations are formed by the exchange of metadata containing digital certificates and other information about the entities in the circle of trust.

VII. PROPOSED SCENARIO

This work shows a cloud-based PACS solution, allowing exams to be performed in various equipment and transmitted over wireless networks, so that from any connected mobile device a doctor may have access to PACS exam images. Figure 2 shows diverse equipment such as cellphones, tablets and laptops accessing the cloud-based PACS over wireless connections, such as Wi-Fi, 3G, 4G or WiMax.

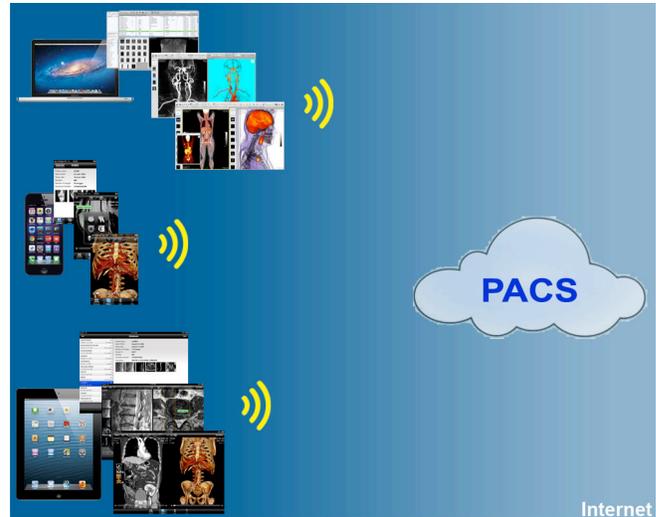


Figure 2 – Accessing PACS through wireless networks.

The PACS system in the implementation is DCM4CHEE [23] [24], with a PostgreSQL database. The devices for test access were a laptop, a smartphone and a tablet, all of them with the Osirix application. All of them are using publicly available 3G and Wi-Fi connections. In every case the connection is successful, showing a mean time of 20 seconds to download a DICOM exam of a lumbar spine containing 100 images.

Table 1 shows the comparison of mean times for accessing a picture, based on different mobile devices and wireless networks. The sizes of medical images vary depending on the matrix used in its acquisition. In this case as the examination of the lumbar spine is the predominant matrix, the matrix is 512 X 512, resulting in an image of 1.6 Mbyte.

TABLE I. DOWNLOAD TIME OF A APPLE DICOM IMAGE

Device / Transmission	Wi-Fi	3G
iPhone	12s	33s
iPad	11s	28s
MacBook	10s	26s

All the mobile devices uses, are Apple products (MacBook, iPhone a iPad) and they present a flexible platform for the development and practice of telemedicine with modern high-resolution equipment and high processing capacity, besides providing the usability of touchscreen [25] [26].

Other mobile devices, which are not support by the Osirix software, may access PACS directly through a web interface. Users accessing the system through the web will be authenticated using their federated identities.

In this case, the user accesses the PACS via browser. The system detects that the user is not authenticated and redirects him to the WAYF of the federation. The user selects his identity provider and informs his credentials, and then is automatically redirected to the PACS system where he can browse and view the patient exam images.

These images can be viewed in two distinct ways. One is to download or view the images in their browser, more in Joint Photograph Experts Group (JPEG) format. The second way is to have the java environment installed on your computer or device, so that the user can perform the download a jnlp file that will be executed at the opening of the viewer, and an extension of the PACS server dcm4chee called Weasis.

The Weasis is a DICOM image viewer developed as plugin for dcm4chee PACS, where the user has some functions to manipulate images of medical examination of the patient. Table 2 shows the opening time of the lumbar spine quoted in Table 1 using the two ways of viewing described above. Appliances are 2 notebooks with Core i5 2.53GHz with 2 cores, 4 Giga Bytes of Random Access Memory (RAM) and operating system Windows 7 Professional 64-bit and Ubuntu 12.04 LTS 64-bit.

TABLE II. DOWNLOAD TIME OF A IMAGE

Device / Transmission	JPEG		Weasis (DICOM)	
	Wi-Fi	3G	Wi-Fi	3G
Notebook (Win7)	2s	4s	14s	38
Notebook (Ubuntu)	2s	4s	13s	37

In comparison with Apple devices that use a communication protocol direct dcm4chee PACS through the method retrieve WADO [24], others equipment also show up more efficient with a longer time considering opening DICOM images. The display mode of JPEG images is much faster but the doctor can see static image, without the power of compounding. A JPEG image is not acceptable for diagnosis. The processor that owns the computer or device has an influence only after the examination to be loaded into memory ready to be manipulated.

Identity Management Implementation

The PACS system was installed and configured on a Windows 2008 Server R2 virtual machine. On this same machine the Shibboleth, Apache 2.2, Tomcat 7, OPENLDAP, JASIG CAS, JDK 1.7 and PostgreSQL 9.2 services were configured.

Apache was configured to allow the use of Secure Sockets Layer (SSL) connections, and also to proxy its requests to Tomcat. Tomcat was configured to run the authentication and identity management applications.

The authentication server enables JASIG CAS SSO authentication via a web interface passing authenticated users to Shibboleth. CAS was set up to search for users in a Lightweight Directory Access Protocol (LDAP) directory.

With the server configured, Shibboleth was installed. The IDP application was installed and configured in Tomcat. To act as the service provider, the Shibboleth application must belong to a federation. TestShib was chosen as the federation, because it was created to test Shibboleth configurations (SPs and IDPs). An IDP was registered informing the hostname and digital certificate created earlier. Shibboleth was then configured to use TestShib's metadata. The CAS Client receives shibboleth

authentication. With this process ready Shibboleth is configured and authenticating users from JASIG CAS.

VIII. CONCLUSION

The need for finding more precise diagnostics allowing effective treatment for patients pushes for a constant technological evolution in medical equipment, as well as smartphones, tablets and laptops that are used to access the images and the communication of these devices with the cloud.

The presence of communication resources in daily life over Wi-Fi, 3G, 4G or WiMax reaching high data transmission rates, enables access to medical diagnostics at a distance using Internet-connected mobile devices, downloading DICOM images with an appropriate application.

The use of a cloud-based PACS has the goal of showing the archiving of medical exam images from different locations in a centralized repository, lowering the investments on storage and processing infrastructure for hospitals and clinics. On the cloud, doctors and patients may visualize these images through any connected mobile device that provides Internet access.

The test results were satisfactory, considering data transmission rates, showing that mobile devices and a cloud-based PACS present a viable solution for the practice of telemedicine.

REFERENCES

- [1] World Health Organization. <http://www.who.org>. [retrieved: December 2008].
- [2] H. A. Franke, F. L. Koch, C. O. Rolim, C. B. Westphall and D. O. Balen, "Grid-M: Middleware to Integrate Mobile Devices, Sensors and Grid Computing," Third International Conference on Wireless and Mobile Communications, March 2007, pp. 19-25.
- [3] E. M. B. Junior. "Teleradiology: Central Remote Diagnostic Imaging Digital Integrated Portal to a Distributed Medical Information. Application of public," Federal University of São Paulo, São Paulo, 2009.
- [4] L. He, X. Ming, W. Ding and Q. Liu, "A Novel Approach To Remote Access Picture Archiving And Communication System On Mobile Devices Over Wireless Networks," Biomedical and Health Informatics (BHI), 2012 IEEE-EMBS International Conference on January 2012, pp. 581-583.
- [5] R-D. Berndt, et al., "SaaS-Platform for Mobile Health Applications," Systems, Signals and Devices (SSD), 2012 9th International Multi-Conference on March 2012, pp. 1-4.
- [6] E. Brown, Published final definition of Cloud Computing. <http://www.inovacaotecnologica.com.br/noticias/noticia.php?artigo=definicao-computacao-em-nuvem&id=010150111128> [retrieved: December 2011].
- [7] S. Chaves, R. Uriarte and C. B. Westphall, "Toward an architecture for monitoring private clouds," Communications Magazine, IEEE. December 2011, pp. 130-137.
- [8] F. Schubert, C. Rolim and C. B. Westphall, "Application Provisioning Algorithms Based on Service Level Agreements for Cloud Computing," IX Workshop Clouds, Grids and Applications. 2011.
- [9] A. R. Bakker, "HIS and RIS and PACS," Picture Archiving and Communication Systems (PACS) in Medicine. NATO ASI Series Volume 74, 1991, pp 157-162.

- [10] R. L. Arenson, "Picture archiving and communication systems," *Western Journal of Medicine* on March 1992, pp. 298–299.
- [11] J. M. Bueno, F. Chino, A. J. M. Traina, A. J. M. Traina, P. M. Azevedo-Marques, "How to Add Content-Based Image Retrieval Capability in a PACS," *IEEE International Conference on Computer Based Medical Systems – CBMS*. 2002, pp. 321-336.
- [12] L. A. B. Silva, C. Costa, A. Silva and J. L. Oliveira, "A PACS Gateway to the Cloud," *Information Systems and Technologies (CISTI)*, 2011 6th Iberian Conference on June 2011, pp. 1-6.
- [13] YJ. Ni, et al., "Implementation and Evaluation of A PQRM-based PACS System," *Information Technology Applications in Biomedicine, 2007. ITAB 2007. 6th International Special Topic Conference* on November 2007, pp. 323-326.
- [14] C. Rolim, et al., "Cloud Computing Solution for Patient's Data Collection in Health Care Institutions," *eHealth, Telemedicine, and Social Medicine, 2010. ETELEMED '10. Second International Conference* on February 2010, pp. 95-99.
- [15] R. Baldemair, et al., "Evolving Wireless Communications: Addressing the Challenges and Expectations of the Future," *Vehicular Technology Magazine, IEEE*. March 2013, pp. 24-30.
- [16] H. Furtado and R. Trobec, "Applications of wireless sensors in medicine," *MIPRO, 2011 Proceedings of the 34th International Convention*. May 2011, pp.257-261.
- [17] W. Lou and K. Ren, "Security, privacy, and accountability in wireless access networks," *Wireless Communications, IEEE*. August 2009, pp. 80-87.
- [18] M. Ameen, J. Liu and K. Kwak, "Security and Privacy Issues in Wireless Sensor Networks for Healthcare Applications," *Journal of Medical Systems, Springer*. 2012, pp. 93-101.
- [19] M. Leandro, T. Nascimento, et al., "Multi-Tenancy Authorization System with Federated Identity for Cloud-Based Environments Using Shibboleth," *The Eleventh International Conference on Network*. 2012, pp. 88-93.
- [20] Z. A. Khattak, S. Sulaiman and J. Manan, "A Study on Threat Model for Federated Identities in Federated Identity Management System," *Information Technology (ITSim), 2010 International Symposium* in June 2010, pp. 618-623.
- [21] M. Erdos and M. SCantor, "Shibboleth Architecture Technical Overview Working Draft 02[Z]," *Shibboleth Project*, 2005.
- [22] W. Ying, "Research on Multi-Level Security of Shibboleth Authentication Mechanism," *Third International Symposium on Information Processing* in October 2010, pp. 450-453.
- [23] Open Source Clinical Image and Object Management. <http://www.dcm4chee.org>. [retrived: May 2013].
- [24] M. J. Warnock, C. Toland, D. Evans, B. Wallace and P. Nagy, "Benefits of Using the DCM4CHE DICOM Archive," *Journal of Digital Imaging* in November 2007, pp. 125-129.
- [25] A. F. Choudhri and M. G. Radvany, "Initial Experience with a Handheld Device Digital Imaging and Communications in Medicine Viewer: OsiriX Mobile on the iPhone," *Journal of Digital Imaging*, 2011, pp.184-189.
- [26] G. Shih, P. Lakhani, et al., "Is Android or iPhone the Platform for Innovation in Imaging," *Informatics. Journal of Digital Imaging*, 2010, pp. 2-7.